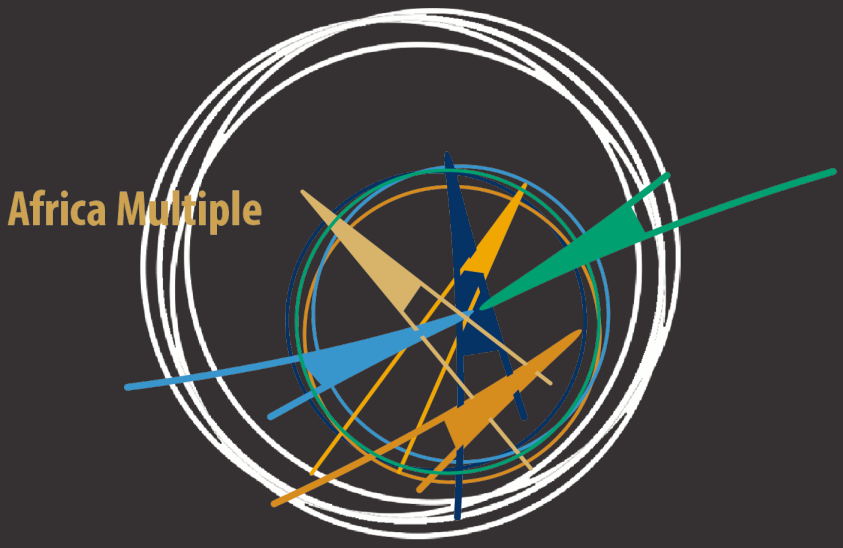
9

University of Bayreuth
African Studies Online



Frontiers in African Digital Research

Anja R. Dreiser and Cyrus Samimi, 2022

UNIVERSITÄT BAYREUTH
INSTITUTE OF AFRICAN STUDIES
africamultiple CLUSTER OF EXCELLENCE

9

University of Bayreuth African Studies Online

Frontiers in African Digital Research

Conference Proceedings

Anja R. Dreiser, Cyrus Samimi, 2022

Institute of African Studies (IAS)

The *University of Bayreuth African Studies Online* are published by the Institute of African Studies (IAS) in Bayreuth, Germany.

The IAS promotes African Studies at the University of Bayreuth by supporting scholars from a wide range of academic disciplines from almost all faculties. It facilitates cooperation between researchers and institutions engaged in Africa-related projects, as well as teaching, both on campus and around the world. The IAS consists of three central bodies: the Iwalewahaus, the Africa Multiple Cluster of Excellence, and the Africa Research Center (*Forschungszentrum Afrika*), which is currently under construction.

University of Bayreuth African Studies Online presents results of Africa-related research at the University of Bayreuth. Publications include articles, edited collections, and single authored monographs – including but not limited to PhD-theses. Submitted papers are subject to external double-blind review. Contributions may be submitted to the editor-in-chief: Dr. Lena Kroeker (mailto:Lena.Kroeker@uni-bayreuth.de)

The *University of Bayreuth African Studies Online* is registered under ISSN 2628-1791 on the EPub document server at the university library:

2005 - 2018

https://epub.uni-bayreuth.de/view/series/Bayreuth_African_Studies_Online.html

2018 - ongoing

https://epub.uni-bayreuth.de/view/series/University_of_Bayreuth_African_Studies_Online.html

Institute of African Studies

Director: Prof. Dr. Eva Spies
Vice-Director: Prof. Dr. Thoko Kaime

University of Bayreuth
Wölfelstr. 2
D-95440 Bayreuth

Phone: +49 (0)921 554511
Fax: +49 (0)921 554502

www.ias.uni-bayreuth.de
IAS@uni-bayreuth.de

University of Bayreuth African Studies Online

Title	Editor(s)	Year of Publication	Issue
Gesundheitssicherung im Kontext sozialer Differenzierung. Wer tritt Krankenkassen im ländlichen Mali bei?	Alexander Schulze	2020	8
Being In-Between. Middle Income Groups in Uganda and their Forms (and Absences) of Political and Social Mobilization	Anna Fichtmüller	2019	7
Competing and Conflicting Power Dynamics in Waqfs in Kenya, 1900-2010	Suleiman A. Chembea	2018	6
Die Finanzierung von Eigentumswohnungen in Addis Abeba. Eine diskursanalytische Annäherung an Wohneigentum	Nadine Wenzel	2008	5
Die Spur der Glasperlen. Akteure, Strukturen und Wandel im europäisch-ostafrikanischen Handel mit Glasperlen	Ulf Vierke	2004	4
Globalization in a Local Context: Perspectives and Concepts of Action in Africa: An Introduction	Roman Loimeier; Dieter Neubert and Cordula Weißköppel	2005	2
From an Anthropology of Astonishment to a Critique of Anthropology´s Common Sense: An Exploration of the Notion of Local Vitality in Africa	Peter Probst and Gerd Spittler	2005	1

Africa Multiple Cluster of Excellence

The Africa Multiple Cluster of Excellence was established in January 2019 through the Excellence Strategy of the German Federal and State Governments. Building on the well-established programme of African Studies at the University of Bayreuth, the Cluster pursues an innovative agenda as expressed in its subtitle – *Reconfiguring African Studies*. It has almost one hundred and sixty members, from three continents, who represent a diverse range of academic disciplines and cooperate with partner institutions in Africa, Europe, Asia, and the Americas. Our understanding of the reconfiguration of African Studies focuses on stimulating new theoretical approaches and the creation of new forms of interdisciplinary collaboration. In addition to the University of Bayreuth, the Cluster is made up of four further African Cluster Centres (ACCs): the Universities of Lagos (Nigeria), Joseph Ki-Zerbo (Burkina Faso), Moi (Kenya), and Rhodes (South Africa).

Our key concepts are *multiplicity, relationality,* and *reflexivity.* We employ them to capture the dynamic interrelationship of diversity and entanglement that characterize African and African diasporic ways of life and world-making. In the "Knowledge Lab", we connect our theoretical, epistemological, and methodological issues. Our "Digital Research Environment" integrates analogue and digital data into a common, digital research platform. Through the Bayreuth Academy of Advanced African Studies, the Cluster coordinates an international fellowship programme for junior and senior researchers, including artists. Bayreuth International Graduate School of African Studies (BIGSAS) offers research-oriented training for doctoral students.

In addition to *University of Bayreuth African Studies Online*, there are the *University of Bayreuth African Studies Working Papers*. The working paper series present insights into ongoing research projects in the field of African Studies at the University of Bayreuth and beyond.

Africa Multiple Cluster of Excellence

Spokesperson: Prof. Dr. Rüdiger Seesemann
Deputy Spokesperson: Prof. Dr. Ute Fendler

University of Bayreuth
Nürnberger Str. 38, ZAPF Haus 3
D-95440 Bayreuth

https://www.africamultiple.uni-bayreuth.de

africamultiple-international@uni-bayreuth.de

About the Authors

Khuyagbaatar Batsuren received the B.S. degree in information technology from National University of Mongolia in 2012, the M.S. degree in computer science from Chungbuk National University, South Korea in 2015, and then Ph.D degree in computer science from University of Trento, Italy in 2018. He worked at University of Trento as a Postdoctoral Researcher. He is currently an Associate Professor at National University of Mongolia. His research interests focus on building multilingual linguistic resources such as morphology, cognates, and colexification. https://orcid.org/0000-0002-6819-5444.

Oliver Baumann is a PhD student and research assistant within the Excellence Cluster "Africa Multiple". He studied computer science at the Ludwig-Maximilians-University of Munich and is now part of Prof. Dr. Mirco Schönfeld's group at the Professorship for Data Modelling & Interdisciplinary Knowledge Generation. His research interests include topics related to contextualized ontologies, natural language processing, data mining, and linking and visualizing latent structures in research datasets. Of particular interest to him is the interdisciplinary application of his research. Contact: Oliver.Baumann@uni-bayreuth.de, https://orcid.org/0000-0003-4919-9033.

Gábor Bella is Associate Professor at the Department of Computer Science and Information Engineering of the University of Trento. He is researching computational systems that deal with the diversity of languages and their speakers, with two goals in mind: to facilitate cross-lingual communication, but also to maintain what is unique and different in each language. As an example of the former, his research on the semantic interoperability of healthcare data aims to allow patients to be cured regardless of linguistic and administrative barriers. His research on linguistic diversity and bias in language processing systems, on the other hand, tries to help capture “the local”: words and ideas that tend to get lost in translation, yet that tend to matter to us all the more. Having a knack for implementing things, he enjoys building actual systems and participating in projects where his research is applied. http://sites.google.com/site/gaborbellaphd Contact: gabor.bella@unitn.it, https://orcid.org/0000-0002-3868-1740.

Jody Butterworth is the Curator for the Endangered Archives Programme based at the British Library, London / UK. Jody has been working for the Endangered Archives Programme, based at the British Library, since December 2012. She first joined the Library in 2006, working in the Prints, Drawings and Photographs section and then worked in an administrative role within the Collection Care department. Her eclectic experiences prior to the Library include working for the National Trust, the Ashmolean Museum and the Royal Collection. Contact: Jody.Butterworth@bl.uk https://orcid.org/0000-0001-7251-7269.

Markus Coester is a cultural anthropologist and ethnomusicologist at the University of Bayreuth/Germany and a project director in the African Studies’ Cluster of Excellence Africa

Multiple: Reconfiguring African Studies. His work and research has centred on popular music and culture, mainly Highlife, preservation, digitisation, and repatriation of important music archives and collections, like the Olive Lewin Collection in Jamaica and the GBC's Gramophone Library in Ghana, as well as the practice and politics of safeguarding and re/creating intangible cultural heritage. Contact: Markus.Coester@uni-bayreuth.de, https://orcid.org/0000-0002-1357-122X.

Mirjam de Bruijn is Professor in 'Contemporary History and Anthropology of Africa' at Leiden University and the African studies Centre Leiden (ASCL). Her specific fields of interest are: mobility, youth, social (in)security, marginality/exclusion, human rights, Information and Communication Technologies (ICTs). She received several grants for research programmes on these topics. Her latest project is DDMAC: Decoding Digital Media in African Conflicts. It was funded by the Norwegian Research Council. In 2016 she founded the organisation Voice4thought (www.Voice4Thought.org). Mirjam teaches 'innovative methods' and 'Media and Power' in the programmes African Studies and History of Leiden University. She is chair of the BA programme African Studies at Leiden University. Contact: m.e.de.bruijn@hum.leidenuniv.nl, https://orcid.org/0000-0002-0499-8758.

Mireille Djenno is the African Studies Librarian and library liaison to the African Studies Program at Indiana University Bloomington. Her research interests include collection development, archives and special collections, and the digital humanities. She oversees several archival collections related to Africa including the Liberian Collections, the Digital Somali Collection, and the H.K. Banda Archive. She is responsible for the development and management of collections of materials from and about the continent of Africa, with particular emphasis on sub-Saharan Africa. Djenno received an Master of Arts in Comparative Literature and Master of Arts in Library and Information Studies from the University of Wisconsin-Madison, and a Bachelor of Arts in Philosophy, French, and Spanish from the University of Wisconsin-Stevens Point. Prior to joining the IU Libraries in 2016, she served as Librarian for African, African American and Diaspora Studies at the University of North Carolina at Chapel Hill. Contact: mdjenno@indiana.edu, https://orcid.org/0000-0003-0790-2915.

Anja Dreiser is an Art Historian and Archaeologist of the Arts and Material Culture of the Islamic World (University of Bamberg/Germany, 2006) and a trained Information Specialist (University of Applied Sciences Potsdam/Germany, 2010). She has been active in both her professions: as Art Historian & Archaeologist she was a research associate in archaeological fieldwork projects in Oman and Mongolia with the German Archaeological Institute and at Bamberg University. As an Information Specialist she worked inter alia with the German broadcasting company "SWR" and in the GLAM department of the "Competence Centre Baden-Württemberg" (BSZ Konstanz). From 2019-2021 she was Data Curator of the Digital Solutions unit of the "Africa Multiple Cluster of Excellence" and in this position organizer of the international conference "Frontiers in African Digital Research". As Information Specialist her interest lie in data modeling, information retrieval and topics of information management, especially transparency in data collections and topics of information bias. Anja Dreiser is currently working as Digital Curator for the "State Gardens and Palaces of Hesse." Contact: anru.dreiser@outlook.de,

Øyvind Eide is a Professor in Digital Humanities at the University of Cologne. He holds a PhD in Digital Humanities from King's College London/UK (2013). He was the chair of The European

Association for Digital Humanities (EADH) from 2016–2019 and actively engaged in international organisations including ICOM's International Committee for Documentation (CIDOC). His research interests are focused on transformative digital intermedia studies, using critical stepwise formalisation as a method for conceptual modelling of cultural heritage information. This is used as a tool for critical engagement with media differences, especially the relationships between texts and maps as media of communication. He is also engaged in theoretical studies of modelling in the humanities as well as beyond. Contact: oeide@uni-koeln.de, https://orcid.org/0000-0002-7766-6287.

Jan Philipp Eisenhuth studied Computer Science at the University of Bayreuth, where he is also currently doing his Ph.D. at the Chair for Databases and Information Systems. Since early 2019 he is working at the Africa Multiple - Cluster of Excellence. He is part of the Digital Solutions team, with the focus of building and managing the infrastructure of a novel digital research environment. His research interest lies within the broad field of heterogeneous data management carried out on various types of database systems. Contact: Philipp.Eisenhuth@uni-bayreuth.de.

Myriel Fichtner has studied Computer Science at the University of Bayreuth, Germany. She received a Bachelor of Science degree with focus on robotics and image processing in 2017. In 2019, she completed a Master of Science degree with focus on business processes. Since June 2019, she has been a research assistant and PhD student at the Chair for Databases and Information Systems at the University of Bayreuth under the supervision of Prof. Dr.-Ing. Stefan Jablonski. As part of this function, she is in charge for the development and implementation of an ontology-based digital research environment for the Cluster of Excellence “Africa Multiple” of the University of Bayreuth. She is a member of the cluster's Digital Solution Team, led by Prof. Dr. Cyrus Samimi, and discusses system requirements with focus on conceptual and technical decisions. Contact: Myriel.Fichtner@uni-bayreuth.de, https://orcid.org/0000-0003-0412-1881.

Matthias Fuchs after studying and working in the fields of transport and traffic in Germany, Matthias Fuchs has been working at the Saxon State and University Library Dresden as project leader of the Specialised Information Service Mobility and Transport Research since 2016. The Mobility Compass is an important part of this project. Contact: Matthias.Fuchs@slub-dresden.de, https://orcid.org/0000-0003-3365-3158.

Fausto Giunchiglia is Professor of Computer Science at the University of Trento, EURAI fellow, AAIA fellow, member of the Academia Europaea. Previously studied or had positions at the Universities of Genoa, Stanford, Edinburgh, and at the research center IRST (now FBK, Trento). Current research interest: computational models of the mind and implications on how the known is grounded in the unknown. Around 10 Best Paper Awards; 50+ invited talks in International events; chair of 10+ international events, among them: ICDCS, KSEM, ODBASE, IJCAI, KRR; editor or editorial board member of around 10 journals, among them: Journal of Autonomous Agents and Multi-agent Systems, Journal of applied non Classical Logics, Journal of Software Tools for Technology Transfer, Journal of Artificial Intelligence Research. He was member of the IJCAI Board of Trustees, President of IJCAI, President and Advisory Board member of KR, Inc., Steering Committee of the CONTEXT conference. Fausto has covered all the spectrum from theory to technology transfer and innovation. He was involved in more than 20 international R&D projects,

including several EC projects. More details can be found at: http://fausto.disi.unitn.it/ https://orcid.org/0000-0002-5903-6150.

Jonas Huisl first studied European Studies with a focus on Hispanic Studies and Geography at the University of Passau. There he came into contact with the diverse possibilities that arise when applying computer science methods to questions in the humanities. Therefore, he subsequently completed the Master's programme Computing in the Humanities at the Otto Friedrich University of Bamberg. He is particularly interested in digitisation, data management and the applications that arise from structured data. Contact: jonas.huisl@web.de, https://orcid.org/0000-0002-5443-2576.

Stefan Jablonski is Full Professor at the University of Bayreuth, Germany. Since 2006 he is head of the Chair for Databases and Information Systems. Stefan Jablonski holds a Doctoral Degree (Dr.-Ing.) in Engineering Science from the University of Erlangen-Nuernberg. Stefan Jablonski's research focus lies firstly on process management. Therein, the conceptual foundation of process management, modelling and agile processes execution and mining build the current core areas. His second research focus lies on specific areas of Data Science. Besides the management of research data, natural language processing for analyzing unstructured text build the two main pillars of these research activities. Contact: Stefan. Jablonski@uni-bayreuth.de.

Temuulen Khishigsuren is M.Sc student in linguistics at National University of Mongolia. Her broader research interests are related to language and cognition, cognitive linguistics, and also psychology and neurobiology of language. Her current research interest is analysing language in order to answer or draw attention to important psychological and cognitive science questions. https://orcid.org/0000-0001-5251-5733.

Wolfgang Kraus is Associate Professor at the Department of Social and Cultural Anthropology, University of Vienna, Austria. He has done extensive ethnographic research on tribal identity and organisation and on oral historical knowledge among Imazighen/Berbers in central Morocco. Other research interests include ethnographic methods, kinship, visual and multimedia anthropology, data management and research ethics. At his home department he is in charge of planning and organising the teaching programme in Social and Cultural Anthropology. Currently he is leading the Ethnographic Data Archive (eda) that he established together with Igor Eberhard since 2017 at the University Library Vienna. Contact: Wolfgang.Kraus@univie.ac.at.

Michael Kwet is a visiting fellow at Yale Law School's Information Society Project and a postdoctoral fellow in Sociology of the Centre for Social Change at the University of Johannesburg. He hosts the Tech Empire podcast at the Yale Podcast Network. Dr. Kwet is editor of the forthcoming book, The Cambridge Handbook of Race and Surveillance on Cambridge University Press, due out in 2022. His work has been published at the academic journals Race & Class, First Monday, and Teaching in Higher Education, and Markets, Globalization & Development Review, as well as media outlets like The New York Times, VICE News, Al Jazeera, Wired, Mail & Guardian, CounterPunch. He received his Ph.D. in sociology from Rhodes University in South Africa. Contact: Michael.Kwet@yale.edu, https://orcid.org/0000-0002-3304-5649.

'Tunde Ope-Davies (formerly Opeibi) is a Full Professor of Digital Discourse, English, Digital Cultures and Discourse Studies at the University of Lagos, Nigeria. He is the founder and director of the Centre for Digital Humanities, University of Lagos (www.cedhul.com.ng). He has won a number of visiting positions and fellowships such as the Commonwealth Professional Fellowship, University of Westminster, London; the Alexander von Humboldt Foundation's Georg Forster Fellowship for Experienced Researchers, at Chemnitz University of Technology, Germany; DAAD Scholarship to attend the University of Leipzig's Summer University in Digital Humanities, the Visiting Research Fellowship of the Institute for Advanced Studies in the Humanities (IASH-SPSS), University of Edinburgh, United Kingdom, and the Humboldt's Visiting Research Fellowship at the University of Hamburg, Germany. He is the immediate past Dean, Faculty of Humanities, Anchor University, Lagos. Professor Ope-Davies is the convener of the Lagos Summer School in Digital Humanities (LSSDH) and a grantee of the Volkswagen Foundation, Germany, and the Nigerian Tertiary Education Trust Fund (TETFund). He is currently the founding President of the Digital Humanities Association of Nigeria (DHAN). Contact: bopeidi@unilag.edu.ng, https://orcid.org/0000-0002-1772-6356.

Stefan Ouma holds the Chair of Economic Geography at the University of Bayreuth. He is working at the intersections of economic geography, political ecology, science and technology studies, and agrarian studies. A defining feature of his research has been the ambition to unpack how global economic processes manifest themselves in the socialites, materialities and spatialities of everyday economic life across the Global South. He holds a BA in Human Geography from the University of Erlangen-Nuremberg, an MA in African Development Studies from Bayreuth University and a PhD in Economic Geography from the University of Frankfurt/Main. Contact: Stefan.Ouma@uni-bayreuth.de, https://orcid.org/0000-0002-3180-1754.

Cyrus Samimi holds a professorship of climatology at the University of Bayreuth where he also serves as Vice Dean of Digital solutions in the Cluster of Excellence Africa Multiple. Before joining the University of Bayreuth, Prof. Samimi held the professorship of geoecology at the University of Vienna first as visiting then as full professor. His research projects on climate and environmental change mainly in Southern Africa, the Sahel and Central Asia have strong statistical and modelling approaches partly working with big data. Most of his projects are interdisciplinary and partly transdisciplinary, the later with the aim of joint knowledge generation. He published more than 60 papers and book chapters. Contact: Cyrus.Samimi@uni-bayreuth.de, https://orcid.org/0000-0001-7001-7893.

Mirco Schönfeld is currently an assistant professor for Data Modelling and Interdisciplinary Knowledge Generation at University of Bayreuth. Before, he worked as a postdoctoral researcher at the Computational Social Science & Big Data group at Bavarian School of Public Policy at the Technical University of Munich. He received his PhD in computer science from the Institute of Informatics and the Mobile and Distributed Systems group at the Ludwig-Maximilians-University Munich. His research interests include data modeling and algorithms for network analysis, text mining, and context awareness. He enjoys working at the intersection of computer science with the humanities and social sciences. He is also a contributor to open source projects. Contact: Mirco.Schoenfeld@uni-bayreuth.de, https://orcid.org/0000-0002-2843-3137.

Lee Watkins is the Director of the International Library of African Music, a music archive affiliated with Rhodes University. His research interests include the applied studies of music in as far as music can assist with developing rural cultural and heritage economies, migration and refugeeism, nationalism, the discontinuities of power, and hip hop. He has published several articles on hip hop and rap style in South Africa, the archive and decolonisation, and migrant Filipino musicians in Hong Kong. He is a member of the new Cluster project entitled, "Beyond the "Digital Return": "New Heritage/s, Sustainability, and the Decolonisation of Music Archives in South Africa, Kenya, and Ghana." Contact: l.watkins@ru.ac.za, https://orcid.org/0000-0003-2980-6043.

Oliver Watteler is a senior researcher at GESIS – Leibniz Institute for the Social Sciences, where he works in the department Data Services for the Social Sciences (DSS). He holds a master's degree in history and political science and is responsible for data acquisition. Together with colleagues he advises and trains on topics of research data management. His focus is on legal conditions of RDM, especially data protection, on which he also publishes. He is a member of the Leibniz Association's Data Protection Working Group and vice member of the GESIS Ethics Committee. Contact: Oliver.Watteler@gesis.org, https://orcid.org/0000-0002-1634-9229.

Stefan Wolff is a programmer at the Saxon State and University Library Dresden, where he is substantially responsible for the developments of the Mobility Compass. https://orcid.org/0000-0003-0015-9671.

Table of Content

Frontiers in African Digital Research

Anja R. Dreiser, Cyrus Samimi (University of Bayreuth)[1]

Preface

This volume of the *University of Bayreuth African Studies Online* series comprises most of the talks given at the conference "Frontiers in African Digital Research" organized by the "Digital Solutions" portfolio of the Cluster of Excellence Africa Multiple that was held in the Iwalewahaus of the University of Bayreuth from 30 January to 1 February 2020.

But what is the Digital Solutions portfolio of the Cluster of Excellence and why are we talking of frontiers for our field? We consider some preliminary contextualizing remarks appropriate to start with.

The implementation of research data management units at German universities is a fairly recent development. The need for specialized units which, in view of the data deluge, manage and curate e-science processes according to the FAIR principles[2] was in fact articulated in the field of natural sciences, such as astronomy, physics or climate studies, as the "fourth paradigm" back in 2009.[3]

[1] Contact: anru.dreiser@outlook.de.

Contact: Cyrus.Samimi@uni-bayreuth.de, https://orcid.org/0000-0001-7001-7893.

[2] The well-known four FAIR principles should now be supplemented by the four metadata guiding principles *"compatible, complete, credible, curated"* that have been articulated by the "Metadata 2020 project", an international community of stakeholders from across the scholarly field. Cf. Kaiser K, Kemp J, Paglione L, Ratner H, Schott D, Williams H (2020), Methods & Proposal for Metadata Guiding Principles for Scholarly Communications. Research Ideas and Outcomes 6: e53916. https://doi.org/10.3897/rio.6.e53916.

[3] Cf. Christine L. Borgman, The Conundrum of Sharing Research Data, in: Journal of the American Society for Information Science and Technology, 63 (6), 2012: 1059. Equally in German language: Stephan Büttner, Hans-Christoph Hobohm, Lars Müller, Research Data Management, in: ibidem., Handbuch Forschungsdatenmanagement, Bad Honnef 2011: 13-14.

While these disciplines could already draw on well-structured databases run by well-established, well-equipped and well-organized research institutions, these framing conditions also evolved in the field of social sciences (e.g. the SowiDataNet | datorum by the GESIS Institute)[4] but still do not exist for most of the humanities, arts and cultural studies.[5] Furthermore, research data management, as the existing digital data repositories show, is mostly found to be discipline driven, usually not linked to further possibly existing repositories in other countries let alone designed under interdisciplinary considerations.

In view of the developments of e-sciences and accompanying infrastructures in German academia, the Cluster of Excellence Africa Multiple with its Digital Solutions portfolio has in this respect set the ambitious aim of establishing a digital research repository for its disciplinary heterogenous research projects ranging from economics to climate studies to linguistics, history and media studies,[6] the "only" common ground being the studied area "Africa" and "African diasporas" (see also the contributions by 'Tunde Ope-Davies and Jonas Huisl et al. for more information on the Cluster and its digital aims). But not only for Bayreuth based researchers: the vision is to link this repository in a second step to repositories of the African Cluster Centres (ACCs) [7] and to other long-standing cooperation partners in African Studies at Bayreuth University, like the Federal University of Bahia in Brazil. It will not come as any surprise that in setting up such a repository, communication with the researchers to build mutual trust for their data delivery and much critical discussion about what is understood by "data" on the part of the Digital Solutions team and that of the involved disciplines was and is still under way, as this issue has been and is still being discussed in many publications, like the *Debates in Digital Humanities* series.[8] Especially at the time the conference was organized, these "*frontiers*", understood as delimitations and points of contact, were being negotiated for the envisioned digital research data environment in African Studies.

We therefore structured the conference in four thematic panels to be found here in the four thematic sections that all touch upon eSciences, or put more generally, the digital transformation of the academic world and society from very different perspectives and disciplinary diversity. Be it knowledge management as a big data business model for academic services (Stefan Ouma), digital neo-colonialism (Michael Kwet), the different legal aspects (Oliver Watteler, Markus Coester), problems of bias in semantic data processing (Gábor Bella et al.), digitalization projects

[4] See for example the available information on the "Kinship Network Questionnaire (KNQ) and associated systems": https://data.gesis.org/sharing/#!Detail/10.7802/83 (last accessed 11.12.2021).

[5] The DARIAH-DE platform as part of the DARIAH-EU projects makes the exception by offering different services and research tools for the arts and humanities, amongst others de DARIAH-DE Repository: https://repository.de.dariah.eu/colreg-ui/ (last accessed 9 May 2022).

[6] For an overview of the research projects and disciplines involved, see the descriptions on the Cluster's website: https://www.africamultiple.uni-bayreuth.de/en/Projects/index.html (last accessed 9 May 2022).

[7] The Cluster of Excellence Africa Multiple has four ACCs, Moi University, Eldoret (Kenya). Rhodes University, Makhanda (South Africa), Université Ki-Zerbo, Ouagadougou (Burkina Faso) and University of Lagos (Nigeria).

[8] For ongoing discussions and perspectives on the field of the Digital Humanities, see the hybrid book/online publication series "Debates in Digital Humanities" edited by Matthew K. Gold, first published in 2011: https://dhdebates.manifoldapp.org/ (last accessed 11.12.2021).

in Africa (Jody Butterworth) or digitization projects and collections built up in Europe (Wolfgang Kraus) – to name just a few aspects – the contributions will not only reach a readership already highly familiar with the field, but especially those who had only random contact with these topics hitherto. We hope this publication will be an incentive to dive deeper into the field and might even stimulate or open up new perspectives or ideas for one's own research – done more digitally.

Unfortunately, not all our speakers were able to contribute their talks to this publication: we deeply regret that Alex Makulilo, law specialist at the Open University of Tanzania and author of the book "African Data Privacy Laws" (Springer 2016), had to drop out of the publication due to his several other commitments. Equally our speaker Yunusa Zakari Ya'u from the Centre for Information Technology and Development (CITAD) in Nigeria, who held a talk on the gender digital divide concerning access to ICT technologies for women in Northern Nigeria (the topic is also briefly addressed in Mirjam de Bruijn's contribution), was prevented by his other several duties. Peter Ochieng, who presented a project about an ontology based knowledge system on poultry and crop diseases as an aid for farmers in Kenya, who can retrieve information through consultation via a mobile compatible chatbot, was kept from contributing due to his ongoing project obligations. Matthias Fuchs and Stefan Wolff as well as Øyvind Eide were understandably impatient with our publication speed but thankfully delivered extended abstracts in the section on "Capturing Semantics". Instead, we ourselves contributed to the publication – not having been that far at the time of the conference – by presenting the digital environment that is being developed with *WissKI*[9] for the research data processing and management of the projects funded in the Cluster of Excellence Africa Multiple (Jonas Huisl, Myriel Fichtner et al.).

Despite the delay in publication and the constantly changing and evolving field of research data management tools and practices and pressure groups that promote the field,[10] the editors believe that the topics and issues presented in these conference proceedings are as up to date as they were two years ago at their time of presentation. At the same time, we are aware that our conference in no way covers the field in its entire complexity and extent.

We thank our speakers and additional authors for their contributions to the volume and their patience with the publication process. We also deeply thank all our peer reviewers for their support and critical feedback. Finally, we are most grateful to the editorial team of the series for their time and accuracy with which this volume could be finalized. The conference was organized within the Africa Multiple Cluster of Excellence at the University of Bayreuth, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2052/1—390713894.

[9] https://wiss-ki.eu/ (last accessed 9 May 2022).

[10] e.g. the "Association German National Research Data Infrastructure" (NFDI) in Germany, first launched in 2016 by the German Council for Scientific Information Infrastructures (Rfii) and officially founded as the Association in October 2020, cf. https://www.nfdi.de/association/?lang=en (last accessed 9 May 2022); https://rfii.de/en/the-council/ (last accessed 9 May 2022).

Four Perspectives on Digital Transformation

Section 1

The Double Enclosure of Co-Produced Knowledge

Moving Towards Open-Access Scholarly Infrastructure and then All Good?

Stefan Ouma (University of Bayreuth)[11]

1 Introduction[12]

This paper engages with various types of knowledge infrastructures that have proliferated in the world of academic research and publishing over the past two decades, a process that has been thoroughly reshaped by the power of digitization. One way to go about the research and publication process is to conceive of it as a largely technical affair. Such a stance has been convincingly refuted by numerous works since at least the 1970s, which have exposed the social and political situatedness of research, writing and publishing. Not only the personal is political, but research is, too! These critical insights can be further extended to incorporate the very infrastructures of publishing, which, today are largely digital infrastructures. Often, these are part of the ever-growing 'ecosystems' of commercial publishing empires. Imperial analogies seem justified given the sheer size of large publishers and their value extraction strategies, which are fundamentally built on capitalizing on the benefits from free or low-paid labour (authors, reviewers, editors, technical support staff, often working from Global South countries) and from largely publicly funded-research (Monbiot 2018).

This contribution proceeds as follows. The first section (1) problematizes a purely technical and assumingly apolitical take on the production and 'dissemination' of academic knowledge. It shows

[11] Stefan.Ouma@uni-bayreuth.de, https://orcid.org/0000-0002-3180-1754.

[12] I thank Prof. Tanu Biswas (University of Stavanger) for earlier comments on this paper, as well as one anonymous reviewer. I am indebted to Julia Blauhut, who helped with the visuals. I also thank one anonymous reviewer for great feedback.

how the sociological concept of embeddedness can be utilized to engage with the multiple relationalities underpinning the 'traditional' production of research knowledge, which is already characterized by several moments of scholarly enclosure. In the next section (2), I show how since the 1990s, scholarly enclosures are increasingly being followed by a second round of enclosures through the corporately controlled process of 'knowledge dissemination'. This process has become governed largely by a few commercial publishers, which have also acquired controlling stakes in other nodes of the 'knowledge value chain', including its very infrastructure. I then (3) differentiate such enclosed knowledge infrastructures from open-access (OA) scholarly infrastructures, which are lately being promoted as an alternative to the corporate publishing model and as a more general way of how research data should be stored, made accessible, used and re-used. Despite the promises read into such infrastructures, there is also a range of risks attached to their uncritical promotion. I shall outline these risks by drawing on perspectives from critical data studies. I close (4) by finding that a research agenda poised to reconfigure 'African Studies', to which this intervention seeks to contribute in the spirit of the Cluster of Excellence "Africa Multiple" at the University of Bayreuth[13], must also reconfigure the way that contemporary knowledge infrastructures work, as well as what is being fed into them.

2 From disembedded to embedded knowledge production

Even though scores of papers and books have made the point that knowledge production is socially and politically situated (for a classic, see Haraway 1988), such views are still finding a hard time with many social and natural scientists who still think of research as an "an innocent pursuit of knowledge" (Ndlovu-Gatsheni 2017). Self-assuring statements on being objective, neutral, apolitical or deploying a merely positive approach to the world are indicative of this stance (Friedman 1953) and resurface from time to time (for a recent example, see Basedau 2020). Yet they are not as apolitical as they appear at first sight, because they are usually mobilized to deem those who are committed to a version of theory that challenges the status quo more radically as being unscientific, normative, or ideological. For these scholars, it is also quite easy to conceive of the knowledge production and dissemination process in purely technical terms (see Figure 1).

Those who claim that all research is socially embedded, including this author, posit that it is shaped by structural features, i.e. a historically grown coloniality of being, knowledge and power (Ndlovu-Gatsheni 2020; Kellecioglu 2020; Smith 2012) that shapes how scholarly knowledge is produced and 'distributed'. This can also be conceived as a specific kind of structural embeddedness, a term originally coined by economic sociologist Mark Granovetter (1985) and more recently reworked by (Hess 2004). This structural embeddedness also shines through in the *relational embeddedness* of knowledge production, which signifies the social networks into which knowledge production is embedded, variously comprising of funders, research participants,

[13] See https://www.africamultiple.uni-bayreuth.de/pool/dokumente/Africa-Multiple_Overview-1.pdf (last accessed 7 September 2021).

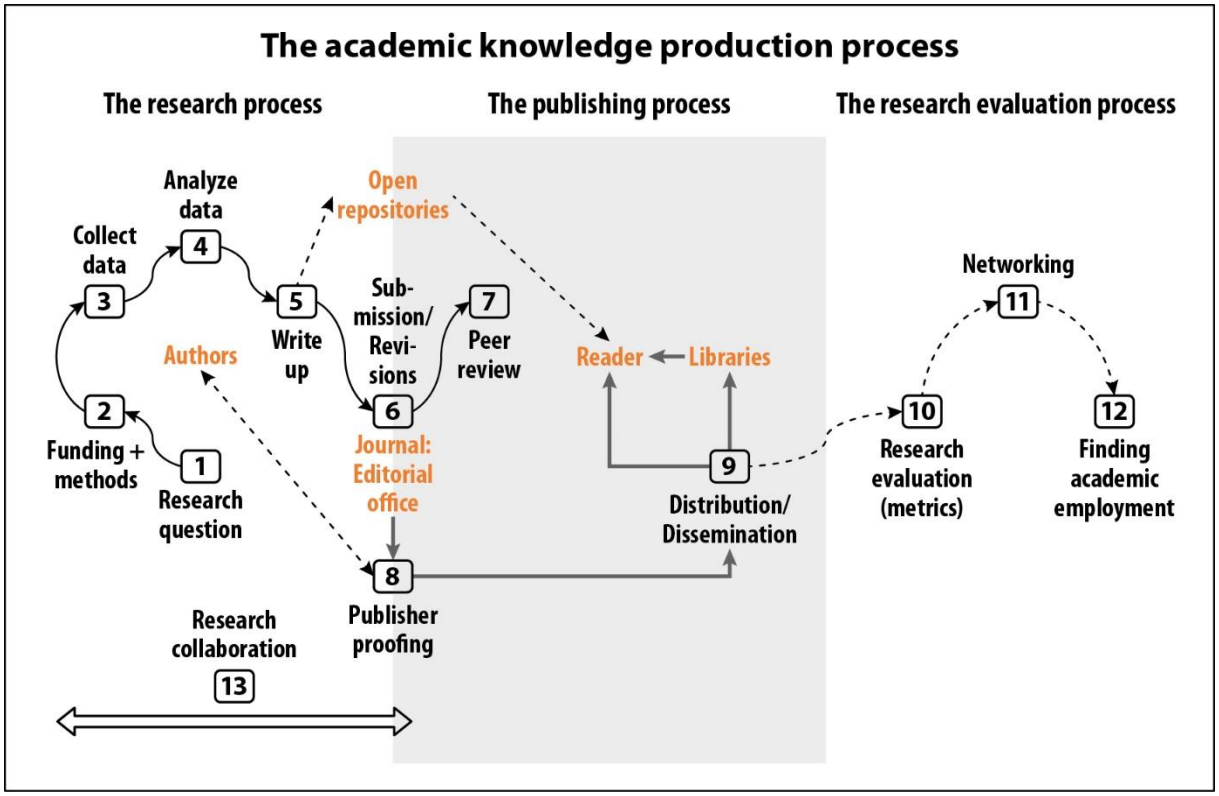


Figure 1: The academic knowledge production process and the traditional role of publishers. Source: Redrawn from Posada and Chen 2018.

donor- and recipient-researchers, research assistants, and publishers/editors of books and journals.

Contrary to what the term network suggests, the social structures of knowledge production in question are not heterarchical (Deane and Stevano 2016; Marchais et al. 2020). Among many others, Judith Buhenwa Nshobole (2020) emphasizes the often-asymmetrical relationship between either senior scholars (who often serve as principal investigators and money holders) and junior scholars (PhD students and postdocs) in the North and South alike and between Northern scholars and their Southern counterparts. In other words, knowledge co-production[14] is co-structured by the intersectional parameters of gender/sex, race, seniority, institutional affiliation, language, and material endowment, all of which have a history and geography to them.

Much of the unevenness in the global publishing landscape is a product of the interplay between the specific structural and relational embeddedness of knowledge production (see for example

[14] I am speaking of co-production here because the knowledge produced by academics could not have been produced without the help of others. As I have argued elsewhere, "[k]nowledge production must always be conceived of as situated – as being dependent on the position of the researcher in local/translocal, materially entangled, and power-laden social relations. What is often staged as an act of autonomous knowledge production by realists is *de facto* an intersectional, often asymmetrically structured co-generation of knowledge" (Ouma 2015: 84). Research participants, field assistants, co-researchers, and prior accumulated knowledge and experiences made by the researcher are firmly part of that co-production (Schwartz-Shea and Yanow 2012: 80).

Addis and Villa 2003 for economics; Bański and Ferenc 2013; Schurr et al. 2020; Jazeel 2016 for geography; Fox et al. 2016 for ecology and evolution). Lastly, *territorial embeddedness* shapes the production and dissemination of knowledge, and adds a further piece to the unevenness. Researchers need to stick to the laws of the country they are working in while carrying out research. Certain type of data may or may not be in the public domain, depending on the regulations of the country (property data would be one prominent example). Furthermore, much of the dissemination of knowledge via journals and books is governed under global copyright arrangements that favour corporate publishing houses and can quickly cause trouble for those who have moved against the prevailing intellectual property regime (for a recent example, see Banka 2020).

Once we have adopted an embeddedness view of the knowledge production and dissemination process, several issues come to the fore. First, this process happens in a field of forces, whereby issues of power, context, subjectivity/positionality, language, and practice are important entry points for a debate on how research works across *all* disciplines. Figure 2 untangles these entry points even though in practice, they cannot be separated (see Figure 2, e.g., all knowledge is contextual, produced via practices, shaped by power relations. language and positionality). On the 'input side', what knowledge can be produced and how it is being produced is further shaped by the policies and politics of funding. For an illustrative controversy on the latter issue, see the push of the later Trump administration to re-organize the funding going to Geography as a diverse discipline that covers both natural and social sciences into the funding docket "Human-Environment and Geographical Sciences Program (HEGS; previously known as Geography and Spatial Sciences) (Hamilton 2020). This move was deliberately designed to marginalize qualitative constructivist, political economy or otherwise critical geographical work. While neo-Positivism may be quickly blamed as the force at work here, it is clear that this was part of a larger attack on critical voices by the Trump administration, which also included attacks on otherwise largely neo-Positivist fields such as climate change science.

On the 'output side', the knowledge produced is packaged into publishable formats, most important of which are still journal articles and books (however, datasets and codes have become new important outputs). These are usually only accessible at a cost and are often hidden in the abysses of the publishing industry. This contrasts sharply with blog posts, open-eds or other forms of public engagement, which are meant to bring scholarly work in dialogue with society.

Second, both the production of small and big data, can be placed within this field of forces. While traditionally, humanities and social science research has been deemed to produce 'small data', recent technological advances across fields relevant for the humanities, social sciences and natural sciences increasingly produce, or at least make use, of 'big data'. The latter is often very large, covers entire populations, has a tight resolution and strong indexicality, offers strong insights into relationality (e.g., social network data), has a fast velocity, covers a wide variety of variables, and offers a high flexibility and scalability (Kitchin and Lauriault 2015). Big data includes data that researchers use, but was produced by others, such as the data that can be extracted from satellite images, or social network data (e.g., Tweets) or certain forms of government data. Big data can also become big when traditional data sets, including images or field notes, are digitized and become part of a larger platform that allows for new ways of

researching, and makes the content open "to the application of big data analytics such as data mining, pattern recognition, data visualization, statistics and modelling" (Kitchin and Lauriault 2019: 86).

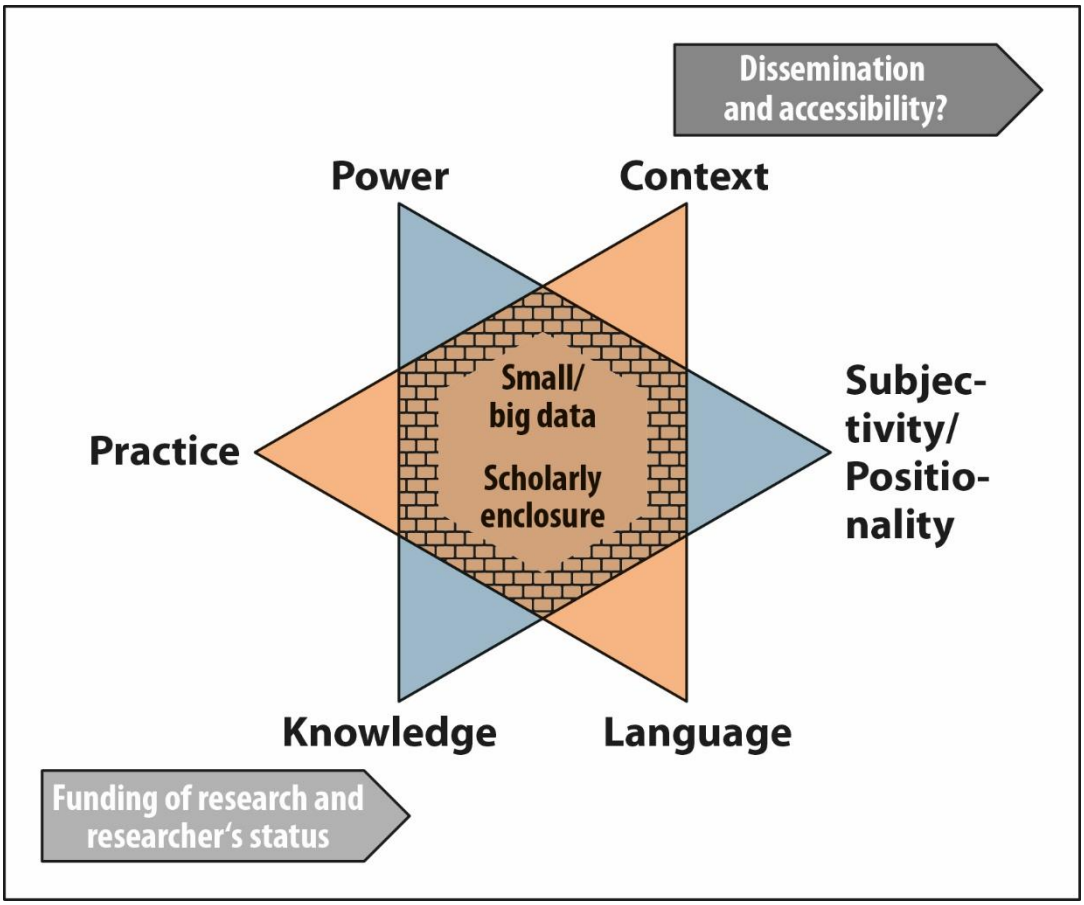


Figure 2: An embedded model of scholarly knowledge production and dissemination. Source: Author's design, with inspiration from Zienkowski 2017.

'Big data' contrasts sharply with traditional 'small data', which, besides some "exceptions, such as population censuses or meteorological data collection", were "typically characterized by limited volume, sampled collection, small geographic extent, and narrow variety and framing" (Kitchin and Lauriault 2019: 84).

Third, and central to this contribution, the interplay of forces in the field of knowledge production as we know it, leads to several forms of scholarly enframing:

- a theoretical enframing, whereby scholars develop a frame that allows them to organize the knowledge co-produced in a certain way while excluding other theoretical framings, including those that might be offered by the research participants themselves.

- a detachment of the knowledge produced from its larger context, which transforms something that was co-produced through interpersonal social relations into impersonal 'data';
- an exclusion of all those players, who may have played a crucial role during the research process, including other scholars, survey/interview respondents, research assistants and enumerators.

While we could be satisfied with describing these processes as acts of framing – "the tracing of a boundary between relationships and events which are internalized and included in a decision or, by contrast, externalized and excluded from it" (Callon 1998: 15), they can be likewise described as scholarly enclosures. The enclosed product of knowledge work can be easily commodified through market transactions precisely because of previous acts of enframing. Scholarly enclosures here refers to "efforts to secure and reinforce specific scholastic territories" (Biltoft 2019: 232). While deeply enmeshed with the logic of 'modern' knowledge production, such enclosures have further accelerated over the past 20 years or so due to the operational logic of the neoliberal academy ('publish or perish'; 'the rank and yank academy', Berg 2012):

> Academics these days are mainly trained to write for each other and not the general reader. It wasn't always like this but over the years, there has been a kind of scholarly enclosure, especially in the West […]. A kind of scholarly enclosure has advanced as academics are encouraged to address whatever conversation seems to be in vogue in a particular moment, and this is often the one that others can't understand, and all of this becomes further validated through the inwardly looking practices we perpetuate of recognition, citation and promotion. (Shah 2019)

What becomes clear here is that in relation to scholarly work, enclosure can be conceived as a product of specific knowledge practices as well as of commodification processes. Both are intertwined. While most researchers cannot directly transform their research outputs into ready cash (though some do, e.g. via patents or lucrative book contracts), scholarly enclosures still help them advancing their careers, while those who have crucially contributed to the underlying research are lucky if they are mentioned in the acknowledgements section of a paper, book or dataset. Any project poised to reconfigure African studies, or any other field of social and natural science inquiry, must firmly reflect on this first act of enclosure and overcome it in favour of a model of "socially just publishing" (Batterbury 2017). This call is pertinent because what is problematized here has been the classic mode of how research has operated over centuries, which was further entrenched by the publish-or-perish ethos of the neoliberal academy:

> Decolonising methodology must begin with unmasking the modern world system and the global order as the broader context from which re-search and methodology are cascading and are influenced. It also means acknowledging and recognising its dirtiness. Our present crisis is that we continued to use re-search methods that are not fundamentally different from before. (Ndlovu-Gatsheni 2017)

3 The corporate enclosure of research knowledge

Scholarly enclosures are a long-standing product of the academic knowledge co-production process, which too often masquerades as 'innocent research'. However, the emphasis on historical continuity can easily make us forget the novel elements that co-structure the process of academic knowledge production and circulation. After this first, 'traditional' round of academic enclosure, a more recent and more direct form sets in: the publication of academic work via the oligopolistic structures of the publishing world. Today a few "megapublishers" (Peekhaus 2012: 577) – a fairly recent development (Berg 2012; Buranyi 2017; Larivière et al. 2015), exert control over several nodes in the 'knowledge value chain' rather than merely having stakes in the dissemination of journal articles and books. In other words, rephrasing Brian Larkin, one of the most prominent critical thinkers on infrastructures, these publishers increasingly control the architecture for the circulation of scholarly knowledge, literally providing the infrastructure of the publishing world of the early 21st century, generating "the ambient environment of everyday [academic] life" (Larkin 2013: 328) – including for the making of academic careers (e.g., via metrics such as the *Impact Factor*).

Figure 3 highlights how the publication landscape has become increasingly concentrated since the mid-1990s, resulting in a shift from society- or university-owned and small-publishing houses (all deemed "small publishers" in Figure 3) to "big publishers". This shift has been most striking for the natural and medical sciences, which promise larger turnover volumes due to their sheer size.

The result of this secular development has been that by 2015, the five large publishers (RELX Group including Elsevier, Wiley-Blackwell, Springer, Taylor & Francis) published around 70% of social science articles and 20% in the humanities (Larivière et al. 2015). Their oligopolistic market power has manifested itself "in skyrocketing journal subscription costs and a drastic lockdown of content through strict application of copyright and licensing restrictions" (Peekhaus 2012: 577). Berg (2012) highlights the consequences of the pay- and copyright wall for both relatively well-endowed institutions in the Global North and their less endowed counterparts in the Global South:

> This profit-taking has significant impacts on the relatively well-to-do academic libraries of major research universities in the global North, many of which have had to reduce their serials subscriptions in order to meet their budgets […]. At the same time, this profit-taking also operates to prevent scientific knowledge circulation, especially in the Global South […]. (Berg 2012: 261)

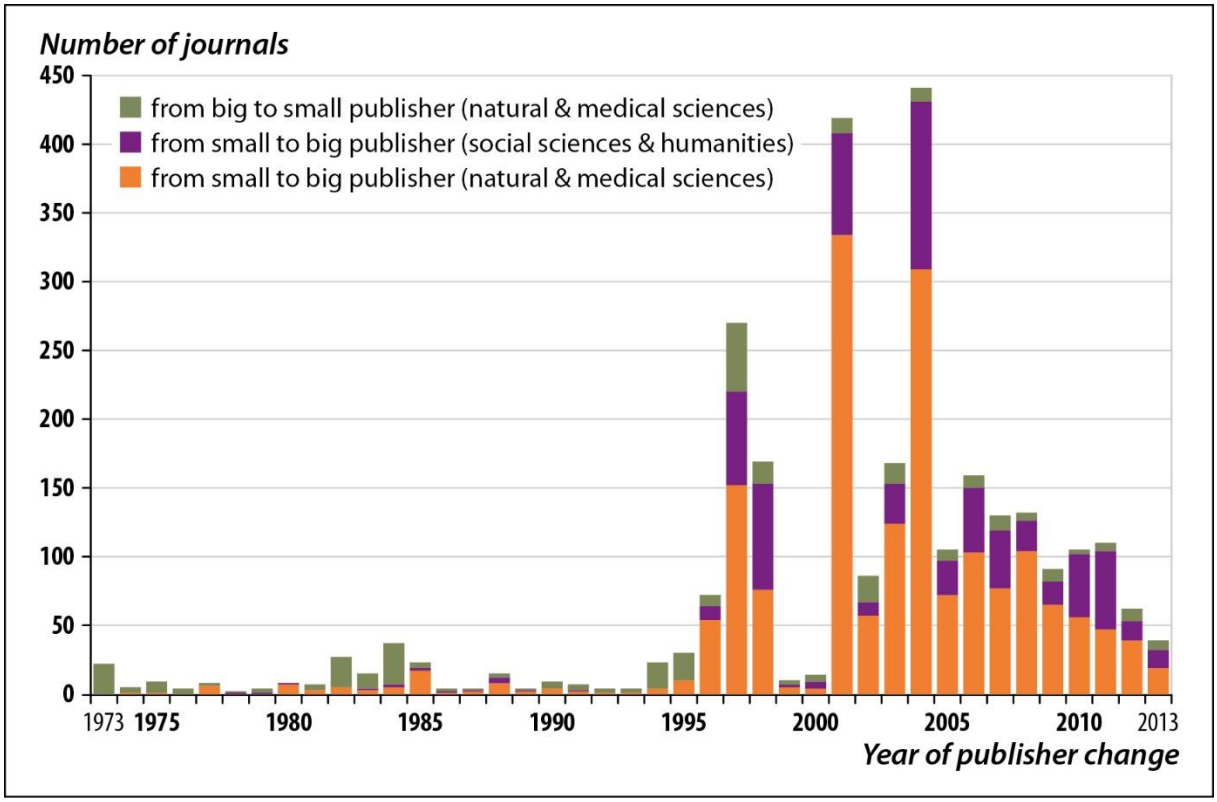


Figure 3: Number of journals changing from small to big publishers, and big to small publishers per year of change in the Natural and Medical Sciences and Social Sciences & Humanities. Source: Redrawn from Larivière et al. 2015. [15]

Oligopolistic competition in publishing has resulted in staggering operating profits and profit margins for the mega-publishers. Larivière et al. (2015) document this for the case of the publishing giant Elsevier (see Figure 4), but older data on other large publishers suggests similar levels of profitability (Fuchs and Sandoval 2013).

[15] The category "from small to big publisher" includes journals that moved between small publishers (i.e. any publisher but the five) and big publishers (i.e. those top five that are presented in Figure 3). Personal information provided by Vincent Larivière, 09 July2021.

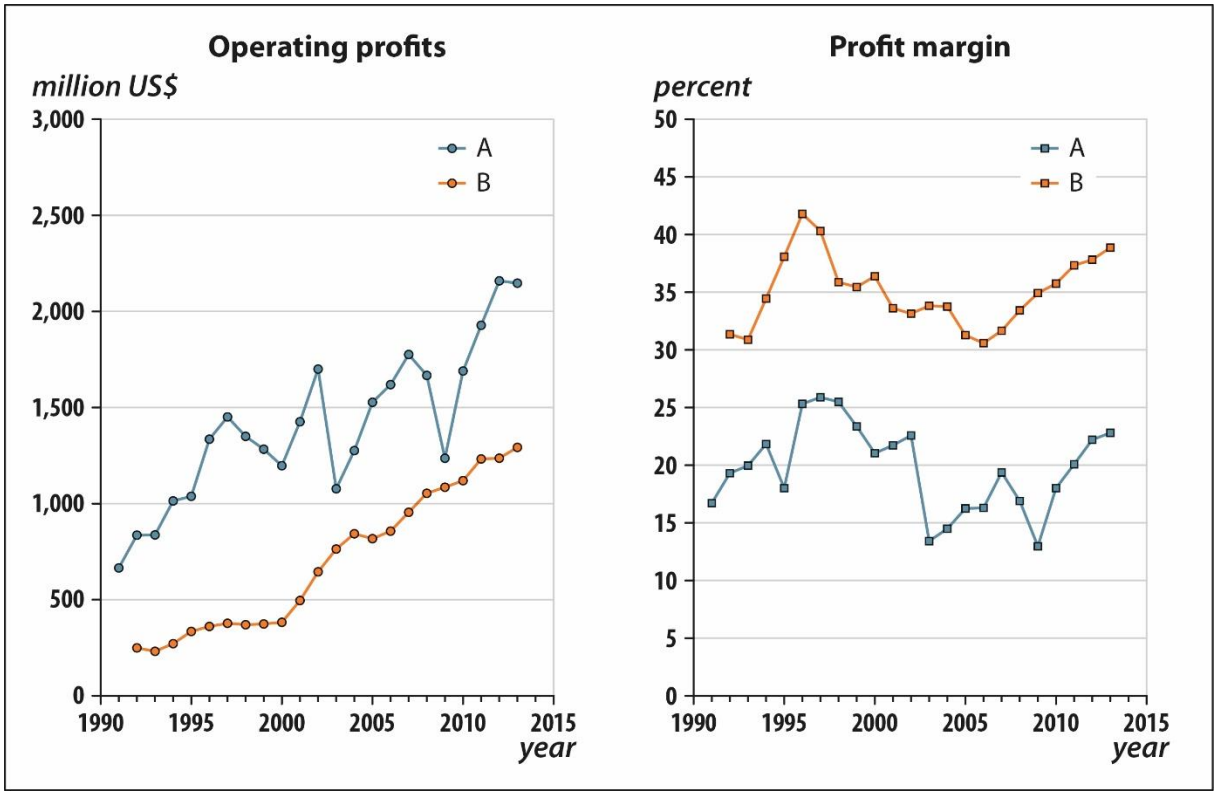


Figure 4: Operating profits (million USD) and profit margin of Reed-Elsevier as a whole (A) and of its Scientific, Technical & Medical division (B), 1991–2013. Source: Redrawn from Larivière et al. 2015.

As indicated above, publishers also increasingly try to acquire control over the very infrastructure of academic knowledge production, circulation and evaluation, which, again, can be well documented for the case of Elsevier (see Figure 5). Thus, they increasingly pose as digital landlords (Sadowski 2020) or 'knowledge rentiers', who generate rents by owning the very infrastructure that allows researchers to collaborate, offer their scholarly 'crops', or lets users purchase them and evaluate their quality (Posada and Chen 2018). Rentiers, here, are defined in line with Sadowski (2020: 565) drawing on Felli, "by their 'ownership of the access to a condition of production' (Felli 2014: 269) and their ability to derive income (rent) from access to assets." Because in general, access to the sources of rent is either limited by nature (e.g. in the case of land) and/or by state regulation (e.g. in the case of intellectual property), those who have property rights over the sources are usually able to generate at least a part of their income from non-productive activities. Income without corresponding investment often leads to above-market rate of returns. Since publishers do usually process the 'inputs' (= research) paid for by the state (Monbiot 2018), and can rely on the largely free labour of editors, editorial board members, authors and reviewers, they are able to extract significant amounts of surplus value (Berg 2012). Historically, infrastructure has played a central role in extending imperial value relations (Cowen 2018), and the quest of mega-publishers to accumulate ever more infrastructural power, must be read in this light. Infrastructure is things, but "also the relation between things" (Larkin 2013: 329).

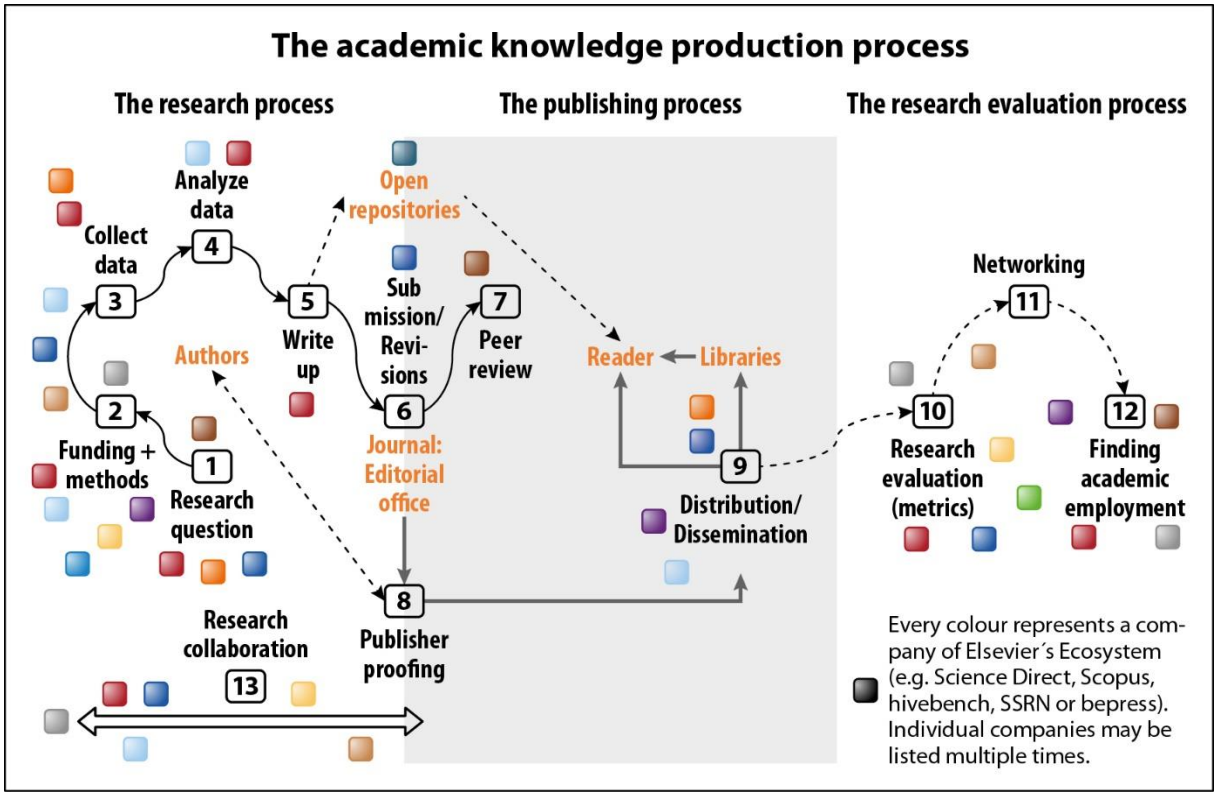


Figure 5: The publishing ecosystem of Elsevier. Source: Redrawn from Posada and Chen (2018).

4 The Promises and Perils of OA Data Infrastructures

As Batterbury (2017: 175) argues, for "some, it is perfectly acceptable to cede author copyright to companies that prepare and sell them, thereby losing ownership and management of that intellectual property." For him and others, including this author, research funded with public money should be a public good. It should be added that both social justice and ecological reasons, generous intellectual property protections must also be dismantled in favour of true OA regimes of knowledge (Fuchs and Sandoval 2013; Raworth 2017).

As part of an "academic spring", governments, research councils, librarians, university leaders, academics and journalists have come out to criticize the corporate publishing model and called for its overhaul. Corporate publishers have reacted by increasingly offering OA models. These were able, often supported by governments, to push for a 'gold OA model' (Fuchs and Sandoval 2013) which requires authors to pay still rather high article processing charges (APC), while making content available to readers immediately and at no cost. As Vega et al. (2021) show, this model is not the dominant one in the global publishing landscape, even if publishers are quick to invoke this impression (Fuchs and Sandoval 2013)[16]. Paper repositories (so-called 'green OA') hosted by universities or other ways to make research findings accessible (e.g., via blogs or

[16] In Vega et al.'s data base, 72% percent of all journals listed do not attract APC.

platforms such as *THE CONVERSATION*) are supplementing the paper-focused OA drive. Ironically, a prominent definition of OA as "digital, online, free of charge, and free of most copyright and licensing restrictions" (Suber 2013, in Kitchin et al. 2015: 665), seeking to demolish both "price barriers (subscriptions, licensing fees, pay-per-view fees) and permission barriers (most copyright and licensing restrictions)" (ibid.), obscures the significant differences between for-profit and not-for-profit OA models (which are non-corporate) (Fuchs and Sandoval 2013).

There are a variety of OA models[17], including assumingly credible publishers such as MDPI that have become more predatory in their business practices (Crosetto 2021) or OA models that neither charge authors, nor readers, nor institutions and have maintained a high level of academic integrity at the same time (e.g., the *Journal of Political Ecology*). Many, however, see the future of producing and disseminating research in OA digital data repositories (Kitchen et al. 2015), a.k.a "Open Scholarly Infrastructures" (Bilder et al. 2015). There are significant differences in how OA data repositories can be organized and funded, but as a rule of thumb these can be largely considered as OA infrastructures meant to reshape the way knowledge is produced, used and reused, with proponents making both scientific and financial arguments for why such infrastructures should be the future of research (Kitchin et al. 2015; see : Benefits of data repositories/infrastructures, source: Kitchin et al. 2015: 666.Table 1).

Table 1: Benefits of data repositories/infrastructures, source: Kitchin et al. 2015: 666.

Direct benefits	*Indirect benefits (costs avoided)*
New research opportunities Scholarly communication/access to data Re-purposing and re-use of data Increasing research productivity Stimulating new networks/collaborations Data available for teaching and student projects Knowledge transfer to industry Improves skills base Increasing productivity/economic growth Verification of research/research integrity Fulfilling mandate(s**)**	No re-creation/duplication of data No loss of future research opportunities Lower future preservation costs Re-purposing data for new audiences Re-purposing methodologies Use by new audiences Protecting return on earlier investment Tools and standards have potential to increase data quality Reduces ad hoc queries concerning data
Short-term benefits	*Long-term benefits*
Value to current researcher and students No data lost from researcher turnover Widens access where costs are prohibitive for researchers/institutions	Secures value to future researchers and students Adds value over time as collection grows and develops critical mass

[17] For current trends data on access journals, see Vega et al. (2020) and European Commission (2021a).

Short-term re-use of well-curated data Secure storage for data-intensive research Availability of data underpinning publications	Increases speed of research and time to realise impacts Stimulates new research questions, especially relating to linked and derived data
Private benefits	*Public benefits*
Benefits to sponsors/funders of research/archive Benefits to researchers and institutions Fulfils grant obligations Increased visibility/citations Commercialising research	Input for future research Motivating new research Catalysing new companies and high-skills employment Transparency in research funding

Over the past years, a number of these infrastructures have been launched, benefiting from "new archiving technologies, a massive expansion in digital data storage, data standards, and open science and open data movements" (Kitchin and Lauriault 2019: 86). Usually, these infrastructures are operated at national or regional scales. For instance, the European Union published its first position papers on open science and data in 2012 (Kitchin et al. 2015). Since then, it has developed an Open Science Policy, centring around eight ambitions. Among others, this includes calls for Open Data, setting up the European Open Science Cloud (EOSC), developing a new generation of metrics to measure impact, and strengthening of Citizens Science (European Commission 2021b).[18] However, it is more appropriate to call such infrastructures 'living digital archives' in order to underline that these not only combine heterogeneous, searchable data related to various dimensions of the knowledge co-production process, but also that they need initial *and* ongoing social and material investments, including *curation*:

> Archives are formal collections of data that are actively structured, curated and documented, are accompanied by appropriate metadata, and where preservation, access and discoverability are integrated into technological systems and institutions designed to last the test of time (Lauriault et al. 2013). Archives explicitly seek to be long term endeavours, preserving the full record set—data, metadata and associated documentation—for future reuse. (Kitchin and Lauriault 2015: 467)

For sure, big data and data infrastructures in the form of repositories or even archives offer unseen and welcome possibilities, especially if curated as 'open access' This notwithstanding, setting up such archives is fraught with all sorts of financial, managerial and technical challenges (for a good overview, see Kitchin et al. 2015). Another way to put it is that there are challenges

[18] I cannot comment in detail on this initiative from the perspective of 'socially just publishing'. Suffice it to say, the European Commission has largely failed to offer a more radical and reflexive imagination on how the future of knowledge production and circulation could look like due ongoing corporate influence (Elsevier has been a major subcontractor of the project).

related to "running the infrastructure (governance), funding it (sustainability), […] preserving community ownership of it (insurance)", and implementing it (Bilder et al. 2015).

Moreover, there is the risk of eclipsing the knowledge politics underpinning the call for digital data repositories. One of those risks is to subject small data to a big data epistemology, as for instance the European Commission does, when it argues that open data facilitates the replication of research, which is usually not a quality criteria for most qualitative research. Such a stance clearly risks that, "knowledges that are not so easily encapsulated within big data frameworks might become devalued" (Graham and Shelton 2013: 257). Other associated risks include:

- the concentration of power in the hands of new gate keepers ('data scientists') (Donati and Woolston 2017).
- decontextualization, that is when one piece of data becomes just one element in a giant data heap.
- that sound and engaged secondary research gives way to quick data mining.
- that people assume that large (quantitative) data is the "end of theory" (Anderson 2008).

This is why data *infrastructures*, as aggregators of data through which this data becomes big (of course, it can also further aggregate data that already was big), must be subjected to the same critical gaze as 'born-big data' itself. There are promises and perils of both big data and big data infrastructures (Graham and Shelton 2013), regardless of whether they are OA or not. For critical data scholars such as Dalton and Thatcher (2014: n.p.), among others this includes accepting that big is not everything, that big data is never raw or technologically innocent, as well as the need to reflect on the crucial questions such as "who controls 'big data,' its production and its analysis? What motives and imperatives drive their work?" and who the subjects of 'big data' are and what knowledges are they producing?

On top, even if one has solved the technical, managerial and financial challenges, and successfully tackled the epistemological risks just outlined, it does not mean that an archive magically emerges. As Stuart Hall noted in his classic definition of the knowledge archive,

> Constituting an archive represents a significant moment, on which we need to reflect with care. It occurs at that moment when a relatively random collection of works, whose movement appears simply to be propelled from one creative production to the next, is at the point of becoming something more ordered and considered: an object of reflection and debate. (Hall 2001: 89)[19]

While one can barely expect Hall's idea of the archive to materialize in national-scale initiatives, this is different for more institution-based archives, where it is possible to craft the intellectual

[19] This is why the distinction between the "anarchic totality of the repository" and "the curated and ordered logic of the analogue archive" seems useful. I thank the anonymous reviewer for this point. Of course, the curation of the analogue archive can also be translated into the digital archive.

philosophy of an archive in Hall's sense into its technicalities - if only those involved are willing do so. Kitchin and Lauriault (2015) are right when they argue that realizing the idea of comprehensive digital data repositories would need a cultural change among researchers, but so would overcoming of the classic scholarly enclosures highlighted above. A common, more radical philosophy underpinning a given digital archive should also further be extended to an engagement with the "deep coloniality" (Kellecioglu 2020) that shapes why and how research is conducted in the very first place, and document the steps that were taken to produce and circulate knowledge in a substantially different way. Suffice it to say, one can have quite a radical take on the archive, without undoing the "dirty history" of research (Ndlovu-Gatsheni 2017; Smith 2012). In fact, this risk should be of significant concern when revisiting the history of 'techno-solutionism'. The language of a technological fix used to change undesirable states of the world has lent itself too often too well to only a superficial critique of what is to be changed. One also should be wary of the fact that even projects that have aimed at explicitly undoing research and its dirty history have struggled to do so in practice (Marchais et al. 2020).

Another risk with the promotion of digital data infrastructures is that nothing speaks against corporate publishers further appropriating the idea (Bilder et al. 2015), especially when considering the imperial qualities of 'infrastructure' (Cowen 2018). In fact, they have already done so, as I outlined above, and as platforms such as *Sage Data Planet*, the *Sage Methods Platform* or *ResearchGat*e and *Academia.edu* show. Years of infrastructural thinking on the side of major publishers, reinforced by the platform gaze of the tech economy, make it relatively easy to turn accumulated published knowledge and associated services into a platform product. The fact that large publishers managed to appropriate OA via APC schemes, largely with the approval of academics and states, serves as a warning here.

Reconfiguring African Studies requires a proper engagement with the politics and political economy of a shifting publishing landscape. The framework text that outlines the research programme of the Cluster of Excellence "Africa Multiple" hints at this, and the plans to create a digital research environment that makes the research produced more accessible can be read as a response to the problem of double enclosures in the world of digitally mediated corporate publishing.[20] This contribution serves a reminder that more work needs to be done in order to come to terms with the question of how knowledge is co-produced, and how that knowledge is stored and made accessible to the broader public in a socially just manner.

5 References

Addis, Elisabetta, and Paola Villa. 2003. The Editorial Boards of Italian Economics Journals: Women, Gender, and Social Networking. *Feminist Economics* 9 (1): 75–91. https://doi.org/10.1080/1354570032000057062.

[20] See https://www.africamultiple.uni-bayreuth.de/en/Digital-Solutions/index.html (7 July2021).

Anderson, Chris. 2008. The editorial boards of Italian economics journals: Women, gender, and social networking. *Wired Magazine*. https://www.uvm.edu/pdodds/files/papers/others/2008/anderson2008a.pdf (last accessed: 25 August 2021).

Banka, Richa. 2020. Delhi HC asks Libgen, Sci-Hub to stop uploading articles as they face copyright infringement charges. *Hindustan Times,* December 24. https://www.hindustantimes.com/india-news/delhi-hc-asks-libgen-sci-hub-to-stop-uploading-articles-as-they-face-copyright-infringement-charges/story-cRWCB1sGs1yMqR3TCpuvmL.html (last accessed: 25 August 2021).

Bański, Jerzy, and Mariola Ferenc. 2013. "International" or "Anglo-American" journals of geography? *Geoforum* 45:285–295. https://doi.org/10.1016/j.geoforum.2012.11.016.

Basedau, Matthias. 2020. Rethinking African Studies: Four Challenges and the Case for Comparative African Studies. *Africa Spectrum* 55 (2): 194–206. https://doi.org/10.1177/0002039720945328.

Batterbury, Simon. 2017. Socially just publishing: implications for geographers and their journals. *Fennia - International Journal of Geography* 195 (2): 175–181. https://doi.org/10.11143/fennia.66910.

Berg, Lawrence D. (2012). Knowledge enclosure, accumulation by dispossession, and the academic publishing industry. *Political Geography* 31 (5): 260–262. https://doi.org/10.1016/j.polgeo.2012.05.011.

Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. 2015. *Principles for Open Scholarly Infrastructures-v1*. https://cameronneylon.net/blog/principles-for-open-scholarly-infrastructures/ (last accessed: 25 August 2021).

Biltoft, Carolyn N. 2019. Against Scholarly Enclosures: Reconsidering the Art and Economics of Review. *Capitalism: A Journal of History and Economics* 1 (1): 231–240. https://doi.org/10.1353/cap.2019.0000.

Buranyi, Stephen. 2017. Is the staggeringly profitable business of scientific publishing bad for science?, June 27. https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science (last accessed: 25 August 2021).

Callon, Michel. 1998. Introduction: The Embeddedness of Economic Markets in Economics. *The Sociological Review* 46 (1_suppl): 1–57. https://doi.org/10.1111/j.1467-954X.1998.tb03468.x.

Cowen, Deborah. 2018. The jurisdiction of infrastructure: Circulation and Canadian settler colonialism. *The Funambulist* (17): 14–19.

Crosetto, Paolo. 2021. Is MDPI a predatory publisher? https://paolocrosetto.wordpress.com/2021/04/12/is-mdpi-a-predatory-publisher/ (last accessed 25 August 2021).

Dalton, Craig, and Jim Thatcher. 2014. What does a critical data studies like, and why do we care? *Society+Space*. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care (last accessed 25 August 2021).

Deane, Kevin, and Sara Stevano. 2016. Towards a political economy of the use of research assistants: reflections from fieldwork in Tanzania and Mozambique: Reflections from fieldwork in Tanzania and Mozambique. *Qualitative Research* 16 (2): 213–228. https://doi.org/10.1177/1468794115578776.

Donati, Gaia, and Chris Woolston. 2017. Data domination. *Nature* (548): 612–614.

European Commission. 2021a. Trends for open access to publications. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en (last accessed 25 August 2021).

European Commission. 2020b. Open Science. https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en (last accessed 25 August 2021).

Felli, Romain. 2014: On Climate Rent. In: *Historical Materialism* 22 (3-4): 251–280. https://doi.org/10.1163/1569206X-12341368.

Fox, Charles W., C. Sean Burns, and Jennifer A. Meyer. 2016. Editor and reviewer gender influence the peer review process but not peer review outcomes at an ecology journal. *Functional Ecology* 30 (1): 140–153. https://doi.org/10.1111/1365-2435.12529.

Friedman, Milton. 1953. *Essays in positive economics.* Chicago: Univ. of Chicago Press.

Fuchs, Christian, and Marisol Sandoval. 2013. The Diamond Model of Open Access Publishing: Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-Commercial, Non-Profit Open Access Serious. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 11 (2): 428–443. https://doi.org/10.31269/triplec.v11i2.502.

Graham, Mark, and Taylor Shelton. 2013. Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography* 3 (3): 255–261. https://doi.org/10.1177/2043820613513121.

Granovetter, Marc. 1985. Economic action and social structure: The problem of embeddedness. *The American Journal of Sociology* 91 (3): 481–510.

Hall, Stuart. 2001. Constituting an archive. *Third Text* 15 (54): 89–92. https://doi.org/10.1080/09528820108576903.

Hamilton, Aretina R. 2020. The white unseen: On white supremacy and dangerous entanglements in geography. *Dialogues in Human Geography* 10 (3): 299–303. https://doi.org/10.1177%2F2043820620966489.

Haraway, Donna. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14 (3): 575. https://doi.org/10.2307/3178066.

Hess, Martin. 2004. 'Spatial' relationships? Towards a reconceptualization of embedded ness. *Progress in Human Geography* 28 (2): 165–186. https://doi.org/10.1191/0309132504ph479oa.

Jazeel, Tariq. 2016. Between area and discipline. *Progress in Human Geography* 40 (5): 649–667. https://doi.org/10.1177/0309132515609713.

Kellecioglu, Deniz. 2020. The deep coloniality of economics – a (r)evolutionary appraisal. *Review of African Political Economy* 47 (165): 484–493. https://doi.org/10.1080/03056244.2020.1763288.

Kitchin, Rob, Sandra Collins, and Dermot Frost. 2015. Funding models for Open Access digital data repositories. *Online Information Review* 39 (5): 664–681. https://doi.org/10.1108/OIR-01-2015-0031.

Kitchin, Rob, and Tracey P. Lauriault. 2019. Data and data infrastructures. In *Digital Geographies*, ed. James Ash, Rob Kitchin and Agnieszka Leszczynski, 83–94. Los Angeles: Sage.

Kitchin, Rob, and Tracey P. Lauriault. 2015. Small data in the era of big data. *GeoJournal* 80 (4): 463–475. https://doi.org/10.1007/s10708-014-9601-7.

Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. 2015. The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE* 10 (6): e0127502. https://doi.org/10.1371/journal.pone.0127502.

Larkin, Brian. 2013. The Politics and Poetics of Infrastructure. *Annual Review of Anthropology* 42 (1): 327–343. https://doi.org/10.1146/annurev-anthro-092412-155522.

Marchais, Gauthier, Paulin Bazuzi, and Aimable Amani Lameke. 2020. 'The data is gold, and we are the gold-diggers': whiteness, race and contemporary academic research in eastern DRC. *Critical African Studies* 12 (3): 372–394. https://doi.org/10.1080/21681392.2020.1724806.

Monbiot, George. 2018. Scientific publishing is a rip-off: We fund the research – it should be free. *The Guardian,* September 13. https://www.theguardian.com/commentisfree/2018/sep/13/scientific-publishing-rip-off-taxpayers-fund-research (last accessed 25 August 2021).

Ndlovu-Gatsheni, Sabelo J. 2020). *Decolonization, development and knowledge in Africa*. London: Routledge.

Ndlovu-Gatsheni, Sabelo J. 2017. Decolonising research methodology must include undoing its dirty history. https://theconversation.com/decolonising-research-methodology-must-include-undoing-its-dirty-history-83912 (last accessed 25 August 2021).

Nshobole, Judith B. 2020. 'Donor-researchers' and 'recipient-researchers': Bridging the gap between researchers from the Global North and Global South. https://www.gicnetwork.be/donor-researchers-and-recipient-researchers-bridging-the-gap-between-researchers-from-the-global-north-and-global-south/ (last accessed 25 August 2021).

Ouma, Stefan. 2015. *Assembling export markets: The making and unmaking of global food connections in West Africa.* Chichester: Wiley-Blackwell.

Peekhaus, Wilhelm. 2012. The Enclosure and Alienation of Academic Publishing: Lessons for the Professoriate. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 10 (2): 577–599. https://doi.org/10.31269/triplec.v10i2.395.

Posada, Alejandro, and George Chen. 2018. *Inequality in knowledge production: The integration of academic infrastructure by big publishers.* 22nd International Conference on Electronic Publishing: OpenEdition Press.

Raworth, Kate. 2017. *Doughnut economics: Seven ways to think like a 21st-century economist.* White River Junction, Vermont: Chelsea Green Publishing.

Sadowski, Jathan. 2020. The Internet of Landlords: Digital Platforms and New Mechanisms of Rentier Capitalism. *Antipode* 52 (2): 562–580. https://doi.org/10.1111/anti.12595.

Schurr, Carolin, Martin Müller, and Nadja Imhof. 2020. Who Makes Geographical Knowledge? The Gender of Geography's Gatekeepers. *The Professional Geographer* 72 (3): 317–331. https://doi.org/10.1080/00330124.2020.1744169.

Schwartz-Shea, Peregrine, and Dvora Yanow. 2012. *Interpretive research design: Concepts and processes.* New York: Routledge.

Shah, Alpa. 2019. Good research is vital. *The Patriot,* November 15. https://thepatriot.in/2019/11/15/good-research-is-vital/ (last accessed 25 August 2021).

Smith, Linda Tuhiwai. 2012. *Decolonizing Methodologies: Research and indigenous peoples*, 2nd edn. London: ZED Books.

Vega, Alejandra Manco; Shen, Cenyu; Armengou, Clara; Mitchell, Dominic et al. 2021. The directory of open access journals. https://www.doaj.org/ (last accessed 25 August 2021).

Zienkowski, Jan. 2017. Reflexivity in the transdisciplinary field of critical discourse studies. *Palgrave Communications* 3 (1). https://doi.org/10.1057/palcomms.2017.7.

Digitalisation of the Academe in Africa

Interrogating the Role of Technology in Reconfiguring African Studies

'Tunde Ope-Davies (University of Lagos)[21]

1 Introduction: The Reality of the African Digital Experience

It is a truism that the rapid emergence of new technologies, as well as their proliferation and deployment for social and intellectual activities, continues to impact our thoughts, conversations, interactions and actions in contemporary society. Popularly described variously as 'the Age of Technology', 'the Era of Digital Revolution', 'the 4th Industrial Revolution', 'the Age of Big Data', 'the Age of Smart Technology', 'the Age of Information', 'the Dispensation of Artificial Intelligence', 'the Internet of Things' and as 'Cloud Computing', among others, today's world is witnessing a truly profound and defining moment in human history.

Whichever description is preferred, it is important to note the overarching idea of the profound nature and influence of digital technologies in sociocultural conversations, as well as in socioeconomic cum political engagements and scholarly endeavours. Mapping the role of digital technologies in human socio-cyber space reveals that democratisation of ideas, enlarged participatory and interactive spaces, global networking, increased accumulation of data, technology-driven analytical techniques and greater accessibility to data – as well as their

[21] Contact: bopeidi@unilag.edu.ng, https://orcid.org/0000-0002-1772-6356.

usability – form the cornerstones of submissions in technology-driven scholarship. Digitalisation of society, including the academe, offers a new and creative perspective on knowledge creation, preservation and dissemination. In fact, cyber technology may be described as one of the most crucial achievements of the present generation owing to its capacity to help in recovering and utilising endangered or lost epistemic bases.

As Ahnert (2020) suggests, the more the people who have access to digitised and digitalised data, "the more likely we are to be able to recover lost voices and rebalance scholarship in favour of the marginalized" (https://www.cdh.cam.ac.uk/ruth-ahnert-digital-landscapes). This submission explains one of the priorities in the reconfigured, technology-driven African Studies. To be sure African histories, civilisations, epistemes, cultural heritage, oral narratives and performances have for long been under-studied and under-presented in global scholarly communities. It may be argued that technology (e.g. social media platforms) is one way of changing the narratives about Africa and her people through African Studies by giving this discipline a new life and a new voice (Mkono 2018, n.p.). Citing an instance where technology can play that critical role, Mkono notes as follows:

> Social media presents a powerful platform for creating multiple stories about Africa. Embracing the accessibility of modern technology, African social media bloggers and commentators are using Facebook, YouTube and other platforms to undermine longstanding "Afro-pessimistic" stereotypes – the backward continent characterised by strife and poverty. By giving ordinary people the space to share their "everyday" experiences, African bloggers are capturing positive "human moments". These moments of joy, play, celebration, love and human interaction, create new narratives of Africa. (https://theconversation.com/changing-the-african-narrative-through-social-media-platforms-97097)

Such new media technologies and online platforms have the capacity to globalise natural and authentic data emanating from Africa, thereby narrating and presenting the world with authentic stories about Africa. The impact of these developments in other parts of the world has continued to inspire measured efforts to bridge the gap in Africa. In fact, Africa is now consistently reported to be one of the continents with the fastest growing rates of Internet penetration and social media population, with the potential to provide greater opportunities for digitalisation in the nearest future. For instance, works by Schelenz and Schopp (2018) and Mohammed (2013) suggest that digital technologies are 'transforming' Africa and impacting growth and development in sectors such as politics, commerce, education and healthcare delivery. The rapidity with which Africa's digital landscape is expanding thus provides a significant leverage for African scholarship and research initiatives.

It is against this backdrop that this chapter provides a general overview of the digitalisation of academia in Africa, with particular reference to the prospects, problems and possibilities that this revolution offers to the research and policy communities. For instance, the digital revolution is inspiring new initiatives and ideas that are being developed and deployed to impact positive socio-economic transformations in sectors such as banking and commerce as well as telecommunications and democratic practice. This study takes up some theoretical and practical

issues and draws insights from Digital Humanities (DH) and Digital Scholarship initiatives in Africa.

The study seeks to answer the following questions: (i) In what ways are new technologies transforming scholarship and research or advancing learning in African Studies? (ii) How can scholars working within this space utilise new technologies to enhance their works? (iii) What role(s) can Digital Humanities play in enhancing a digital research environment towards greater collaboration within and outside academia, with particular reference to the works at the Centre for Digital Humanities, University of Lagos?

The chapter thus argues that new technologies offer huge resources that will enable humanities scholars to redefine and reconfigure their research orientations, activities and outputs. With the availability of tools and techniques in DH and the creation of a viable Digital Research Environment (DRE) and the researcher's experience at the University of Lagos, it is proposed that scholars and researchers in African Studies must take advantage of this growing digital revolution to reconfigure their theoretical and methodological approaches. They can also re-engineer their research efforts and outcomes in order to expand the frontiers of knowledge in research and innovation for more global impact and increased public good.

2 Theoretical Issues: Digitalisation of the Academe

In the last decade, the African academic ecosystems have been witnessing significant reconfigurations. While efforts are being made towards achieving the goals of full digitalisation of academia, one of the visible marks of digitalisation is the setting up of institutional repositories for the research activities of African scholars in the bid to preserve and globalise their contributions. Apart from the gradual automation of some administrative and academic structures such as admission processes, course registration, as well as examinations and results processing, most institutions have also been developing and deploying digital solutions that will help to fully digitalise research activities and research outcomes. Indeed, concerted efforts are being made to take advantage of the digital revolution in the effort to overhaul academia.

2.1 Digitalisation vs Digitisation

Given that they demonstrate how technologies are being applied to transform the works of academics and researchers beyond the general application of digital tools to transform society, digitalisation and digitisation are interrelated concepts, hence the confusion of the terms in certain quarters.

Gbadegeshin summarises the views of scholars on these concepts thus:

> *Digitalization* is an organization of several and diverse social life spheres via digital communication technologies, whereas digitization is a conversion of analogue information into digital forms (…). digitalization is a structuring of those technologies across: infrastructural, terminal, functional and rhetorical, and market convergence dimensions (Brennen & Kreiss, 2016), process, organization, business domain (Parviain-en et al., 2017), industries (Tihinen & Kääriäinen, 2016), and the entire

> economy and society levels (Degryse, 2016; Parviainen et al., 2017). Gbadegeshin (2019: 50)

In the context of this study, digitisation is seen as the use of technology to process and transform non-born digital materials (text, images and audio files), analogue data, social data and cultural collections on Africa into electronically processed digital forms for research, policy and preservation purposes. Asogwa (2011) describes it as the transformation of analogue information in whatever form and from whatever support to a digital code using computing technology. It also includes the process of archiving born digitals into institutional collections. It is anticipated that a substantial component of the activities that will engage the attention of scholars within the Bayreuth Multiple Cluster Excellence's initiative will largely focus on the use of technology to transform and process the contents of the materials and research data in African Studies into digital forms. Commenting on the frame of the Cluster of Excellence project, the University of Bayreuth submits that:

> The African Cluster Centres (ACCs) are located at the University of Lagos (Nigeria), Moi University (Kenya), Rhodes University (South Africa), and Université Ki-Zerbo (Burkina Faso) and are equipped with the financial means for own research, thus enabling them to make their own contributions to the four Clusters' agenda. They serve as the cluster's on-site research partners in Africa, facilitating the exchange of ideas; the mobility of researchers; the collection of and access to data; and above all offering a platform for reflexive research on the continent itself. The ACCs' responsibilities include initiating and conducting research projects linked to the cluster; connecting its members to the large cluster network in Africa and beyond; organising and hosting international conferences; and providing a platform for academic exchange at doctoral and postdoctoral levels. The overall aim is to promote a generation of young academics whose objective, like that of the cluster, is to reconfigure African studies through transdisciplinary approaches that create new insights into the multiplicity of Africa and its diasporas (https://www.africamultiple.uni-bayreuth.de/en/african-cluster-centres/index.html).

Given this onerous task, the Cluster's initiative incorporates the critical role of digital technology embedded within the digital research environment to facilitate the seamless accomplishment of the objectives. With the availability of gargantuan amounts of pristine data in Africa, researchers are often perplexed by a sea of the research information as well as cultural materials and collections that must be digitised, processed, analysed, preserved and made available for use by other researchers. A well-articulated, well-structured and well-funded technology-driven research on cultural and linguistic information can unveil startling discoveries about the histories, peoples, cultures, civilisations, epistemes and worldviews of traditional and contemporary African communities.

Extant on digitisation initiatives in Africa has highlighted some of the digitisation projects which have been ongoing on a much slower scale in Africa since around the year 2000. Adeleke (2007), for example, reports that although digitisation is still a novelty in Africa, some digitisation initiatives and programmes and small-scale projects are ongoing, for instance, in institutions in

southern Africa, Egypt and sub-Saharan Africa countries such as Kenya and Nigeria. Tsebe (2005) identified some of such projects as follows:

- The German Colonial Society's collection of 55,000 photographic impressions from Africa – a project completed in 1999.
- The West African Research Centre's complete digitisation of 150 colonial reports in 2002.
- Sabinet Online's digitisation of 40 scholarly journals in 2002 and another 141 titles in 2004.
- The University of Cape Town's partnership with the National Library of South Africa which led to the digitisation of 345 drawings of the Black collections in 2003.
- Michigan State University's digitisation of 10 African journals by 2004.
- The National Library of Egypt's digitization of 100,000 pages by 2004.

In a similar context, the University of Lagos, Nigeria is currently digitising a number of records and manuscripts in its holdings. However, none of these previous digitisation initiatives can be compared with the proposed ambitious projects that the Africa Multiple Cluster Centres are set to undertake.

No doubt, digitisation may be described as one major area of the digital research enterprise that holds the brightest future for the Cluster. Indeed, digitisation projects lie at the heart of the reconfiguration concept (i.e., the transformation and processing of hard data into digital formats) being promoted by the Cluster. As referenced above, African digitisation has so far been under-studied, undervalued and under-resourced. While many cultural and memory institutions in other parts of the world are making huge investments in digitising their collections and holdings, very little is being done to preserve African cultures and memories in digital forms. No doubt, it is the conviction of this study that when digitisation is properly executed, it will help in preserving rare and endangered materials, provide greater access to the digitally archived materials to more people across the globe through online platforms such as the internet, social media networks (SNS) and dedicated websites. By the same token, such digitisation projects will also reduce over-handling of materials such as manuscripts and published documents that might lead to defacement and destruction of rare materials and will assist in promoting the wider circulation and transmission of the collections while making the institutions and their works more visible to the global community.

On the other hand, digitalisation deals with a broader framework for the application of digital technologies to all human activities. Morze and Strutynska (2021) observe that digital technologies underpin the phenomenal digital transformation and its impact in different of hman activities in modern society. They argue that "modern digital technologies, services and systems are extremely important for social development (p.1)" In other words, this new phenomenon deals with how computer methodologies, digital technologies and Information and Communication Technology (ICT) are transforming the way we live, work, communicate, teach and learn, as well as conduct research and communicate its results in modern societies.

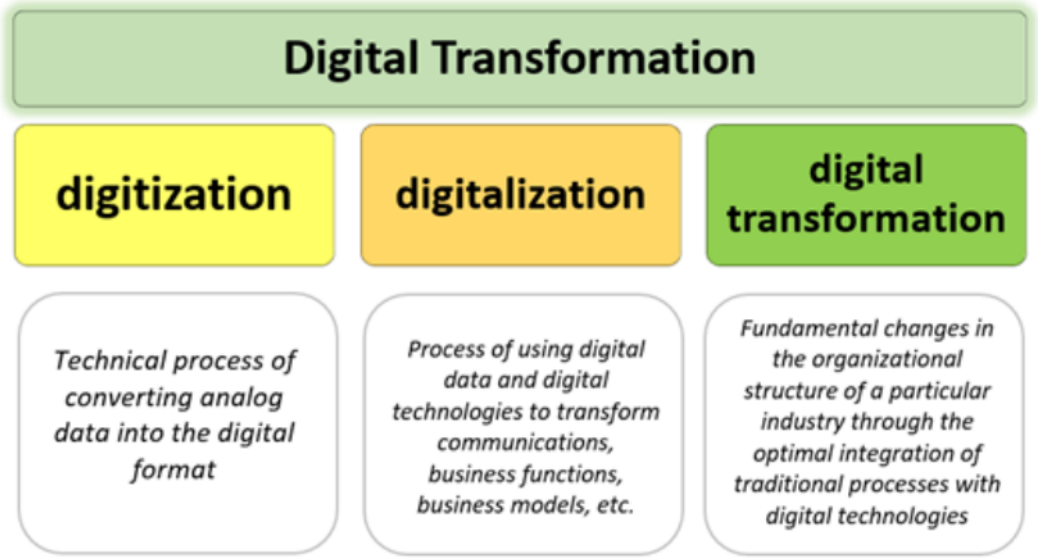


Figure 6: Key Stages of digital transformation as espoused in this study. (Source: Morze and Strutynska 2021: 3).

2.2 Digital Technology and African Studies

Over the years Africa has been misrepresented in western media and its potential grossly undervalued, yet the continent cannot be ignored by the global community. Whether it is believed that the continent's plight is self-inflicted or imposed by external forces, the stereotypical misrepresentation of the continent cannot be sustained with empirical sociological scholarship. Over the years, Africa has been presented as a continent plagued with wars, conflict, poverty, illiteracy and underdevelopment. Sadly, very little attention has been paid to the positive achievements of Africa and Africans at home and in the diaspora. The bustling potential and excellent contributions of African academics and professionals have been continually underreported in western media. Believing and sustaining these narratives has grossly undervalued the contributions of the continent to global development. However, as presently configured, the Global North continues to benefit immensely from Africa in terms of huge migrant students, scholars and professionals that are contributing in no small measure to the educational and scientific development of their host countries. The emerging economic aand educational markets in Africa have also become business destinations for many international corporations and foreign institutions while the continent has continued to sustain productions of goods and services in international business ecosystem with large deposits of natural resources, and human capital. The ongoing digital revolution has upscaled further the relevance of the continent in global socio-economic, political and technological affairs.

It is noteworthy however, that the global digital technology community has recently begun to realise that Africa remains the next "biggest miracle" to happen (i.e., Africa hosts some of the most thriving emerging tech markets and is home to a huge young population with untapped potential). Africa's major cities are thus key in the quest to build global digital hubs, reliable supply chains and destination points as ready markets for their products and innovations. With the huge

number of thriving start-ups and digital initiatives in Nigeria, Kenya, South Africa and Egypt, a number of these emerging digital hubs are set to redefine and reconfigure research and innovation in Africa that will impact the global community.

In 2015, Shapshak had rightly observed thus:

> Africa's internet growth is faster than anywhere else, as worldwide international internet capacity has slowed, according to new research, which is good news for African internet users who will start to see not only increased speeds but hopefully cheaper prices as the network effect suggests. African Internet bandwidth grew 41% between 2014 and 2015, and 51% compounded annually over the last five years (Shapshak 2015, n.p.).

It may be argued here that with the growing access to mobile telecommunication services and internet penetration, as well as availability of new digital technologies, African scholars are better equipped to deploy these technologies to reconfigure Africa Studies.

3 Reconfiguring Research and Scholarship in Africa through Digital Technology: Implications for African Studies

Before the establishment of the Africa Multiple Cluster initiative with a strong focus on creating a digital research environment that will drive the research projects, a number of digitisation initiatives driven by librarians and archivists had been developed. As explained below, digitalisation and digitisation are interconnected because both lend credence to the roles of technology in transforming research activities and research data. While the former is a process, the latter may be viewed as a product. McKay (2003) observes that the availability and use of digital technology has become a fundamental part of the institutional efforts to improve and digitize archives, museums and libraries, and these technologies have resulted in museums and archival institutions transforming their services and re-evaluating their target audience. (McKay 2003:2; also cited in Asogwa, 2011, n.p) In addition, an ongoing project at the University of Lagos focuses on the use of digital tools to construct a specialised monitoring corpus of online discursive practices. In other parts of Africa and in the diaspora, researchers and scholars have been exploring technology-driven projects focusing on Africa, such as the Atlantic Slave Trade projects by some Nigerian-born scholars in U.S. institutions and the digitisation of some cultural practices in eastern and southern Africa. The Digital Imaging South Africa (DISA) project has equally contributed to the numerous small-scale initiatives. Easterbrook states that:

> Over the last decade, numerous foundations in the United States, including those forming the Partnership for Higher Education in Africa, have provided unprecedented support for the revitalization of African universities and their libraries, especially through digitization of African materials and the development of associated information technologies. North American and European institutions, some in partnership with African institutions, have also undertaken a range of library and archival projects, many focused on digitization. (Easterbrook 2007 n.p.)

However, Easterbrook advises that more attention should be paid to what is being digitized as well as those who are involved and how the digital information will be used and disseminated.

Regardless of the potential challenges that may confront digitisation projects in Africa, technology will continue to play a crucial role in reconfiguring African Studies, given the growing expansion of digital technologies on the continent. On that note, Javed asserts thus:

> the penetration of internet in Africa will triple by 2025 to exceed 50%. This represents 600 million regular Internet users. The study also predicts a strong potential for the Internet of Things in developing countries. By 2020, these countries could represent 40% of the global market value of the IoT.
>
> Currently 15% of the world population lives in Africa. More than half of global population growth between now and 2050 is expected to occur in Africa. Therefore, the deployment of a connected system is essential to this (…). The potential of the Internet of Things in Africa is unlimited. As technology advances and integrates daily life of most of citizens, we always expect more from IoT solutions to solve the problems. (Javed 2016 n.p.)

Schelenz and Schopp observe that in Africa

> much has been done to advance digitalization with improved telecommunications, regulations, electricity, and the integration of ICT in the banking and health industries, education, entertainment, and politics (…). Moreover, tech hubs are emerging across the continent with currently 50 percent of hubs located in the digital frontrunners South Africa, Kenya, Nigeria, Egypt, and Morocco. (Schelenz and Schopp 2018: 1412)

It simply implies that the rate of technological development on the continent bears regional variations. This might have underpinned the choice of the centres within the University of Bayreuth's Africa Multiple Cluster's initiative. The vision for digital transformation of the works at the ACCs through the Digital Research Environment component forms the cornerstone of the cluster's digitization and internationalization strategy.

Although a few universities in some of the regions where the Cluster Centres are located have made some progress in enhancing their operations and services through new technologies, Mohammed (2013) argues that the rapid progress being made in the application of digital technologies in other sectors in Africa has not been fully realised in academia. With over 900 higher education institutions in sub-Saharan Africa and more than 150 research institutions, very little is known about the important work that scholars and researchers are undertaking in these institutions.

Fortunately, in the last one deacade several universities in Africa are now setting up open-access institutional repositories for storing African scholars' intellectual properties and research enterprises, thus making them more accessible online. Postgraduate theses, research articles, manuscripts and scholarly academic publications are also being uploaded on the repositories. Consequently, the works of research communities in Africa are becoming more visible and

accessible both locally and internationally. Thus, some of the Centres within the "Africa Multiple" Cluster have some existing digital frameworks on which to leverage. Molteno's (2016) historical narrative on 'African Institutional Repositories' provides an interesting exposé. He submits that South African universities started massive digitilastion programmes in 2000, followed by other institutions in sub-Saharan Africa. In his words, "[w]hat is exciting is how fast this dimension of Africa's digital landscape is expanding." (Molteno 2016: 172-3)

The second major component of the digitalisation initiative has been the establishment of open universities in some African countries. Both digitalisation and digitisation initiatives are going on hand-in-hand in many of the institutions, especially in South Africa, Nigeria, Kenya and Egypt. However, the processes of automating academic, administrative and research activities have enjoyed greater successes than digitisation projects. For instance, a number of these universities now use online platforms to deliver course contents and training. The University of Lagos for instance, operates a vibrant Distance Learning Institute (DLI) that enrolls more than 20,000 students, while the National Open University in Nigeria (NOUN) provides learning opportunities for more than 40,000 students. A significant number of graduates of the open universities are using this system to secure employment opportunities and/or advance their careers. Many of them including the University of Lagos now deploy and sustain different online learning platforms for teaching and learning following the COVID-19 public health crisis.

How does this connect with the topic of discussion and the preoccupations of the digital research group within the Cluster? There are two broad undertakings that speak to our role as digital humanists helping other researchers. We help to identify relevant digital tools and techniques and assist researchers and scholars to locate their bearings and create the enabling digital environment for them to map out their research trajectory in this new approach to conducting research in African Studies. It is the deployment of these digital solutions, situated within digital scholarship orientation, that will redefine research and scholarship in Arican Studies.

Digital scholarship has been gaining momentum as the technology-driven approach to reconfigure academia and its relevance to the society. However, it requires serious commitment and investment. Weller (2016), in 'The Digital Scholar', suggests that 'digital scholarship' is really shorthand for the intersection of three technology-related developments, namely: (i) digital content (ii) networked distribution and (iii) open practices.

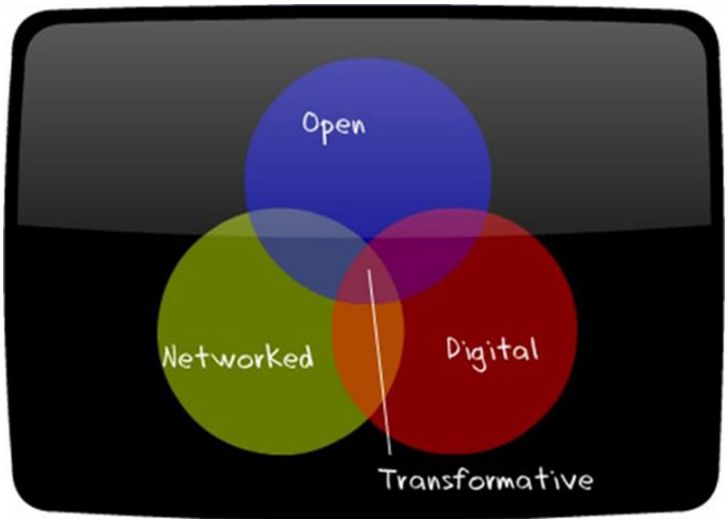


Figure 7: Integrative Digital Scholarship (Source: Weller 2016).

Weller argues that it is only when digital, networked and open practices intersect that transformational practice can occur. The three concepts are germane to the philosophy of the digital research environment undergirding the digital research framework of the Cluster. They underlie the vision to reconfigure African Studies through technology. When digitalisation is fully pursued and implemented, it will promote easy and rapid knowledge production, information sourcing and management, big data creation and curation, as well as data redistribution and reuse. With technology at the heart of the research process, content production, content distribution and content sharing based on open practice will galvanise African scholars' efforts for maximum global impact.

Weller (ibid.) presents some of the possibilities which the application of digital technology affords scholarly practice. These include: (i) building a digital collection of information for further study and analysis, (ii) creating appropriate tools for data collection and building, (iii) creating appropriate tools for the analysis and study of collections, (iv) using digital collections and analytical tools to generate new intellectual products, and (v) creating authoring tools for these new intellectual products, in either traditional or digital form (Weller 2011: 32). These are obviously profound insights encapsulating a set of visions and missions worth pursuing within the framework of the African Centres' initiative in the Cluster.

Rademacher's (2019) proposition that the "digital revolution" in Africa is taking place successfully because it is based on important values originating in indigenous cultures, including African cultures rather than only Western principles, helps to localise and contextualise these perspectives within the framework of African Studies. Rademacher suggests that since many African countries have leapfrogged into the mediated world of smartphones and applications without passing through the stage of wired telephone use, the necessary preparatory platform has been created to drive the digital agenda of the Clusters. The digital research environment component bridges the critical gap by providing digital infrastructure and local experts who can provide the back-end support for the digitalisation and digitisation of research activities at the centre. This digitisation agenda stands to redefine, reconfigure and retool studies that utilise

Africa's historical, epistemological, social and cultural data. Within the context of the African Multiple Cluster therefore, digitisation is taken as the process that will systematically shape the transition from analogue to digital work (Seesman and Fendler, 2019, n.p.). They assert that through virtual communication, data sharing and co-creation of research contents among other technology-driven collaborative activities, these noew orientation will create important impetus for the conceptual as well as structural realignment of African studies.

The coalition of these activities supports the technology-driven collaborative and translational component which will escalate the transformative power of the DRE as summarised below.

3.1 Some Benefits of Digital Technologies in the Lagos Cluster

Within the last couple of years that the University of Bayreuth initiated the Africa Multiple Cluster of Excellence, we have witnessed some significant improvement in collaborative research works across the Clentres as well as among researchers within the Lagos Centre, most of which works are largely facilitated through the use of technology and the digital research environment. At the institutional level, it has become much simpler to respond to researchers' and students' demands for and shared data. In a way this has enhanced access to educational facilities, with more students enrolling for postgraduate programmes in African and Diaspora Studies at the Lagos Centre. It has been proved that the deployment of digital technologies for research in the humanities can deliver additional benefits that are directly related to the goals and objectives of the cluster: **(i) Integration of qualitative and quantitative approaches** – Researchers in different fields in African studies can present and interlink digitized text, images, and time-based media with maps, timelines, data, and visualizations. **(ii) Content management and data analysis**. Researchers and scholars are becoming active creators (not passive consumers) of digital contents through enhanced skills in digital humaniites. They can mine, map and re-organize the resources. New technologies and web applications now deliver techniques that can be deployed to uncover and discuss trends, themes and key learnings. **(iii) Quicker access to information through digital access** – By promoting the principles of Findable, Accessible, Interoperable and Reusable (F.A.I.R) in data management, it means more people can review, see and learn from the project. Researchers within and outside the projects and in other parts of the world are able to more easily search through the data, combine different data sources, hyperlink to relevant background materials, and more. **(iv) Enhanced teaching** - By adopting techniques and tools in Digital Humanities, the works will help students and mentees to learn by being able to see more, experience more, and collaborate together. **(v) Improved collaboration** – This remains one of the biggest benefits of digital technologies for research purposes. The use of digital resources and digital research environments within the framework of the initiative can provide a common platform for project development and group-sourcing of materials, and facilitate local, regional and global partnerships. **(v) Public impact** – The translational dimension in digital humanities projects shows forth more profoundly in the potential of technology-driven humanities projects to deliver social good. The impacts of such project will extend beyond the classroom and make a public impact. Such research outcomes with significant public impact demonstrate the value of the study of Humanities and prove that digital projects can also help to inform and engage those outside the university setting. As a corolarry, the aggregate of these benefits can enhance the fiancial sustainability of the institute by using technologies to create new sources of revenue,

attract industry and stakeholders‘ patronage as a result of an enhanced institutional brand awareness(https://digitalhumanities.duke.edu/benefits-digital-humanities) .

Working within the same digital research space, the Centre for Digital Humanities, University of Lagos has been able to provide and promote digitally mediated environments for learners and researchers. The Lagos Summer School in Digital Humanities organised by the Centre for Digital Humanities, University of Lagos – now in its third edition (2017, 2018, 2021) – as well as symposia and seminars being organised by the Centre have created more awarenss about the potential of the new technology-driven field and digital tools to improve research and scholarship among scholars in the humanities and social sciences. Learners and researchers at the Centre and those affiliated with us are increasingly developing collaboration and commuication skills by utilising existing and new digital platforms. The deployment of technologies within these research environments has also helped to enhance a range of computer methodologes and analytical skills that will be useful within the work environment in professional careers outside academia.

The focus of the digital research group within the Cluster would be to create both the digital environment, techniques and skills that can help to promote open scholarship while providing and preserving access to digital content and ensuring support for research data management.

While ethical issues have often been debated in this space, efforts have been made to ensure that local and international social and legal frameworks are respected and complied with. Suffient information on metadata is often stressed, while subject identity anonymisation and subject express permission in data collection are often incorporated in the course of data collection. Communities where sets of specific data are elicited are often encouraged to sign up as joint owners of the data. Accurate references and citations are also included in the various African Studies-based research projects being explored and executed at the Centres.

3.2 The Digital Humanities Approach: Pushing the Frontiers of Digital Research in Africa

The last few years have witnessed the emergence of Digital Humanities (DH) initiatives in some parts of Africa. Basically, DH is an interdisciplinary orientation that espouses the application of computer methods and digital technologies for research and studies in the humanities. The Africa Multiple Cluster of Excellence, with its strong emphasis on the technology-driven component, may be described as “a child born in due season.” While a lot of infrastructural investments may be required to achieve the full objectives of this laudable initiative by the University of Bayreuth African Studies Programme, a number of existing DH programmes, digitalisation initiatives, facilities and skills in some of these universities such as Lagos in Nigeria, Rhodes in South Africa, and Moi in Kenya can be exploited and developed towards the realisation of the goals of the Cluster.

The DH initiatives now receiving wider embrace in sub-Saharan Africa and South Africa are a springboard that can help to leapfrog this initiative. For instance, at the University of Lagos, our Centre for Digital Humanities is the first full-fledged stand-alone centre for DH in Africa. We started as a small research unit in 2015 on my return from a Humboldt fellowship in Germany. Recently, we have now been upgraded to a full research and academic unit with full institutional authority. The Lagos Summer School in Digital Humanities (LSSDH) has become the flagship of

our training programmes. We have trained over a hundred junior and middle-level scholars and researchers on courses such as (i) Programming for the Humanities/Minimal Computing, (ii) Text Mining and Annotation, (iii) Social Media Analytics: Processes and Productions, (iv) Introduction to Digital Mapping and Visualisation, and (v) Digital Archives for Humanities.

4 The Digital Research Environment as a Tool for Transforming African Studies

As the application of digital technologies impacts research and scholarship, the creation of a digital research environment remains a laudable initiative that must be strengthened to help drive the digital research component of the Cluster's vision and to keep such questions on the front burner.

Unarguably, the application of digital technologies in African studies will re-equip the researchers and add a new intellectual impetus to our research activities and outcomes. The Digital Research Environment (DRE) is a concept rooted in the application of technology in research projects to enhance and amplify multidisciplinarity and transdisciplinarity or the forming of research Clusters for knowledge production, knowledge sharing, knowledge processing and knowledge publishing. The emphasis is on collaboration, networking, data processing and management, as well as a multidirectional exchange of information. DRE throws up the relevance and centrality of the use of digital technology and computer-based applications in translational and collaborative research directed towards achieving scientific and social impact.

Digital research has assumed an interesting dimension even in the humanities in research years. Although the use of technology may drive humanities-based projects, digital research is not defined by a particular perspective, method, tool, or unit of analysis. Rather it is defined by the degree to which the digital is centrally relevant in the phenomenon, context, or focus of analysis and the project itself (Markham (2020). It enhances the capacity of researhers to innovate and make new discoveries in traditional humanistic data. Some of the perspectives shared in some on digital research speak to the Cluster's philosophy and must continue to guide the preoccupations of the centres. Some of these digital reseach perspetives in the humanities are :

- Such technology-driven projects are situated in studies of people, societies, and sociotechnical relations that are digitally-saturated, or somehow impacted by transformations wrought by digitalization, widespread internet connectivity, and the largescale datafication of society.
- The use of multidisciplinary, multi-institutional and often international consortia approaches to promote collaborative agenda and activities.
- The projects must emphasise greater socially-impactful outcomes based on stronger evidence of the collection of empirical data from actual communities seeking to solve real-life challenges and must welcome inputs from more diverse partnerships.
- The use of web-based applications and digital tools, the consolidation of research equipment and the need to manage and share data effectively, technology-driven analysis of data in order to promote greater efficiency, and transparency in data management that

> will encourage open access to research findings; and drive the research projects (e.g. Markham, ibid. University of Nothingham's digital research centre)

Within the context of the Africa Multiple Cluster's initiative, it was observed elsewhere that the creation of a viable DRE will help to build and maintain the multidirectional connections between the Digital Media Laboratory, the Research Units and the ACCs. For instance, at the Lagos Centre, the Digital Media Laboratory is the digital research space provided by the Centre for Digital Humanities with digital infrstructures that support the works at the Centre whie the research units are individual research project groups on African Studies. Both components work within the African Cluster Centres (ACCs) to promote the goals of rhe Cluster. The main objective would be to provide digital solutions that enable these affiliates to function optimally and deliver on the mandates of the centres. Technology-driven academic activities such as live streaming of lectures, symposia, workshops and conferences will enhance our goal of reaching out and connecting with colleagues, researchers, students and the larger society. The DRE will also handle the entire Cluster's data management and digitisation projects. As the digital solutions provider, it will ensure that resident scholars and visiting researchers, fellows, faculty and the ACC members contribute their data and research outputs to the Cluster's common database. The DRE will be the 'virtual' administrative and research hub for connecting the Cluster's theoretical, epistemological and methodological issues, eliciting debates and intellectual exchange, generating novel ideas, collaborating and networking to stimulate new theoretical advances, and engineering a thriving research environment for the Cluster's researchers.

Given this general digital outlook, different concepts, practices and ideas that transcend academic materials will be integrated into its overall framework of practices and deliverables. The traditional, conventional, born-digital and digitised data will be warehoused into the online repository, which will be accessible for research, pedagogical and policy purposes. Qualitative and quantitative data from social media and other online platforms will be archived into a common digital research platform. Moreover, we shall adopt standard and novel procedures of the entire chain of data management from creation to consumption. Because the DRE also recognises other forms of knowledge production and transmission, it will overcome the existing limitation and structures in the conventional databases by leveraging on the ongoing cross-disciplinary collaborations and contributions from other academic disciplines and means of knowledge production. The DRE team will also work to ensure that the internal databases are connected to all the relevant data sources and work seamlessly with all the connecting research groups within the Centre.

Since the Clusters will be dealing with Big Data, we have proposed the use of Apache Spark Technology for data management among other useful DRE data management digital infrastructure. Broadley provides some insight on this technology:

> Apache Spark™ is a unified analytics engine for large-scale data processing. In Human terms, Spark is a distributed computing framework that has a common interface across multiple languages for things like SQL & MapReduce when querying in-memory datasets. Being distributed, Spark can work with extremely large datasets across a cluster of machines.

> From an Engineering perspective, Spark is available in multiple languages: Scala, Java, Python and R. This seemed like a pretty big win, as it doesn't limit what we can hire for a specific language. Spark has its own DSL (Domain Specific Language) that's the same across all implementations, meaning there's a common language despite the choice of implementation language. (Broadley 2018 n.p.)

From users and researchers' perspective, 'Spark Archives' will be a useful facility for storing our data and making them available for public consumption. It provides a data management platform enriched by various functionalities, including design and record management processes (Transfer, Communication and Consultation, Processing, Preservation, e-discovery adaptable for African Studies) and a logistical monitoring of storage facilities. With tools that will provide a rich yet flexible user experience networking across multiple channels, it will enable a virtual environment that will generate innovative ideas and research topics, as well as help in creating new knowledge and new concepts. The role of the DRE will thus significantly scale up the impact of digital tools in the reconfiguration of African and Diaspora Studies. It will cascade on individual scholars' efforts, helping to inspire new thinking and encouraging digitally-driven approaches in transdisciplinary traditional research contexts.

The significance of the Digital Research Environment (DRE) for the research projects at the African Cluster Centres is further amplified below:

Another important component of the Africa Multiple Cluster is its innovative Digital Research Environment (DRE). All cluster institutions, whether in Bayreuth, Africa, or elsewhere on the globe will be interconnected through the DRE. The DRE not only ensures communication and virtual collaboration, *but also creates the conditions for integrating the highly heterogeneous qualitative and quantitative data material into a common digital research platform, thus making it accessible to all participating researchers*. (Spektrum, 2019: 9)

Essentially, one other major goal of developing the DRE is to enhance the Centres' ability to handle, prioritise and analyse huge data in a way that such data can be repurposed, reprocessed, reproduced and reused by other scholars. It does adopt the EU's FAIR Policy for Open Science, which requires that data must be findable, accessible, interoperable and reusable.

5 Benefits of the Digital Research Environment

As explained above, the digital framework is being put in place in most of the Cluster's locations. Humanities scholars in Africa can benefit from the ongoing upsurge in Internet penetration, expansion in mobile network services and availability of free digital and corpus tools to transform African Studies. Experts believe that our ability to manage, share and curate our data in a digital research environment can make considerable difference in our works. According to the policy postion espoused on Libguides(www), it is believed that the application of technologies for research data management will help to (i) ensure research integrity and validation of results, (ii) increase research efficiency, (iii) facilitate data security and minimise the risk of data loss, (iv) ensure wider dissemination and increased impact, (v) enable research continuity through secondary data use, and (vi) promote international collaborative research among other benefits.

Keating (2012) states that digital technology can escalate the impact of the humanities' possibility of generating new tools that can be used and adapted to mediate knowledge within humanities scholarship. These tools can be deployed to interact with humans and communities to achieve set objectives. Additionally, new knowledge is generated, together with new kinds of community discourses arising from multidisciplinary and interdisciplinary research relationships. It is also possible to gain new perspectives on old problems or to use old perspectives in viewing new problems.

For the purpose of emphasis, Digital Reaserch E nvironment (DRE) promotes the following, among other benefits:

> (i) Visibility: Open-access content attracts more attention than content that is not open-access.
>
> (ii) Openness: Open content promotes increased citation and usage as well as greater public engagement.
>
> (iii) Collaboration: DRE supports the Clusters' aspiration to achieve transdisciplinary synergies.
>
> (iv) Digitisation – meaning 'systematically shaping the transition from analogue to digital work (Seesemann and Fendler 2019), which increases access to and impact of our research products.
>
> (v) Produces new discoveries: Open-access data and papers accelerate the pace of scientific enquiry with faster impact, wider collaboration and increased interdisciplinary conversation. It thus creates new knowledge, and facilitates the growth of fast-paced technology-driven collaborative projects in the humanities.

It should be noted that DRE is not only about the development of a virtual research environment for the project but also about the future use of digitisation for scientific practice in Africa and for African Studies. We are therefore laying a solid foundation for the digitisation of data on Africa and for African and Diaspora Studies.

Roger's (2015) view on digital methods as a research practice is equally helpful in guiding the works at the centres. She assers that digital research and its methods is "part of the computational turn in the humanities and social sciences, and as such may be situated alongside other recent approaches, such as cultural analytics, culturomics, and virtual methods, where distinctions may be made about the types of data employed (natively digital and digitized) as well as method (written for the medium, ormigrated to it)(p.1)"

6 Digital Scholarship in African Studies: Towards Translational and Collaborative Impact

Digitalisation initiatives have often shown the transformative and collaborative power that these new technologies are able to deliver. From the most simple and most affordable technology (e.g.,

the e-mail) to the most sophisticated (e.g., cloud computing), the speed, breadth and power to reach the global audience at just a click of the button remains the most unbeatable collaborative benefits of these technologies. Christensen's (2016) model of disruptive technology considers collaboration as one of the biggest advantages of new technologies.

Technology provides the basis for translational research, which has become a new approach to knowledge production and processing. Academic and policy notes the new challenge to academia in Africa as society seeks evidence of our relevance and contributions to social problems. The use of technology remains the only viable option for the academy to answer back more powerfully with a positive impact.

Therefore, it is incumbent on African scholars to reset their research agenda and activities in the light of the impact of these technologies, especially by using them to transform traditional ways of thinking and acting in today's knowledge industry. Consequently, African scholars are forced to reappraise their methodologies and how to communicate their research outcomes. This will enhance Africa-based democratisation of knowledge and open science, thereby improving the global visibility and impact of African research.

In creating the digital research environment, then, African researchers must focus on creating an enabling research space that is accessible, robust and efficient. Relevant equipment, as well as highly trained experts and knowledgeable researchers who understand the objectives of the Cluster, must therefore be available.

With clear emphasis on the role of technology in redefining scholarship and innovation in African Studies, Lubar's (2014) descriptive conceptualisation of the benefits of the disciplines of the humanities to society is germane. He captures this in five adjectives that clearly speak to new ways of working beyond the traditional methods, viz: applied, translational, open, digital, and public humanities.

7 Pedagogical Approaches to Technology-Driven African Studies

With the right infrastructure and human capital in place, the role of teaching and learning within the scope of the digital research environment cannot be underestimated. The approaches of digital humanities and digital scholarship, which are now gaining some momentum across Africa, must be maintained. To be sure, the University of Lagos Centre for Digital Humanities is engaged in raising a new generation of scholars and researchers in Africa. The annual Summer School in Digital Humanities has been training these scholars in the use of digital tools for research and innovation. By 2020/2021, the University of Lagos will be commencing a Master's degree programme in Digital and Public Humanities.

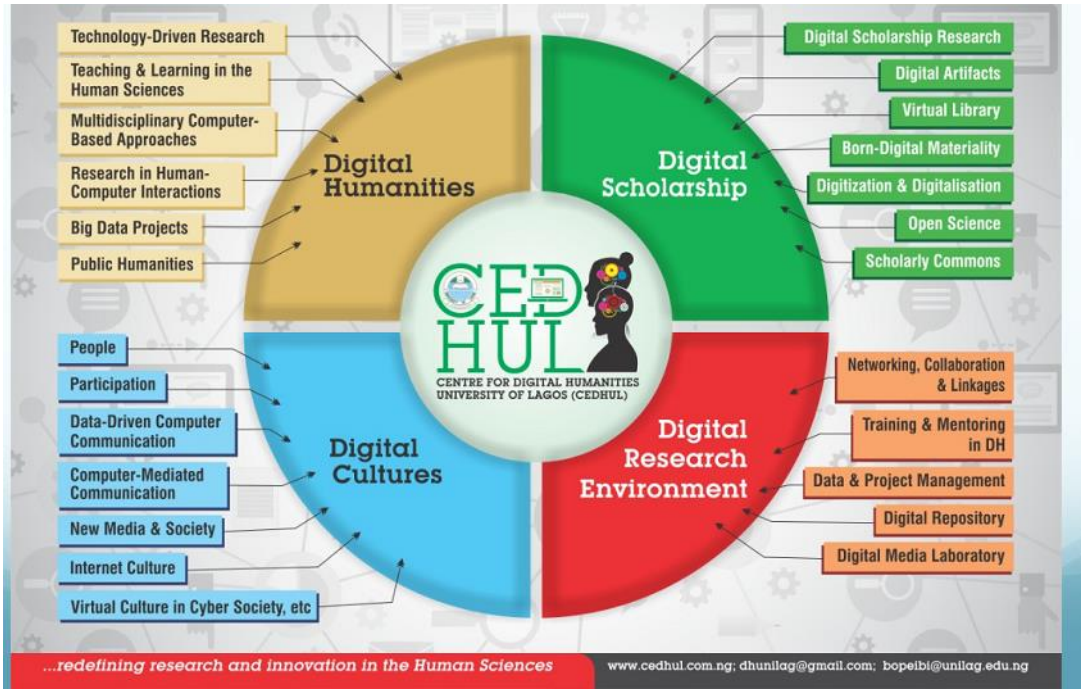


Figure 8: Illustration of the activities at the Centre for Digital Humanities, University of Lagos (CEDHUL).

The Centre for Digital Humanities, University of Lagos (CEDHUL) provides a framework—and the expertise—that supports the Lagos Centre. By pioneering the promotion of the principles and practice of Digital Humanities and digital scholarship in sub-Saharan Africa, CEDHUL has begun to contribute towards reconfiguring African Studies through the application of digital technologies.

With minimal training, researchers in the different sub-disciplines of African Studies can focus more directly on key issues in digital transformation as they affect cultural forms and practices, societies, archival issues, as well as knowledge production and dissemination. Museums, archives and cultural heritage institutions may adopt crowdsourcing strategies as a means of accessing and curating digital records and engaging members of the public with digitised archival collections that will be published on dedicated websites.

Other strategies may include leveraging on the growing interrelationship between data and culture. Hence, efforts should be made to bring together artists, historians, philosophers, literary scholars, social scientists and computer/data scientists to engage with one another and with the public on sociocultural data. The digital research initiatives must focus on how scholars can utilise digital tools and explore technology-based analytic techniques in order to unveil new understandings of African cultural practices. Since the ubiquity of digital technologies, social media, and the network, in general, have influenced almost every domain in everyday life, production and distribution of culture is being transformed and reshaped. The application of technology in the academia is enhancing systematic research which shows that physical space and cyberspace interpenetrate as people actively surf their networks online and off-line. The huge

data being generated daily thus impacts how culture is [re-]produced, [re-] distributed and influenced (Levin & Mamlick, 2021: 2; Wellman, 2001)

This study submits that a serious and consistent engagement with different aspects and sub-disciplinary areas in digital humanities (DH) will help African scholars and researchers to explore the continent's rich social, epistemological, historical and traditional heritage, as well as its cultural-memory data, for the repositioning of Africa studies and the benefit of the global community. Besides, technology-driven research, which is at the very heart of DH, will transform their research projects into cultural commodities and globalise research outcomes emanating from the continent and from scholars in the diaspora working on Africa. Construction of digital repositories and creation of a digital research environment will promote preservation of the data and research work while encouraging co-creation alongside collaborative and translational research outcomes. Application of digital technologies can also help to engage with younger generations of Africans and online citizens, in addition to helping to preserve research projects on Africa for posterity. By the same token, digital humanities projects can help these scholars to retell their stories in a more compelling way via new technologies and approaches to research in the humanities.

Against this backdrop, the underlisted areas of Digital Humanities, among others, can therefore help to reconfigure African Studies as explained above:

(i) Computational Humanities, (ii) DH and Decolonisation, (iii) Digital Postcolonial Analysis, (iv) Big Data and Cultural Studies + Visual Culture, (v) Digital Musicology, (vi) Digital Humanities and New Media, (vii) Data Modeling, (viii) Data, object, and Artefact Preservation, (ix) (x) Digital Activism and Advocacy, (xi) Digital Archiving, (xii) Digital Art Production and Analysis, (xiii) Digital Ecologies and Digital Communities Creation Management and Analysis, (xiv) Digital Libraries Creation, Management and Analysis, (xv) 3D Printing, Critical Making, (xvi) Attribution Studies and Stylometric Analysis and Cultural Analytics, (xvii) Data, Object and Artefact Preservation, (xviii) Digital Archiving, (xviii) Digital Activism and Advocacy, (xix) Digital Art Production and Analysis, (xx) Digital Ecologies and Digital Communities Creation, Management and Analysis Digitisation (2D & 3D), (xxi) Electronic Production and Analysis (Ethnographic Analysis), (xxii) Manuscripts Description, Representation and Analysis, (xxiii) Media Archaeology, (xxiv) Public Humanities Collaborations and Methods, (xxv) Natural Language Processing, (xxvi) Music and Sound Digitisation, Encoding and Analysis, (xxvii) Social Media Analysis and Methods, (xxviii) Text Mining and Analysis and TEI-XML for Textual Editing, etc. (source: ADHO 2019, with personal input).

In that regard, this chapter advocates guided and customised training in Digital Humanities and digital scholarship in order to train researchers and scholars on how to deploy digital technologies to facilitate, enhance and change scholarly practices and knowledge production in African Studies. With the availability of digital tools, scholars can now adopt established and emerging methods for data modelling, data mining and text analytics and other approaches that can help to work with linked open data, reproducible research and platforms for sharing research and teaching resources across the Clusters.

Digital Humanities training enables scholars in Africa to learn how to use network analysis, social media and virtual ethnography to advance their scholarly work and to reflect upon ethical and legal issues in digital research. The digital research team in all the centres must have continued and sustained conversations on developing a robust and functional database to warehouse all the digitisation projects in Africa and on African studies being undertaken in Africa and globally, not only at the Cluster centres. Our websites must be interconnected and have digital contents that can showcase the works of scholars and researchers.

7.1 Proposal for Regular Training in Digital Research and Digital Scholarship

Given the current knowledge gap in the use of digital technologies by the majority of humanities scholars in most African universities, there is an urgent need to design aggressive, structured and tailor-made training programmes in digital research and digital scholarship. My experience during our Lagos Summer School in Digital Humanities confirms the necessity and usefulness of regular training. For instance, many of the attendees of the Summer School are already deploying the skills and experience gained for teaching and research purposes. Recently, one of them worked with an ICT expert to develop an application for language documentation in his local area. No doubt, then, training will help to raise and equip a critical mass of scholars that can truly transform African Studies through new technologies. Consequently, I would like to propose a one-week Summer School in Digital Research and Digital Scholarship in African Studies.

8 Problems/Challenges Confronting Digitalisation of the Academe in Africa

It bears noting that as laudable as this digital research initiative is to the work of the Clusters, there are some challenges that must be confronted squarely in our quest for a successful implementation of this ambitious project. Issues of the 'digital divide', political and social inequality, limited and inadequate accessibility to digital infrastructure and Internet service, as well as erratic power supply, etc., are serious obstacles to the commencement of technology-driven projects targeting the development of African Studies beyond the previous traditional paradigms. In sum, these challenges include the following: (i) Digital infrastructure Deficit in some African Centres, (ii) Low Internet penetration and inadequate bandwidth, (iii) Energy problem/Power supply dysfunctionality, (ii) Dearth of ICT Experts, (iii) Non-availability of Data Curators, (iv) Absence of proper and well-articulated regulations on Data Management Ethics/Data Protection, etc.

However, it should be noted that some of these problems are beyond the capacity of the Clusters. We must anticipate and prevent them from impeding the realisation of the noble objectives of the Cluster initiative. The directors of the centres must begin to work round some administrative challenges in their domains and secure the support of university administrators and policymakers. Industry players, companies and organisations interested in sponsoring Africa-based cultural research may also be educated about the initiative, with their support solicited in relevant areas. We must equally encourage owners of private museums, galleries and libraries holding precious manuscripts, images, pictures, audio recordings and traditional performances to support the vision of the Africa Multiple Cluster. They must be assured of joint ownership of the data and the research outcomes.

It also bears reasserting that the absence of well-funded projects on digitisation of data and materials from Africa exposes the continent's researchers to the danger of losing generations of collective memory and information that can influence the present and future generations. The University of Bayreuth's Africa Multiple Cluster of Excellence, with its emphasis on creating a viable Digital Research Environment to technologically drive the projects, is thus a welcome, timely initiative that will change the narrative of African Studies.

The urgency of the digitisation projects is highlighted in the story narrated below. Olayinka (2019 n.p.) writes on his work on the national archives and concludes that "Nigeria's memories are brittle and falling apart without digitization." He further reports his experience with a member of staff at the National Archive, Abuja:

> A senior official with the National Archives of Nigeria who spoke… lamented the poor state of collections and manuscripts in the holdings. He said: "My heart bleeds as an archivist when I see the deterioration of these archive materials. "Some are brittle and disintegrating before our eyes, and there's not much we can do due to lack of funds," this official explained that funding from the federal government is very limited, and acknowledged that digitization would be a very effective way of preserving some of the endangered materials.
>
> They also pointed out that physical infrastructure needs improvement. "Our cooling units don't work," they complained, "most of the materials are not being kept under the right temperature to preserve them." (Olayinka 2019: n.p.) Most scholars agree that digitazation projects must play a key role in preserving archival and manuscript heritage.

9 Conclusion

As rightly cited in the report 'Innovation with an Impact: The Digital Africa Initiative: 2015-2017', produced by the German Federal Ministry for Economic Cooperation and Development, "…modern information and communication technologies (ICT) can help make processes more efficient, quicker and more cost-effective. ICT can link people who would usually not come into contact with each other. Solutions can emerge that previously seemed unthinkable…. (n.d, n.p)"

It is therefore imperative that the Clusters must use digital technologies and methods to develop and strengthen cooperation across all sectors and countries. Digital technologies may also be developed to address specific needs within each Cluster.

As optimistic as we are, we must bear in mind that ethical considerations are contested issues in digitalisation initiatives all over the world. Some of the questions we must begin to address at this early stage will be the following: How will it provide inclusive and harmonious development of the African and international information society? How will this initiative be sustained? In what ways can African scholars be trained on a more consistent basis to develop tailor-made applications and technologies that can best tackle the local challenges that are peculiar to different research and cultural contexts?

Zhu (2018) in Digital Maturity, suggests that "digital transformation is a long journey, and the path for digital transformation can be iterative, evolutionary, revolutionary, or disruptive" (cited in Gbadegesin 2019: 49). Seeseman, and Fendler (ibid. n.p) more aptly emphasise this point on the key role that digital research component will play within the Afican Multiple Cluster's projects. They assert that the application of relevant technologies stands at the centre of the Cluster's quest to facilitate the three key concepts of 'multiplicity', 'relationality' and 'reflexivity'. Technology has the capacity to open up new perspectives on African ways of life, to unearth new discoveries on old materials, to facilitate exchange of ideas and mobility of researchers and to support collection and access to data. It will serve as a catalyst to equip researchers in Africa and on African Studies and its various disciplinary multiplicities to engage in innovative and translational research projects.

As stated elsewhere,

> digital research environment (DRE) integrated into the Africa Multiple Cluster of Excellence will be of great use for the necessary reorientation of African Studies in a number of ways. Heterogeneous sets of research data are brought together here in such a way that they can be linked and used in a variety of contexts (…). The openness of the DRE to data from different sources and formats, and the transdisciplinary cooperation of participating research institutions can make an important contribution to the democratization and decolonization of knowledge about Africa. (Ope-Davies 2019: 9)

With the introduction of DRE and the application of relevant technologies, we have begun a steady and sure journey to truly reconfigure African Studies in all its multidisciplinary, multi-theoretical and multinational and multidimensional thrusts.

This conclusion is further solidified by the following observations that this study wishes to present. One, the continent's dynamic and resilient communities, which are rich in cultural heritage and oral traditions, have not yet been fully digitised. Two, the continent has the largest and biggest bits of unexplored data in virtually every sphere of human endeavour. These huge sets of data are steeped in rich histories, languages, civilisations, epistemes, philosophies, native intelligence, art and culture, oral traditions, as well as narratives and oral performances that can enrich global scholarship and become a catalyst for unimaginable innovations, etc. Three, Africa has the largest untapped human resources endowed with natural talents and creative abilities. Four, Africa's huge population—including in the diaspora—remains a great advantage for researchers exploring any area of African Studies. Five, African students and scholars all over the world are one of the most hard-working groups of people who can serve as research assistants, data curators and data analysts as well as research partners.

The systematic and critical deployment of technologies in African Studies, as is gradually being encouraged and undertaken through the Bayreuth Africa Multiple Cluster of Excellence's initiative, is a timely and welcome development. It will help in accelerating the process of exploring, curating, analysing, documenting and preserving these huge data sets in African indigenous knowledge/epistemologies and cultural studies, as well as in making them more easily available and accessible to the global scientific, social and research communities.

10 Acknowledgements

(i) I thank the anonymous reviewer(s) and Ayodeji Adedara (mentee) for the comments and support.

(ii) The Alexander von Humboldt Foundation, Germany is equally acknowledged for the support provided towards the establishment of the Centre for Digital Humanities, University of Lagos under its Alumni Sponsorship and Network programmes.

11 References

Adeleke, Akinniyi A. Digitization Capacity and Skills of Academic Librarians in Nigeria. https://wiredspace.wits.ac.za/jspui/bitstream/10539/20521/1/ICADLA%203%20Akinniyi%20A.%20Adeleke.pdf. (last accessed 31 May 2022).

ADHO, 2019. Alliance for Digital Humanities Organisations: Digital Humanities 2019: "Complexities" Call for papers. https://staticweb.hum.uu.nl/dh2019/dh2019.adho.org/call-for-papers/cfp-english/index.html. (last accessed 31 May 2022).

Ahnert, Ruth. Digital Landscapes: Understanding Archival Representation with Digital Methods. https://www.cdh.cam.ac.uk/events/digital-landscapes-understanding-archival-representation-digital-methods. (last accessed 16 August 2021).

Asogwa, Brendan. (2011). Digitization of Archival Collections In Africa For Scholarly Communication: Issues, Strategies, and Challenges. Library Philosophy and Practice (e-journal.) https://digitalcommons.unl.edu/libphilprac/. (last accessed 16 August 2021).

Broadley, Ashley. 2018. A Journey Into Big Data with Apache Spark: Part 1 Part 1. The first in a series of posts about getting know Apache Spark for Big Data Processing. https://towardsdatascience.com/a-journey-into-big-data-with-apache-spark-part-1-5dfcc2bccdd2. (last accessed 31 May 2022).

Christensen, Clayton. 2016. The Clayton M Christensen Reader. Harvard Business Review.

Gbadegeshin, Saheed A. 2019. The Effect of Digitalization on the Commercialization Process of High-Technology Companies in the Life Sciences Industry. Technology Innovation Management Review 9 (1): 49–63. https://doi.org/10.22215/timreview/1211.

Javed, Aisha. 2016. Internet of Things: A Booming Connectivity For Africa. https://www.xorlogics.com/2016/08/14/internet-of-things-a-booming-connectivity-for-africa/. (last accessed 31 May 2022).

Keating, John. 2012. The Importance of Digital Humanities. An Foras Feasa The Institute for Research in Irish Historical and Cultural Traditions.

Levin, I.; Mamlok, D. 2021, Culture and Society in the Digital Age. Information 12, 68. https://doi.org/10.3390/info12020068.

Lubar Steven. 2014. Applied? Translational? Open? Digital? Public? New models for the humanities. https://stevenlubar.wordpress.com/2014/06/05/applied-translational-open-digital-public-new-models-for-the-humanities/. (last accessed 31 May 2022).

Markham, Ania. 2021. Doing Ethnographic Research in the Digital Age. In Leavy, Patricia(ed) Filed of Qualitative Research. London: Oxford University Press (Pre-print version: https://doi.org/10.31235/osf.io/hqm4g.

McKay, Sally. 2003. Digitization in an Archival Environment. E-JASL: The Electronic Journal of Academic and Special Librarianship (31): 1–12.

Mkono, Mucha. 2018. Changing the African Narrative Through Social Media Platforms. https://theconversation.com/changing-the-african-narrative-through-social-media-platforms-97097. (last accessed 31 May 2022).

Mohammed, Abubakar. 2013. Institutional Digital Repository: An Option for Scholarly Communication in Nigeria. International Journal of Education and Research 1 (6): 1–10.

Molteno, Robert. 2016. Digital Repositories: Making Africa's Intelligentsia visible? Bulletin of the National Library of South Africa 70 (2).

Morze, Natalia V. and Strutynska, Oksana V (2021). Digital transformation in society: key aspects for model development. Journal of Physics.: Conference Series, 1946 01 . https://doi.org/10.1088/1742-6596/1946/1/012021.

Olayinka, Wole. 2019. National Archives: Nigeria's memories are brittle and falling apart without digitisation. https://techcabal.com/2019/08/20/national-archives-nigerias-memories-are-brittle-and-falling-apart-without-digitisation/. (last accessed 31 May 2022).

Ope-Davies, 'Tunde. 2019. New synergies for African Studies: The Digital Research Environment (DRE) of the Cluster of Excellence. Spektrum. The Science Magazine of the University of Bayreuth. https://epub.uni-bayreuth.de/5802/1/issue-02-2019.pdf. (last accessed 31 May 2022).

Rademacher, Ute, and Terri Grant. 2019. Out of Africa: A new perspective on digitalisation in Africa. TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis) 28 (2): 41–47. https://doi.org/10.14512/tatup.28.2.s41.

Research Data Management: Benefits of Data Management, n.d. Leabharlann UCD. UCD Library https://libguides.ucd.ie/data/benefits. (last accessed 31 May 2022).

Rogers, Richard, 2015. Digital Methods for Web Research. In Robert Scott and Stephan Kosslyn. (eds) Emerging Trends in the Social and Behavioral Sciences. Amsterdam: John Wiley & Sons, Inc. 1-22

Schelenz, Laura, and Kerstin Schopp. 2018. Digitalization in Africa: Interdisciplinary Perspectives on Technology, Development, and Justice. International Journal of Digital Society 9 (4): 1412–1420.

Seeseman, Rüdiger and Fendler, Ute (2019). A new chapter in Africa-related research: Interview with the two chairpersons of the new Cluster of Excellence for African Studies at the University

of Bayreuth. https://epub.uni-bayreuth.de/5802/1/issue-02-2019.pdf. (last accessed 31 May 2022).

Shapshak, Toby. 2015. African Internet Growth Continues to Outstrip the World. https://www.forbes.com/sites/tobyshapshak/2015/09/02/african-internet-capacity-growth-continues-to-outstrip-the-world-2/. (last accessed 31 May 2022).

Spektrum 2019. Spektrum. The Science Magazine of the University of Bayreuth. https://epub.uni-bayreuth.de/5802/1/issue-02-2019.pdf. (last accessed 31 May 2022).

Tsebe, John. August 14th - 18th 2005. Networking cultural heritage: Africa. Oslo, Norway.

Weller, Martin. 2011. The Digital Scholar: How Technology Is Transforming Scholarly Practice. London: Bloomsbury Academic.

Wellman, B. 2001, Physical place and cyberplace: The rise of personalized networking. Int. J. Urban Reg. Res. 2001 (25): 227–252.

Zhu, Pearl, 2018. Digital Maturity: Take a Journey of a Thousand Miles from Functioning to Delight. Goodreads.

Webography

ADHO.https://staticweb.hum.uu.nl/dh2019/dh2019.adho.org/call-for-papers/cfp-english/index.html. (last accessed 31 May 2022).

Benefits of Digital Humanities. Duke Trinitiy College of Arts and Sciences. Digital Humanities. Available at: https://digitalhumanities.duke.edu/benefits-digital-humanities. (last accessed 31 May 2022).

About Digital Research. University of Notthingham. https://www.nottingham.ac.uk/jobs/currentvacancies/digital-research-opportunities/about-digital-research.aspx. (last accessed 31 May 2022).

German Federal Ministry for Economic Cooperation and Development: Innovation with an impact: The digital Africa initiative. First Report 2015-2017. https://hobbydocbox.com/Art_and_Technology/73867560-Innovation-with-an-impact-the-digital-africa-initiative.html. (last accessed 31 May 2022).

Research Data Management: Benefits of Data Management, https://libguides.ucd.ie/data/benefits. (last accessed 31 May 2022).

Digital Colonialism and Infrastructure-as-Debt

Michael Kwet (Yale Law School / Centre for Social Change, University of Johannesburg)[22]

Abstract

This article takes a look at the issue of digital colonialism in South Africa. First, it provides a snapshot of the global situation, characterized by the dominance of US empire. Second, it provides a snapshot of digital colonialism in South Africa. This includes a breakdown of how foreign corporations dominate some industries, products, and services, as well as how they invest in infrastructure to capture market share for long-term dominance – what we can call *infrastructure-as-debt*. Third, it looks at the concept of tech hegemony – neomissionary doctrines pushed through proponents of the so-called Fourth Industrial Revolution and the reformist orientation of the US "techlash". Finally, it outlines the broad contours of a "People's Tech for People's Power" agenda and a socialist Digital Tech Deal required to see it through.

1 Digital colonialism: a Brief Snapshot

Digital colonialism is the use of digital technology for political, economic, and social control of a foreign territory or nation (Kwet 2019a). It is principally achieved through ownership and control of the digital ecosystem – software, hardware, and network connectivity – which is then designed by the owners for the purposes of profit and plunder, including the extraction of rents and data. It is supplemented by and integrated into more traditional sources and circuits of power, such as financial wealth, the domination of legal systems, and labor exploitation along global commodity chains (Starrs 2013; Rikap 2021).

The United States is the central aggressor. It exercises hegemony over the global society by imposing the capitalist mode of production and distribution as a world system. US-based

[22] Contact: Michael.Kwet@yale.edu, https://orcid.org/0000-0002-3304-5649.

transnational corporations dominate most industries and functions in the digital ecosystem, including search engines [Google]; web browsers [Google Chrome]; smartphone and tablet operating systems [Google Android, Apple iOS]; desktop and laptop operating systems [Microsoft Windows, macOS]; computer chips [Intel, AMD, Nvidia]; office software [Microsoft Office, Google Workspace]; cloud infrastructure and services [Amazon, Microsoft, Google]; social networking platforms [Facebook, Twitter]; transportation [Uber, Lyft]; business networking [Microsoft LinkedIn]; streaming entertainment [Google YouTube, Netflix, Hulu]; and online advertising [Google, Facebook] – among other products and services.

Of the *frightful five* Big Tech corporations, sometimes called GAFAM (Alphabet/Google, Amazon, Meta/Facebook, Apple, Microsoft), three – Google, Meta, and Apple – generate more revenue outside of US borders than inside (Johnson, 2021; Jain 2020; Vailshery 2021), while Microsoft generates just over half its revenue inside the US (Microsoft Corporation 2021). Amazon is the only one to generate a substantial majority inside the US (Rathore 2019), and even then, its AWS cloud division takes 33 percent of the global cloud market (Richter 2020) – a core component of the digital economy.

China is the nearest competitor, but its tech corporations are nationally contained and dominant inside Chinese borders (Starrs 2018). Outside of the mainland, China has substantial market share in a few areas, such as 5G [Huawei], CCTV cameras [Hikvision and Dahua], smartphone handsets [Huawei, Xiaomi], and smart cities contracts, with market share varying by country. Yet on the whole, China is well behind the US, posing little threat to American dominance in the global digital economy.

2 Foreign Domination in the South African Market

Within this context, South Africa is primarily subject to US dominance. While President Cyril Ramaphosa embraces 5G technology from China, in South Africa, the core technologies listed above are still predominately owned by the United States and other countries in the North. Even where products are Chinese, the more interesting parts are often from the US or other global North countries. For example, in Johannesburg, a corporation called Vumacam announced a mass CCTV surveillance network to "blanket" Johannesburg and other metropolitan areas with 100,000+ cameras. The network can be used by private security guards to watch over neighborhoods and provide footage to police in the event of a crime (Kwet 2019c).

The most interesting parts are the software systems that make their system a "smart network" based on video analytics: iSentry for "unusual behavior detection" derived from the Australian military, Israeli-based BriefCam for searching through video footage using advanced analytics, Denmark-based Milestone Systems for the Video Management System that organizes the network into a coherent whole. Ricky Croock, the CEO of Vumacam, uses his own companies for storage and automatic number plate readers - Calabash Systems and Plate ID, respectively. The only Chinese component is the Hikvision "dumb" cameras that do not presently perform video analytics. Similar component parts exist in other projects for the City of Cape Town and the City of Johannesburg CCTV networks (Kwet 2021a).

A similar dynamic exists for Chinese smartphones. According to StatCounter (2021), the big three smartphone providers in South Africa are Samsung (48%), Huawei (16%), and Apple (15%). However, for Samsung phones, Google Android is the operating system. Huawei also used Google Android until President Donald Trump forced Google in 2020 to stop working with Huawei. This forced them to replace Google Android with a *non-Google* version of Android called Harmony OS, and it remains to be seen how it will impact sales (Amadeo 2021). Nevertheless, to date, Google, Apple, and other US corporations dominate the software, which gives them access to user data and shapes their user experiences around American firms.

American transnationals also dominate other major products and services. Forty percent of South Africans are active social media users (Lama 2020), while Google services are dominant in advertising, smartphones, and its major app offerings. In 2017, Anton Harber reported at the Financial Mail that Google and Facebook are the "nemeses" of the South African press. Google, Harber explained, takes 70% of local online advertising, and social media – led by Facebook – another 12%. South Africa's major media groups take just 8% of the pie (Harber 2017). This follows broader global trends where Big Tech giants undermine local media by sucking away the advertising revenue. Given that local media doesn't have an alternative model in place, they are being starved of revenues. Over the last ten years, the number of community newspapers has decreased from 585 to 200, in large part due to the power of Big Social Media (Daniels 2020). During 2020, several news outlets, including City Press and News24, began locking up content to try to force audiences to pay for access.

In the area of transportation, Uber has major inroads into the local taxi market, beginning in 2013. The controversial firm relies on a deregulatory and anti-competitve business model built to drive competitor services out of business. The company has operated at a loss for most of its existence; after cornering the market, they can drive prices back up and eventually turn a profit. Working conditions are often brutal. In South Africa, many exposés tell stories of drivers working for long hours – often sleeping in their cars – for little pay. By 2016, "taxi wars" broke out between metered taxi drivers and Uber drivers. The results were tragic: a man in Manenberg was attacked by assailants, and his car set alight (Abbas 2016). A year later, Lindelani Mashua, an Uber taxi driver, was burned to death, ostensibly by competitor taxi drivers.

Other industries, such as entertainment, are likewise being colonized by foreign tech giants. In December 2017, MultiChoice announced that 115,000 Premium subscribers defected to other services, with many probably switching to Netflix. The subscriber base grew from 44,000 in 2016 to 338,000 by mid-2020 (Statista 2021). By Q1 2019, Netflix took 68% of subscription video-on-demand (SVOD) originals in South Africa (Parrot Analytics 2019). The behemoth is also commissioning original African shows through its African Originals division (Kazeem 2018; Vourlias, 2020). Apple Music, Google Play Music, Spotify (Sweden), and Deezer (France) are leading players in the streaming music space (Vermeulen 2017; 2018), while YouTube generates massive amounts of traffic in the country.

3 Infrastructure-as-Debt

In the colonial era, Europeans seized the land, exploited laborers, developed and monopolized advanced industrial processes and infrastructure, and broadly exercised economic and military

domination over indigenous peoples across the world, for the purpose of profit and plunder. Yet this form of rule did not last; it evolved and adapted to broader changes in the global society. Over the course of the twentieth century, countries in the South successively threw off the yoke of formal colonization. As direct European rule began to recede in the face of popular resistance, a new form of colonialism, neocolonialism, replaced it. As Kwame Nkrumah wrote, "In place of colonialism, as the main instrument of imperialism, we today have neo-colonialism" whereby the economic systems of nominally sovereign states in the South are "directed from the outside" (1966: ix). "The result of neo-colonialism," Nkrumah explained, "is that foreign capital is used for the exploitation rather than for the development of the less developed parts of the world. Investment under neo-colonialism increases rather than decreases the gap between the rich and the poor countries of the world" (Nkrumah 1966: x).

Within the neocolonial context, debt plays a critical role reinforcing inequality and the North/South divide (Ndlovu-Gatsheni 2021). In response to the 1970s oil crisis, poor countries borrowed many billions of dollars from lenders in the North. When they couldn't pay it back, the North restructured the IMF to demand repayments rather than forgive the debt, provoking the Third World Debt Crisis. To help alleviate debt payments, the IMF imposed Structural Adjustment Programs – a set of policies which compelled slashes to social spending, the privatization of public services, and forms of deregulation beneficial to the creditors. Instead of offering to write off the debts, the rich countries undercut developmentalist reforms and commanded a massive transfer of wealth from ordinary people in the South to rich elites in the North. Today, Jason Hickel notes, "poor countries pay over $200 billion each year on interest alone to foreign creditors, much of it on old loans that haven been already paid off many times over, and some of it on loans accumulated by greedy dictators" (Hickel 2018: 25).

The digital era offers a similar dynamic. In the digital economy, colonial powers own and control digital property in the form of knowledge – such as intellectual property, data, platforms, and digital intelligence – and infrastructure, in the form of software, hardware, and internet connectivity. Just as the settlers once built railroads for economic and military purposes, today the tech colonizers build digital infrastructure and fence off knowledge for rent extraction, surveillance, and imperial control. Just as loans have been a source of power that perpetuate domination and dependency on the whims of foreign lenders, offering up foreign-owned digital infrastructure is now a means to impose the domination of Big Tech in key emerging markets, locking the South into their products and perpetuating dependency.

In neo-apartheid South Africa, Microsoft has been a market leader with business deals across the country ranging from contracts with provincial educational departments to training programs on Microsoft software. In March 2017, Google announced that it had reached its goal to train one million young people in digital skills across Africa, but only 70,000 were trained in South Africa, compared to 450,000 in Nigeria and 400,000 in Kenya. Google said the low number in South Africa "is due to the lack of buy-in from government and universities" (Rawlins 2017). Disappointed with the figures, the head of policy and government for Google SA, Fortune Mgwili-Sibanda, said, "[w]e are no longer going to knock at the door, we are going to bang" (Ventkess 2017).

Historically, South African policymakers tried to stop foreign Big Tech corporations from "banging at the door" and taking over their economy. In the early 2000s, South African policymakers

understood that the digital revolution was rapidly advancing, and they commissioned a series of reports to assess the ideal software policy for the country. Working groups were formed, and several reports spanning hundreds of pages concluded that South Africa should adopt Free and Open Source Software (FOSS) in the public sector, in order to develop an open tech ecosystem based on the sharing of knowledge and software code rather than proprietary ownership (see Kwet 2019b: 179-229). In 2007, the South African Cabinet passed a Free and Open Source Software (FOSS) policy preference mandating the use of FOSS in the public sector – including schools – so long as no considerably superior proprietary software components were needed.

The FOSS policy preference was formulated during a time period where a lot of software was still executed inside individual computers rather than on the cloud. The growth and shift towards cloud-based services owned and controlled by giant transnational corporations changed power relations in the tech ecosystem and effectively nullified the freedoms that copyleft free and open source software licenses granted users for non-cloud software. The free software community tried to adapt to the change by creating the Affero GNU General Public License (AGPL), which mandated the disclosure of source code when using the software for a network service, but it did not catch on widely enough to prevent an enclosure and exploitation of the commons by Big Tech.

South Africa's FOSS policy preference was never widely implemented, and the tech ecosystem changed to coalesce around Big Data surveillance and platform ownership. By July 2016, the City of Johannesburg (2016a) announced an R200 million, five-year deal with Microsoft "to empower the poor". Then Mayor Parks Tau notified the public that "800 000 youths aged between 18 and 34 and 200 000 residents aged 35 or older would be provided with free training to equip them with computer skills needed for entry-level jobs" (City of Johannesburg 2016a). Computer skills include training on Microsoft Office and are available without charge to the participants.

Big Tech also made inroads into the university and startup space. In September 2016, the Wits University-owned Tshimologong Precinct, dubbed "Johannesburg's own Silicon Valley in Braamfontein", opened for business, with Microsoft one of its founding partners (City of Johannesburg 2016b). A year earlier, IBM undertook a ten-year program of investment in partnership with the Department of Trade and Industry (DTI). The investment, priced at R700 million (about $47 million dollars), included a brand-new Research Africa facility in the Tshimologong Precinct in Braamfontein, Johannesburg. The Microsoft App Factory, an apprenticeship program that provides "deep training" on how to develop apps for Microsoft software, is housed in Tshimologong. Cisco, for its part, added R12 million in the Precinct, and J.P. Morgan provided financial support. South African investors included Data Centrix and Telkom, among others (see Kwet 2019b: 239-243).

A leading founder of the Tshimologong Precinct, Barry Dwolatzky, also ran the Joburg Centre for Software Engineering (JCSE), for which Microsoft was a founding partner. According to the City of Johannesburg (2009), JCSE collaborations with Microsoft were "aimed at instilling business principles" in students and developing innovation for business. In 2020, Dwolatzky and Mark Harris, CEO of ICT networking firm Altron Nexus, wrote, "The current international software scene is dominated by large multinational platform vendors such as Microsoft, Google, IBM and Amazon. There is the potential for domestic software developers to create applications which run on these platforms" (Dwolatzky and Harris 2020).

These investments are a form of infrastructure colonization whereby state-of-the-art facilities are built for a resource-strapped country in need of help to build their digital economy. As a quid-pro-quo, centers like the JCSE and Tshimologong are embracing the Big Tech colonizers that the FOSS policy preference intended to resist. There is a conflict of interest in taking money from Big Tech, as well: conversations about tech ethics and digital colonialism – as well as organized actions against them – will almost certainly be put off the table, as the tech corporations have no reason to support programs that criticize them and subvert their dominance.

4 Tech Hegemony

Another feature of digital colonialism is *tech hegemony*, whereby those with power are so dominant that they set the agenda, preferences, and parameters of thinking in those they dominate. As scholars like Edward Herman and Noam Chomsky (2002) have pointed out, it is *unthinkable* for the US intellectual classes to consider that their own country is a "rogue state" which is guilty of war crimes and exploits the global population through military and economic force. The mainstream "left" and right concur that that US military partakes in "humanitarian" interventions, even if they sometimes "strategically blunder" in their decisions to wage war on the poorer nations. The notion that the US would invade Iraq in order to control their oil is beyond the bounds of acceptable thought.

In the digital domain, the hegemonic perspective features a limited spectrum that serves to manufacture consent. The tech industry and its supporters form the most conservative position, arguing that humanity benefits from Big Data, intellectual property, centralized clouds, the "internet of things", "smart" cities littered with surveillance, automation, algorithmic decision-making, Big Tech corporations, and surveillance capitalism. The best we can do is put a few rules in place to constrain its excesses, but there is no stopping it.

The mainstream of thinking in South Africa embraces this position. For years, political conversations about digital technology were framed around the so-called "Fourth Industrial Revolution" (4IR), a phrase coined by World Economic Forum (WEF) leader Klaus Schwab (2016). The concept serves a useful purpose at the periphery of empire: it steers inquiry about tech into the WEF agenda. Reflecting the interests of its elite board of trustees, its ideology privileges the private sector and treats the current digital order as if it is the only one possible.

At the policy level, the African National Congress embraces the 4IR (Kwet 2019b). It now has an official 4IR Commission, and it seeks to develop the above-listed technologies in partnership with foreign (primarily US-based) tech giants, local corporations like MTN, Multichoice, and Naspers, who are seeking to get in on the action, and universities like the University of the Witwatersrand, the University of Cape Town, and the University of Johannesburg.

In the education sector, Microsoft is especially dominant. It has extensive influence in the training of teachers to use technology on Microsoft software, and it has contracts with at least five provinces at the Basic Education level. Google is trying to catch up and score its own contract (Kwet 2019b: 190). The implementation of corporate software in schools form part of the agenda to capture market share and indoctrinate students into accepting the Silicon Valley's framing of digital capitalism.

The South African startup sector also embraces the 4IR doctrine. While the tech sector is not particularly booming – the most "successful" startup to date is GetSmarter, which was acquired by US-based 2U for $103 million in 2017 – the sector replicates the Silicon Valley model of digital capitalism. FinTech service providers are widespread, and some of them, such as micro-loans provider JUMO and insurance-provider Fo-Sho, use surveillance to regulate the behavior of users.

Pushback against the exploitative practices of Big Tech is just beginning to bubble among the South African press, and will likely accelerate in due time. Yet there is a major question about what form this will take.

In the United States, tech criticism has, to date, predominantly revolved around reforms that leave private property, capitalism, and US state-corporate hegemony intact (Kwet 2019b). The *tech ethics* conversation, sometimes called the *techlash*, is dominated by an elite set of intellectuals situated within high-ranking universities in the North (the Ivy League plus a few other universities like Oxford and Cambridge), rich (mostly corporate) media outlets, think tanks, NGOs, and even the Big Tech corporations themselves. The non-profits and university departments producing the ethics research are typically funded by Big Tech, wealthy foundations, and rich philanthropists (Katz 2020).

The solutions of this American School of tech ethics revolve around antitrust designed for competitive capitalism (Kwet 2020b, 2022); content moderation that fails to challenge capitalist social media and real-world socialist alternatives; a small set of bans on things like facial recognition and predictive policing (often after activists make the issues politically popular); the unionization of imperialist corporations; more representation and better treatment of marginalized workers within Big Tech; and weak privacy laws like the European Data Protection Regulation that have, to date, failed to curb mass state-corporate surveillance.

Taken together, these reforms are much like the Sullivan Principles under apartheid. The Sullivan Principles, initiated by Reverend Leon Sullivan of General Motors, was a set of six principles for US corporations operating in apartheid South Africa that would allegedly improve the treatment of black workers. They included desegregation of the workplace, fair employment practices, equal pay for equal work, job training and advancement within the corporate hierarchy, and improvement of workers' lives outside of the work environment.

The Sullivan Principles were embraced by corporate America and government leaders. Anti-apartheid activists and intellectuals opposed them, arguing that the Principles were designed to keep US profits flowing from South Africa and extend the life of apartheid (Larson 2020). The reforms threatened to blunt the antiapartheid movement's more aggressive strategies, such as Boycott, Divestment, and Sanctions (BDS), designed to strangle and bring about the hasty demise of the indefensible apartheid *system* (Schmidt 1980).

In today's context, the need to replace the capitalist *system* is an emergency, in large part because it is speeding the world towards ecological collapse (Satgar 2018; Bond 2019; Hickel 2020). Options to replace intellectual property with a knowledge commons, socialize private ownership of the means of computation, and beat back an American-centered tech empire are seldom addressed by the "techlash" intellectuals, let alone centered in the conversation, despite the

centrality of these issues. These solutions are needed to drastically reduce global inequality and fit into broader struggles for justice, equality, and environmental sustainability (Kwet 2022; Baker 2016, 2018; Perelman 2002).

5 Solutions: People's Tech for People's Power, a Digital Tech Deal, and Direct Action

If we're going to defeat digital colonialism, a new set of Sullivan Principles for "responsible" Big Tech will not work. This time around, we need to replace digital capitalism with a digital socialism model initiated from below (Hind 2019; Hanna and Lawrence 2020; Liu 2020; Muldoon 2022; Kwet 2022).

During the 1980s, the anti-apartheid movement formed a "People's Education for People's Power" movement to counter the Bantu education system. In the 21st century, a "People's Tech for People's Power" movement could counter the system of digital colonialism, rooted in a digital capitalism which forms a world system under the control of US empire (Kwet 2020a).

A People's Tech model envisions a digital ecosystem based on free software and internet decentralization, supported by socialist legal solutions, critical education, grassroots movements, and bottom-up democracy. The private ownership of knowledge as private property would be abolished, in order to equalize access to knowledge and technology transfer from rich to poor. Software would be licensed as a copyleft digital commons to ensure anyone can use, study, modify, and distribute software freely. Wherever possible, internet services would be decentralized. Internet infrastructure would be under the democratic ownership of local and international communities. Strong privacy laws would ban the existence of mass surveillance, be it through end user devices or in public and work spaces. Technology corporations – and corporations themselves – would be phased out of existence. Workers would be protected against exploitation, with wealth and income re-distributed within and across borders. Privacy-by-design would rein in Big Data and secure privacy, while communities would determine how to share and control the data that is legitimate to collect. The poor would be provided access to fast internet, data, and quality devices so they can participate on equal footing in the global digital society.

To make this a material reality, we need a socialist Digital Tech Deal that dovetails with a Green and Red Deal (Kwet 2021a, 2022; The Red Nation 2020; Ajl 2020). The financial, intellectual, and material wealth currently monopolized by the rich and powerful would be seized by the masses and redistributed broadly to stabilize the world situation. A #BigTechBDS movement targeting things like research support from Big Tech and public sector contracts could help strangle the largest and most destructive corporations. Such a movement would fit with a Digital Tech Deal by providing a just transition for tech workers as the tech giants are dissolved and the digital society transformed along socialist principles.

The need for a People's Tech movement is urgent. At the moment, we are facing ecological collapse due to both global heating *and* material overconsumption. According to environmental scientists and environmental economists, the middle and upper classes must reduce their material footprint, the global poor must grow their material well-being, and humanity must enter into a steady-state economy – within a few decades (Hickel and Kallis 2020). Because digital capitalism

and colonialism creates inequality, perpetuates growth, and pushes frivolous consumerism, it is fundamentally incompatible with 21st century economics and social justice.

6 References

Abbas, Gadeeja. 2016. Attack on Uber driver a warning. https://www.iol.co.za/news/attack-on-uber-driver-a-warning-2044206 (last accessed: 22 December 2021).

Ajl, Max. 2021. *A People's Green Deal*. London: Pluto Press.

Amadeo, Ron. 2021. Huawei's Harmony OS: "Fake it till you make it" meets OS development. https://arstechnica.com/gadgets/2021/02/harmonyos-hands-on-huaweis-android-killer-is-just-android (last accessed: 22 December 2021).

Baker, Dean. 2016. Rigged: How Globalization and the Rules of the Modern Economy Were Structured to Make the Rich Richer. https://deanbaker.net/images/stories/documents/Rigged.pdf.

Baker, Dean. 2018. *Is Intellectual Property the Root of All Evil? Patents, Copyrights, and Inequality*. Working Paper. The Great Polarization: Economics, Institutions, Salt Lake City, UT. https://cepr.net/report/is-intellectual-property-the-root-of-all-evil-patents-copyrights-and-inequality/ (last accessed: 20 June 2022).

Bond, Patrick (ed.). 2019. *Degrowth, Devaluation, and Uneven Development from North to South.* In Ekaterina Chertkovskaya, Alexander Paulsson, and Stefania Barca, Eds., 2019, *Towards a political economy of degrowth*, 137-156, Rowman & Littlefield International.

City of Johannesburg. 2009. Course boosts ICT students' skills. https://web.archive.org/web/20170319025719/https://joburg.org.za/index.php?option=com_content&view=article&id=4155&catid=118&Itemid=199 (last accessed: 20 June 2022).

City of Johannesburg. 2016a. City, Microsoft in R200m deal to empower the poor. https://web.archive.org/web/20170422224245/https://joburg.org.za/index.php?option=com_content&id=10741&Itemid=266 (last accessed: 20 June 2022).

City of Johannesburg. 2016b. Tshimologong hub open for hi-tech business. https://web.archive.org/web/20170422230138/https://joburg.org.za/index.php?option=com_content&id=10790&Itemid=266 (last accessed: 20 June 2022).

Daniels, Glenda. 2020. Power and Loss in South African Journalism: News in the age of social media. https://www.dailymaverick.co.za/article/2020-08-24-power-and-loss-in-south-african-journalism-news-in-the-age-of-social-media (last accessed: 22 December 2021).

Dwolatzky, Barry and Mark Harris. 2020. Fast tracking the critical task of building South Africa's digital economy. https://www.wits.ac.za/future/stories/fast-tracking-the-critical-task-of-building-south-africas-digital-economy.html (last accessed: 22 December 2021).

Hanna, Thomas M. and Mathew Lawrence. 2020. Ownership futures: towards democratic public ownership in the 21st century. https://www.common-wealth.co.uk/reports/ownership-futures-towards-democratic-public-ownership-in-the-twenty-first-century (last accessed: 22 December 2021).

Harber, Anton. 2017. Nemesis of the Press. https://www.businesslive.co.za/fm/features/cover-story/2017-11-16-how-google-and-facebook-are-the-biggest-threat-to-south-african-news-media/ (last accessed: 22 December 2021).

Herman, Edward S. and Noam Chomsky. 2002. *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.

Hickel, Jason. 2018. *The Divide: Global Inequality From Conquest to Free Markets.* New York: W.W. Norton & Company.

Hickel, Jason. 2020. *Less is More: How Degrowth Will Save the World.* London: William Heinemann.

Hickel, Jason, and Giorgos Kallis. 2020. Is Green Growth Possible? *New Political Economy* 25 (4): 469–486. https://doi.org/10.1080/13563467.2019.1598964.

Hind, Dan. 2019. The British Digital Cooperative: a new model public sector institution. https://thenextsystem.org/bdc (last accessed: 22 December 2021).

Jain, Virang. 2020. Facebook: Revenue by Region 2020. https://businessquant.com/facebook-revenue-by-region (last accessed: 22 December 2021).

Johnson, Joseph. 2021. Revenue distribution of Alphabet from 2015 to 2020, by region. https://www.statista.com/statistics/266250/regional-distribution-of-googles-revenue (last accessed: 22 December 2021).

Katz, Yarden. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence.* New York, Chicester, West Sussex: Columbia University Press.

Kazeem, Yomi. 2018. Netflix is finally investing in original African content. https://qz.com/africa/1482172/netflix-will-commission-original-african-shows (last accessed: 22 December 2021).

Kwet, Michael. 2019a. Digital colonialism: US Empire and the new imperialism in the Global South. *Race & Class* 60 (4): 3–26. https://doi.org/10.1177/0306396818823172.

Kwet, Michael. 2019b. Digital Colonialism: South Africa's Education Transformation in the Shadow of Silicon Valley. PhD dissertation, Rhodes University, Grahamstown, South Africa. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3496049 (last accessed: 22 December 2021).

Kwet, Michael. 2019c. Smart CCTV Networks Are Driving an AI-Powered Apartheid in South Africa. https://www.vice.com/en/article/pa7nek/smart-cctv-networks-are-driving-an-ai-powered-apartheid-in-south-africa (last accessed: 22 December 2021).

Kwet, Michael. 2020a. A Digital Tech New Deal to break up Big Tech. https://www.aljazeera.com/opinions/2020/10/26/a-digital-tech-new-deal-to-break-up-big-tech (last accessed: 22 December 2021).

Kwet, Michael. 2020b. Social Media Socialism: People's Tech and Decolonization for a Global Society in Crisis. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3695356.

Kwet, Michael. 2021a. The City Surveillance State: Inside Johannesburg's Safe City Initiative. *SAIIA Policy Briefing* 231. https://saiia.org.za/research/the-city-surveillance-state-inside-johannesburgs-safe-city-initiative (last accessed: 22 December 2021).

Kwet, Michael. 2021b. A Digital Tech Deal: Digital Socialism, Decolonization, and Reparations for a Sustainable Global Economy. https://giswatch.org/node/6225 (last accessed: 22 December 2021).

Kwet, Michael. 2022. The Digital Tech Deal: a socialist framework for the 21st century. *Race & Class* 63 (3): 63-84. https://journals.sagepub.com/doi/full/10.1177/03063968211064478.

Lama. 2020. Social media statistics and usage in South Africa. https://www.talkwalker.com/blog/social-media-stats-south-africa (last accessed: 22 December 2021).

Larson, Zeb. 2020. The Sullivan Principles: South Africa, Apartheid, and Globalization. *Diplomatic History* 44 (3): 479–503. https://doi.org/10.1093/dh/dhaa002.

Liu, Wendy. 2020. A*bolish Silicon Valley: How to Liberate Technology from Capitalism*. London: Repeater.

Microsoft Corporation. 2021. Microsoft Annual Revenue by Geography: Fiscal 2002 – 2021. https://dazeinfo.com/2019/11/11/microsoft-revenue-by-geography-by-year-graphfarm (last accessed: 22 December 2021).

Muldoon, James. 2022. *Platform Socialism: How to Reclaim Our Digital Future from Big Tech*. London: Pluto Press.

Ndlovu-Gatsheni, Sabelo J. 2021. 'Moral evil, economic good': Whitewashing the sins of colonialism. https://www.aljazeera.com/opinions/2021/2/26/colonialism-in-africa-empire-was-not-ethical (last accessed: 22 December 2021).

Nkrumah, Kwame. 1966. *Neo-colonialism: The Last Stage of Imperialism.* New York: International Publishers.

Parrot Analytics. 2019. South Africa SVOD TV market share Q1 2019. https://www.parrotanalytics.com/insights/south-africa-svod-tv-market-share-q1-2019/ (last accessed: 22 December 2021).

Perelman, Michael. 2002. *Steal This Idea: Intellectual Property Rights and the Corporate Confiscation of Creativity.* New York: Palgrave.

Rathore, Itu. 2019. Amazon Annual Revenue By Region: FY 2013 to 2020. https://dazeinfo.com/2019/11/06/amazon-annual-revenue-by-region-graphfarm (last accessed: 22 December 2021).

Rawlins, Lauren Kate. 2017. Google trains 1m young Africans in digital skills. https://www.itweb.co.za/content/APero37Zg3dMQb6m (last accessed: 22 December 2021).

Richter, Felix. 2021. Amazon Leads $150-Billion Cloud Market. https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers (last accessed: 22 December 2021).

Rikap, Cecilia. 2021. *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered.* London: Routledge.

Satgar, Vishwas. 2018. *The Climate Crisis: South African and Global Democratic Eco-Socialist Alternatives.* Johannesburg: Wits University Press.

Schmidt, Elizabeth. 1980. *Decoding Corporate Camouflage: US Business Support for Apartheid.* Washington D.C.: Institute for Policy Studies.

Schwab, Klaus. 2016. *The Fourth Industrial Revolution.* Geneva: World Economic Forum.

Starrs, Sean Kenji. 2013. American Economic Power Hasn't Declined-It Globalized! Summoning the Data and Taking Globalization Seriously. *International Studies Quarterly* 57 (4): 817–830. https://doi.org/10.1111/isqu.12053.

Starrs, Sean Kenji. 2018. Can China Unmake the American Making of Global Capitalism? *Socialist Register* 55.

StatCounter. 2021. Mobile Vendor Market Share South Africa: Oct 2020 - Oct 2021. https://gs.statcounter.com/vendor-market-share/mobile/south-africa (last accessed: 22 December 2021).

Statista Research Department. 2021a. Estimated number of active streaming subscribers to Netflix in South Africa from 2016 to 2020. https://www.statista.com/statistics/607684/south-africa-netflix-subscribers (last accessed: 22 December 2021).

Statista Research Department. 2021b. Number of monthly active Facebook users worldwide as of 2nd quarter 2021. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide (last accessed: 22 December 2021).

The Red Nation. 2020. The Red Deal: Indigenous Action to Save Our Earth: Part One - End the Occupation. http://therednation.org/wp-content/uploads/2020/04/Red-Deal_Part-I_End-The-Occupation-1.pdf (last accessed: 22 December 2021).

Vailshery, Lionel Sujay. 2021. Share of Apple's revenue by geographical region from the 1st quarter of 2012 to the 3rd quarter of 2021. https://www.statista.com/statistics/382288/geographical-region-share-of-revenue-of-apple (last accessed: 22 December 2021).

Ventkess, Kyle. 2017. Google upskill programme receives lukewarm SA response. https://www.news24.com/fin24/Tech/News/google-upskill-programme-receives-lukewarm-sa-response-20170315 (access unavailable).

Vermeulen, Jan. 2017. The massive rise of music streaming in South Africa. https://mybroadband.co.za/news/business/208418-the-massive-rise-of-music-streaming-in-south-africa.html (last accessed: 22 December 2021).

Vermeulen, Jan. 2018. Spotify vs Apple vs Google vs Deezer vs Simfy in South Africa. https://mybroadband.co.za/news/internet/252091-spotify-vs-apple-vs-google-vs-deezer-vs-simfy-in-south-africa.html (last accessed: 22 December 2021).

Vourlias, Christopher. 2020. Netflix's Head of African Originals Lays Out Streamer's Plans for the Continent. https://variety.com/2020/digital/news/netflix-head-african-originals-lays-out-plans-for-continent-1203518648/ (last accessed: 22 December 2021).

Mediatization of ethnographic Research Practice in the digital Age

An Essay about Research and Sahel Dynamics since 1990s

Mirjam de Bruijn (Leiden University)[23]

Abstract

This essay departs from my search to understand how (new/digital) media interact with society and vice versa, especially in relation to political communication and political change. These 'effects' of changing media technology, however, not only touch the societies that we try to understand but also our own practice of 'doing' research. Especially in area studies, the ethnographic method, the relations with the field, with people is at the center of 'knowing'. Hence changes in these relationships, such as caused by (digital/new) Information and Communication Technologies (ICTs), will also influence the 'data' gathering, the insights, in short, the knowledge production processes and outcomes. In this essay I compare my own presence as a researcher in the Sahel in the 1990s and now, and I reinterpret this presence in relation to the changing communication ecology in the Sahel. My lens to analyze this is the concept of mediatization and mediation. It offers an interpretation of the changing relations in the field that has expanded with

[23] Contact: m.e.de.bruijn@hum.leidenuniv.nl, https://orcid.org/0000-0002-0499-8758.

the use of digital media. Hence, I try to understand changes in knowledge 'construction' in ethnographic practice in the digital age.

1 Introduction

Anno 2021, we live in the post-global world, as my colleague Rijk van Dijk and I stated in the introduction to the book 'The social Life of Connectivity' (de Bruijn and van Dijk 2012). What we then meant is that we already had gone beyond the world in process of globalization. Instead, the world is globalized, and people are connected. Connectivity, meaning the digital means that enable to connect, has taken a huge leap over the past 15 years. Also, the areas in Africa that are considered marginal and remote, such as the Sahel where I did most of my research, have at least the possibility to be connected through simple mobile phones and increasingly smartphones. The latter have also introduced the internet. In a world where people are connected, information from all sides and levels can flow, and people can transgress borders being in-situ; our perception of space and place changes (see Gupta and Ferguson 1992).

The relationships in ethnographic research, between the researcher and the researched have been under discussion as long as ethnography exists. It is a relationship that is subjective, in which intersubjectivity defines the outcome of the relationship and the way it reveals insight in some societal processes. Part of this debate is the discussion of power relations between researched and researcher, where the ethical concerns around inequality in these relationships are highly problematic. The introduction of new communication possibilities, that open a possibility to change the relationships between researcher and researched ties into this discussion. In this essay I try to understand through my own experiences what the potential is of the access to new communication technologies such as the mobile phone and smartphone have, for a further change of relationships in the field, and also in what direction these changes steer us in knowledge production. What has become the relationship between the researcher and the researched?

The connectivity in the digital as a fact also brings the mediatization in the global with a question mark. With this connectivity new means of communication or mediation have developed. For the regions discussed in this paper these are mainly mobile phones, social media and wireless internet. In mediatization studies (Hepp 2013) a relation is sought between changes in media technology, their effect on forms of mediation; and how these inform and frame socio-political processes and discourses[24]. Also interesting in such studies is the very old but still relevant question put on the agenda by McLuhan[25]: if and how the technology itself shapes the content of the message and hence the ways of communication.

[24] Mediatization research investigates the interrelation between media communicative change and sociocultural change, understood as a meta-process (a conceptual construct designating long-term processes of change). Media do not necessarily 'cause' the transformations but they have become co-constitutive for the articulation of politics, economics, education, religion, etc.

[25] An interesting compilation of the thoughts appears in of McLuhan, Marshall, and Quentin Fiore. 2008. *The medium is the massage.* London: Penguin.

In studies on these relationships the main direction in social change hinges to the democratization effect of new ICTs, in that they make access to information and communication more equal. Castells (2013) interpreted this as the power of communication. Those who control communication techniques can exert power. Hence if the technology to communicate is in the hands or is controlled by the ordinary man this would give him/her another position in the power hierarchies (Castells 2013). This implies that people can search their own information and create relationships in easier ways, avoiding or even erasing power relations. Which can of course also have the opposite effect and reinforce existing power relations based on age, gender, political hierarchies (de Bruijn et al. 2010). Or in case that the base of the technique is in the hands of the State, it can also be used to cut off communication (e.g. Dwyer and Molony 2019). Another often hailed advantage that is mentioned is the fact that such new ICTs overcome distance and bring people together. Such ideas are criticized with a reference to the effect of algorithms (Thorson et al. 2021) or the way information may also divide, i.e. the spread of misinformation and hate speech (Gagliardone 2019). If the effects are 'negative' or positive', it is clear that these changes profoundly influence societies in which the ethnographer interacts; and it would as well change the relationship of the researcher with his/her field and access to information. A first study on such effects was done by Lotte Pelckmans (2009), who interviewed Anthropologists on their use of mobile telephony in their relation with the field. It is clear that also then at the beginning of the mobile phone epoch the researchers experienced that distance was shrinking, and that worlds of difference were bridged. They also were positive about the contact they could keep over long distances. However, the field 'at home' was considered deficient and not representing the 'real' field. I would like to add that we also have understood that information from such distance relations change in character and also in content. It might steer the discussions and field findings in different directions (cf. Mano 2015).

In this essay I will put my own research practice as an ethnographer forward as a case study to relate to the question: How does the (digitalized) mediatization of the researched world, the field, influence the practice of doing research and hence its outcomes? The search is for changing research practices and its consequences in a digitized communication environment. The search is also if the researcher and the researched enter a new power balance in which the production of knowledge changes and probably becomes more equal. But also, how the lens of mediatization gives us different insights that affect knowledge production. What does this mean for the practice of research and its outcome(s)?

Instead of producing a philosophy of digitalization and decolonization I would like to share some of my empirical experiences with the digital and organization of research over the past three decades. I am an ethnographer. Ethnography is 'embodied research practice of immersion within the field site and the use of ethnographic sensibility' (Krause 2021). Ethnography generates data that tell us how people make meaning and exert agency; and how this interferes with processes of socio-political change. It is especially the 'ethnographic sensibility' and the 'research field in relation to agency' that I review here through the lens of mediatization.

2 A short intro to the case study for this paper

I started as a MA student to work in Northern Cameroon and moved for PhD research to Mali. In both cases I happened to find myself in (semi) nomadic communities, people who herd cattle, the Fulani[26]. Also, in both cases I was confronted with the crisis these types of livelihoods live through: ecological (drought), land access (in relation to land grabbing), cattle access, as more and more cattle is in the hands of urbanites, and violent conflict as is now the case in West-Sahel (Mali, Niger, Burkina Faso). In this essay I especially relate to the contacts I established in Mali. After the PhD-project in Mali that ended with the publication of the thesis in 1995, I never lost contact with the region, instead I worked in different research projects connecting back to Mali. I deliberately tried to keep contact with the people who received us in the 1990s. Then, beginning of the 1990s, there was no other way to be in contact with the region then through travelling. There were hardly fixed telephone lines. Also, the drought migrants who were in the South of Mali or in Ivory Coast, could only connect home through news that came by travellers. This did change in the course of the 21st century, when gradually mobile telephony and wireless internet became available also in central and northern Mali (Keita 2015). Today Mali's connectivity is registered in the ITU statistics as being more than 100% of GSM connections (see ITU 2021). The nomads who have become our friends during fieldwork also embraced mobile telephony to connect.

3 A note on ethnography

In this essay on the influence of new ICT on research I focus on ethnography in the definition of Jana Krause (2021) who emphasizes the immersion in the field. The immersion is shaped by the person of the researcher and his/her interactions with people in the 'field' and the environment. It makes a difference to be a female, of a Western society how people in Mali receive the researcher. However, this is not the complete story. It is also important what personality one has, which is related to emotions, and (inter)subjectivity. The knowledge that is generated then is also shaped by how people, both at the end of the researcher as the researched, make meaning and exert agency. In a world of digital communication, the question is raised if we can have similar relationships in a digitally steered communication as compared to a world with more physical communication. Can the ethnographic practice of immersion in the field as a person being there, be replaced by 'remote' communication? This is an important question specifically for my own practice of research in which I combine the ethnographic method with the biographical interview, and biographical reconstruction, to understand changes over time (cf. Jackson 2013; de Bruijn 2014). Such biographies can be constructed indeed in interviews, but they can also be compiled from the various meetings and encounters that appear during the practice of the ethnographic method. Can communication on distance contribute, or even be primordial in such research that is focused on the personal stories? How did I manage to continue this research in Mali, while the region of research is closed for physical visits due to the conflict in the region? A conflict that is phrased in the war on terrorism and goes with kidnapping of elites and people who work in NGOs

[26] Some of the publications of this research that was first done for the PhD diploma and after: de Bruijn and van Dijk 1995, 2003; de Bruijn 2007, 2015.

and also with ugly violence. I could simply no longer go there. This was a situation and is a situation that has pushed me into the distanced relationship with the field.

4 Practising ethnography in Mali

The practice of ethnography is an experience that is also related to the specificities of the 'field', and its relationships. Therefore, in the following I need to explain some of the ethnographic particularities of the environment in which I did my fieldwork in Mali in the 1990s and after. I will explain the way we interacted, the political development and the relationships in the field as a consequence of that.

The start of my encounter with the (semi) nomads in central Mali:

Understanding the relationship between histories of ecological crisis and societal changes was the first entry point in our study of semi-nomadic society in Central Mali. The 'we' here, is me and Han van Dijk. We worked together in this PhD research and later on some programmes in Mali. Accessing the world of these nomads was complicated. The nomads were part of a hierarchy of small chiefdoms. In the structure of these chiefdoms the elites would guard the pasture areas for the nomads who herded their cattle together with their own. We negotiated our entry to society through the chiefdoms in the region. This was also a first discovery of the hierarchies in which these nomads had to make their lives. We were probably naïve to think that nomads have a free life (cf. Riesman 1977). Their lives are marked by rules of access to pastures and water, that in this part of Mali are negotiated in a hierarchy that has its foundation long before the 20th century. The role of the elites, however, changed under colonial rule and the independent state into intermediaries with these governance structures. These hierarchies kept the nomads away from direct relationships with the State, except via the elites. Processes of democratization and decentralization did not change this landscape profoundly, instead it led to more power for the elites, who could also (mis)use the service of the nomads. Access to the wider world as the mobile telephone will do, could change this power relation. The elites as the mediators to that outside world, and as such were also able to keep nomads outside 'modernity' and exploit them in different ways. In our research we were introduced to these systems and have written about them.

4.1 Communication landscape

The life of the (semi-)nomads in central Mali is mobile following the rhythm of their animals. In the early 1990s we stayed with a group of nomads in the Seeno, a vast area with pastures where water was scarce. The nomadic pastoralists would move to follow water and pasture. They were also cultivating fields. In their mobility they depended on sedentary farmers with whom they exchange space for manure and milk for grains. The droughts of the 1980s had been a moment of wide dispersal of the nomads to other regions, the South of Mali, Ivory Coast etc. This mobility out of crisis however is not an uncommon phenomenon for nomadic people. Instead, they find themselves spread over long distances. Communication in these circumstances, without new technologies of communication, is mainly the mobility itself. Travel to meet, travel to get information, travel to search for lost cattle. Also travel to gather in the rainy season when marriages are concluded, or travel to bring condolences, are important moments of

communication and exchange. Moving to the weekly regional markets is another way of connecting and information gathering. Nomads need to be informed to be able to move with their cattle, to know prices, to understand ecological fluctuations, etc. But also, and probably especially, to know how the others are doing, to be a community. There is no community without communication.

If (semi) nomadic society is a society that, par excellence, constructs around mediation over large distances, and communication is a central element of society, then it seems relevant to delve into the effect of new technologies for mediatization in such a community. The region where we did our studies was traversed by one road that was tarmacked only in the mid-1980s. The region had various weekly markets, where also people from Burkina Faso would come to sell their products. However, technologies of communication, like TV and telephones were absent in the area. Radio was introduced in the course of the 20th century and widely used, also by the nomadic groups. The arrival of mobile telephony was a gradual process in the 21st century. It landed first in the small towns of the region, and only gradually the rural areas. The first mast that was enabling the nomads to also communicate directly from their region (but only from some spots where they could reach the signals) was planted in 2006. The signals got stronger, and more masts were planted in the region. Around 2010 we were in regular contact with people from the region, also from their camps and not only from the small towns. The mobile phone became for these (semi) nomads in Mali a new form of being connected, of travel, as is suggested in the work of Hahn (2020) who researched this relationship for nomads in Asia.

With the changes in communication ecology, we question changes in society, but as we concluded in the research programme 'Mobile Africa Revisited'[27] that run from 2007 to 2013 and focused on the relation between mobility in society and the introduction of new communication technologies, these technologies land in a landscape of communication in which the new technology integrates. It is both continuity and change (de Bruijn, Brinkman, and Nyamnjoh 2013). The balance between continuity and change is decided upon by the actions of the people who are confronted with these new mediums.

How did it affect nomads' lives? It is important to relate here to the fact that since 2000 other changes had hit the region: the return of people from Ivory Coast where conflicts pushed the Fulani back home. They also brought phones and other 'modern' ideas about the world. The arrival of television in small towns where the arrival of permanent electricity made television a possible medium. The process of decentralization of politics, and the local elections that also made the nomads more aware of politics not only because they were increasingly considered important by the representatives of the region in national politics, but also as a growing awareness of participation in local politics as a strategy to access services etc. To summarize: the nomads were already increasingly entering a digitized world. The mobile phone has certainly contributed to this development. The fact that they could call to their (politicians) benefactors in towns, that they could relate directly to people in Bamako, the capital of Mali, and that they could now also directly

[27] www.mobileafricarevisited.com; the 'we' here refers to Mirjam de Bruijn, Francis Nyamnjoh, and colleagues.

relate to the researchers that were in their region (amongst others myself) did give them access to a new sociality.

The research programme 'Connecting in Times of Duress'[28], that I run from 2012 to 2018, was taking place in the communication ecology where the mobile phone became increasingly a point of contact, of information and of negotiation with the semi nomads in central Mali. The PhD student in this programme who continued the studies in central Mali, Boukary Sangaré[29], was in daily contact with his main interlocutors in central Mali, from Bamako and so was I.

These developments do affect the ways of researchers who connect to these nomads and vice-versa. It is not only a research topic in itself (which it became for me), but it will directly affect the forms of relationships that can be established. The mobile phone as a technology contains this change in its technique of being a medium that directly influences the relationships in the field. It also invites for a reflection on the difference between the way the research communicated before and then after the introduction of the mobile telephone. Participation in the community was first done by participation in the space of society: hence travelling a lot, being mobile; but also, being disconnected when not in the region. Since 2006 the mobile phone became part of the communication space and hence, we need to question how this changed our relationship, how the messages changed and if and how this influenced knowledge-production about the region and its people.

5 Communication and ethnographic encounters

5.1 Our communication in the 1990s

After a month of negotiation with the chiefs we were allowed to visit the region of our research. We soon were welcomed by the lineage leader who took us in as his friends. And we could start to build our relationships. As people who would not be able to fully live a nomadic life (without falling ill) we decided to work with a car, so that we could bring water for ourselves, but as soon became clear the car became a research and communication assistant in itself. Then, our fellow anthropologists would tell us that the car was too much of an alien to the nomad's society, and it would place us high on the scale of the hierarchy of Malian society, giving us too much the jacket of being the other, may be associated to Development work. We experienced being with the car differently. Besides that we did not have a choice, as we choose to bring water and other things with us, the car became a mediator, a way to access people. It was the aid to transport food, to transport people, to help the sick. It was also our camel. The car was probably, without us being consciously aware, also a mediator in a different way. It stood for the world we come from, it sends messages about possibilities. In a way it relates a story on modernity. As Walter Nkwi (2011) has shown in his study on the history of communication in Cameroon, cars were symbols of wealth, but they were also the messengers from another world. They were telling stories. In our case the

[28] www.connecting-in-times-of-duress.nl

[29] De Bruijn & Sijsma, 2019: https://voice4thought.org/hope-less/; watch this film Hope(less) in which Boukary Sangaré shows some elements of his research in Central Mali; with film footage from the research of Boukary Sangaré (Mali) and Souleymane Adoum (Chad).

car was increasingly a connection to a wider world. We brought food when the need was high, we helped to transport ill people when they needed access to the hospital. We brought people in short time to the markets, etc. It hence created communication and produced relations of trust. Maybe it was the first story line in which the mediation of the elites was not necessary. In hindsight we really wonder if indeed our car and our presence with it in the 1990s, were a first step towards a new self-realization for the nomads (Guyer 1993). It is only since 2010 that the camp became part of a route between markets and cars were then frequently passing the area. Other than that cars had always belonged to the elites, people from development projects, that were in the local hierarchy not at a similar level as the nomads. Our car however was theirs.

We would also speak the language Fulfulde, travel with the nomads, stay in the camps during the rainy season, etcetera. For us this was a period of isolation from 'our' own world, with which we were only in contact via a phone-line that was at least two hours' drive from a nearby small town where we had hired a small house in a compound of an Arab family. Also, this house became a point of contact for the nomads who had to go to the hospital or who were travelling somewhere. We became part of the communication landscape. Were we also in a way a technology of communication for the nomads? This became crystal clear when we were asked to carry cassettes with audio messages from the camp to other parts of Mali where they expected that family had gone during the droughts.

The rudimentary connectivity by modern technologies, meant that after we left the region after two years in 1993 there was no easy contact possible, we were completely disconnected. And we wrote our dissertation then without contacting the nomads again. We wrote the ethnography![30]

The reinterpretation of this time through the lens of mediation and mediatization now brings other questions to the fore, than the ones we formulated then. I now wonder what were the technologies we as researchers brought to the world of the nomads that changed the communication landscape? Were we ourselves as well a technique of mediation, hence were we mediators, who brought in ideas and things about another world that impacted on this society?

5.2 New media technology and socio-political changes

The influence of the mobile phone and mobile internet got a strong influence on the region of research in central Mali in 2009. Why is this a turning point? 2009 was the year that mobile phones became carriers of wireless internet and social media. Wireless internet allows the installation of platforms such as Facebook, WhatsApp and Telegram. Facebook has become synonymous to internet for many youths in these regions. It is clear that the nomads in rural camps, and often on the move with their cattle are not continuously connected and might not have smartphones. But the arrival of these possibilities did not go unnoticed, and young nomads started using these technologies, accessing new information, and connecting to the world as they had not been able to do before.

[30] 'Arid Ways', published in 1995 (de Bruijn and van Dijk 1995).

Since ICTs have become part of the life of (first male) nomads, our communication started to change. I can no longer write my texts without hearing A., or some other people from the area. A. is our friend in the region, the son of our former host who passed away in 2006. A. calls when he needs money, but also when there are things happening in his environment that he wants me to know. Until 2011 I could still go to the region, and also my students (PhD and MA) with whom I worked in different projects could go and bring back their observations. These students who lived in Bamako would have intensive contacts via phone as part of their work. I received phone calls in Utrecht, where I live in the Netherlands, at any moment of the day. The nomads are sitting at my dining table through the phones. I then could not imagine how this would become the only reality to relate with my 'field' in Mali in the near future. At the beginning of this move to phone communication I was not so much seeing it as a form of research. It was simply contact with old friends.

I consider A. as my brother. He is also the herder of our cattle, born from one cow we bought in 1992. Another close contact in the region is B. who is the son in an elite family of the region. He, however, had very good relations with the nomads. A. is his good friend. Another young man recently joined this group. H. tried to contact me via Facebook messenger first, then he tried WhatsApp. As I was not so eager to answer, he also started calling. At first, I had no good idea who he was, but then he made clear who he was to me, the son of my research assistant in the 1990s. Through him I received the phone number of his mother. I can continue to lengthen this list. There are important differences in this communication: A. calls, B. and H. send audio messages through WhatsApp.

B. and I rediscovered each other when things started to go wrong in the camp and small town where he lives. He became one of my phone interlocutors helping me to understand the situation as the conflict advanced.

5.3 Central Mali 2012 to date: a warzone

Central Mali has become a warzone. Since 2012, when the Tamacheq started a rebellion in the North and at the same time a coup d'Etat put political landscape in chaos, central Mali started an episode of its history that has surprised in its violence and persistence. The Malian state retreated from the region. Jihadi groups hijacked the rebellion of the Tamacheq and took control of the centre. Instead of chaos this turned into a relatively calm period that until today is appreciated by the nomads. An interesting story that was told by A. is that the Jihadi group would offer them protection by giving their phone numbers that could be called in case of a problem or urgency. It seems that this system did work, and it created trust between the population and these Jihadists.

After the French defeated the Jihadi groups in 2013 the region did not return to calm, on the contrary it became a chaotic situation where protection of villages etc. was in the hands of self-defense groups. The state did not return. Hospitals and schools are understaffed until today, and in the rural areas not present. The situation became increasingly problematic and chaotic, ending with the return of the Jihadi and now relatively well organized (so it seems) and increasingly incorporating the self-defense (youth) militia. These are often Fulani from the region. It has developed into a Jihadi occupation of the rural areas in the region. Near the camp where A. dwells

an important cell of the Jihadi is based: Kaatiba Serma, joined by the sons of our friends. In this region the Malian army undertakes actions against what they think are Jihadists, as do the French troops. Human rights violations are committed at all sides. Many people fled the region heading to the urban areas of Bamako, or Mopti. The region is closed for foreigners and local elites who risk being kidnapped.

Central Mali is a region where travelling has become very difficult. It is a red zone not only for the western researcher, but also for many people from the region. The Fulani elites are no longer welcome. They are accused by the Fulani militia/Jihadi groups as being the exploiters of the nomads and others. Such a region becomes a no-go area, where no government services are present. Military and gendarmerie are restricted to urban sites. Boukary Sangaré, a researcher and PhD, was able to go to the area in the period from 2012 to 2016, but from 2016 he also stopped going. A few months ago, B. fled the region, because he is accused to support the Jihadi, while others flee because the Jihadi accuse them of being traitors.

How can we know what happens in this zone? Is here a role for mobile telephony? How did the conflict deeply reconfigure our 'field' and the form of data that we are still able to collect in the region? What about the 'ethnographic sensibility'? How does 'distanced' research lead to a reconfiguration of the relationship between researchers and researched?

5.4 The 'Field-relations' in war and the digital environment

How did the 'field' change? In this case mobile telephony and wireless internet enabling social media platforms are means of communication that have been added to the communication landscape of the region. It has turned the (semi) nomads' society into an ever more networked or mobile society. These networks enable information to flow without interference of the old hierarchies. The hierarchies in the social media are re-constructed. For instance, in the sense that those who have telephones and access to new information and communication can have a different role. They can for instance bypass the old communication system that would go via the elites. Or they can get information from far where in the past they would not have access to. In the case of the nomads, it is interesting to see how A. was able to use the smartphone of his son to access information about the occupation of the area by MNLA (Tamacheq) in 2012. He sent me a message about this discovery while he was watching the videos. Later we have seen the appearance of the audio taped preaches of a Jihadi preacher from the Innerdelta, who later became the leader of the Jihadi groups in the region, and who explains in his sermons that the Malian state/cum elites have exploited the nomadic population for too long, and in this preach he also calls for revolt against the state (Galy Cissé 2019). The union under the flag of a Muslim state is one of the answers offered to the population. This is reinforced by the stories of power in pre-colonial times when the Jihad was invading the Sahel and established empires. Such information spread in audio form travels in social media platforms and reaches many young nomads who feel that it is time for a change. Promises that are also picked up in social media where people portray urban life or show how wealthy people in other parts of the world live for instance. The absence of the state services such as health care, schools, etc., reinforces the idea among the nomads that the state has not been there for them. Such messages can be categorized as messages that inform

people about who they are, but with a very binary vision of the world, and a tendency of othering where often violence then becomes the only way to act (Walby 2013).

During this period travel was still a technique to connect, but the direction of travel has changed: our former research partners from the center travel to Bamako on our invitation. This is how I was able to talk to A., B., and others since 2015. But the increasing violence also along the roads makes such travel impossible today (since 2019). Bamako and southern Mali have become a refuge for nomads (and others of course) who cannot stay in the center for safety reasons. Some of them live in quarters of Bamako where their family or co-villagers did migrate in earlier times. Others ended up in camps for displaced people, that were in most cases spontaneously developing. There are some spaces for the nomads created by their elites who live in Bamako and have the means to help the people from their home areas. We do relate to the displaced who are old friends, and listen to their stories, that will be colored after their travel and influenced by their fear. Our only 'real' connection to the region itself is access through calling. This has also become a form of participant observation, to 'feel' the situation. For instance, in recent times calling, or being connected from the region with outside, has become one of the inflictions on the rules of the Jihadi groups and of control by the military and other militias in the region. It is dangerous for A. to be intercepted while calling with me. Hence, the phone calls with A. are about cattle, and sending money. He hardly refers anymore to the situation in the area. They are also about how he is doing; and in these small exchanges he will reveal some parts of his life that are information about the changes he is living through. These 'snippets' help me to get an impression of his life and what it means to live the situation of central Mali today. The form of the calling contact, the non-said, the snippets are also a form of participant observation.

Another example of such phone contacts are the audio messages of P., an acquaintance of long date, a young human rights student who decided to return to his home village to follow and report on what is happening. He provides me with information, about events, killings, negotiations, kidnapping, etc. He also links me to nomads who not yet have my phone number, and in case A. cannot reach me A. tries to link through P. P. has decided to keep me updated about the region and indeed the flow of information that he sent me are a story of how a population is caught in between the army/state and the Jihadi, that are locally called, the people from the bush. I do regularly call him, over WhatsApp when I am in the Netherlands, or over the phone when I am in Bamako. P.'s willingness to inform me is for him also a way to get the stories out. He is very concerned with the region and its future.

In this situation the mobile phone has become the only entry to the world of the nomads in central Mali under conflict. The ethnographic encounter is summarized in WhatsApp messages, through phone calls and through the violent events that are reported in various WhatsApp groups of which I have become a member. Also, the PhD students who were able to go to these regions in the recent past are now depending on these information channels. This is of course a limited access to the world of the nomads. On the other hand, the deep relationships that we have been able to build up over three decades, do help even in these circumstances, have the 'ethnographic sensibility'. And after all this field of connectivity has become our world and hence the field as well. One could argue that therefore the information that we receive, the emotions that we feel through the

communication channels are part of ethnography and as such ethnographic information that can help construct the ethnographic narrative.

6 Is the media the message

As we can read in the theory of mediatization: Media do not necessarily 'cause' the transformations, but they have become co-constitutive for the articulation of politics, economics, education, religion, etc. A comparison of the period of 1990s and the period after 2012, shows also continuity in such relationship. The technologies of communication that were brought into the nomadic camps gave access to the world and new forms of information. Such happened already in 1990 with the car and the researchers that opened the road to understand the world out there, and hence entering 'modernity'. In retrospect such an analysis through new technologies of communication and what they bring and how they co-create society is important. This has become very clear with the advancement of new communication technologies after 2009, when access to mobile phones and also smartphones (wireless internet) became normal. Access to other flows of information, maybe the first steps towards understanding exploitation, the role of elites was already entering the world of the nomads.

An example here is the new forms of Muslim religion that were present in the region in the early 1990s, a period in which Wahhabism took roots. It was also the period that the sons of our interlocutors in the camps would be sent to itinerant Koranic schools, that were probably at least partially influenced by these new ideas. From the camp that we frequented then, a few boys were sent to such schools. They were the first to enter this itinerary that became an option for a future path for the nomads. Muslim scholars have always been traveling through the region and have been messengers. Now the nomadic children started to enter that itinerary. It is probably not a surprise that they are among the leaders of the Kaatiba's (the camps of Jihadists) in the region today. In a way they have followed the future paths as promised in the social media messages.

Another example of the opening of the world for the nomads is A. and his 'self-realization'. He is the son of a lineage leader, and after the death of his father he inherited this role. He was among the first to start connecting to political party leaders in the region. He became even a member of the party of businessmen from Douentza (small town in the region) and proposed this party as the choice for nomads, where in the past they would have followed their own elites. It went together with him buying land and constructing a house in a nearby small town. He was transforming his life and introduced also his family to another world. This is also shown in the demand of one of the girls from the family to bring her high heeled shoes that she saw on television at a house in that small town.

The introduction of new technology of communication did inscribe in these tendencies (that were itself part of technological changes in a way) and emphasized and intensified them. The flows of information that could then emerge, and that were also sent in the right platforms affected the minds of the nomads who were underway to no longer accept their situation. The ways they did this have now become history.

These processes are directly related to the mediatization of the nomad's world. That parallels the mediatization of the researcher's world. Together we entered a new communication landscape that transgresses borders that could not be done by the old communication landscape. Here the researcher would also have its place, as a form of media/medium. As I showed the car that we brought, our own persons, were like techniques of communication that defined our mutual relationship. It was not an equal relationship, but it was a relationship based on mutual understanding, of friendship. It allowed us insights in many parts of the world of these nomads. But we did also transfer messages. It is difficult to evaluate what we as researchers become when we stay for so long in a community, that is transcended in the new media landscape. But we were also part of the birth/creation of that ecology and hence influenced the ideas and reflections of the people with whom we interacted.

7 Reflections

MacLuhan stated that the media is the message; I learn from him that the technology in itself has the property to steer and form the message, which means that the technology of communication (medium) defines what kind of content can be transferred. In comparison between the 1990s and the period after 2009 we, the researchers, and the researched were in very different ethnographic encounters when it concerns the media of communication. The mediatization of the ethnographic encounter changed with changes in the field. The forms of the narratives that resulted are indeed different, but is that a consequence of the medium or of the changes in communication landscape itself, i.e. the changes of the field? My research field moved with the medium, and the mediatization that transformed with it also fit the changes that I made in my research: from the ecological crisis and its consequences to the introduction of mobile telephony, to the causes of conflict and today the interrelation between social media and conflict.

Another query with which I started this essay was if the mediatization changes the ethnographic experience: If relationships find new forms and new expressions, I certainly think that this is the case. The adding of mobile telephony to our repertoire of communication that allows us to travel in situ, to be informed whenever the urgency is felt, has made a difference in connectivity that is also a way of being together. In my specific case the changes in communication are placed in a continuum in time. I started to be in the region in the late 1980s, continued going to the region till 2012; then continued with the eyes of PhD and MA students, and finally had to turn to the mobile phone communication and travel to Bamako. I can imagine the feelings when A. calls me. And I still have the possibility to call P. or B. to understand better what the encounter with A. told me. The ethnographic experience has become more limited, in the sense that we are not there. But as I said in the previous paragraph, we are also there, because we do share this world of WhatsApp, telephone, internet. We do share these spaces, that are spaces of social being and of being as the village and the camp are. Information that we gather in these exchanges does have value and add to narratives that we construct from various sources as the ethnographer does. I do not say that it is the same as being in the camp, travelling with our car to the next pasture area, etc. but experiences in social media and on the phone are also shared experiences that represent a space of social realities.

I think that indeed the medium does define the form of message and also the content of the message. There is certainly a transformation of content, a steering of the narrative. This is what we could label media biases, or mediatization environment specificities. These media biases are important to decipher. But then we also need to reflect on the pre-new ICT periods where we mediated in different ways with other techniques of mediation. These had their own biases and informed us in other ways to tell a narrative of a shared space.

This new communication landscape, dominated by mobile telephone and social media, is of course not exactly the same as the village, but in the social interactions sometimes it comes close. We do not enter a marketplace where we happen to meet people, although social media may sometimes have such an effect. Also, the person who calls does have a specific idea in his/her mind to call. Hence communication is more planned. Serendipity, such a highly cherished part of fieldwork, the unexpected, the not-known, is less present (Rivoal and Salazar 2013). There will also be a bias in who accesses the phone. Women have less access to phones than men; and the youth will be more eager to use phones than older people.

An interesting characteristic of the mobile telephony and social media platforms is its horizontal way of communication. It can therefore circumvent old hierarchies. A. who became member of the political party of the opponents of the elites is such an example. Costs of communication may still hamper access, as may government interventions do. But I experienced the two way communication that I have with A., B. H. and P. as an interesting way to continue research. They call me when they need me, and no longer only the other way around. They call me when they want to share a story, or they do so through WhatsApp sending messages and images. P. for instance explicitly told me that he hopes I will take up the role of being their advocacy person. A role I could ignore, or not see, if I was not connected to P. This is just an example, that I use here to share my experience of these relationships as less unequal, not me demanding and wanting to know, but also the other way around. And these moments of communication via the wireless techniques have a certain immediacy, a mediation in urgency of in the instant that is also hard to refuse. It has changed my relationship with these old friends and acquaintances, and it has changed my role in these relationships. Inevitably choices here are made both in sending and in receiving information, and decisions that are made what is relevant. It is still to the authors of narratives to decide on what information is considered relevant. There it is inevitable to still allow hierarchies in knowledge production, but it is unavoidable and necessary to allow multiple authors to decide.

8 By way of conclusion

I am not sure if there is a conclusion to this essay. It is really work in progress. The essay relates to discussions in anthropological circles on hierarchies in the ethnographic fieldwork. What I have shown in this essay is how transfer of information changes and becomes also a deliberate effort of the people with who we interact via social media and telephones. Next to communication to get each other's news, there is also a feeling of need to disclose information. A more equal urge of knowledge exchange. Is this a form of de-hierarchization of knowledge production? Probably it is the beginning, as both sides of the communication lines see an importance of the transfer. It is

however the researcher who is in charge of the transformation of this information into 'relevant' knowledge.

The questions of intersubjectivity, biases, inequality in knowledge production will never disappear, but with the digital environment and the forms of mediatization that it has brought and will bring in the future, we have to reflect and redefine the relations between researchers, researched in the disciplines of area studies, especially where we connect over long distances. These discussions start with the reflection on the changes that we are living through as researchers in our own practices. In this essay I have tried to start this reflection for my own knowledge production relationships. The reality of the new mediatization forms, that can be described as more equal, is an everyday experience. Conversations about the everyday do no longer know time or place distances. This brings new responsibilities for the researcher and researched, but it also forces us to rethink the relations in knowledge production, and to allow for new forms of knowing.

After all in the ethnographic encounter we create stories, and narratives. Fieldwork is a subjective encounter between different worldviews, lifestyles etc. The narrative constructed through fieldwork experiences is full of choices (but well guessed choices) of the author(s) of these narratives. The flows of information and their variations and diversity is then very important. The internet fieldwork that increasingly becomes the practice of ethnographic research is then also a confrontation with the world that is streamlined through those people who do have access to mobile (smart) phones and who want to voice their concerns; and also, who want to be in contact. For me the experience from being ethnographer *in* the field, to the ethnographer sharing the communication landscape with the people I worked with, has forced me to rethink the relationship and the information flows, in a way it unveiled an important question that we should ask in all ethnographic encounters: whose information do we use, and who wants us to use their information to create knowledge about a people, about a situation, or a socio-political development.

9 References

Castells, Manuel. 2013. *Communication power*. Oxford: Oxford University Press.

de Bruijn, Mirjam, and Han van Dijk. 2003. Changing population mobility in West Africa: Fulbe pastoralists in Central and South Mali. *African Affairs* 102 (407): 285–307. https://doi.org/10.1093/afraf/adg005.

de Bruijn, Mirjam, and Dijk J.W.M. van. 1995. *Arid ways: Cultural understandings of insecurity in Fulbe society, central Mali.* Wageningen: CERES. https://openaccess.leidenuniv.nl/handle/1887/9333.

de Bruijn, Mirjam. 2014. Connecting in mobile communities: an African case study. *Media, Culture & Society* 36 (3): 319–335. https://doi.org/10.1177/0163443714521088.

de Bruijn, Mirjam, and Rijk van Dijk (eds.). 2012. *The social life of connectivity in Africa.* New York, NY: Palgrave Macmillan.

de Bruijn, Mirjam. 2007. Mobility and society in the Sahel: An exploration of mobile margins and global governance. In *Cultures of migration*, ed. Hans Peter Hahn, 109–129. Berlin, Münster: LIT.

de Bruijn, Mirjam. 2015. Quest for Citizenship of the Fulbe (Semi)Nomads in Central Mali. https://mirjamdebruijn.wordpress.com/2015/10/12/quest-for-citizenship-of-the-fulbe-seminomads-in-central-mali/.

de Bruijn, Mirjam, Inge Brinkman, and Francis Nyamnjoh. 2013. *Side@Ways: Mobile Margins and the Dynamics of Communication in Africa.* Mangon: Langaa RPCIG.

de Bruijn, Mirjam, Francis Nyamnjoh, and Tseghama Angwafo. 2010. Mobile interconnections: Reinterpreting distance, relating and difference in the Cameroonian Grassfields. *Journal of African Media Studies* 2 (3): 267–285. https://doi.org/10.1386/jams.2.3.267_1.

de Bruijn, Mirjam and Sjoerd Sijsma. 2019. Hopeless, biographies of radicalization, film published by Voice4Thought, https://www.voice4thought.org/watch-online-hopeless-biographies-of-radicalization/

Dwyer, Maggie, and Thomas Molony. 2019. *Social Media and Politics in Africa Democracy, Censorship and Security*.

Mano, Winston. 2015. From analogue to digital social media in Africa. *Journal of African Media Studies* 7 (2): 105–107. https://doi.org/10.1386/jams.7.2.105_2.

Galy Cissé, Modibo. 2019. Hamadoun Koufa: Spearhead of radicalism in central Mali. In *Biographies of radicalization: Hidden messages of social change*, ed. Mirjam de Bruijn, 152–173. Berlin, Boston: De Gruyter Oldenbourg.

Gagliardone, Iginio. 2019. Defining Online Hate and Its "Public Lives": What is the Place for "Extreme Speech"? *International Journal of Communication* (13): 3068–3087.

Gupta, Akhil, and James Ferguson. 1992. Beyond 'Culture': Space, Identity, and the Politics of Difference. *Cultural Anthropology* (7 (1)): 6–23. https://doi.org/10.2307/656518.

Guyer, Jane I. 1993. Wealth in People and Self-Realization in Equatorial Africa. *Man* 28 (2): 243. https://doi.org/10.2307/2803412.

Hahn, Allison. 2020. Nomadic Digital Ethnography and Engagement. *Nomadic Peoples* 24 (2): 299–311. https://doi.org/10.3197/np.2020.240209.

Hepp, Andreas. 2013. *Cultures of mediatization.* Cambridge: Polity Press.

ITU 2021. Digital trends in Africa 2021, Information and communication technology trends and developments in the Africa region, 2017-2020. United Nations

Jackson, Michael. 2013. *The Wherewithal of Life: Ethics, Migration, and the Question of Well-Being.* Berkeley: University of California Press.

Keita, Naffet. 2015. Téléphonie et mobilité au Mali. Mangon: Langaa RPCIG.

Krause, Jana. 2021. The ethics of ethnographic methods in conflict zones. *Journal of Peace Research* 58 (3): 329–341. https://doi.org/10.1177/0022343320971021.

McLuhan, Marshall, and Quentin Fiore. 2008. *The medium is the massage.* London: Penguin.

Nkwi, Walter, G. 2011. Kfaang and its technologies: Towards a social history of mobility in Kom, Cameroon, 1928-1998. *African Studies Collection* (30).

Pelckmans, Lotte. 2009. Phoning anthropologists. In *Mobile Phones: The New Talking Drums of Everyday Africa*, ed. Mirjam de Bruijn and Francis Nyamnjoh, 23–49. Mangon: Langaa RPCIG.

Riesman, Paul. 1977. *Freedom in Fulani social life: An introspective ethnography.* Chicago, Ill.: University of Chicago Press.

Rivoal, Isabelle, and Noel B. Salazar. 2013. Contemporary ethnographic practice and the value of serendipity. *Social Anthropology* 21 (2): 178–185. https://doi.org/10.1111/1469-8676.12026.

Thorson, Kjerstin, Kelley Cotter, Mel Medeiros, and Chankyung Pak. 2021. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society* 24 (2): 183–200. https://doi.org/10.1080/1369118X.2019.1642934.

Walby, Sylvia. 2013. Violence and society: Introduction to an emerging field of sociology. *Current Sociology* 61 (2): 95–111. https://doi.org/10.1177/0011392112456478.

Programme websites:

www.connecting-in-times-of-duress.nl.

www.mobileafricarevisited.com.

The Legal Framework of and Ethical Considerations in Data Processing

Section 2

GDPR and Research Data in the European Union

An Overview

Oliver Watteler (Leibniz Institute for the Social Sciences)[31]

1 Introduction[32]

Human subject research projects generate personal data about the individuals participating in these projects. Although this type of research data is primarily used for scientific purposes, the processing of data from or about research participants might also have unintended or negative consequences for those participants. These consequences might for example be harms inflicted on research participants like identity theft or fraud, financial loss, or damage to the reputation, and it is the aim of data protection to minimize the risk of such unintended or negative consequences from happening and thus to protect fundamental rights of the data subjects.

Since May 2018 the General Data Protection Regulation (GDPR)[33] governs data protection in the European Economic Area. The GDPR is part of a legal framework between the EU Charter of Fundamental Rights (CFR)[34], other EU regulations, national constitutions, and national data protection and specialized laws. It covers rules and regulations for the processing of personal data by a controller or a processor established in the Union, of data subjects who are in the Union, or

[31] Contact: Oliver.Watteler@gesis.org, https://orcid.org/0000-0002-1634-9229.

[32] I thank one anonymous reviewer for very helpful comments.

[33] https://eur-lex.europa.eu/eli/reg/2016/679/oj (last accessed 9 May 2022).

[34] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT (last accessed 9 May 2022).

by a controller not established in the Union, but in a place where Member State law applies (Art. 3 GDPR). The 99 articles and the 173 so called 'recitals' of the GDPR cover a great variety of data protection, also including data processing for research purposes. This paper gives an overview of data protection in the EU with a focus on the GDPR and the impact of the Regulation on human subject research.

The first section of the paper introduces the connection of human subject research and data protection. Here we describe a research data lifecycle with six stages in order to locate various data protection tasks and obligations in a research context. The second section presents an overview of data protection and positions the GDPR in the frame of the CFR.

The following main part of the paper covers the GDPR, its position in a hierarchy of norms and its main features. We look at central concepts like 'personal data', 'special categories of personal data', and what is meant by 'controller', 'processor' and 'data processing'. All processing of personal data needs a legal ground like a law or a contract, and the ground probably most prominent in human subject research is informed consent. Irrespective of what is agreed in an informed consent or in other legal grounds, the 'data subjects', research participants in our case, have a number of basic rights which need to be respected in processing operations. One principle idea behind these rights is the transparency of the data processing.

The GDPR follows a risk-based approach, which means that the possibilities of harms or harmful events need to be considered. If the risk of such harms or harmful events is high, the controller needs to undertake a Data Protection Impact Assessment which we look at in the next section. Regardless of the risk level technical and organizational measures, so called TOMs must be implemented in order to safeguard for the fundamental rights of the data subjects. Two of the most prominent TOMs are anonymization and pseudonymization. The TOMs concern all kinds of processing, but the GDPR also makes exemptions for research.

The paper concludes with a look at the research data lifecycle again, this time collating basic data protection obligations and TOMs to the six stages. Overall, this paper can only offer a glimpse into data protection and human subject research, but the references to publications by the European Data Protection Supervisor, the European Data Protection Board and its predecessor, the 'Article 29 Working Party', as well as other EU institutions help the reader to go into detail on the field of data protection in human subject research.

2 Human subject research and data protection

Human subjects research is an area of scientific investigation about living individuals in which a researcher "(i) obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (ii) obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens" (45 CFR 46.102).[35] One possible outcome of this research is data, in our case based on information collected from or about human beings. 'Research data' can generally be defined "as factual records

[35] https://www.law.cornell.edu/cfr/text/45/46.102 (last accessed 9 May 2022).

(numerical scores, textual records, images and sounds) used as primary sources for scientific research, […] commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated" (OECD 2007:13).

Now, this type of research data is primarily used for scientific purposes but the processing of data from human subjects, 'personal data' for short, might also have unintended or negative consequences for the humans under investigation. These consequences might be harms inflicted on research participants like identity theft or fraud, financial loss, or damage to the reputation, and the protection against those harms is an important categories of research ethics and legal regulations. Therefore, voluntary commitments by research associations and legal regulations of data processing, meaning data protection are generally acknowledged (European Commission 2010). Since 2018, the GDPR is the principle piece of data protection regulation in the European Union, which also proposes technical and organizational means to enforce the safeguard of personal data.

Human subject research projects collecting personal data from research participants need to follow the rules laid out in the GDPR. In the context of such a project data protection can be considered an ongoing task which should be planned in an early stage of the undertaking. The process of a research project can be separated in stages from planning to sampling, to data collection, data preparation, analysis, and result publication. Research data also have a second life if they are persevered, documented and published via a repository or data archive for secondary use. A 'research data lifecycle' (see figure 1) helps us to organize the stages of this process, and to structure the technical and organizations means of data protection throughout and after a research project.

Figure 9: Data Lifecycle figure, Source: Ball 2012.

In general, we can differentiate six stages in a 'research data lifecycle'. Based on research questions, the first stage marks the planning and design stage during which important decisions

on the methodological approach, the sampling of individuals to be researched, and the manner and amount of data to be collected are defined among other things. During the second stage the actual data collection and data preparation for analysis take place. The third stage sees the analysis and interpretation of the data. This stage would be the place for the publication of results in journal articles, book chapters or books. It also marks, in a way, the end of the first halve of the life cycle, because after that researchers will look into the possibilities of preserving their data either for re-using them again in other research undertakings, or in order to release the data for the use by others. Thus, stage four is the stage during which activities like format migration, data documentation and archiving take place. The result of stage four will be a package of data, metadata and documents, which allows third parties to re-use the data. The distribution or dissemination of the data is usually done by data repositories or data archives in stage five. These infrastructures take over tasks like, for example, long-term preservation, access control if necessary, and promotion of the data. After the release by data repositories or data archives stage six offers opportunities for other researchers to re-analyze published results, to answer new research questions, or to use the data for training and teaching activities (see Corti et al. 2020 for an introduction on research data management).

The planning for data protection should ideally take place in an early stage of research and the technical and organizational means of data protection are carried out throughout the entire lifecycle. We come back to the six stages at the end of the text in order to logically sort the data protection topics as well as the technical and organizational means presented in this paper. Next, we give an overview of what data protection means.

3 Data protection – an overview

Data protection can be considered a part of the personality right to privacy. "Privacy is a personal condition of life characterised by seclusion from, and therefore absence of acquaintance by, the public" (Neethling 2005:233), and the right to privacy covers possible harms like the intrusion into a private sphere like a home as well as the disclosure of personal information. The prevention of an unwanted disclosure of personal information or the misuse of such information are at the core of data protection, which is "the protection of a person (data subject) with regard to the processing of his personal data by another person or the state" (ibid.:234). In the European Union the protection of this personal data is governed by various legal acts on different legislative levels.

On the European level Art. 8 (1) of the CFR[36] states that "(e)veryone has the right to the protection of personal data concerning him or her." Art. 8 (2) then highlights core principles of data protection in general, which are:

- The 'fair processing' of personal data,
- the processing of personal data guided by a 'purpose',
- the processing based on 'consent' or another legitimate legal basis,

[36] With the Lisbon Treaty of 2009 the CFR was officially adopted and became binding EU law (FRA 2020).

- freedom of information concerning one's own personal data, and
- the right to have one's own data rectified if necessary.

These principles of data protection are embraced in the GDPR (Recital 1). According to Art. 1 (1) the Regulation "lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data." Furthermore, the Regulation "protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data" (Art. 2). In this context the fundamental right to data protection is guided by six principles (Art. 5) which are directly related to the aforementioned Art. 8 (1) CFR:

(1) 'Lawfulness, fairness and transparency' of data processing: Data must be processed in a way transparent for natural persons called 'data subjects' in the GDPR. The personas would be research participants in our case.

(2) 'Purpose limitation': Data may only be collected "for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". Here, a first exemption is made when it comes to scientific research, because research as a purpose is seen as in line with the initial purposes for which the data was collected. We get back to these exemptions toward the end of the text.

(3) 'Data minimization': Data should be collected which are adequate and relevant to what is necessary. This principle is meant to limit the amount of data collected.

(4) 'Accuracy': Data collected for a given purpose should be kept correct and deleted or corrected without delay if necessary.

(5) 'Storage limitation': Data collected for a given purpose should be kept no longer than is necessary. Here again GDPR makes an exemption for research purposes. Personal data may be kept longer, if "appropriate technical and organizational measures" are implemented to protect the data.

(6) 'Integrity and confidentiality': Data should be protected against "unauthorized or unlawful processing and against accidental loss, destruction or damage", again using "appropriate technical or organizational measures".

On the same level of fundamental rights there is "(a)nother right to balance against the rights to respect for private life and to data protection [which] is the freedom of the arts and sciences, explicitly protected under Art. 13 of the CFR" [37] (FRA 2018:74). The two fundamental rights might

[37] In most European countries science and research are also protected by constitutional law. On the freedom of research and constitutional law see Santosuosso (2012).

come into conflict if researchers try to deceive individuals, invade their privacy through covert research methods (e.g. through observation in non-public places), or by exacerbating individuals' vulnerabilities (e.g. when doing research with refugees or migrants) (European Commission 2010). Now, "where several fundamental rights are at issue, it is necessary to reconcile the requirements for the protection of those different rights and achieve a fair balance between them" (FRA 2020:76).

This demand to find a balance between the fundamental rights to data protection and to the arts and siences is incorporated in Art. 89 of the GDPR. This article covers the safeguards and derogations relating to the processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. It therefore acknowledges science and research as legitimate purposes of data processing, but only if they are subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. These rights and freedoms are on the one hand the fundamental rights and freedoms guaranteed by the CFR and on the other hand the individual data subjects' rights in Art. 12 et seq. Art. 89 (2) then concludes how these individual rights may be restricted by the European and national legislators. But we will get back to this issue at a later stage.

A practical solution to resolve the tension between researchers' rights and research participants' rights is to find a situation in which researchers and research participants agree on what the research endeavor is aiming at and what conditions the participation. This practical solution is called 'informed consent', and we will look at it when we discuss the lawful grounds for the processing of personal data. Next, we look at the GDPR itself and its surrounding legal framework.

4 The General Data Protection Regulation

The GDPR was adopted on 14 April 2016 and is applicable as of 25 May 2018. It was meant, among other things, to superse the Data Protection Directive dating from 1995. The Regulation comprises 99 articles and 173 so called recitals, which help to interpret the meaning of the articles. The Regulation holds about 150 so called "opening clauses" or exemptions, and they are leaving room for extensions or refinements by the European and national legislators. The GDPR is thus integrated into a framework of norms. If we talk about science and research the GDPR includes some exemptions which we will look at throughout and especially at the end of this paper.

It is important to recognize that while the GDPR applies directly and was aimed at harmonizing data protection in the European Union it is not the only law that applies when it comes to protection of personal data. The up to 150 opening clauses in the GDPR allow European and national legislators to pass laws or acts which link other laws or acts to the GDPR. This way for example national exemptions and traditions are perpetuated, and the result can be assessed as a "co-regulation and cohabitation of Union law and national law" as Roßnagel et al. (2018:10) found in their comparative evaluation. Others also find „significant differences" (Custers et al. 2017:1) and a „complicated nature of national adaptions" (Tambou 2019:26). [38] Nonetheless, the core of

[38] Gabel and Hickman (2019) present an overview of the variations between national implementations of GDPR in the EEA member states.

GDPR applies to all member states of the EU, and we briefly look at the legal situation in Germany in order to take a glimpse at the national level of data protection.

In the German case, the GDPR is supplemented by the Federal Data Protection Act (BDSG 2019) and 16 State Data Protection Laws. Several specific laws also apply like the Federal Statistics Act [39]when it comes to the processing of official statistical data, or the Telemedia Act [40]when it comes to the processing of personal data in online services.

Besides the harmonization of data protection in the European Union the GDPR in Art. 1 (1) and (3) also aims at free data traffic and at strengthening accountability of the people or institutions responsible for data processing (data 'controller' and 'processor'; Art. 5 (2)). The GDPR is committed to technological neutrality (Recital 15), which means that it does not mention or define any technologies which might be related to the processing of personal data. [41] And the controller and processor have to designate a data protection officer that is an independent body taking charge of compliance with data protection rules and regulations (Art. 37-39).

Another feature of the GDPR is that it envisages the concepts of 'privacy by design' and 'privacy by default' (Art. 25), which mean that the preservation of privacy should be an integral feature of any personal data processing activity from the very beginning on. The term 'privacy by design' was first coined in the 1970s and then formalized in the early years of the World Wide Web (EDPB 2018). Today it stands for mechanisms that should prevent unwanted and unauthorized processing of personal data without the knowledge of the persons concerned, such as tracking and profilingibid.; see EDPB 2020a for examples).[42]

One feature of the GDPR which has gained repeated attention by the media is the increase in the scope of penalties (Chapter VIII - Remedies, liability and penalties). According to Art. 83 (4) supervisory authorities[43] in the EU can impose administrative fines up to 10,000,000 EUR, or "in the case of an undertaking, up to 2 % of the total worldwide annual turnover of the preceding financial year, whichever is higher". This new instrument has widened the scope of action for Data Protection Agencies, and there are some spectacular recent cases[44] of fines at the upper level of the scale:

---

[39] https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/010a_BStatG_Engl.pdf?__blob=publicationFile (last accessed 9 May 2022).

[40] https://www.gesetze-im-internet.de/tmg/BJNR017910007.html /last accessed 9 May 2022).

[41] One exemption is the mentioning of "information society services" in Articles 8, 17 (1) (f), and 21 (5). But these services are not explicated any further.

[42] https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf (last accessed 9 May 2022).

[43] Examples of supervisory authorities are the Information Commissioner's Office (ICO) in the UK, the French data protection authority CNIL, or the Federal Commissioner for Data Protection and Freedom of Information in Germany.

[44] These and other cases can be found on the EDPB's news page at https://edpb.europa.eu/news/national-news/ (last access: 9 May 2022). That these fines might be contested in court is a different story.

- A fine of 20 million Pounds against British Airways for data breach affecting more than 400,000 customers by the British Information Commissioner's Office (October 2020).
- A fine of 35.3 million Euro for data protection violations against the fashion label H&M by the Data Protection Agency in Hamburg, Germany (October 2020).
- A fine of 12 million Euro against Vodafone for aggressive telemarketing practices by the Italian Data Protection Agency (November 2020).

4.1 Personal data

By adopting and expanding the definition of 'personal data' from the Data Protection Directive 95/46/EC the European legislator opted for a broad definition of what must be considered 'personal data'. This type of data is defined in Art 4 (1) of the GDPR:

> '(P)ersonal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

The breadth of what falls under 'personal data' is reflected in the term 'any information', which can refer to "objective" and "subjective" information (Article 29 2007:6).[45] A 'natural person' means a living person (Recital 27) and excludes legal persons like companies or associations (Recital 14). This 'natural person' might be 'identified' or 'identifiable' and "account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify [a] person" as the European Court of Justice (ECJ) ruled in its decision whether IP addresses have to be considered personal data or not (Breyer vs. Federal Republic of Germany; CJEU 2016, Case C 582/14).[46] This relative position is further defined in the next sentence, because a 'natural person can be identified 'directly or indirectly'. And here a non-exclusive list of factors is named which might be used for the purpose of identification. These factors are of interest for the purpose of research data management, because they must be handled with care and eventually must be removed temporarily or permanently. The factors are:

- names,
- identification numbers (e.g. social security numbers or tax numbers),
- location data (e.g. addresses or geo-codes),
- online identifiers (e.g. static IP addresses; Recital 30 of the GDPR also mentions "cookie identifiers or other identifiers such as radio frequency identification tags"), or

[45] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (last access: 9 May 2022).

[46] https://curia.europa.eu/juris/document/document.jsf?docid=184668&doclang=EN (last access: 9 May 2022).

- one or more factors "specific to
  - the physical,
  - physiological,
  - genetic,
  - mental,
  - economic,
  - cultural or
  - social identity of that natural person"

Only some of these 'specific' factors are directly defined in the GDPR while the meanings of others must be deducted e.g. from the Recitals. In the context of the GDPR physical, physiological or behavioral characteristics are mentioned in connection with "biometric data" like fingerprints and facial images (Art. 4 (14)). "'Genetic data' means personal data relating to the inherited or acquired genetic characteristics of a natural person" (Art. 4 (13)) and is related for example to biological samples (also see Recital 34). Issues of physical or mental health of a natural person are subsumed under "data concerning health" (Art. 4 (15)). This is related to information needed for health care services, or information on "disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state" (Recital 35) of a natural person. Finally, the economic, cultural or social identity of a natural person can be related to the concept of "profiling". In the context of the GDPR this means "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" (Art. 4 (4)). This "automated processing" is related to automated decision-making and is particularly sensitive to what is called 'special categories of personal data' in Art. 9.

As we see the list of information or data considered 'personal data' is quite broad and thus potentially covers a lot of research areas. While the GDPR is quite specific in some cases like the processing of fingerprints it remains rather vague in other cases like the economic, cultural or social identity of individuals. But there are also 'special categories' of personal data we look at next.

4.2 Special categories of personal data

Apart from personal data like names, addresses, ID's etc. the GDPR mentions "special categories of personal data" in Art. 9. These special categories relate to personal data revealing one or more of the following issues:

- racial or ethnic origin,
- political opinions,

- religious or philosophical beliefs,
- trade union membership,
- the processing of genetic data,
- the processing of biometric data for the purpose of uniquely identifying a natural person,
- the processing of data concerning health, and
- the processing of data concerning a natural person's sex life or sexual orientation.

We can easily recognize major scientific fields that deal with these 'special categories' of personal data like research on political and religious opinions, beliefs and attitudes, medical research, or demographic research in ethnically diverse cultures or political systems. The list of 'special categories' of personal data is closely related to Art. 21 of the CFR which talks about the prohibition of discrimination on any of these grounds. This is also made clear in Recital 75 of the GDPR which names most of the issues and explains some of the harms that might be inflicted on individuals to a varying degree, such as identity theft or fraud, financial loss, or damage to the reputation.

The collection and processing of personal data is only allowed when certain requirements are met. In the case of 'special categories' it is even prohibited unless there is a legal ground ('prohibition subject to authorization'). The requirements are subsumed under the concept 'lawfulness of processing'. But before we go into detail on this concept, we briefly look at three other topics that we have already mentioned: 'data controller', 'data processor' and 'processing'.

4.3 Controller, processor and data processing

The legal definitions of what the GDPR means by 'controller', 'processor' and 'data processing' are listed under Art. 4, but again we also have to turn to the recitals and other articles in order to find out e.g. about the obligations of the controller.

> "'(C)ontroller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data" (Art. 4 (7)).

The basic responsibilities of the controller are laid out in Chapter IV, and here basically in Art. 24(1): "the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation", based on the possible risks to the rights and freedoms of natural persons. He / she is also responsible and can be held accountable for compliance with the principles of processing personal data, like 'lawfulness', 'transparency' and 'confidentiality' in Art. 5 (Art. 5 (2)). We assume here that the researchers or the institutions on whose behalf they act are the controllers, and there might be two or more controllers who jointly determine the purposes and means of processing (Art. 26 (1)). But the challenge to determine the controller might become complicated

when it comes to e.g. health science in which we must also consider sponsors or commercial enterprises (EDPS 2020:14-16).[47]

The controller should consider privacy by design and by default, and many controllers like research institutes or universities are also expected to appoint a data protection officer as "a person with expert knowledge of data protection law and practices [who] should assist the controller or processor to monitor internal compliance with this Regulation" (Recital 97). Furthermore, the controller should run a record of processing activities in order to document his/her compliance e.g. with the principles in Art. 5 (Art. 30).

Externally, the controller has to file a notification to the supervisory authority in case of a privacy breach (Art. 33), meaning "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed" (Art. 4 (12)). Data subjects also have the right to file complaints with one of these supervisory authorities (Art. 77). And these authorities can for example impose fines as mentioned earlier.

> "'(P)rocessor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller" (Art. 4(8)).

The processor is thus in a situation dependent on the controller and his / her obligations are often laid out in a contract. The classical case of a processor is an IT company taking care of the controller's IT infrastructure. Processors in our context might also be fieldwork companies carrying out interviews, or repositories handling the long-term preservation of personal data.

> '(P)rocessing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction (Art. 4(2)).

This rather broad conception of 'processing' is due to the manifold digital environments researchers are working in. Processing is also based on the principles laid out in Art. 5. It is important to note that this definition refers to non-automated means of personal data processing, too, which means that a hand-written list of addresses for other than private purposes might also fall under the regulations of the GDPR.[48]

---

[47] https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf (last accessed 9 May 2022).

[48] See for example case C-25/17 of the ECJ (2018) on the collection of addresses and names by members of the religious community of Jehovah's Witnesses in Finland for preaching purposes. The EJC ruled that this collection must be considered processing of 'personal data', because the collection of this information does not necessaitly have to include data sheets, specific lists or other search methods (ibid. par. 62).

4.4 Lawfulness of the processing

According to the GDPR the processing of personal data is only allowed if there is a legal justification. This so called "lawfulness of processing" is laid out in Art. 6 of the GDPR and in this section we give an overview of the six legal justifications. 'Lawfulness of processing' is defined as follows:

> Processing shall be lawful only if and to the extent that at least one of the following applies:
>
> a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
>
> b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
>
> c) processing is necessary for compliance with a legal obligation to which the controller is subject;
>
> d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
>
> e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
>
> f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Examples for b) would be a work contract; for c) the processing of official statistical data by an statistical agency based on a legal act governing this processing; for d) any processing "which is essential for the life of the data subject or that of another natural person" (Recital 46); for e) public services of public interest like public health and social protection (Recital 45), situations of humanitarian emergencies (Recital 46), or the keeping of public registers (Recital 76); for f) processing for legitimate interests e.g. of the controller like fraud prevention, direct marketing, or IT security reasons (Recitals 47 to 49). In case of legitimate interests, the controller needs to balance his / her interests against the interests or rights of the individual subject to the processing. We look at consent more closely in the next section.

As we have seen, processing is conditional on the existence of one of these justifications. And this means that personal data processing not justified in any way is either subject to strict limitations or, in the case of the 'special categories' of personal data (Art. 9), even prohibited. These restrictions on data processing also have impacts on research every time that researchers possibly violate research participants' fundamental rights. Internet research and research using social

media data without researched individuals' consents are other areas (European Commission 2018:5-10).

4.5 Informed Consent

The lawful basis for the processing of personal data from natural persons in research probably used most often is informed consent (see Art. 6 (1) (a)). According to the GDPR this means the "freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" (Art. 4 (11). This kind of consent is important because it is also the major bridge to research ethics and was basically defined as early as 1947 in the Nuremberg Code although not being called "informed consent" at that time. This Nuremberg Code was later the basis e.g. for the Declarations of Geneva and Helsinki for medical research as well as the Belmont Report on biomedical and behavioral research (Faden et al. 1986).

The issue of informed consent is captured very broadly in the GDPR and we can only highlight some aspects here. We first focus on the four core attributes of informed consent as mentioned above: "freely given", "specific", "informed", and the "unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action", signifying, "agreement to the processing of personal data relating to him or her".

"Free" means that individuals have a choice to participate in a research project or not. Individuals should not be coerced to participate or be afraid to suffer disadvantages if they do not participate. Negatively defined, "consent will not be considered to be free if the data subject is unable to refuse or withdraw his or her consent without detriment" (EDPB 2020b:7).[49] In this context researchers also must consider possible imbalances of power between researchers and research participants (ibid.:7-13).

"Specific" means that the consent relates to one or more particular purposes. This specification is meant to limit what sometimes is called 'function creep', the limitation of unanticipated use of personal data. And the information for the processing of personal data should clearly be separated from information about other matters (ibid.:13-15).

"Informed" refers to the information provided by the controller. This information should be accessible in order to facilitate the control of research participants over their data. Based on the GDPR itself, one can distinguish six required items for a consent to be 'informed' (ibid.:15-16):

1. the controller's identity,
2. purpose of each of the processing operations for which consent is sought,
3. what (type of) data will be collected and used,

[49] https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf (last accessed 9 May 2022).

4. the existence of the right to withdraw consent,

5. information about the use of the data for automated decision-making in accordance with Art. 22 (2) (c) (3) where relevant, and

6. possible risks of data transfers due to absence of an adequacy decision and of appropriate safeguards as described in Art. 46.

This information must be presented in clear and plain language, which can be easily understood. The controller must be easily identifiable, and he/she needs to consider whom he/she is providing the information (e.g. children) to. While Art. 7 provides the conditions for consent Art. 13 and 14 mention information that needs to be provided where personal data are collected from the data subjectand where they have not been obtained from the data subject. Information like "the contact details of the data protection officer, where applicable", "the recipients or categories of recipients of the personal data, if any;", or "the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period" could be part of a document for informed consent. The European Data Protection Board believes that a valid consent can also exist without mentioning all elements listed in Art. 13 or 14 (ibid.:17). But considering the trust relationship between researchers and research participants one should broaden the scope of information in order to inform participants more extensively (see e.g. European Commission 2018:13-14).

Lastly, the informed consent should be an "unambiguous indication of the data subject's wishes", which means that it should be given by ways of a statement or by a clear affirmative act. Such an act precludes "silence, pre-ticked boxes or inactivity" (Recital 32). The GDPR does not prescribe any form on how this affirmative act shall take place, which means that an informed consent does not have to be given in writing. But independent of the form the controller "shall be able to demonstrate that the data subject has consented to processing of his or her personal data" (Art. 7 (1)). In any case, "the burden of proof will be on the controller" (EDBP 2020:22).

The GDPR places an emphasis on gathering data from children. The challenge when doing research with children is that they "may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data" (Recital 38). Thus, in order to protect this group of vulnerable individuals the GDPR includes children (and adolescents) in Art. 8, if the processing of their personal data is based on informed consent. In these cases, consent is valid if the individuals are 16 years or older. If they are below the age of 16 the parents must be involved to back the consent. National legislation may move the threshold to as low as 13 years or below, and controllers still need to adhere to national legislation which might set different thresholds.

The area of 'special categories' of personal data is another focus of the GDPR when using informed consent for data processing. The research participant must give "explicit consent" (Art. 9 (2) (a)) in case this type of data is gathered and processed. One of the other legal exemptions for the processing of 'special categories' of data is for "scientific or historical research purposes or statistical purposes" in accordance with the general exemption for science and research in Art. 89. Therefore, if researchers want to collect 'special categories' of data, they should mention them in the information provided to research participants.

Now, since human subject research is a very broad area there are different needs for informed consent depending on the disciplines involved, the research designs used, and the type of data being processed. While for example data for clinical trials requires a written consent on the basis of the EU Regulation on clinical trials on medicinal products for human use (Regulation 536/2014, here Art. 29 (1))[50], social researchers using qualitative methods might reject this idea as not feasible and counterproductive to their research settings (see for example, Wynn and Israel 2018). The principle problems and possible solutions of human subject research in many areas of the world, which also means gaining informed consent from potential participants, are reflected in the Ethics Guidelines of the European Commission for the Horizon2020 program (European Commission 2018b; especially pp.16-21).[51]

In any case, the GDPR does not prescribe any form of "informed consent" controllers such as researchers "are free to develop methods to comply with this provision in a way that is fitting in their daily operations" (EDPB 2020b: 22).[52] This may include electronic means, oral statements, ticking a box when visiting an internet website, or "another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data (Recital 32). Summers et al. (2020: 177-190) give a very good overview of this topic and examples of how consent is handled in social research.

One very important point for research projects that are based on informed consent is that a participant has the right "to withdraw his or her consent at any time" (Art. 7 (3) (1)). This means that researchers might face the fact that participants' data must be deleted also after the project has ended if the research data can be attributed to the data subjects. In this way and through other means, the GDPR has explicitly strengthened the rights of the 'data subjects', in our case the research participants. They do not only give a one-off consent or remain passive by-standers when it comes to the processing of their data. The next section looks at these rights of individuals.[53]

4.6 Rights of the data subjects

The data subjects, research participants in our case, have extensive rights in the GDPR. These are laid out in Articles 12 to 22. The main concept behind the rights of these individuals is transparency. This means that on the one side the controller has the duty to provide information "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" (Art. 12 (1) (1)). Furthermore, he must facilitate the exercise of data subject rights, provide information on action taken on a request, and inform the data subject without delay, free of charge (Art. 12). One the other side the data subjects have the right to access their information, demand

---

50 https://ec.europa.eu/health/sites/default/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf (last accessed 9 May 2022).

51 https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf (last accessed 9 May 2022).

52 https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf (last accessed 9 May 2022).

53 References to Art. 29 WG and EDPB were made explicitly in order to give readers with a novice level on data protection the chance to read up on the GDPR more easily.

rectification, corrections, and the deletion of their data. Thus, a system of checks and balances is set up in principle, which is supported by several instruments strengthening those rights.

In detail, the data subjects have the rights:

- to receive information from the controller where personal data are collected from him / her (Art. 13),
- to receive information from the controller where personal data have not been obtained from the data subject (Art.14),
- to access the data (Art.15),
- to rectification (Art.16),
- to erasure ('right to be forgotten') (Art. 17),
- to restriction of processing (Art. 18),
- to receive information regarding rectification or erasure of personal data or restriction of processing (Art. 19),
- to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format (Art. 20),
- to object (Art. 21), and
- not to be subject to a decision based solely on automated processing, including profiling (Art. 22).

While these rights empower the data subjects in general, they might be restricted by the Union or Member State law for example to safeguard national security, defence, or public security (Art. 23 (1)). More importantly for us, further restrictions come into play when it comes to research: "Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21" (Art. 89 (2)). This concerns the rights to access, rectification, restriction of processing, and to object.

Furthermore, where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 19 and 20 (Art. 89 (3)), meaning receiving information regarding rectification or erasure, and receiving a copy of personal data on request.

So, the data subjects may have extensive rights in the GDPR, but these might be restricted and might also vary from country to country. This underlines the point we made earlier: concerning the complex legal situation created by the GDPR. It is therefore important to closely look at national legislature besides the Regulation when planning research projects.

4.7 Risk-based approach and Data Protection Impact Assessment

The GDPR follows a 'risk-based approach' (von Grafenstein 2018:79-82). While this concept is not uniquely defined in the Regulation, von Grafenstein (ibid.) names Art. 24 (1), which defines the responsibilities of the controller, as the "central provision" for this approach. The "varying risks" mentioned here relate to the possible harms that might be inflicted on individuals, and the GDPR distinguishes between different levels or a continuum of risk. While the Regulation talks about "possible risks to the rights and freedoms of natural persons", it also knows a "high risk to the rights and freedoms of natural persons" (Art. 35 (1)) on the one side and supposedly little or no risk on the side of 'anonymized' data (Recital 26). For a clarification of the differences between 'risk' and 'high risk' we must rely on the different sources.

The Article 29 Working Party (2017:15) defines 'risk' as "a scenario describing an event and its consequences, estimated in terms of severity and likelihood."[54] The definition follows the classical view on the issue (see for example Gigerenzer 2014) and includes as components of risk possible events, the likelihood of occurrence of these events as well as the consequences of these events, meaning the possible harms and damages. Some of these like identity theft, fraud, financial loss, or damage to the reputation are mentioned under Recital 75. Of special interest is the category of 'high risk', because this makes it necessary for a controller to carry out a Data Protection Impact Assessment (DPIA).

A DPIA "is a process designed to describe the processing [of personal data], assess the necessity and proportionality of a processing and to help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data (by assessing them and determining the measures to address them)" (Article 29 2017:4).[55] The DPIA is thus categorized as a process which assesses a risk with relation to the rights and freedoms of natural persons and proposes protective means. Based on Art. 9 and 35, and Recitals 71 and 91 of the GDPR the Working Party lists nine categories of personal data processing which can be taken as criteria for a DPIA in order to meet the challenges of a high-risk processing situation. In the working Party's opinion, a DPIA becomes necessary if two or more of these criteria are met:

1. Evaluation or scoring, including profiling and predicting,
2. automated-decision(sic!) making with legal or similar significant effect,
3. systematic monitoring such as Closed-Circuit Television (CCTV),
4. the processing of sensitive data (as defined in Art. 9),
5. data processed on a large scale,
6. datasets that have been matched or combined,
7. data concerning vulnerable data subjects,

[54] https://ec.europa.eu/newsroom/just/redirection/document/47711 (last accessed 9 May 2022).

[55] https://ec.europa.eu/newsroom/just/redirection/document/47711 (last accessed 9 May 2022).

8. innovative use or applying technological or organizational solutions, or
9. when the processing in itself "prevents data subjects from exercising a right or using a service or a contract" (Art. 22 and recital 91).

A DPIA is undertaken by the controller who is supposed to get in touch with the appropriate data protection officer. The controller needs to document the assessment as part of his or her responsibility to demonstrate compliance with the principles of personal data processing. According to Art. 7 (7) (a)-(d) the assessment should contain at least a systematic description of the envisaged processing operations and the purposes of the processing, an assessment of the necessity and proportionality of the processing, an assessment of the risks, and the measures envisaged to address the risks. And if personal data is processed, technical and organizational measures must be in place to protect this data from being misused.

4.8 Technical and organizational measures

Data controller and processors are supposed to safeguard the personal data they handle, and risks for personal data and thus the data subjects should be reduced. For this purpose, so called "technical and organizational measures" (TOMs in short) were incorporated into the GDPR which support the processing of personal data. Art. 32 (1) (b) defines that the TOMs should ensure a level of security including among other things "the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services".[56]

Important about the TOMs is that they shed light on what Mackey and Elliot (2013) call the "data environment". This comprises the data itself, agents, and infrastructures, like statistical offices. The data is often processed by several agents and in many research projects also in international environments.[57] If for example an informed consent was used to collect the data, certain commitments were made to the data subject, and these commitments must be implemented as part of the data management. Furthermore, if the data is supposed to be published after the research project has ended, this data must very likely be made accessible in a way which minimizes the re-identification risk for research participants. And this might mean to restrict access to the data.

Art. 32 and Recital 78 give a basic overview of some TOMs. The controller should for example "adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default" (Recital 78). This statement refers to Art. 25 of the GDPR, as mentioned earlier, and means that data protection should be implemented from the very beginning on. The controller should also minimize the processing of personal data, which relates to the idea to reduce data processing to a necessary level (Art. 5 (1) (c)). Furthermore, the functions and processing of personal data should be transparent, enabling the data subject to

[56] Except of the resilience of processing systems and services, this definition is identical to the definition of information security in the appropriate ISO/ IEC standard on information security, which "ensures the confidentiality, availability and integrity of information" (ISO/IEC 27000:2018, p.12). https://www.iso.org/isoiec-27001-information-security.htm (last accessed 9 May 2022).

[57] In international contexts it is necessary to clarify where the and how the GDPR applies (EDPB 2019b).

monitor the data processing. And the controller is supposed to create and improve security features, like encryption (Art. 32 (1) (a)). What does this all mean in practice?

It is obvious that the protective aims of the GDPR and the related TOMs are manifold. In order to get an overview the German Data Protection Supervisory Authorities have put together a "Standard Data Protection Model" (SDM), which "provides appropriate measures to transform the regulatory requirements of the GDPR to qualified technical and organisational measures" (SDM 2020:6).[58] The SDM lists 23 key data protection requirements of the GDPR which can be aligned with seven protection goals that show up e.g. in Art. 5 of the GDPR:

Data minimization;

- availability;
- integrity;
- confidentiality;
- unlinkability;
- transparency and
- intervenability.

The SDM also lists practical implementations for the protection goals, like access restrictions, authorization rules, or controls for processing activities. These are ordered by the protection goals.[59]

Now, if we focus on the data, pseudonymization and anonymization are the two of the TOM's which are very likely to be used in research projects. They aim at reducing the level of information in the data. While 'pseudonymization' basically means to replace some parts of a dataset or database by some sort of key and keeping the remaining part as well as the "truncated" part, 'anonymization' aims at making this replacement permanent. Both concepts are thus related, but while different forms of anonymization of personal data are widely used by researchers across multiple disciplines, the GDPR strongly focusses on pseudonymization.

4.9 Anonymization

Anonymization, meaning the permanent reduction of information in a dataset, can be regarded as a way forward for researchers to handle data collected from individuals, because data which cannot be used to identify a person behind the data is not subject to the GDPR (Recital 26 sent. 5 and 6). But as decades of research into anonymization as a means of research data management have shown, anonymization comes at a cost. The more information is taken out of a dataset or database the more the value of this data for research in general is reduced (Elliot et al. 2016).

---

[58] https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V2.0b.pdf (last accessed 9 May 2022).

[59] For the sake of brevity, we do not go into detail on the practical means proposed in the SDM.

The GDPR defines 'anonymization' in Recital 26 in such a way that the principles of data protection no longer apply to "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". The Regulation does not define precisely what 'anonymous' means, but Recital 26 names "objective factors" which can help to identify means that are likely to be used to identify the natural person. These 'objective factors' might be "the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."

These provisions correspond for example to the approach laid out in the Federal Statistical Act of Germany (last changed 2020). The Federal Statistical Office can make data available off-premise, for example to universities, "if the individual data can only be allocated with a disproportionately large amount of time, costs and manpower (de facto [or 'factually'] anonymized individual data)" (BStatG §16 (6)).[60] This conception of defining anonymity not in absolute terms was perpetuated by the Article 29 Working Party (2014), which preceded the European Data Protection Board until 2018. And the ECJ defined anonymity in the same manner in its decision on whether dynamic IP addresses constitute personal data (CJEU 2016, Case C-582/14).[61]

This lengthy explanation highlights the idea that 'anonymity' might not just mean 'absolute' anonymity but refers to a 'relative' concept. 'Absolute anonymization' would mean that the risk of re-identifying a data subject is basically zero or not existent after the anonymization has taken place. And here we can make a connection to the risk-based approach of the GDPR. In fact, the literature talks about at least three scenarios which are relevant to researchers (for the following see Elliot et al. 2016:17-22 and Höhne 2010):

1. 'Formal anonymization', which means any process that removes or masks direct identifiers from a dataset,
2. 'absolute' or 'guaranteed' anonymization, which means a form of anonymization where the risk of re-identification is zero, and
3. 'factual' or 'statistical' anonymization, which mean that it is basically impossible to reduce the risk to zero and to control or limit the risk of disclosure events instead.

'Formal anonymization' usually means that direct identifiers like names, addresses, and IDs are removed from the dataset collected for analysis. This separation is supposed to be done as soon as the research process allows for it, but the direct identifiers do not have to be deleted at this stage. It is probably enough to pseudonymize the data and delete the identifiers e.g. when the project ends.

But it is often not enough to simply take away direct identifiers. 'Factual' or 'statistical' anonymization go one step further and consider indirect identifiers and contextual issues. Indirect

[60] BStatG = Federal Statistics Act. The concept of 'factual anonymity' was part of the German Federal Data Protection Act (§3 (6)) until 2018. Here, anonymization meant "changing personal data in such a way that individual details about personal or factual circumstances can no longer or only with a disproportionately large expenditure of time, cost and labor be attributed to a specific or identifiable natural person".

[61] https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf (Last accessed 9 May 2022)

identifiers might be occupations, geographic information on residence, or any topic regarded as special category in the GDPR, like political opinions or sexual orientation. Elliot et al. (2016:138-148) present a series of standard key variables combined in 19 scenarios, which might serve as a starting point for considerations on anonymization in the framework of research data management. The important point is that researchers must consider the data environment and existing contextual data or information which might serve to re-identify individuals despite the reduction of information undertaken to reach formal anonymization.

Finally, 'absolute' or 'guaranteed anonymization' is at the extreme end of the anonymization spectrum. It is usually reached by aggregating data into clusters or categories. The presentation is than done e.g. in summary tables based on frequencies of those clusters and categories. This is what is usually done for publicly available tables and publications by statistical offices (see e.g. Elliot et al. 2016:43-60).

As we have seen the GDPR itself does not really define what is meant by 'anonymous' data. But when it comes to data used for research, Art. 89 (1) of the Regulation makes an indirect reference to anonymization. It focusses on TOM's that guarantee for the rights and freedoms of the data subjects and aim at data minimization. Here, pseudonymization is mentioned as a possible TOM, but the article goes on:

> "Where those purposes [of safeguarding for the rights and freedoms of the data subject] can be fulfilled by further processing [personal data] which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner."

If 'personal data' is information relating to an identified or identifiable data subject (Art. 4 (1)) and the personal data can be considered anonymous, because it is processed in a way that the data subject is not or no longer identifiable (Recital 26), then Art. 89 clearly advocates anonymization as a TOM for research data management.

While 'anonymization' aims at permanent changes to the data, 'pseudonymization' leaves open the option to re-join direct or indirect identifiers with the 'pseudonymized' data. The latter therefore must be considered personal data if both parts of the original database exist. Since this concept has become quite prominent in the GDPR we look at it next.

4.10 Pseudonymization

The GDPR strengthens the role of 'pseudonymization'. It is probably the most prominent of the TOMs in the Regulation (see e.g. Art. 25 (1), Art. 32 (1), and Art. 89 (1)). 'Pseudonymization' is defined in Art. 4 (5) as the "processing of personal data in such a way that the personal data can no longer be attributed to a specific data subject without the involvement of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person". The definition is not clear since it does not say what is meant by "processing" in this case. But usually 'pseudonymization' refers to data separation which practically means the replacement of what is called "additional data" above with a pseudonym.

Pseudonyms are "fictitious name[s]" according to Merriam-Webster or, technically speaking, IDs which allow for the re-joining of the reduced data and the "additional data".

This data separation is similar to the notion of 'factual' or 'statistical anonymization' mentioned earlier with the difference that the "additional data" is still available. But while anonymization aims at permanent changes to the data, the concept of 'pseudonymization' only refers to the separation. And this means that in legal terms pseudonymized data, according to the Regulation, remains personal data (STOA 2019:27)[62]. The "additional data" separated from the actual data for analysis should further be protected by TOMs like encryption.

4.11 Research exemptions in the GDPR

The GDPR considers scientific research a special category of data processing (STOA 2019:20) and has privileged scientific research in several ways. Art. 89 looks at "safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" and these safeguards should aim at data minimization as laid out in Art. 5 (3). Six other articles spell out exemptions for historical and scientific research as well as for archiving and statistics. They concern purpose and temporal limitations of data processing, the processing of special categories of personal data, the information of data subjects in case data have not been obtained from them, the right to erasure ('right to be forgotten'), and the right to object.[63] These exemptions are all closely connected to Art. 89 (1), which means that if exemptions are made they are only valid if they also include "safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject" like TOMs. (Other articles of the GDPR may also be relevant in this respect.)

For the purpose of this paper we only look at these exemptions spelled out in these six articles of the GDPR mentioned above.[64] Furthermore, the GDPR covers four concepts relevant for 'research': "archiving purposes in the public interest, scientific or historical research purposes or statistical

[62] https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf (last accessed 9 May 2022)

[63] There are also 17 Recitals referring more or less to research purposes as laid out in Article 89, some of which have been mentioned in this text: No.26 GDPR not applicable to anonymous data; No.33 consent to certain areas of scientific research; No.50 further processing of personal data; No.52 exemptions to the prohibition on processing special categories of personal data; No.53 processing of sensitive data in health and social sector; No.62 exemptions to the obligation to provide information; No.65 right of rectification and erasure; No.71 profiling; 113 transfers qualified as not repetitive and that only concern a limited number of data subjects; No.156 archiving, scientific or historical research or statistical purposes; No.157 information from registries and scientific research; No.158 processing for archiving purposes; No.159 processing for scientific research purposes; No.160 processing for historical research purposes; No.161 consenting to the participation in clinical trials; No.162 processing for statistical purposes; No.163 production of European and national statistics.

[64] Article 89 (2) and (3) furthermore open possibilities for the European and national legislators to restrict data subjects' rights for the purposes of scientific research and archiving in the public interest. Most EEA member states used this opportunity and imposed additional conditions. The nature of these conditions varies substantially, and only Bulgaria, Hungary, Liechtenstein, Lithuania and Romania do not impose further restrictions for the processing in these cases (Gabel and Hickman 2019).

purposes". If not indicated otherwise we always talk about these four concepts when we refer to 'research'.

We have already seen that the Regulation makes exemptions for the scientific processing of 'special categories' of data in Art. 9 (2) (j) and alleviates the principle prohibition of this type of processing when it comes to scientific research and archiving. This exemption for research does not only refer to Art. 89 but also to Union or Member State law.

Art. 5 (1) (b) specifies that personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". An example of a purpose limitation is the use of personal data for a company's communication with its customer, usually laid out in the company's terms of service or data protection statement. The individual should be able to rely on the information he / she received about the original data processing purpose. The only exemption that the GDPR make is for 'research': "further processing […] shall, in accordance with Art. 89 (1), not be considered to be incompatible with the initial purposes ('purpose limitation')". That means for example that data collected e.g. for market research may later be used for other 'research' purposes.

A similar exemption is made in Art. 5 (1) (e) on 'storage limitation'. While personal data should basically only be kept in a form "which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed", research can rely on longer time periods. But in the latter cases, the personal data may only be stored for 'research' purposes and with the appropriate safeguards mentioned in Art. 89 (1).

Sometimes personal data is not collected from the data subject him- or herself, but from a relative, a friend, or a colleague.[65] In these cases, the individuals must be informed about the processing extensively (unless they have been informed otherwise). Art. 14 (5) (b) makes the exemption for 'research' that data subjects do not have to be informed, if "the provision of such information proves impossible or would involve a disproportionate effort […] or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available". According to the Article 29 Working Party (2018)[66] to 'render impossible' or to 'seriously impair' mean that "if a data controller seeks to rely on this exemption it must demonstrate the factors that actually prevent it from providing the information in question" (ibid.:29). That means, the data controller must try to provide the information to data subjects. "One appropriate measure, as specified in Art. 14 (5) (b), that controllers must always take is to make the information publicly available" (ibid.:31) e.g. via a website.

[65] Examples are interviews about household members, personal networks, or workplace situations.

[66] https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51025 (last accessed 9 May 2022).

A similar exemption is made in the case of the right to erasure (or 'right to be forgotten'). Art. 17 (3) (d) states that this right is restricted in cases of 'research', "in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing". Again, this means that the data controller must demonstrate and document that the erasure of data would limit his / her research opportunities.

Finally, Art. 21 (6) restricts the data subjects' right to object to the processing of their personal data for 'research' purposes (excluding 'archiving purposes in the public interest'), if "the processing is necessary for the performance of a task carried out for reasons of public interest". As we have seen before this might mean processing for service purposes of public interest like public health, social protection, or situations of humanitarian emergencies.

To sum up, the GDPR provides some exemptions for 'research', which are all closely related to Art. 89 (1). This means for example that data collected for other purposes might be further processed in 'research' contexts, that personal data might be kept longer for 'research' than for other purposes, or that data subjects rights might be restricted when it comes to the right to be informed, the right to demand the erasure of one's data or the right to object to personal data processing. But as we have seen this is all narrowly defined and does not mean that research is privileged in a way that it can simply overrule individuals' fundamental rights concerning their personal data.

5 Conclusion – the research data lifecycle and data protection

Human subject research data is primarily used for scientific purposes, but there is a risk involved in the processing of this data. The processing might have unintended or negative consequences for the research participants ('data subjects') like identity theft or fraud, financial loss, or damage to the reputation. It is the aim of data protection and especially of the GDPR to minimize the risk of these consequences from happening and thus to protect the fundamental rights of the data subjects. The Regulation is based on the CFR and part of a framework of norms including the fundamental rights on data protection, other EU regulations like that on clinical trials, and numerous national laws. Although it was originally meant to offer one harmonized data protection legislation for all EU member states the many exemptions opened paths for national governments to preserve legal traditions and thus to hollow out the aim of harmonizing. Nevertheless, the GDPR is directly applicable and offers a common data protection core for the EU and the EEA.

We saw that the GDPR offers a broad definition of what must be considered 'personal data' and thus the Regulation potentially covers a lot of research areas. From the rather vague definition of anonymization one can conclude that the European legislator wanted to include all data except for those "which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". The Regulation also knows 'special categories of personal data' like racial or ethnic origin, political opinions, or data concerning health, which have been included based on the prohibition of discrimination in the CFR. Since these special categories of data are the basis of many human subject research areas, like medical, demographic or public opinion research extra safeguards need to be in place in order to protect research participants from potential harms.

It is the controller and potentially one or more processors that bear the responsibilities for safeguarding for the rights and freedoms of data subjects when processing their personal data. All processing activities need a legal ground like a law or a contract, and in the case of 'special categories' of personal data the processing is even explicitly prohibited unless there is such a legal ground. The ground probably most prominent in human subject research is informed consent, the "freely given, specific, informed and unambiguous indication" that a data subject is for example willing to participate in a research project. It is up to the controller to provide the necessary information in clear and plain language, which can be easily understood, and although the GDPR does not prescribe a form for consent, the active approval must be documented. And irrespective of the content and purpose of the consent the data subject is endowed with a range of rights including the one to withdraw his or her consent to the data processing at any time.

As we also saw, the GDPR follows a risk-based approach, which means that the possibilities of harms or harmful events need to be considered and the risks need to be reduced by the controller and the processors. In the case of high risks during data processing, the controller needs to undertake a Data Protection Impact Assessment in order to manage the risks. Technical and organizational measures (TOMs) are the tools to handle and possibly reduce those risks. 'Anonymization' and 'pseudonymization' are of special importance to research, because these TOMs were introduced to temporarily or permanently separate actual research data from all information like addresses which is most likely not needed for scientific analysis. And anonymous data is not subject to the GDPR.

Finally, the Regulation makes several exemptions for research. The exemptions are all closely connected to the personal data processing for "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" in Art. 89. In the case of 'purpose limitation' the further processing of data for 'research' is allowed. The 'storage limitation' is also alleviated and personal data can be processed longer for research than for other purposes. A similar relieve takes effect in the case of the processing of 'special categories' of data, but this exemption also relies on Union or Member State law. If personal data is not directly collected from the data subject, the controller or processors are exempt from informing the data subject in case the information involve a disproportionate effort, or is likely to render impossible or seriously impair the achievement of the objectives of the processing. In any case the controller should make the information about the data processing publicly available, e.g. on a website. A similar exemption is made in the case of the right to the erasure of the data. And finally, the GDPR restricts the data subjects' right to object to the processing of their personal data for 'research' purposes, if "the processing is necessary for the performance of a task carried out for reasons of public interest", like public health, social protection, or in cases of humanitarian emergencies.

If we go back to the six stages of the data lifecycle, we can collate those basic data protection topics and TOMs mentioned above to human subject research projects.[67]

[67] We have not talked about research data management in detail. The reader should refer to Corti et al. (2020), or the 'Data Management Expert Guide' of the Consortium of European Social Science Data Archives (CESSDA; www.cessda.org) for details.

The first stage, we said, covers the planning and design of the research project. Here decisions must be made about methods, sampling of individuals, the kind of data that will be collected etc. This planning stage often covers a data management plan (DMP), as requested e.g. by funding agencies. And this DMP should cover the practical tasks of data protection which need to be carried out throughout and after the research project. It also names responsibilities, workflows and tools. During the second stage the actual data collection and data preparation for analysis take place. This is the stage in which legal grounds like informed consent come into play and the TOMs like the separation of research data and for example contact data takes place. The access to this data should be safeguarded and personal data should be stored securely.

In the third stage researchers work with "their" data and publish results. Data most likely must be pseudonymized and direct identifiers like addresses should be encrypted. During stage four the data is likely to be set up for long-term preservation and publishing or releasing. The DMP should cover the actual data which can be preserved and published, and this stage also covers activities like format migration and data documentation. Most importantly it might comprise a further reduction of the information in the data via anonymization for example. The result of this stage is a package of data, metadata and documents, which allows third parties to re-use the data. The long-term preservation and publishing of the data is usually done by data repositories or data archives. Controllers like researchers must be clear about the rights they hold to the final data which is supposed to be published. Particular rights need to be transferred to data repositories or data archives in order for the data to be published, and the infrastructures also take over tasks to control the access to the data if necessary, and to promote the data for re-use.

After their publication or release the data can be used by other researchers in order to re-analyze published results, to answer new research questions, or to use the data for training and teaching activities. In the case of data collected from human individuals this regularly means that data access should be "as open as possible, and as closed as necessary" – the general approach of the European Commission 'Guidelines to the Rules of Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020'. This guiding principle envisages easy access, but also possible access restrictions, because researchers must always weigh their own fundamental rights against those of the research participants. In this case the reduction of information in a dataset, particularly the possible anonymization of the data is an important way forward when it comes to data access.

Research stage	Legal obligation	Data	TOM
Plan and design	Legal framework (GDPR; possibly other international or national laws)	Sampling-design	RDM plan; Allocation of roles and responsibilities; plan for data protection; set up TOMs
Collect and capture	Lawfulness of processing; informed consent in many cases	Contact data; possibly other direct identifiers; research data	Restrict access to data storage locations and data
Interpret and analyse	Publish only results covered by legal basis of processing	Contact data; possibly other direct identifiers; research data	Pseudonymization; restrict access to data storage locations and data; possibly use encryption for contact information etc.)
Manage and preserve	Reduce information in data; deletion of data not for research (e.g. addresses)	Research data	Reduce information in data (e.g. anonymization); define usage rights and data access (e.g. for repository)
Release and publish	Transfer of usage rights to e.g. repository; informed consent should not prevent data sharing;	Research data for secondary use	
Discover and reuse	Usage regulations; authentification and authorization	Research data for secondary use	Usage regulations; authentification and authorization

Figure 10: Data lifecycle stages – RDM and data protection

6 References

Article 29 Working Party. 2007. *Opinion 4/2007 on the concept of personal data 01248/07/EN WP 136*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf. (last accessed 6 September 2021).

Article 29 Working Party. 2014. *Opinion 05/2014 on Anonymisation Techniques 0829/14/EN WP216*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. (last accessed 6 September 2021).

Article 29 Working Party. 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of

Regulation 2016/679, 17/EN WP 248 rev.01. https://ec.europa.eu/newsroom/just/redirection/document/47711. (last accessed 6 September 2021).

Article 29 Working Party. 2018. *Guidelines on transparency under Regulation 2016/679 17/EN WP260*. https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51025. (last accessed 6 September 2021).

Ball, Alex. 2012. *Review of Data Management Lifecycle Models.* Barth, UK. https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf. (last accessed 6 September 2021).

Bundesministerium der Justiz und für Verbraucherschutz. 2019. *Federal Data Protection Act (BDSG)*. https://www.gesetze-im-internet.de/englisch_bdsg/. (last accessed 9 Juni 2021).

Code of Federal Regulations. 2018. *§ 46.102 Definitions for purposes of this policy. Subpart A - Basic HHS Policy for Protection of Human Research Subjects*. https://www.law.cornell.edu/cfr/text/45/46.102. (last accessed 6 September 2021).

Corti, Louise, Veerle van den Eynden, Libby Bishop, Matthew Woollard, Maureen Haaker, and Scott Summers (eds.). 2020. *Managing and Sharing Research Data: A Guide to Good Practice.* Los Angeles, London: SAGE Publications.

Court of Justice of the European Union. *Case C-582/14, Patrick Breyer vs. Bundesrepublik Deutschland*. https://curia.europa.eu/juris/document/document.jsf?docid=184668&doclang=EN. (last accessed 6 September 2021).

Custers, Bart, Francien Dechesne, Alan M. Sears, Tommaso Tani, and Simone van der Hof. 2018. A comparison of data protection legislation and policies across the EU. *Computer Law & Security Review* 34 (2): 234–243. https://doi.org/10.1016/j.clsr.2017.09.001.

Elliot, Mark, Elaine Mackey, Kieron O'Hara, and Caroline Tudor. 2016. *The Anonymisation Decision-Making Framework.* Manchester: Ukan Publications.

European Commission. 2018a. *Ethics and data protection*. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf. (last accessed 6 September 2021).

European Commission. 2018b. *Ethics in Social Science and Humanities*. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf. (last accessed 6 September 2021).

European Data Protection Board. 2018. *Guidelines 3/2018 on the territorial scope of the GDPR (Article 3)*. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_3_2018_territorial_scope_after_public_consultation_en_1.pdf. (last accessed 6 September 2021).

European Data Protection Board. 2020a. *Guidelines 4/2019 on Article 25, Data Protection by Design and by Default Version 2.0*. https://edpb.europa.eu/sites/default/files/files/file1/edpb_

guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf. (last accessed 6 September 2021).

European Data Protection Board. 2020b. *Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1*. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf. (last accessed 6 September 2021).

European Data Protection Supervisor. 2020. *A Preliminary Opinion on Data Protection and Scientific Research*. https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf. (last accessed 6 September 2021).

European Parliament. 2014. *Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC*. https://ec.europa.eu/health/sites/default/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf. (last accessed 6 September 2021).

European Union Agency for Fundamental Rights (FRA). 2020. *Applying the Charter of Fundamental Rights of the European Union in law and policymaking at national level: Guidance.* Luxembourg: Publications Office of the European Union.

European Union Agency for Fundamental Rights, Council of Europe (FRA). 2018. *Handbook on European Data Protection Law 2018.* Luxembourg: Publications Office of the European Union.

Faden, Ruth R., Nancy M. P. King, and Tom L. Beauchamp. 1986. *A history and theory of informed consent.* New York: Oxford University Press.

Gabel, Detlev, and Tim Hickman. 2019. GDPR Guide to National Implementation: A practical guide to national GDPR compliance requirements across the EEA. https://www.whitecase.com/publications/article/gdpr-guide-national-implementation. (last accessed 6 September 2021).

Gigerenzer, Gerd. 2014. *Risk Savvy: How to Make Good Decisions.* New York: Penguin Books.

Grafenstein, Maximilian von. 2018. *The principle of purpose limitation in data protection laws*. Baden-Baden: Nomos.

Höhne, Jörg. 2010. *Verfahren zur Anonymisierung von Einzeldaten.* Wiesbaden: Statistisches Bundesamt.

Mackey, Elaine, and Mark Elliot. 2013. Understanding the Data Environment. *XRDS: Crossroads, The ACM Magazine for Students* 20 (1): 36–39. https://doi.org/10.1145/2508973.

Neethling, Johann. 2005. Personality Rights: A comparative Overview. *Comparative and International Law Journal of Southern Africa* 38 (2): 210–245. https://hdl.handle.net/10520/AJA00104051_58.

Organisation for Economic Co-operation and Development. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding.* Paris: OECD.

Panel for the Future of Science and Technology. 2019. *How the General Data Protection Regulation changes the rules of scientific research.* Brussels: European Parliamentary Research Service.

Roßnagel, Alexander, Tamer Bile, Michael Friedewald, Christian Geminn, Olga Grigorjew, Murat Karaboga, and Maxi Nebel. 2018. *National implementation of the general data protection regulation: Challenges, approaches, strategies.* Karlsruhe: Fraunhofer-Institut für System- und Innovationsforschung.

Santosuosso, Amedeo. 2012. Freedom of Research and Constitutional Law: Some Critical Points. In *Scientific Freedom*, ed. Simona Giordano, John Coggon and Marco Cappato, 73–82. London: Bloomsbury Publishing.

Standard Data Protection Model. 2020. *The Standard Data Protection Model: A method for Data Protection advising and controlling on the basis of uniform protection goals, Version 2.0b.* Kiel: AK Technik of the Independent Data Protection Supervisory Authorities of the Federation and the Länder.

Summers, Scott, Libby Bishop, Veerle van den Eynden, and Louise Corti. 2020. Legal and Ethical Considerations in Sharing Data. Managing and sharing research data: A guide to good practice. In *Managing and Sharing Research Data: A Guide to Good Practice*, ed. Louise Corti, Veerle van den Eynden, Libby Bishop, Matthew Woollard, Maureen Haaker and Scott Summers, 159–195. Los Angeles, London: SAGE Publications.

Tambou, Olivia. 2019. Opening Remarks: National Adaptations of the GDPR. In *National Adaptations of the GDPR*, ed. Karen McCullagh, Olivia Tambou and Sam Bourton, 24–27. Luxembourg: Collection Open Access Book.

Wynn, L. L., and Mark Israel. 2018. The Fetishes of Consent: Signatures, Paper, and Writing in Research Ethics Review. *American Anthropologist* 120 (4): 795–806. https://doi.org/10.1111/aman.13148.

Problematic Practices and New (Digital) Avenues in Copyright and Intellectual Property Rights in African Music

Markus Coester (University of Bayreuth)[68]

Abstract

This paper focuses on intellectual property rights and copyright in African popular music, and particularly practices associated with it in Ghana and related to Ghanaian music in recent times. It draws on some years of research and direct involvement in arts promotion in Ghana. Based on this I investigate on the ways actors and systems in place have dealt with these rights. Current debates on copyright I reflect on are centred in the fast-changing digital world of music and mainly related to the supposed (mis)management of music rights through the representational national royalty collecting body and copyright society. They help to chart the prevailing issues and concerns in *rights matters* and reveal a new artist sensibility towards them. An ethnographic perspective on these issues substantiates that music business practices related to copyright and intellectual property as experienced by music-makers in Ghana, are ethically and legally problematic and frequently not beneficial to creative artists. This explains why many musicians get the feeling that *rights are not right(s)*. To them, the digital world of music, that is of online music distribution, streaming, and internet-based consumption patterns, however, seems to promise a change for the better while digital accessibility potentially creates new *rights avenues* also.

[68] Contact: Markus.Coester@uni-bayreuth.de, https://orcid.org/0000-0002-1357-122X.

1 Introduction

This paper focuses on intellectual property rights (IPR) and copyright in African popular music, and particularly on practices and issues in Ghana and Ghanaian music related to them. Over the years, my research on popular music, close contact with artists, involvement in music promotion, and music archive work – including digitalisation – have all reemphasised the significance of this topic and made me reflect on it more and more. The scholarly discourse on music copyright took strong roots in the seminal Music and Copyright, edited by Simon Frith and Lee Marshall, first published in 1993. In addition to my own experiences in the field of live music and concert tour organisation and facilitation, radio and record label-related work and research, their revised edition of 2004 has much informed my approach to the issues of copyright and IPR (Frith and Marshall 2004).

Fortunately, there is a growing scholarly awareness of and interest in copyright, intellectual property rights, and related subjects, and how these affects immaterial culture in Africa (Röschenthaler and Diawara 2016a). On popular music in South Africa, Erlmann's analysis of the need for an alternative concept and form of intellectual property as to African music and his conceptualisation of an 'original-cover dialectics' are thought-provoking (id.: 97f.). "IP expansionism" and the "wholesale expansion of IP law" (id.: 97) was prepared during the colonial era but the increasing imposition of Euro-American IP law and practices on Africa since the 1990s mainly came through the so-called TRIPS agreement of 1994 (Trade-Related Aspects of Intellectual Property), a "dramatic paradigm shift in international copyright regulation which has seen copyright become embedded in a wider network of trade relations" (Frithand Marshall 2004: 13; see also Peukert 2016).

Partly, at least, this shift led Ghanaian musicians, artists, and musical entrepreneurs, situated in a music market that like others in Africa (see Tcheuyap 2016) from the 1980s latest suffered from piracy (Wallis and Malm 1984), increasingly claim that their compositions need to be protected somehow for them to gain revenues. The wish for a Collective Management Organisation (CMO) that works and does so by considering local notions of property and ownership different from Euro-American legal regimes is widespread among Ghanaian artists.

My concern in this paper is on the experiences and practices of the current transformative processes of licencing, exploitation, rights management, digitalisation, online sales, and streaming and how they interrelate. The paper seeks to contribute to a better understanding of the ways actors and systems in place have dealt with and fragmented such rights. Current debates on rights must be seen in context with a fast-changing digital world of music. This was obvious in a seminar I attended in Accra, organised by MUSIGA – the Musicians Union of Ghana – the main body representing musicians in Ghana. Disputes in this seminar brought out the prevailing issues and concerns as to IPR and copyright matters – both from the musicians' union side as well as the artists. They also point to the dynamics of a new sensibility on claiming rights.[69]

[69] This "Music Business Seminar", as it was advertised, took place at the British Council in Accra in March 2017. I will get back to this event and my gatherings from it later in this paper.

These issues mainly relate to the supposed mis/management of music rights by the Ghanaian copyright for music society GHAMRO (the Ghana Music Rights Organization). An ethnographic perspective substantiates that common music industry practices on copyright and IPR in Ghana must be considered ethically and legally problematic. How artists and musicians experience them makes clear why many of them are left with the feeling that *rights are not right(s)*.

This phrase and the double meaning it comes with are a result of my own interpretation of many discussions I had with musicians about their problems with the existing copyright system, its principles and Euro-American, colonial preconditions, as well as the malfunctioning royalty payment system in Ghana. In a sense, this points to and implies the need for an alternative, more locally appropriate approach to protect and reward creative work and products – as suggested by Erlmann in the South African context and already prefigured and probably practised for some time in Kenya. According to Wallis and Malm (1984: 145/46), and this must have been an early effort on adjusting European, colonial copyright to an African country and transform it into a locally appropriate and practicable system, the" Kenyan National Music Organizations Treaty" of the 1960-70s was a "remarkable document" which "recognized some of the specific conditions that apply in the Kenyan music industry. For instance, that performers are invariably considered to be the composers of the songs they sing" (id.). This treaty may be useful again in reviewing and revising copyright regulations on the African continent.

My findings corroborate Boateng and others' impression that there is a considerable extent of incompatibility of cultural practices in Africa and copyright as designed and practised in Europe and America (Boateng 2011; Röschenthaler and Diawara 2016b), more and more effective globally after TRIPS (Peukert 2016: 37).[70] Seeger's exhortation that there is music "that is not appropriate to become" a commodity sits uncomfortably in a global pop music world where "any music can become a commodity", as "traditional music" is "reworked" into pop, facilitated by international copyright legislation that treats any music as commodity (2004: 157f.). Thus, copyright needs to be extended to music that has not been commodified (id.).

Not only in Africa but a ubiquitous practice there (observed for example also by Röschenthaler and Diawara 2016b: 21f.), musicians adopt traditional songs and transform traditional musical elements into their own versions of them, to which they subsequently claim copyright. My (admittedly few) reflections on the exploitation and uncredited use of traditional African music by popular artists in this paper relate to this. The mentioned "Kenyan copyright treaty" (Wallis and Malm 1984: 145) and those behind it probably paid tribute to that when linking composer credits to performance and arrangement. Protecting and copyrighting communal and local music ownership is still a pending issue and asks for more attention.

Music has been a profitable business across the world for about a century. Record companies and major labels of the 20th century, like HMV/EMI, Philips, and DECCA can be considered among the earliest global players, operating in transnational and transcontinental networks from the first half of the 20th century, strongly operating on the African continent. In music, thus, capitalist

[70] For an enlightening sketch of the history of copyright in Europe and its expansion in the colonial context, see (Röschenthaler and Diawara 2016: 11f); see (Collins 2006) for an earlier reflection on this and other effects of 'copyright imposition' in Ghana.

business practices obviously have a longer *global* history in Africa than in other fields of the creative arts. Rights and access to music, to own music and to sell it, as a commodity, potentially means making money. Popular music, at stake here and "by definition a commodity" (Seeger 2004: 157f.), has in recent years been transformed into the "digital music commodity" (Morris 2015). This came via new forms of digital music distribution and consumption and a dynamic digital music industry evolving around it, expanding the global music enterprise started by the 20th century transnationals for which Africa was an important market.[71]

In a system like this, music rights are to be contested. This has become a driving force in the digital world of music, too. The extent to which digital music distribution via the internet has made music rights a key issue on the African continent in recent years, the wider context this paper is situated in, cannot be underestimated. Moreover, online music distribution, streaming, and internet-based consumption patterns seem to promise a change for the better while digital accessibility of music potentially creates new rights avenues also.

2 Exploitation of Ghanaian music – lived experiences of *rights that are not right(s)*[72]

Some time ago in Ghana an artist called me with "a problem" and subsequently came to see me. A fellow musician had recommended me as a researcher working on popular music with an interest in artist rights, intellectual property, and good practices in royalties. In the two and a half decades I have also been engaged in arts promotion, I have encountered several ethically and legally problematic agreements for exploiting African music outside of Africa.[73]

This musician's compositions of the 1970-80s are in the Afro-Funk style. Originally released on vinyl records at that time, these originals are rare and highly sought after among music collectors and thus also expensive when sold on the collectors' market. His music has been rereleased, marketed and sold by record companies in Europe who specialise in this field in the vinyl as well as digital formats. Vinyl rereleases mainly target record collectors and DJs. The artist's problem turned out to be that about five years earlier, record companies in Europe *licenced* his songs from him and this involved straight payment of *licencing fees* to him. After making inquiries on royalties,

[71] These activities are of course key in understanding popular music/production and distribution in Africa historically. HMV, for example, started work in Africa in the early 20th century (Vernon 1994). From the late 1920s, there was a link in this via HMV's Indian branch (see https://www.hmv.com/music/100-years-of-hmv-our-story-so-far, last accessed 11 January 2021), see also (Jones 1985) for the early multinational history of the Gramophone Company, who owned HMV; see also (Wallis and Malm 1984) for the influence and entanglement of transnationals in global music business before digitalisation started. In 1933, HMV introduced the *GV series* of mainly Cuban and also South American music which became much liked and reportedly sold very well as it was distributed and marketed as Latin American music in various parts of Africa (The Gramophone Company, ca. 1951; His Master's Voice, ca. 1950s; Vernon 1994, see also afrodisc.com, last accessed 11 January 2021). As discographical and archival records show the companies mentioned did local music productions in West Africa by the 1940s through 1950s. For the commodification of music also see (Straw 1999-2000) and (Taylor 2007).

[72] *To exploit* is the standard term in music business when it comes to using music for business purposes, usually used without the connotation of the verb to take advantage of something for your own ends. To me the latter meaning - closer to abuse - is more than implicit in what I am describing here.

[73] This is not to say that things were or are better on the local market.

one record company recently informed him that they were about to repress his releases, and that this would involve a financial remuneration for him. Yet, six months after this and with his albums back in the market and remarketed and sold on digital platforms all over the world, he was neither paid nor did the company respond to his follow-up inquiries on his share (Coester 2017).

There is no easy way for this musician to claim his rights. Legal steps would be too expensive and complicated. No public office for consultancy or legal advice on rights exists in Ghana. The musicians' union could and should provide this. There is GHAMRO, in copyright diction the Collective Management Organisation (CMO) founded in 2011, which is the official body "to protect, promote, and develop the collection and distribution of music royalties", as well as "foster music creation", and "promote the value of music to the creative, cultural, and business sectors in Ghana" (GHAMRO 2016).[74] GHAMRO's mission could include proper counselling and support.

Trying to create awareness for their job with the slogan "Pay before Play", they are claiming to protect copyright and intellectual property, collect fees related to public use of artistic works and distribute these revenues fairly among rights owners in Ghana, but they are not thought fit to do this. Therefore, GHAMRO has been under critique from musicians for mismanagement and not handling or distributing artist shares adequately. Thus, deals have been done without them which makes it harder to call upon them for support.

The internal problems came up at an 'extraordinary GHAMRO meeting' in Kumasi in March 2017. According to the press, the atmosphere was so tense that the police had to come in to solve the row. "The confusion", wrote the reporter "arose over board's failure to render account of its stewardship to the members over the past two years" (Confusion rocks…) Later that month, a new board was established, headed by the musician Rex Omar – who took over from Kojo Antwi – someone said to have struggled a lot for musicians' rights in Ghana in recent years, apparently promising a change in direction and rights/royalties management (Entertainment Ghana 2017).

Exporting musical works and transfers of exclusive rights attached to them, to foreign music companies/record labels, like in the example above, is done in most cases without the involvement of GHAMRO although the respective artists are members of it. Provided these works are reproduced in Europe or America and distributed worldwide in ways that generate income, this is in most cases to some advantage of the artists, who are paid straight by the companies. In the *ideal case*, a label will account thoroughly to the artist for the commercial use of their music. The copyright society in Ghana might have records of these songs but has no knowledge of these deals. While most artists would not inform them, due to the lack of information exchange with the copyright societies outside of Ghana, who administer the Ghanaian songs for the music companies who licence them abroad, they lose out twice on being on track with what happens to the works of their country's artists.

[74] In 2011-12, COSGA, the Copyright Society of Ghana, legally established and founded in 1985, was split into three different copyright bodies by way of review of copyright regulations and legislative change based on the 2005 Copyright Act. From January 2012 there were The Ghana Music Rights Organisation, GHAMRO, for music, The Audiovisual Rights Society of Ghana, ARSOG, for audio-visual works, and CopyGhana - for literary works (see Tetteh 2015: 4; for a historical sketch of copyright in Ghana see Amegatcher 1993).

On the other hand, more widespread Internet access makes the increasing commercial use of music on the net visible and more obvious for everyone. New ways of digital music distribution have come into place and are prolific, but they are a legally problematic fields that demands more attention from legal experts in Ghana and abroad. How can musical works and individual compositions be legally exploited for which no clear records of ownership have been established?

Since 2005 and established through the *Copyright Act, 2005*, the government in Ghana runs the Ghana Copyright Office whose mandate is "the administration of copyright" and who "shall implement copyright and copyright related laws…" and "investigate and redress cases of infringement of copyright" (The Parliament of the Republic of Ghana, Act 690: 28). Musical works are to be registered there by their creators. Moreover, the mandate has been extended to the establishment and protection of intellectual property rights.[75] It is then in collaboration with potential rights owners and GHAMRO, which is a public representative organisation meant to be closely connected to the Copyright Office, that such records should be established through registration as copyright protected work.[76] The dynamics of music licencing, globalised digital music distribution and digital copyright issues have contributed to a growing awareness among musicians for the need to protect their works and claim IPR. The first step of which is the registration of their compositions.

3 Intellectual property rights in music & digital streaming - a new artist sensibility and prospect

Licences and contracts that transfer music rights have come with ethical and legal problems and the implications of such 'transfers' need to be understood better by artists, composers, and other rights owners. Musicians in Ghana have come to realise the pitfalls of this more and more in recent years. It is then less surprising that a "Music Business Seminar" – subtitled "Talk Music Business – The West African Experience"– , organised by Ghanaian musicians' official representative MUSIGA in March 2017 and, misleadingly, promoted under the theme "The Role of Music in Promoting Tourism", took a completely different turn. The potential relationship of music and tourism was hardly touched, but the predominant theme turned out to be artist rights and potential financial benefits from compositions through proper rights management, which reflects a growing sensibility of Ghanaian musicians to the issues of copy- and intellectual property rights. This trend is partly at least triggered by an interest in Ghanaian popular music abroad, as in the example mentioned above. Licencing fees and royalties are included in deals, and money goes straight to artists while GHAMRO is side-lined. The news on potential revenues gained from re/selling their music *outside* spreads fast.

The issue of rights management and associated problems was strongly evident also in the main talk of the seminar by Thomas Dayan, the general secretary the International Federation of Musicians, FIM, which is the umbrella organisation for musicians' unions worldwide, based in

[75] https://www.copyright.gov.gh/ (last accessed: 9 May 2022). Personal communication Yaa Attafua, Chief State Attorney and Head of the Copyright Office Ghana (March 2016, April 2017).

[76] I think a greater effort by the Copyright Office, as well as by GHAMRO, to promote and encourage copyright registration would facilitate the implementation of copyright and establishment of IPR.

Paris. The most critical and substantial change, as Dayan made clear, and problems arising from it for artists, relates to problems in rights management in digital music streaming, expected to be the main (and only?) form of distribution and consumption of music in the future (Dayan 2017).

Early second-half 2010s figures talk of about 50 million paying subscribers for Spotify, which was founded in 2006, who have established themselves as market-leader as other audio streaming services like Deezer, Napster, Amazon Unlimited and others follow suit. The number of music streams is said to have doubled in 2015 to 300 million, in comparison to 2014, pointing to a rapid and fundamental change in music consumption patterns in recent years, due in part to the expotential increase in streaming speeds.

Spotify subscribers rose by 20 million within twelve months in from 2016 to 2017 to 70 million, and 120 million users in total with non-subscribers getting a cost-free, limited streaming offer. Since then, Spotify has doubled its paying customers to 140 million in 2020 and are currently indicating a total of about 320 million users, offering audio streaming via the net in over 90 countries, but are not yet working south of the Sahara on the African continent except for in South Africa. This means they doubled subscription figures and thus revenues from paying customers alone in the last three years which strongly corroborates the dynamic streaming trend predicted in 2017 and shows a strongly expanding company generally. *Unlimited* music access is offered at relatively low fees.[77]

Looking at the royalties aspect for creators of music, without a change of law, Dayan implied in his talk, there is no way to make digital music exploitation through the existing structures or streaming services beneficial to writers, composers, and least of all to performers (2017). Small financial benefits from their musical compositions and intellectual property did not start with digital streaming for artists in Ghana and royalties is a complicated topic generally. While from a Euro-American perspective on copyright and royalties and also for collecting societies, the digital turn has brought about many issues and a loss in income. For Ghanaian musicians, the new digital outlets yield hope for improvement of revenues and connectedness. Their feeling is that if rights management in the future will involve close collaborations of CMOs on the global scale (as expressed in Dayan 2017), it will be beneficial to them.

In Ghana, musicians' discourse on artist rights and royalties now centres on *logging*, a technology that promises transparent and fair rights management. As several participants at the seminar

[77] https://de.statista.com/statistik/daten/studie/297138/umfrage/anzahl-der-zahlenden-abonnenten-von-spotify/ (last accessed 9 May 2022); https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers/ (last accessed 9 May 2022) The standard subscription fee is around 10 US dollars/euros per month. There are different packages, and the fee varies from country to country. There is also a free service. Standard subscription offers streaming access as Spotify claims to over 60 million songs, compared to 30 million in 2017. Spotify reported an overall turnover of around 6.8 billion euros in 2019. I wrote the main part of this section in 2019, which was based on research done before 2017-18. In 2017, the company was not active on the African continent yet. In 2020 they were operating in Egypt, Algeria, Morocco, South Africa, and Tunisia (www.press.spotify.com/us/about, first accessed 28 April 2017, last accessed 12 January 2021. See also https://de.wikipedia.org/wiki/Spotify , last accessed, 9 ay 2022); Spotify has continued expanding rapidly, and by early 2021 was streaming in over 40 African countries (https://support.spotify.com/de/article/full-list-of-territories-where-spotify-is-available, last accessed 9 May 2022). Like in other newly conquered African markets, where fees are much lower as the amounts mentioned above, they started streaming in Ghana in February 2021.

made clear, they are expecting MUSIGA and GHAMRO to quickly establish an electronic system that allows tracking public use of music on the radio, mobile phones, public venues and elsewhere; they hope that this will provide a record of which music is really circulating and which is not, and thus to whom royalties should and should not go to (Coester 2017).

4 The better choice? – Music registration in Europe or: what is in a record in the hot and contested *digital world of music*?

Due to little confidence in local copyright implementation, Ghanaian and other African musicians with a link to Europe or America have registered their recordings and compositions with copyright and collecting organisations abroad for many years, like PRS in Britain, GEMA in Germany, SACEM in France, and ASCAP in the US. How is this done and what are the benefits for the artist? In reissuing music, the common practice has been to licence compositions and songs to record companies, mainly independent record labels, who as mentioned before usually pay the artist straight for this, and in a next step register this music on behalf of the artist in their own countries.

Without going into detail here, such licencing deals are somewhat problematic generally, since they collide with music rights in countries like Ghana, claimed by the CMO and the producers of music. At the same time, artists take advantage of and contest the pretty unclear situation of who is in charge of managing such rights and the shared ownership of music or music productions respectively, when a clear legal position is not in place or at least not implemented in practice. Likewise, music producers claim ownership, grant licences, and sell rights for their productions of individual artists when approached, and they do not account well for this. The financial deals involved are usually not transparent to the composers and performers of the music who are not considered to be in charge of the business-side of music by most producers.[78]

Directly related to representational problems of property rights and also intertwined with licencing and publishing practices of African music in and for Europe, is the lack of accurate records/data on composers, producers, and other rights owners – like performers and arrangers – especially for music recordings done before the 1970s. Talking to GHAMRO while looking for composers of Highlife created in the 1950s and 1960s, I noticed that such information is missing in their records and causes a big gap in potential rights representation and management.

The GHAMRO precursor, COSGA (Copyright Society of Ghana) was only established in the mid-1980s and thus many earlier compositions were simply not registered, some were with PRS in England, but a migration of these data has apparently not been implemented. Still registration is not a common practice until today because many musicians don't put trust in their country's CMO

[78] The producer-artist relationship in Ghana is complex and must be assessed in its specific socio-cultural context. Claims to create legal and moral correctness –as perhaps implicit in my approach– may not be appropriate in less formally administered settings as in Ghana and as to long-time personal relationships that exist between producers and musicians, often of ongoing informal artist support. But to provide an example here, in 2016 I alerted a well-known musician in Ghana to the recirculation of one of his songs by an American record company. He didn't know of it. The deal was struck by the producer. On enquiry, the producer subsequently paid around US $1,000 to the artist as the latters share in this deal (Coester research notes Ghana 2016).

as a representative. On the other hand, many pre-1970s compositions of well-known artists were published abroad from the 1980s onwards – usually by small, specialist record companies and music publishers – with an interest in African music based on agreements with individual artists or family representatives. Most of them have lost track and at the same time, this meant these songs were not registered locally from the 1980s by heirs of IPR, and still have no official local or national representative when it comes to music rights.

Music publishing in Europe and America is a very sophisticated and long-established system of work protection, distribution, promotion, copyright control, and financial income. This system does not exist in West Africa, which has led to a vacuum when it comes to publishing rights. It has likewise encouraged musicians to sign *publishing agreements* for their works with music companies abroad. This seems reasonable on principle and economically necessary, yet like licencing, it doesn't consider all rights owners back home. Even for those who sign the deal this does not come automatically with proper accountability, as for the musician waiting for his royalties from Europe, mentioned as an example above.

Proper rights management and fair royalties for musicians are desirable. As long as collaborations between collecting societies in Europe and Africa are not in place or effective – for reasons beyond the scope of this paper – the way forward for musicians may be, as I am told, to register their compositions abroad (which has been going on for many years, actually). In practice this means musicians have to become a member of a foreign CMO in order to benefit directly from exploitation of their music. While membership criteria seem to vary from country to country, this is easier said than done, because the knowledge on how to join and deal with the bureaucracy involved in this is not widely known.

Paying registration and annual membership fees constitute an additional cost that may keep musicians from joining, considering the lack of funding many Ghanaian musicians face. Nevertheless, becoming a member would guaranty financial benefits if their works are used on the global, digital music market. On the other hand, this may be in conflict with local understandings of who owns rights to the music – mainly affecting the *share* for the producer – for a long time a major rights claimer in popular African music.

Releases and reissues of African music by contemporary music companies in the Global North often include the (long-term) transfer of exclusive composer rights through licencing and publishing. Subsequently, this leads to the distribution of music that was formerly not available in European, American and global music markets and puts artists on that map, so to say. Artists and their music entering these markets can, potentially, benefit from it. At the same time, other aspects of the music business come with it, arguably less beneficial to them.

In a way, contemporary Euro-American music business practices generate a feeling of unfair distribution and improper use of African intellectual property. Unjust claims to African music has a long and sorted history. For instance, the case of "That Happy Feeling"- in the 1950-60s comes to mind that involved the Ghanaian artist Ghanaba, then known as Guy Warren, and the German

bandleader Bert Kämpfert.[79] They produce a competitive system whose details can only be untangled through a study of the legality of claims for exploitation of African music as pinned down in contracts for transfer of exclusive rights, and through investigating registration and publishing rights as indicated in the existing CMO systems.

Trying to illuminate this more, I will touch here on another example of a Ghanaian musician-composer who has attracted quite some attention outside of Ghana in recent years. Although a number of producers, record labels, booking agents, and other music *exploiters* have been part of the musician's success – and the artist himself has to take on responsibility for some of the problems described here – his story helps to shed light on the problematic practices in the current exploitation of music from Africa.

This artist gave me a written note in which he demanded his record company to take his name and pictures out of the promotion for another artist, who had also been put on contract by the artist's record company. Apparently, the latter artist was picked as the next, viable export from Ghana on what is a Europe-US-niche music market. The British record label and the producers involved with the new artist used the first, better-known artist's name and recognition as an outstanding musician, as a means to promote the new artist – somehow in the vein of a mentor-story – without the former's knowledge and consent. No agreement on his involvement in the promotion of his "successor" hade been made. In practice, this has led to a pretty stiff relationship of two formerly very close musicians in Ghana. Moreover, the licencing of songs for an oeuvre-compilation of the new artist was done incorrectly, critically without involving or paying documented rights holders to some music selected, among them the musician who complained to me. All of this points to ethically problematic – but not uncommon – practices in *securing* African music and legally contestable, incorrect licencing.[80]

[79] A composition Warren had recorded in 1957 in America, named as "Eyi Wala Dong" – probably a traditional Ga melody – was adopted by Kämpfert (often spelled Kaempfert) for his high selling LP *A Swinging Safari*. The record, the sheet music of the song, and airplay on TV in America made the melody known around the world under the title "That Happy Feeling". As Ghanaba remembered, it involved a lawsuit to reclaim the rights for this song (interview with the author 2007). In this case, this would precede the significant case of Solomon Linda's "Mbube" – "The Lion Sleeps Tonight" – where Linda's heirs claimed rights from Disney Enterprises Inc. and won the case (Erlmann 2016). In addition to "Eyi Wala Dong", Kämpfert used South African music for *A Swinging Safari*. The record "was heavily influenced by South African kwela style music, containing versions of 'Zambesi', 'Wimoweh', 'Skokiaan', and 'Afrikaan Beat', as well as the title track, which made Kämpfert an early exponent of world music" (https://en.wikipedia.org/wiki/Bert_Kaempfert, accessed 29 April 2017). Except for "Afrikaan Beat" with composer credits for Kämpfert, the songs mentioned are credited to other composers on the original LP. The title song, "A Swingin' Safari" clearly adopts a recorded *kwela*, the South African township pennywhistle jive, and has touches of "Wimoweh" also. The kwela is from "Tom Hark" by Elias and the Zig-Zag Flutes, a song from 1956 that became known in Britain, as it even made it to number two of the UK charts in 1958 (see Coester 2014). While "That Happy Feeling" is credited to Warren, at least on the 1966 issue originally published 1962). In the liner notes of this release, it still reads, "A bright succession of Kaempfert hits – the most recent of which, and perhaps the most significant being AFRICAN BEAT (sic) together with the title song of this album, and THAT HAPPY FEELING. All these three titles are original Kaempfert compositions." (Liner notes for Bert Kaempfert And His Orchestra, *A Swingin' Safari*, 1966). I wonder whether the "original-cover dialectic", Erlmann brings up in "The Lion Sleeps Tonight"-context, cannot be expanded beyond Africa, when royalties for songs should clearly go there (2016).

[80] In order to avoid the impression of a public charge and to respect the privacy of the artists involved, names are not mentioned in this case.

The most problematic legal and moral aspect of this artist's experience of dubious handling of intellectual property rights by European music agents is that he was lured into a publishing contract by a music exploiter in Europe for lifetime copyright. This case exemplifies the somehow naive but widespread practice of African artists to sign contracts and assign music rights (perhaps based on trust in good business practices) without legal advice (which is not available easily) and properly understanding the long-term implications of such deals. Cases like this of taking advantage of fragile copyright and IPR in African music ask for a legal revision of contracts based on wrong assumptions. Ghanaian and other African musicians cannot be satisfied with *deals* like this nor with how copyright and intellectual rights are managed at home. Their rights are not worth much, as not only this musician holds, as long as such ethically problematic practices in copy- and intellectual property rights *here and there* continue.

This occurs in a globalised system of music business, commodification, and circulation that has changed the human experience with music over the last century, in which it evolved, substantially. The age of the mechanical reproduction of artworks, as Walter Benjamin alerted us propelled capitalist music exploitation through the technical possibilities of mechanical music reproduction, songs became phonograms (1968). Interrelatedly, music is now *old* consumption media. Records, record players, tapes, radios, cassettes, and other anaologue media forms propelled music's local and translocal circulation, made phonograms reproducible basically everywhere, and reinforced a capitalist music business with crude *acquisition*, ownership and control practices.

Music distribution and consumption via the Internet –in the age of the digital reproduction of artworks – represents a major structural shift as it moves along with the digital commodification of music. It comes with new technical, technological, medialized, and aesthetic options – simultaneously accelerating music's global circulation. While in this digital world, music will probably continue to be a lucrative business for record companies, and securing control over rights will increasingly shape it, the *digital shift* promises to be financially more beneficial to more music creators in the long run than it currently looks to the established industry.[81]

For African music, this requires technological adjustments, a number of digital music platforms specialising on Africa have started up in recent years, as well as better practices on copyright and intellectual property rights than the ones reported on. One of the digital music companies is Africori, the self-acclaimed "Home Of African Music", offering various kinds of music services including marketing and collecting royalties, and with a strong focus on music rights.[82] The

[81] For several reasons which I can't discuss here, the established music industry sees the current phase of digital music distribution, especially streaming services, as involving a considerable decline in royalties for performers and producers of music, and thus as disadvantage to the creators of music (Dayan 2017).

[82] More information about Africori can be found at the website, Africori.com. According to their website, Africori offers "digital music solutions", and two of their marketing themes are "The Pan-African Music Company for Artists and Labels" and "Collecting revenues for artists, writers, composer, producers". Africori was founded in 2009 and relevant to issues on music rights in Africa, the Swedish mobile telecom company Millicom came on board in 2014. This started their *AMR venture*: "We're pleased to announce the launch of a new partnership with Swedish telecoms giant Millicom to fund, acquire and manage music rights through a new venture, Africa Music Rights (AMR). Under Africori's management, AMR will acquire catalogues from record labels across Africa and provide worldwide distribution services" (http://africori.com/#section-our-service, last accessed 9 May 2022). The dynamics of digital services and local awareness of music rights – or rather copyright and IPR – in Ghana, I reported on earlier on, possibly owes quite some impetus to this initiative. Contemporary Ghanaian popular music artists, like Becca, are now also represented by

scramble for African digital music rights has started as music, unique and outstanding as an aesthetic and performative art form, immerses in a market-driven business of digital cultural re/production and distribution. Music exploitation has become a standard mode of thought and action globally, often to the disadvantage of creators on the African continent.

The foundational and potentially transnational agreements of the Rome Convention of 1960, to protect "performers and producers of phonograms" (WIPO n.d.),[83] and the so called TRIPS agreement of 1994 on IPR, which took in these key points of the Rome Convention, may be useful for the benefit of creators on the African continent. Especially TRIPS has come to be seen to represent the "wholesale expansion of IP law" (Erlmann 2016: 97). This is despite its promises of copyright and related rights protection. While strengthening copyright by establishing "minimum levels of protection and enforcement", it does so along the lines of "Western-style copyright protection" (Frith and Marshall 2004: 12-13), and the link to trade sanctions, arguably represents the neo-colonial imposition of Euro-American economic and legal regimes on Africa.[84]

By establishing minimum levels of protection and enforcement, they could be useful still in locally adjusted form, demanding governments and their agents – like the Copyright Office in Ghana – to respond to them, probe their copyright laws and regulations, which most have, and deal with copyright and IRP enforcement more in their respective countries and for their own artists. As these agreements have not been implemented fully in countries like Ghana, and the need of creating locally appropriate copyright systems seems to have slowed this process down, currently their supposed usefulness, *authority*, designs, and legal frameworks struggle with the digital world of music. This may lead to reviewing TRIPS on the WTO level while on the African continent working out appropriate and collaborative copyright regulations and protection is at stake. Hence, that music registration at home becomes the first choice.

Moreover, varying national legislation on copyright and IPR as well as collective management challenges impede unanimous rights concepts for protecting creative and intellectual work and their musical products in digital Africa. Claims, possibilities, and practice are far apart, on the local as well as the global scale. They are still – and probably increasingly – shaped by transnational and complex music business structures and relations, and enduring if reconfigured, sometimes conflicting interests of composers, performers, producers, publishers, record companies, CMOs, the new streaming companies. Other digital music managing and marketing companies, like

Africori. On the other hand, this touches on the question of who owns rights to music and what is the role of local and national CMOs in this, a key issue as to the findings presented in this paper.

[83] http://www.wipo.int/treaties/en/ip/rome/summary_rome.html (last accessed 9 May 2022).

[84] The "Agreement on Trade-Related Aspects of Intellectual Property Rights" came about as part of the GATT negotiations of the WTO. The WTO website reads: 'The TRIPS Agreement, which came into effect on 1 January 1995, is to date the most comprehensive multilateral agreement on intellectual property. The areas of intellectual property that it covers are: copyright and related rights (i.e. the rights of performers, producers of sound recordings and broadcasting organizations)' (https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm, last accessed 9 May 2022; TRIPS works on the condition of WTO membership: "TRIPS requires WTO members to provide copyright rights, covering content producers including performers, producers of sound recordings and broadcasting organizations" (https://en.wikipedia.org/wiki/TRIPS_Agreement, last accessed 9 May 2022).

Africori, are also caught and act inside. To many musicians, composers, arrangers, and performers in Ghana, the hot, digital world of music promises a change for the better.[85]

5 Digital music repositories and rectifying rights - who owns the music?

A new chance of refiguring copyright and establishing IPR for African music comes with large digital repositories and databases of popular music created over the last decades. Efforts to digitise and capture information on major analogue collections have led to migration of this music and attendant catalogue and discographical records, as metadata, formerly only available in analogue formats, into the digital world of music; including an enormous amount of potential and relevant rights data. In Ghana this concerns the so-called "Gramophone Library" of the national broadcasting station, Ghana Broadcasting Corporation, one of the largest archives of analogue-born and now digitally accessible recorded music of the 1950s through the 1980s on the African continent. Digitisation and re/documentation of this repository started from 2008 and is ongoing at GBC.

The archival content as well as the physical site are not only significant parts of the nation's musical and cultural heritage; exemplarily documented in *Ghana Muntie – Recordings from the G.B.C. Gramophone Library and Radio Ghana, 1947-62* (compiled by Coester 2012), a 2CD-set educational and research publication. The archive dates back in time and during the time of Ghana's first president, Kwame Nkrumah, for instance, a lot of gramophone records from other parts of Africa were acquired also, and kept until today. Memorable radio programs were the result, like the 2 hour-weekly *Way Down South* on Radio Ghana which brought popular music from South East and Central Africa to West Africa for the first time in the 1950s. This extends the potential for refiguring heritage, transcontinental circulation of music, and rectifying music rights via new, reconstructed metadata to other parts of Africa. Currently, the digitalised collection comprises about 100,000 songs and attendant discographical and contextual documentation, crucial in this endeavour.

To creators and potential rights holders of this music, the existence of such digital repositories as well as new access to their musical legacy is significant. It yields creative potential as well as the potential, provided the necessary research is done, to rewrite the foundational script for new IPR and copyright management of music rights especially for Ghanaian music prior to the 1990s.

To achieve this, the rights to access and recirculation of music and documentation data are to be negotiated between the stakeholders of such repositories when it comes to IPR and copyright. These are the national custodian GBC, who owns and maintains the collection and archive

[85] "Composers and performers" keen on exposing their music as widely as possible, as Frith and Marshall pointed out around two decades ago, "see the Internet as a wonderful tool for the promotion of their music" (2004: 5). Rather, record labels and publishers, "who hold massive amounts of copyright", are interested in keeping "tight control of their markets" (id.). The tension implicit here has been enhanced by understanding that those who "create copyright" in the first place, composers and performers, do virtually not hold it in this system (id.). And by the lack of effective copyright laws on African digital music rights and digital music sales, which as mentioned have become the main mode of selling music today, so that CMO's can collect respective royalties for creators, as eventually musicians will ask for their share.

materially, the digital masters and the metadata; the music rights organisation, GHAMRO; the potential custodian of IPR and copyright, the Copyright Office; as well as individual rights owners identified and accredited as creators of works.

While the institutional bodies are expected to act responsibly to preserve national cultural heritage, they seem to follow disparate agendas, claims and outlooks on this repository, partly because of different interpretations of Ghanaian copyright law. Over the years, individual composers and producers have been provided by GBC on request with copies of their music/works and this has had quite an effect on cultural re/creation and ownership questions, yet to be studied; as probably had the 'digital return' of many music on air in the weekly GBC radio programme *Gram Time - Music brought back from the GBC Gramophone Library 1950s-1970s*, started in 2010.

Agreed accessibility for recirculation as well as rights research, however is still a goal to achieve and the basis for rectifying copyright and fixing IPR for the benefit of creators. Closed-up archives are conjuring up colonial practices of cultural control, ownership, and closure which do not ask, neither answer the question, "Who owns this music?", beyond the confines of material property. Historically, the custodian GBC came out of the Gold Coast Broadcasting Service, the media apparatus installed during British colonial rule, and modelled on the BBC. With it came the colonial notion of corporate authority and ownership.

Digitisation and opening up archives have triggered the digital return, music recirculation, and has had effects on reclaiming copyright already as well as on re/dissemination of an extensive cultural repository, a process of freeing the creative legacy of many. Opening up digital repositories, like the GBC's, are thus a key practice to decolonize archives, perhaps most significantly through the reconstruction and rectification of rights. For this, the stakeholders, GBC, GHAMRO, and individual rights owners need to negotiate and clarify ownership and copyright, possible and overdue now, and create a mutual, legal and ethical framework based on the Copyright Act the Ghanaian government has provided – and most likely will have to adjust more to the digital world of music – to avoid not only the continuation of colonial custodianship but also the proliferation of the neo-colonial buyout and exploitation of music nationally and globally, based administratively at least on the lack of copyright and IPR information sketched above. This would be an altered approach to the creative legacy in music and potentially *de-problematize* IPR and copyright practices as it streamlines the global community's interest in African music ethically for the benefit of artists and their country, so that *rights can be rights*.

6 References

Amegatcher, Andrew O. 1993. *Ghanaian Law of Copyright.* Accra: Omega Law Publishers.

Benjamin, Walter. 1968. *Illuminations.* Originally published in 1935. New York: Harcourt, Brace & World, 219-254.

Boateng, Boatema. 2011. *The Copyright Thing Doesn't Work Here: Adinkra and Kente Cloth and Intellectual Property in Ghana.* Minneapolis: University of Minnesota Press.

Coester, Markus. 2014. Revisiting Britain's 'Afro Trend' of the 1960s and 1970s: Musical Journeys, Fusions, and African Stereotypes. In *Black Popular Music in Britain since 1945*, 47–66. London, New York: Routledge Taylor and Francis Group.

Coester, Markus. 2016/2017. Field notes Ghana (unpublished field notes).

Confusion rocks GHAMRO meeting. *Graphic Showbiz* (Ghana). 2.3.2017, 8.

Collins, John. 2006. Copyright, folklore and music piracy in Ghana. *Critical Arts* 20 (1): 158–170.

Cowley, John. 1994. uBungca (Oxford Bags): Recordings in London of African and West Indian music in the 1920s and 1930s. *Musical Traditions* 12:13–26.

Dayan, Thomas. 2017. *Acting Together in a Globalised World: Talk given at the MUSIGA Music Business Seminar.* Accra: British Council; Recording transcribed by the author.

Entertainment Ghana. 2017. TAKEOVER! Rex Omar Is Chairman For New GHAMRO Board. http://www.entertainmentgh.com/news/takeover-rex-omar-chairman-new-ghamro-board/. (last accessed 9 May 2022).

Erlmann, Veit. 2016. Lion's Share: Intellectual Property Rights and the South African Music Industry. In *Copyright Africa: How Intellectual Property, Media and Markets Transform Immaterial Cultural Goods*, ed. Ute Röschenthaler and Mamadou Diawara, 86–102. Canon Pyon: Sean Kingston Publishing.

Frith, Simon and Lee Marshal (ed.). 2004. *Music and Copyright*, 2nd edn. New York: Routledge.

GHAMRO (ed.). 2016. *Constitution of Ghana Music Rights Organization*.

Ghanaba, Kofi. 2007. Interviewed by the author. Nsawam/Ghana, 4/2007.

His Master's Voice. n.d. ca. 1950s. Lista De Discos. Musica Latino Americana (HMV GV Series). Catalogue (in Portugese). Hayes, Middlesex, England: The Gramophone Company, Ltd.

Jones, Geoffrey. 1985. The Gramophone Company: An Anglo-American Multinational, 1898–1931. *Business History Review* 59 (1): 76–100. https://doi.org/10.2307/3114856.

Morris, Jeremy Wade. 2015. *Selling Digital Music, Formatting Culture.* Berkeley: University of California Press.

Peukert, Alexander. 2016. The Colonial Legacy of the International Copyright System. In *Copyright Africa: How Intellectual Property, Media and Markets Transform Immaterial Cultural Goods*, ed. Ute Röschenthaler and Mamadou Diawara, 37–68. Canon Pyon: Sean Kingston Publishing.

Röschenthaler, Ute, and Mamadou Diawara. 2016a. *Copyright Africa: How Intellectual Property, Media and Markets Transform Immaterial Cultural Goods.* Canon Pyon: Sean Kingston Publishing.

Röschenthaler, Ute and Mamadou Diawara. 2016b. Introduction: African Intellectual Worlds in the Making. In *Copyright Africa: How Intellectual Property, Media and Markets Transform Immaterial Cultural Goods*, ed. Ute Röschenthaler and Mamadou Diawara, 1–34. Canon Pyon: Sean Kingston Publishing.

Seeger, Anthony. 2004. Traditional Music Ownership in a Commodified World. In *Music and Copyright*, 2nd edn, ed. Frith, Simon and Lee Marshal, 157–170. New York: Routledge.

Stewart, Gary. 2003. *Rumba on the river: A History of the Popular Music of the Two Congos.* London: Verso.

Straw, Will. 1999-2000. Music as Commodity and Material Culture. *Repercussions* Spring-Fall: 147–171.

Taylor, Timothy D. 2007. The Commodification of Music at the Dawn of the Era of "Mechanical Music". *Ethnomusicology* 51 (2): 281–305.

Tcheuyap, Alexie. 2016. Music for Everyone.: The Dynamics of Piracy in Cameroon. In *Copyright Africa: How Intellectual Property, Media and Markets Transform Immaterial Cultural Goods*, ed. Ute Röschenthaler and Mamadou Diawara, 270–291. Canon Pyon: Sean Kingston Publishing.

Tetteh, Evelyn O. Apronti. 2015. The Awareness of the Copyright Law among Library Users in Tertiary Institutions in Ghana. *Globus* 3: 30–48.

The Gramophone Company. n.d. ca. 1951. *Latin-American Rhythm: G.V. Series* (Magenta Lable). Catalogue. Hayes, Middlesex, England: The Gramophone Company, Ltd.

The Parliament of the Republic of Ghana. 2005. Copyright Act, 2005 (Act 690). https://bcp.gov.gh/acc/registry/docs/COPYRIGHT%20ACT,%202005%20(ACT%20690).pdf (last accessed 20 December 2020).

The Parliament of the Republic of Ghana. 2009. Copyright (Amendment) Act, 2009 (Act 788). https://ictpolicyafrica.org/fr/document/qgnjzm9hz7m?page=1 (last accessed 20 December 2020).

Vernon, Paul. 1994. Savannaphone: Talking Machines Hit West Africa. FolkROOTS 122, 21-25. Published online in 1997, http://www.bolingo.org/audio/texts/fr122savanna.html. (last accessed 9 May 2022).

Wallis, Roger and Krister Malm. 1984. *Big Sounds from Small Peoples: The Music Industry in Small Countries.* London: Constable.

WIPO. n.d. Summary of the Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations (1961). http://www.wipo.int/treaties/en/ip/rome/summary_rome.html (last accessed 9 May 2022).

7 Discography

Bert Kaempfert And His Orchestra. 1966. *A Swingin' Safari*. Polydor, 33rpm, 12", Catalogue Number: 46384, London (UK Release).

Elias And His Zig-Zag Jive Flutes. 1958. *Tom Hark/Ry-Ry*. Columbia, 45rpm, 7", Catalogue Number: 45-DB 4109, London (UK release).

(edited by) Coester, Markus. *Ghana Muntie - Recordings form the GBC Gramophone Library & Radio Ghana 1947-1962. 2012.* Accra: Ghana Broadcasting Corporation/Frankfurt: popular african music.

Capturing Semantics for Knowledge Management Systems

Section 3

Modelling in the Humanities

Productive Tension between Disciplines, Language, and Media

Øyvind Eide (University of Cologne)[86]

Extended Abstract

Creating and using models in scholarly and scientific work has many purposes. The use of computers in the humanities has always been based on modelling of data and of modelling as part of the development of computer system. As part of the Volkswagen Foundation funded project "Modelling Between Digital and Humanities: Thinking in Practice" an interdisciplinary modelling workshop was organised in 2017, reported in Ciula et al. (2018b). A general overview of modelling from a digital humanities perspective, with some links to modelling more generally, can be found in the introduction to the workshop volume Ciula et al. (2018a). For modelling in the sciences see, e.g., Frigg and Hartmann (2012). In this short article I will offer some thoughts on modelling, also as a learning strategy (a process of coming to know) based on experiences from digital humanities research, teaching, and practice.

[86] Contact: oeide@uni-koeln.de, https://orcid.org/0000-0002-7766-6287.

Models have a long history in the humanities, with historical models of raise and fall of nations or empires (e.g., Herodot), models of society types such as feudalism, and the use of models in early modern natural philosophy. The 19th century saw the rise of novel and creative forms of visual models, as seen exemplified in Figure 11 (Rendgen 2018). In the 20th century yet other types of models, such as networks, were put to use. In literary studies, formalistic models were used to examine the structure of literature (Propp 1958), leading up to structuralist models in narratology (Chatman 1978). This also traversed over into computer-based studies (Meister 2003). A seminal work on modelling in digital humanities is the first chapter of McCarty (2005), in which a large number of traditions are combined, including the classical period of modelling in social/cultural anthropology, from the 1950s to the 1970s.

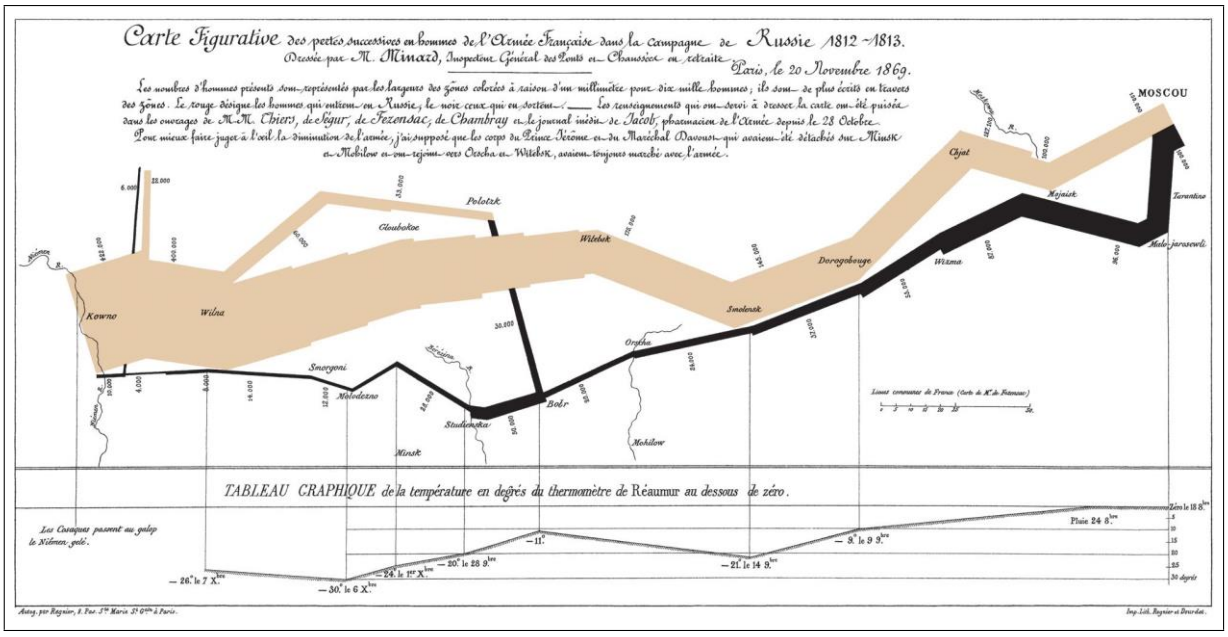


Figure 11: Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path. Lithograph, 62 x 30 cm. Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Minard.png (last accessed 20 September 2021).

Modelling in digital humanities can be encircled through a number of oppositions. The digital is formal, rule based, structured, and discrete, and the use of computers is linked to a practice-based form for research and teaching. This is contrasted to the humanities, being seen as analogue, continuous, nuanced, and hermeneutical; what Geertz (1973) called thick descriptions, The bridge between the digital and the humanities is the special computer-oriented form for operationalisation, where concepts of humanities inquiry are made measurable and computable. In this process, data modelling plays a key role in the transforming from what one in the humanities knows as sources to the data used in statistical analysis and other computer-based methods (Flanders and Jannidis 2018). When using modelling as part of a critical inquiry, however, what cannot be measured and computed are also central areas of interest (Eide 2015). What escapes formalisation can give us insight in the objects and processes we study.

Models as they are discussed here are media products in the sense of Elleström (2021), they are created as part of communication processses. In the humanities they are also often created on the basis of media products: a digital map based on a book, a set of charts based on some analysis of a collection of books, a network modelling relationship between characters in a drama text. Some models are seen to the user as a visual item, but are also based on a mathematical expression, such as 3 dimensional models of historical artefacts (based on a set of geometric expressions) and a network visualisation (based on a mathematical graph) (Eide 2021).

Models in digital humanities, as in the sciences, are pragmatic in the sense that someone (a person or a group) creates and uses the model for some purpose (Gelfert 2016, 113; Ciula and Marras 2016). This does not mean that modelling cannot be scholarly, rule based, or rigorous. It is rather in line with the basic semiotic understanding that signs are connected to what they signify for someone and is thus a basic feature of modelling as an example of a meaningful human activity. Indeed, seeing models as icons in the sense used by Peirce is an interesting line of analysis still in the process of being developed in digital humanities (Kralemann and Lattmann 2013; Ciula and Eide 2017). While models can have quite different forms, whether they are carried by computers or not, their relationship with the modelled objects and processes is iconographic, operating in the span between image like, structural, and metaphorical similarity, and is always dynamic and in movement, as all meaning production is.

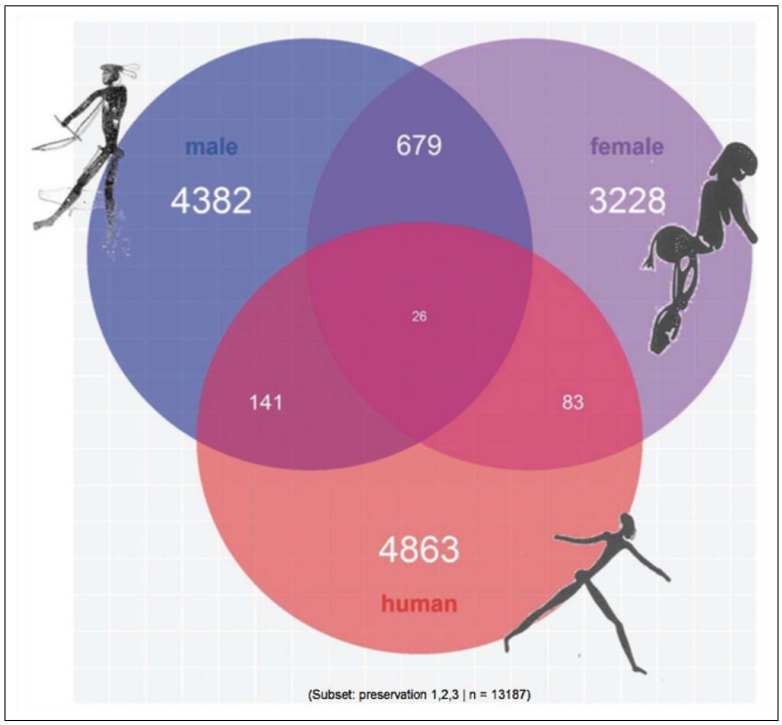


Figure 12: Gender types in Daureb/Brandberg as modelled by Vogels and Fäder (2018).

'Model' is a word well known from art. Many artists paint based on their models, as photographers photograph their models. This concept of modelling is different, but not totally disconnected from scholarly modelling (Eide and Eide 2016). Maybe the figures in the rock art of Dâureb/Brandberg in Namibia are models of something (this I am not competent to claim), but modern research on these painting surely create models such as the one showing gender types in Figure 12. The close connections between art and modelling, and the pedagogical power of using these connections, are suggested by Fishwick (2017; 2018).

Modelling in digital humanities introduce, or re-introduce (Bod 2018), a type of practice-based reasoning in the humanities, a playful form of experimentation. Learning by doing is at the core of processes of making, which are also processes of coming to know. This is a basis for our study programmes at the Department for Digital humanities at the University of Cologne.[87] This is also underlying our research. In this way we try to connect to the best traditions for practical teaching and inquiry, which includes Teach Kids Code,[88] Girls Make Games,[89] and Forensic Architecture.[90] This enables our students not only to know the functionality of the systems they study, but also to know how to open up some of the black boxes, peep inside, and see what happens when things are changed. That combines other literacies, including data literacy, with the modelling literacy needed to be a citizen in a modern society.

References

Bod, Rens. 2018. Modelling in the Humanities: Linking Patterns to Principles. *Historical Social Research Supplement* 31: 78-95.

Chatman, Seymour. 1978. Story and Discourse. Narrative Structure in Fiction and Film. Ithaca and London: Cornell University Press.

Ciula, Arianna and Cristina Marras. 2016. Circling around texts and language: towards 'pragmatic modelling' in Digital Humanities'. In *DHQ: Digital Humanities Quarterly* 10(3).

Ciula, Arianna and Øyvind Eide. 2017. Modelling in digital Humanities. Signs in Context. *Digital Scholarship in the Humanities* 32: i33–i46.

Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle. 2018a. Modelling. Thinking in Practice; An Introduction. *Historical Social Research, Supplement* 31: 7–29.

[87] https://dh.phil-fak.uni-koeln.de/studium-lehre/studieren-am-idh (last accessed 20. September 2021).

[88] Teach kids code | Torgeir Waterhouse | TEDxOslo, https://www.youtube.com/watch?v=-TP7Zjqmj_4, cf. Olari et al. (2021). (last accessed 31. October 2021).

[89] https://www.girlsmakegames.com (last accessed 20. September 2021).

[90] https://forensic-architecture.org (last accessed 20. September 2021).

Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle (eds.). 2018b. Models and Modelling between Digital and Humanities. A Multidisciplinary Perspective. *Historical Social Research*, *Supplement* 31.

Eide, Øyvind. 2015. *Media Boundaries and Conceptual Modelling. Between Texts and Maps*. Basingstoke: Palgrave Macmillan.

Eide, Øyvind. 2021, Modelling and Networks in digital Humanities. In *Routledge International Handbook of Research Methods in Digital Humanities*, eds. Kristen Schuster and Stuart Dunn, 91-108. Oxford and New York: Routledge.

Eide, Øyvind. and Oda Bakkeli Eide. 2016. Modeller av mennesker, mennesket som modell [Models of Humans, the Human as a Model]. In *Digital Humanities in the Nordic Countries, Conference Abstracts*, 65–66. Oslo.

Elleström, Lars. 2021. The Modalities of Media II. An Expanded Model for Understanding Intermedial Relations. In *Beyond Media Borders: Intermedial Relations among Multimodal Media*, ed. Lars Elleström, 3-91. Cham: Palgrave Macmillan.

Fishwick, Paul A. 2017. Modeling as the Practice of Representation. In *2017 Winter Simulation Conference (WSC)*, 4276–4287.

Fishwick, Paul A. 2018. A Humanities Based Approach to Formally Defining Information through Modelling. *Historical Social Research, Supplement* 31: 154– 162.

Flanders, Julia and Fotis Jannidis. (eds). 2018. *The Shape of Data in Digital Humanities. Modeling Texts and Text-Based Resources*. London: Taylor & Francis.

Frigg, Roman and Stephan Hartmann. 2012. Models in Science I. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford: The Metaphysics Research Lab, Center for the Study of Language and Information.

Geertz, Clifford. 1973. The Interpretation of Cultures. Selected Essays. New York: Basic Books

Gelfert, Axel. 2016. How to Do Science with Models. A Philosophical Primer. Basel: Springer International.

Kralemann, Björn and Claas Lattmann. 2013. Models as icons. Modeling models in the semiotic framework of Peirce's theory of signs. In *Synthese* 190(16): 3397– 3420.

McCarty, Willard. 2005. Humanities Computing, Basingstoke: Palgrave Macmillan.

Meister, Jan Christoph. 2003. Computing Action: A Narratological Approach. Berlin: De Gruyter.

Olari, Viktoriya, Kostadin Cvejoski, and Øyvind Eide. 2021. Introduction to Machine Learning with Robots and Playful Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* 35(17): 15630-15639.

Propp, Vladimir. 1958. Morphology of the Folktale, Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Rendgen, Sandra. 2018. The Minard System. The Complete Statistical Graphics of Charles-Joseph Minard. New York: Princeton Architectural Press.

Vogels, Oliver and Eymard Fäder. 2018. Using Digital Data to review Gender Construction in the forager Rock Art of the Brandberg/Daureb, Namibia. In *20th International Rock Art Congress*.

Towards an Integrated and Contextualized Research Data Management at the Cluster of Excellence *Africa Multiple*

Jonas Huisl, Myriel Fichtner, Philipp Eisenhuth, Oliver Baumann, Mirco Schoenfeld, Stefan Jablonski, Cyrus Samimi[91]

Abstract

In large collaborative, transdisciplinary and potentially geographically distributed research projects, joint data management enables not only the collection of research artifacts; it also fosters discovery and generation of new knowledge in the form of previously invisible interconnections between data. At the same time, a successful system for Research Data Management (RDM) offers immediate benefits to individual researchers: it helps save time finding, sharing, or editing relevant data objects, and can help prevent data loss through centralized backups. Following generations of researchers might benefit from a smartly maintained data collection. Furthermore, funding bodies increasingly require a concept for RDM as part of a grant proposal. As such, a resilient concept for RDM should be considered part of good scientific practice. With these aspects in mind, we present the Digital Research Environment at the Cluster of Excellence *Africa Multiple* and outline how it can foster best practices for data collection and discovery. Specifically, we introduce *WissKI* (*Wissenschaftliche Kommunikationsinfrastruktur*, "scientific communication-infrastructure"), the system where research data is collected and new interconnections between data items can be determined programmatically.

[91]Affiliation of all authors: University of Bayreuth.

Contact: jonas.huisl@web.de, https://orcid.org/0000-0002-5443-2576. Contact: Myriel.Fichtner@uni-bayreuth.de, https://orcid.org/0000-0003-0412-1881. Contact: Philipp.Eisenhuth@uni-bayreuth.de. Contact: Oliver.Baumann@uni-bayreuth.de, https://orcid.org/0000-0003-4919-9033. Contact: Mirco.Schoenfeld@uni-bayreuth.de, https://orcid.org/0000-0002-2843-3137. Contact: Stefan. Jablonski@uni-bayreuth.de. Contact: Cyrus.Samimi@uni-bayreuth.de, https://orcid.org/0000-0001-7001-7893.

1 Introduction

Managing research data is more than collecting bibliographic entries. It is an integral component of scientific work and collaboration covering both technical infrastructure as well as motivational capability. Research Data Management (RDM) helps researchers save time finding, sharing, or editing relevant data objects. Also, it adds a layer of security to individual data practices through periodic and centralized backups, as well as systematic redundancy, significantly reducing the risk of data loss or corruption. A well-maintained data collection can provide benefits for generations of researchers to come, as it clearly documents the origin of data items and can convey the "big picture" on which different areas of research and their results are laid out. A good data practice increases visibility, reputation, transparency, and verifiability of research outcomes. And, finally, a resilient concept for RDM is part of good scientific practice (Deutsche Forschungsgemeinschaft 2019).

This paper aims to add another layer to the task of RDM: We see RDM as a means to reveal novel inter- and transdisciplinary links and interconnections between researchers. With this paper, we describe our concept of an integrated and contextualized RDM which will serve as a key component of the Digital Research Environment of the Cluster of Excellence *Africa Multiple*.

The Cluster of Excellence *Africa Multiple* is a large international consortium devoted to reconfiguring African studies, interconnecting many researchers from all over the globe and from various scientific disciplines.

At the heart of the Cluster lies the Digital Research Environment (DRE), saving, organizing, and managing heterogeneous research data. Figure 1 depicts the envisioned data life cycle comprising both classical tasks of data management (curating and storing data) as well as data mining efforts to establish a network mapping of data objects, thus revealing novel relationships between data and researchers.

With this paper we outline our efforts in building such an integrated and contextualized RDM that supports this concept of data life cycles as well as the idea of interconnecting research from various disciplines enabling them to pose novel research questions transcending borders of scientific disciplines.

The remainder of the paper is organized as follows. Section 2 highlights important related work from the field. Section 3 outlines core aspects of the technical infrastructure of the Digital Research Environment. Subsequently, Section 4 describes our concept of contextualized ontologies which are the building block of establishing interconnections between data objects. Section 5 concludes the paper.

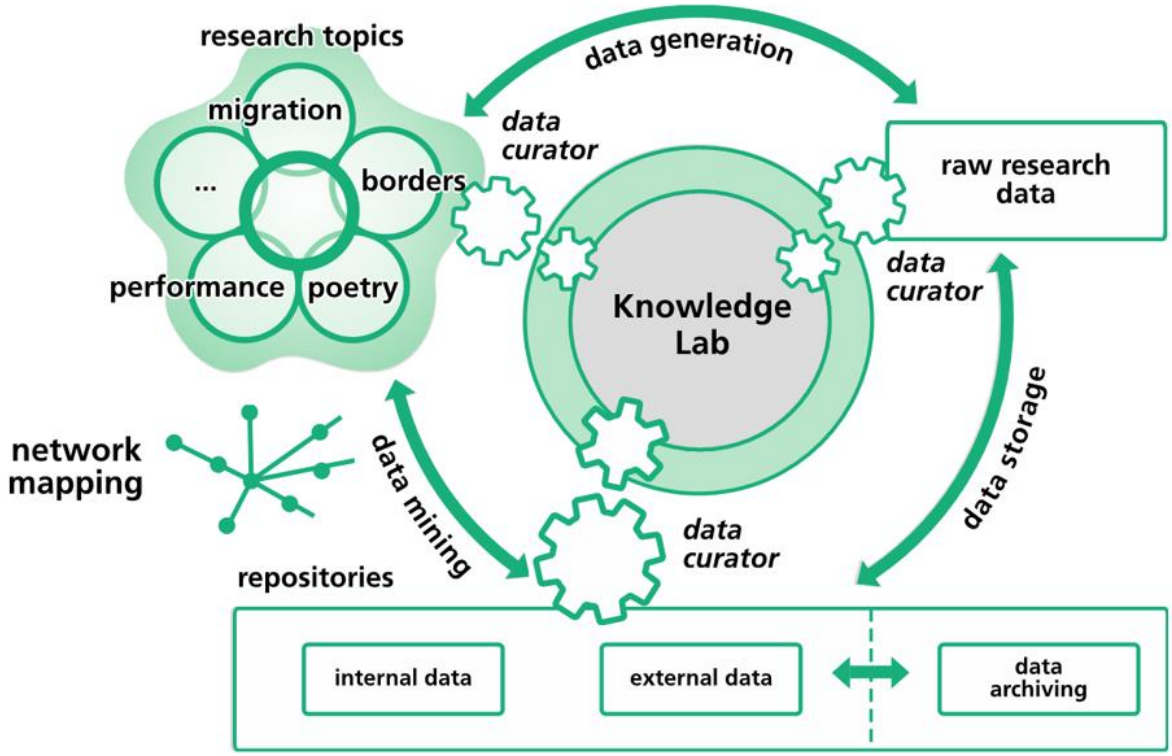


Figure 13: The data life cycle of the Cluster of Excellence Africa Multiple in the Digital Research Environment (DRE).

2 Related Work

Digital Research Environments are typically used to store, present, and manage research data coming from potentially multiple different research areas. Oftentimes, the overarching goal is to serve the needs of a specific research community, for example by helping them find an interdisciplinary solution to a common research question. Typically, DREs are very flexible and can be adjusted to the specific needs of the corresponding research community. Due to their web-oriented conception, they also promote easy collaborative research, which is not limited by geographical location or time zones of the participating users (Candela et al. 2013).

The reuse of research data is promoted by different funding organizations such as the DFG. To ensure the research data meets certain standards to be reused efficiently, different principles have been proposed by the *GO FAIR* initiative (GFISCO 2021). These principles suggest that the data should be described by sufficient metadata to be *Findable*, *Accessible*, *Interoperable* and *Reusable*, especially by automated processes (Wilkinson et al. 2016). *GO FAIR* provide a list of properties for the metadata which should be satisfied to achieve these four requirements.

A promising approach to ensure the metadata fulfills the above criteria in a sufficient manner is the use of well-designed ontologies for data description. Ontologies in the context of computer science have been defined as "*a formal, explicit specification of a shared conceptualisation*" (Studer et al. 1998), building on the original description from Gruber (1993). The idea of a shared

conceptualization mainly supports reuse i.e. by using the same ontology for describing things in different contexts, whereas the idea of a formal and explicit specification ensures the description is available in a machine-readable way. Through reuse and extension of existing, well established ontologies, a consistent understanding of the described elements can be promoted.

In the following we highlight existing projects or frameworks which are using several of the previously mentioned concepts for research data management. The *DARIAH-DE* (SUB Göttingen 2021) infrastructure provides a collection of different tools for building a DRE in the field of humanities and cultural studies.

A further project with a similar vision to the Cluster of Excellence *Africa Multiple* that also utilize the concept of a DRE in combination with an ontology for describing and connecting their research data is shortly described in (Wang et al. 2016). This is an interdisciplinary project that tries to solve a common research question in the context of sustainable manufacturing by tackling it from different viewpoints; an approach closely related to that taken within the Cluster of Excellence *Africa Multiple*.

The research project *Objekte im Netz* (Andraschke and Wagner 2021) (engl. "objects in the web") was realized by *Friedrich-Alexander-Universität Erlangen-Nürnberg* (FAU) in collaboration with *Germanisches Nationalmuseum* (GNM). Their goal was to build an architecture for managing different research data collections from multiple disciplines by using ontologies. For the technical implementation, they used the system known as *WissKI* (*Wissenschaftliche Kommunikationsinfrastruktur*, "scientific communication-infrastructure") and adjusted it to their specific needs. The *WissKI*-project was also chosen as the fundamental framework for our work. Its architecture and functionality are described in greater detail in Section 3.

Contextualized ontologies have been reasoned about in detail by Srinivasan and Huang (2005), who discuss a set of key points by which to identify systems utilizing these types of ontologies. They argue that "fluid ontologies" foster a tight relationship between researchers' interests and the fabric of the collection of knowledge.

One of the key ideas of contextualized ontologies is the aspect of personalization. Huang et al. (2002) explore a hybrid recommender system for a digital library, leveraging both content- as well as community-based ideas to construct a multilayer, graph-based system. They find that the hybrid approach has a low, but significant impact on precision and recall over purely content- or community-based approaches, suggesting that the synthesis of features present in the content- and community-domains is a valuable approach. Chaudhari et al. (2017) implement a privacy-preserving recommender system based on a knowledge graph of entities present in the consumed and recommended content, and the relations between them. They report a twofold increase in precision of the entity-based recommender over a baseline. These results are promising, as their knowledge graph essentially realizes an ontology of entities, leading to a semantic graph where the edges signify the roles taken on by the nodes.

A further key aspect of contextualized ontologies is that participants of a system should be able to create metaviews of the content in the repository, and subsequently share these views with other

participants. This concept is strongly rooted in the field of collaborative visual analytics. Heer and Agrawala (2008) recognize the beneficial impact visual representation of data has on human perception in analytical tasks, and discuss how visual interfaces can enable and foster social collaboration. They provide a comprehensive framework for reasoning about, as well as designing such systems.

The idea of designing interactive systems that provide visualizations to aid the sensemaking process is also prevalent in Rossi and Ahmed (2016), who propose an "interactive data repository" actively supporting scientific research. While the underlying idea is very related to ours the authors limit their repository to a very specific type of data. In our work, however, we envision a system for truly heterogeneous research data.

3 WissKI

As described in Section 1, the Cluster of Excellence *Africa Multiple* at the University of Bayreuth has a clear vision for a Digital Research Environment. One of the core aspects of its plan is a central repository in order to collect and store structured research data of all cluster related projects. Theses data rank from metadata of pictures or films to annotations of text, the researchers' comments, or bibliographical information of research objects in general. As the Cluster of Excellence includes different research areas, the data can be diverse. Furthermore, this shall secure that the data can be used in other contexts for further research like network analysis and knowledge generation. In general, the research data repository itself shall be subject of subsequent research (Deutsche Forschungsgemeinschaft 2021b).

After considering the two possibilities — developing a system from scratch or using an existing one — the decision to use *WissKI* was reached. *WissKI* is an acronym for the German term *Wissenschaftliche Kommunikationsinfrastruktur* (Hohmann and Schiemann 2009), (Scholz and Görz 2012). Users benefit from a simple web-based interface built upon the widely used Content Management System *Drupal* (Drupal Association 2021). Moreover, it enables the use of semantic web technology in the data repository as it extends *Drupal's* relational database management system; thus, it lends itself well as a Digital Research Environment as it enables the storage and linking of data by structuring the content in a semantically enriched data format (Görz 2011).

WissKI was developed by a group of researchers from several German institutions, and the process was funded by the Deutsche Forschungsgemeinschaft (DFG) (Deutsche Forschungsgemeinschaft 2021a). It is used in a variety of research projects not only from universities but also from memory institutions like museums and archives for data management. One central aspect of the system is that it allows the use of an ontology for structuring research data. This means that all inserted data is represented, grouped, and classified by an underlying ontology. Since the ontology is developed in close cooperation with researchers within the Cluster of Excellence *Africa Multiple*, it reflects both technical expertise as well as domain knowledge of the participants. Moreover, the establishment of an ontology that integrates the various research disciplines is an interesting and ambitious research aspect.

It can be stated that the system allows easy access for researchers from all over the world with its web-based user interface. The only prerequisite is a browser with internet access. Therefore the system can be used in a platform-independent manner, supporting researchers from different disciplines who may be using various operating systems and terminals. This does not only simplify access for researchers in Bayreuth but also for our partners from all over the world, as well as use of the system during field trips. An extensive role management helps in the administration of the system. The accessibility of specific data objects, sections of the system or the use of certain functionality, e.g. the creation of new content, can be restricted to selected users or user groups. So, research projects can decide if they want to share their data with other researchers or the public. Further security aspects - like the anonymization of research data - are in the area of responsibility of the researchers.

Additionally, the system provides intuitive input masks in order to efficiently integrate research data into the system. However, researchers need not be familiar with technological aspects as the mapping of data elements to their corresponding ontological elements happens in the background. This functionality is provided by the *Pathbuilder*, which defines relations between fields of the input masks and their corresponding entities in the ontology. Further details regarding the functionality of the *Pathbuilder* can be found in Fichtner and Ribaud (2012). An example in context illustrates its use: Assume there exists a concept *author* for a research project. This concept has several attributes and interacts with other concepts of our ontology, e.g. an *author* has a name, or writes a *book* which has a *title*. For the researcher the technical aspects of these connections are hidden. However, it is possible to interlink entities in *WissKI*, so that the researcher can follow these links on the website and explore the system. In our example, it is thus possible to explore all links connecting the *author* to their other works, or view all items in the systems that link to the book's *title*. These links between the different entities shall play a central role for the further analysis of the collected data. On the one hand, this can be done by specialized software in the fields of network mapping and data analysis. On the other hand, each user can explore the system and find interesting links between different research projects, which is related to the concept of serendipity (Van Andel 1994).

Furthermore, the underlying software is published under open-source licensing. This aligns with the overall principles of the Cluster of Excellence as it does not incur further licensing costs. This technology enables our partners to set up a system on their own. On top of that, it supports several requirements of a researcher workflow oriented system including extensibility, flexibility, and the integration of existing systems.

3.1 System architecture

As already mentioned, there are several aspects that are expected of the overall structure of the DRE. Firstly, it needs an easily accessible and intuitive user interface that reduces technical access restrictions to improve the collaboration between the researchers in the Cluster of Excellence. Secondly, there shall be an ontology layer that structures the immanent concepts of the research projects. This ontology shall be used for the validation of data and enables further research. Thirdly, the data has to be stored in a system that can interact with other RDM systems.

The *WissKI* module for *Drupal* meets all these requirements. The regular *Drupal* configuration[92] provides amongst other things a highly configurable user interface, detailed possibilities to handle access rights, manage users and the option to interlink different entity types. These entity types form the top-level structure for *Drupal* content, e.g. *nodes* or *users*, whereas bundles specify them in a more detailed way. So, the entity type *node* can be an *article* or a *basic page*. Moreover, these bundles contain several fields that add structure to the content. For example, a *Drupal* article has a title and a text field for its content. Now, how does this structure apply to *WissKI* entities? To enable modules like *WissKI* to align with this structure, they have to be designed appropriately. In the following we explain how this structure is applied to *WissKI* entities.

At first we have a look at the underlying ontology. *WissKI* can use an ontology to structure the data and link the entities with each other, as multiple concepts can exhibit multiple relations. In the backend the ontology is saved as an *.owl*-file and loaded into the system. With this, the system automatically determines hierarchies, domains and ranges in the so-called *WissKI* Store Abstraction Layer Zero (SALZ) adapter. In this module the data concerning the graph is interconnected, e.g. where the data should be stored, or which *owl*-properties should be used.

Subsequently, the specific *WissKI* entities are modelled in the *Pathbuilder*, where fields of the entity are matched to a path, which either corresponds to an object property, or a data property from the ontology. This knowledge is then used to store the data in the RDF triplestore as a semantic dataset in the form: subject, predicate, object (ontotext 2021).

In general, bridging the gap between the structure of *Drupal* entities and their relational data model and the *WissKI* entities in the triplestore is one of the main tasks of the *WissKI* module. The connection and interaction between the different systems and software can be seen in Figure 14. It enables all further analysis of the data as it is now highly structured in a semantic graph.

[92] For the regular setup of a Drupal system and its underlying architecture see https://www.drupal.org/docs/installing-drupal (last accessed: 9 May 2022).

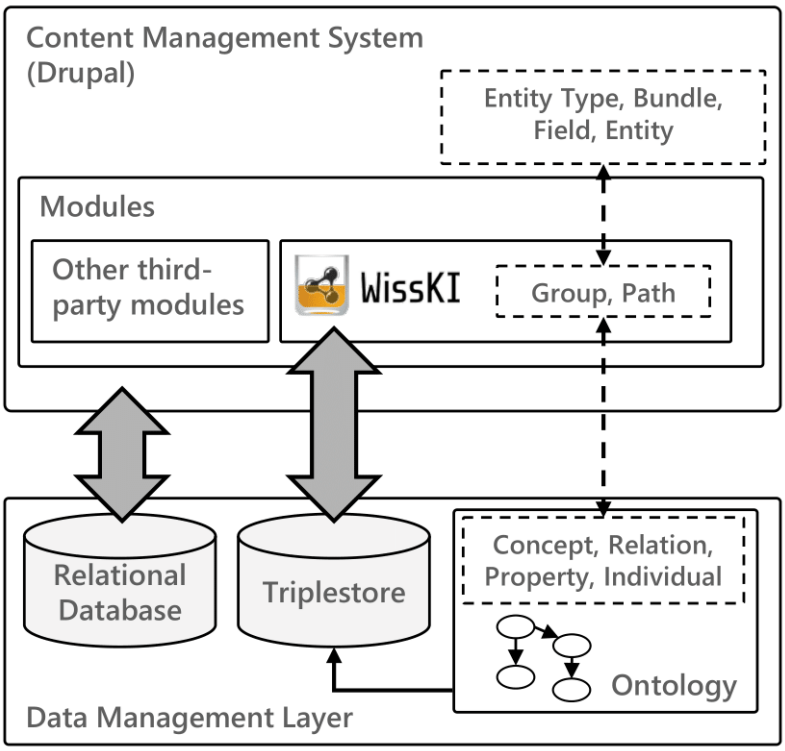


Figure 14: The system architecture of *WissKI*.

3.2 Workflow

3.2.1 System setup

A crucial point of the whole project is the development of a profound ontology that includes all aspects from the different research areas present in the Cluster of Excellence *Africa Multiple*. These are mainly based in the humanities and social sciences, but also include other fields like Law or Climatology. Therefore, it is important to abstract from specialized categories and extract the core aspects of the research, which can be often subsumed in general domains. Here lies one of the most important tasks for the development of the ontology. The regular proceeding for our Digital Solutions team is that we first meet with researchers in order to gain insight into their data. To aid this process, we designed a structured questionnaire with a focus on the following aspects: a general description of the project and its members, how the data gets collected, how the research data is structured, what kind of data is used in the project, what kind of tools or software is used for the organization and analysis of the data, digital preservation, and the sharing of data among the researchers and in the scientific community in general. Afterwards, we also ask for an exemplary data set, which can provide further insights and helps in order to integrate the structure into the already existing ontology.

As the Cluster of Excellence *Africa Multiple* includes various research areas, there are big differences in the use of digital data and tools. Therefore, different tasks have to be addressed, ranging from the creation of an initial data scheme from scratch (as the research data has not been structured previously) to the integration and transformation of already existing data management

systems. After extending the ontology according to a specific research project, this structure also has to be modelled in *WissKI*.

When this procedure is concluded for all entities of the research data, we continue with the establishment of a first prototype for a project in our *WissKI* system. After the presentation of the prototype to the research projects, it is evaluated and possibly reworked. This iterative process can be repeated several times, as new insights are often gained while working with the system.

3.2.2 Extensions and Adaptions

One of the great advantages of *WissKI*, apart from using semantic web technologies, is that it is based on *Drupal*. On the one hand, this allows us to implement modules on our own, as *Drupal* is an open-source software project. On the other hand, we can also extend the basic *WissKI* system with already existing *Drupal* modules to match requirements from the research groups. The great advantage is that we can build upon the efforts of a highly active community from all around the world. Thus, we benefit from this diverse community regarding not only their applications but also their cultural background and technical experience. However, mostly it is a mixture of these two approaches. As *WissKI* extends *Drupal* with certain functionality, it is often necessary to extend existing *Drupal* modules so that their intended behaviour can also be applied to *WissKI* entities.

3.2.3 Extension of *WissKI*

Moreover, we can participate in the further development of *WissKI*. Compared to the *Drupal* community as a whole, the *WissKI* community is rather small. However, it is an open community that encourages users to help increase the usability and functionality of the system. There are recurring events like user meetings or hackathons, where new features are developed collaboratively by mixed teams from the different institutions that are using *WissKI*. New *WissKI* versions are released regularly which include the progress of such events as well as further extensions which are developed by members of the IGSD[93] which is an association founded by the *WissKI* main developers and supporters.

3.3 Restrictions

Several challenges arise from the use of *WissKI* and the structure and organization of the Cluster of Excellence *Africa Multiple*. Firstly, when we compare our approach with other projects and institutions that use *WissKI* one deviation is obvious: most of the other users do not develop their own ontology, but use CIDOC CRM94, which was designed especially for cultural heritage institutions. However, we extend this ontology and integrate special needs from the research groups if the existing schema is not sufficient.

Secondly, it has to be clear that there are also restrictions for the development of new features and the adaptation of existing modules. As stated earlier, it is often necessary to adapt existing

[93] Interessengemeinschaft für semantische Datenverarbeitung e.V., http://www.igsd-ev.de/ (last accessed: 9 May 2022).

[94] http://www.cidoc-crm.org/ (last accessed: 9 May 2022).

modules so that their functionality can also be applied to *WissKI* entities. Although it is possible to modify any part of the software, normally one does not change the Drupal core code. Otherwise, problems with compatibility may arise when new major or minor versions of Drupal are released as this is the central software element and many extensions and additional modules are developed based on this code. Therefore, modifications of the program code are mostly limited to modules, including *WissKI* itself. However, one has to be aware that all these adaptions have to be kept in mind as they potentially cause problems in the future and have to be maintained.

Thirdly, the rather small *WissKI* community has several restricting implications, but it has also positive aspects. Obviously, when fewer people use a system they either have to be very active or there will be less publications, additional materials and discussions groups or fora. However, we were able to benefit from this fact as right from the beginning we were able to get to know the developers and users, participate in community events, and receive support. Additionally, we are directly involved in the development of new features. This mutual support allows us to gain further insights and additionally help in the establishment and improvement of this project.

In summary, *WissKI* offers a wide variety of possibilities for researchers and users. It fulfills the core requirements for the DRE and allows us to contribute to its further development and establishment in the area of digital research environments.

4 Contextualized Ontologies

At the heart of *WissKI* lies its ontology. The categories defined therein serve their purpose in structuring the large and heterogeneous repository of knowledge. The ontology enables researchers to see and specify which data items may be related, e.g. because they share attributes or are descendants of a common parent-class.

The *WissKI*-ontology is defined *a priori*, i.e. it is clear what categories it encompasses before new data items are added to the repository. Researchers contributing new data are involved in the evolution and adaptation of the existing ontology, as outlined in Section 3.2. They contribute samples of their data, based on which the existing categories are evaluated and extended, if required.

This approach leads to an ontology that is tailored to the data flowing into the repository, and that serves as a point of reference for researchers wishing to integrate their data into the system.

4.1 Implications

Several implications on how different personas can interact with *WissKI* arise from this process. We shall distinguish between the persona of the *creator* and the *consumer*; while a creator adds new data items to the repository, the consumer usually poses specific queries against the system to extract information.

A researcher can take on any of these two roles by either integrating their research into *WissKI*, or by conducting research on the items already present.

4.1.1 Fixed system of reference

As the ontology used by *WissKI* is largely defined *a priori* and only extended on a case-by-case basis when required, a system of reference is already in place when new data enters the repository.

For the creators of new content, this means they need to integrate their data into this fixed system and find its place within the larger framework. Creators can, however, rely on this system being established by means of an informed process, guided by the knowledge and evaluation of domain experts. This ensures that the system as a whole remains applicable to the many different data items flowing in. Furthermore, although the ontology is static to a large degree, it can still evolve if the current state clashes with the expectations of creators.

The consumers conducting research with *WissKI* are also confronted with this fixed frame of reference, which should help them navigate the space of available data points. They can thus utilize the ontology as a map of the repository, and navigate the different sub-spaces shaped by the classes making up the current state. These aspects may also help consumers re-formulate existing queries they pose against the system to apply them to other parts of the repository. For instance, a researcher seeking paintings created in a specific socio-political setting should be able to easily adapt their query to also return music and literature by referencing the relevant elements of the ontology.

4.1.2 Informed process

As *WissKI* includes informed decisions by experts into the formulation of the ontology, creators can actively participate in extending and adapting the system by pointing out shortcomings and suggesting new categories. However, their influence on a more nuanced formulation of existing elements may be restricted, to avoid fragmenting the ontology into highly specific branches that become applicable only to single fields of research, and thus compensating generality with specificity.

As a result of this informed process of evolution, consumers can rely on the ontology's applicability to the data present in the repository, and the knowledge of experts having influenced the classification of items. Consumers can trust the current state of the repository to be the result of those contributing data also determining where that data is placed in the larger system of references and referents.

4.2 Augmenting with context

It is important to note that contextualized ontologies do not replace the fixed system of reference. Rather, they help establish an additional mode by which to navigate the repository. This approach is based on the concept of "fluid ontologies" proposed by Srinivasan and Huang (2005). They envisioned a system for personalized digital museums or inventories that allow for novel ways of interaction with the collection. In essence, they provide four key points that aid in reasoning about fluid ontologies:

1. Content creators are *involved* in the definition and extension of the ontology, and can provide guidance on where their data should be located within the space.

2. Users can create and share *metaviews* of the repository, illustrating how they navigate and make sense of its content.
3. The ontology is *adaptive* and *evolves* over time to account for changes in how users experience the content.
4. Artificial bots utilize users' interaction histories and contribute elements of *personalization* that are tailored to what the user has seen in the past and may expect to see in future.

From their point of view, users interacting with a collection might help in revealing novel interconnections between items that would have been impossible to identify from the ontology alone. By letting users create metaviews and by adapting to these metaviews to some extent, the ontology of a digital collection will experience fluidity.

For our system, we build on the idea of an evolving ontology. Since systems of RDM are guided by different principles than those of digital museums, we have to specify how to introduce fluidity in a controllable and well-defined manner. Therefore, we operationalize the concept of fluidity by taking into account *context information*. That is, we enrich the ontological structure with attribute information on entities and relations and use this attribute information to identify novel interconnections.

We call this attributed ontological structure *Contextualized Ontologies*. The following analysis will outline different types of contexts that can help enrich the experience of an integrated and contextualized RDM.

4.2.1 Context from users

From user's interactions with *WissKI*, context attributes can be extracted, i.e. information that directly relates to users, but aims at explaining how or why they are performing an action rather than pertaining to the raw, unprocessed action itself. As an example, consider the user that picks an individual item from a collection to inspect in more detail; while the system can record the simple event that the user expanded on this item, context helps us understand by which path they arrived there, and perhaps project into the future what other items they may inspect. By recommending these to the user, an RDM environment can support uncovering new connections between data.

Context can also be derived from a user's interaction with other users of the system. This lets us envision an architecture where either unary ("subscription", "following") or mutual ("friendship", "collaboration") relationships form among participants. These relationships can again be exploited to guide and improve the experience of the user when engaging with the system as a whole, e.g. by opening up new pathways of collaboration among researchers in the same field.

Referring back to the key points guiding the analysis of contextual ontologies, user-context and especially interaction histories can be valuable instruments for creation of new content. By taking a user's history into account, the system is able to leverage its understanding of previous content published by them to suggest categories, or pre-fill certain values that rarely change in their

contributions, thereby increasing speed and consistency for the data-entry task. It can also help shape the ontology if users can directly contribute suggestions, either when creating or browsing content, to account either for misconceptions present in the current state, or to allow gradually giving more nuance to existing classifications.

Rich interaction histories allow the system to create and suggest metaviews of the repository, including perhaps only the sub-space of items the user has shown interest towards in the past. When including items from the fringe of this sub-space by following references to other spaces, the user may find new trajectories through the knowledge system.

Furthermore, by consulting users' previous experience of the system, it is possible to establish links between sub-spaces if they are frequently explored in conjunction. For instance, if a user regularly contributes to or consumes from the sub-spaces "sociology" and "music", these spaces may be of particular interest to them. These links should have a lower valence than regular links in the ontology, but can nevertheless serve as orientation both in the placement of new content as well as the browsing experience.

Finally, software in the form of crawlers can analyze the user-generated metaviews and thus accumulate and suggest new views, either globally or on a per-user basis. Crawling the network of user-user relationships may further uncover connections between closely and loosely related groups in the social graph by feeding data to community detection algorithms implemented as part of the system. By analyzing commonalities among items viewed by users, collaborative filtering approaches popular in the design of recommender systems can further improve the discovery of similar or related content frequently appearing in the user's cohort.

While the ideas noted in this section can help consolidate and improve the user experience, it must be noted that users should be made aware of the fact that their actions may have a direct consequence on how the system behaves. They should be given the choice of contributing to the collection of interaction histories, or remaining anonymous to these features; perhaps even to an extent of not receiving any personalized suggestions at all, and instead experiencing the system without any notion of user-context.

4.2.2 Context from data

Another type of data we can utilize to construct a more fluid notion of ontologies can be extracted from the actual data present in the repository. Within the Cluster of Excellence *Africa Multiple*, rich and diverse forms of data are collected, such as biographic information, images, but also artistic artifacts such as music, paintings, and literature. With regard to these types of content, metadata offers one way to enrich the system with additional context. Specifically, this relates to metadata in the form of annotations added by a human participant, but also via content-aware crawlers analyzing the repository. As an example, consider perhaps a piece of music that is annotated automatically with a label referencing the person contributing it to *WissKI*, along with a timestamp of when this event occurred. This piece could then be further annotated by a participant with a set of keywords describing the piece (tagging).

While this form of contextual data provides a valuable route, we shall focus on a different type of data in the following section. We believe that rich context can also be extracted from latent metadata which is inherent in the individual items, but not collected or annotated explicitly. Latent metadata refers on one hand to elements of human perception, such as mood, tension, or conflict; on the other hand, it includes properties of the actual data that is not part of the ontology itself. For a piece of music, this may be its tempo or features of its frequency domain; for paintings and images, the distribution of colours, or whether lines or areas of colour dominate; and lastly, for literary texts or written biographies, which topics are present in the work, or which sentiment and emotional valence defines it. We believe that in constructing context from this type of data, items can be interrelated based on common or similar properties, even if they do not share links in the original ontology, and thus a better understanding of the space as a whole can be obtained.

With regard to the key points outlined earlier, context from data can be beneficial in the construction of metaviews, as it allows for discovering new relationships among data items and including them either in views generated by the system, or in user-curated views. Furthermore, this contextual data can be aggregated and used to visualize clusters of items sharing the same context, and in doing so provide insight into how the repository as a whole is structured.

This notion of self-structuring may also prove valuable to the adaptiveness and evolution of the ontology, as it can allow to suggest interrelations during creation time based on how similar items have been placed in the repository. As an example, consider the addition of a piece of music authored by a composer who is already represented in the system. Based on the mood or tempo prevalent in the composer's other works, suggestions can be provided to the user to consider whether the work at hand shares these attributes. In general, exploiting the connections between items and their levels in the ontology, a richer and broader understanding of the system can be obtained.

Finally, automated crawlers can analyze the data to obtain the features mentioned previously, as well as aggregate and relate them. This information can then be fed into the subsystems responsible for constructing metaviews or generating recommendations, leading to a fully automated process. Using context retrieved from the data itself rather than cohort analysis provides a promising route to explore recommendation systems, as it takes into consideration the actual items rather than "flocking" of other users around prominent items.

4.3 Potential Drawbacks

The previous section noted how an RDM environment can leverage user- and data-context to improve the user-experience, e.g. by recommending items of interest and potential collaboration partners; supporting the browsing- and data-entry-tasks; and in general fostering a broader understanding of individual branches of research and how they relate to each other.

While we believe that both forms of context can contribute to a richer experience a brief discussion of potential drawbacks is appropriate.

First, conflating system-generated with user-generated content (metaviews, annotations, etc.) poses a risk to the overall quality of data. Any automated system, as fine-tuned as it may be, should

be expected to make false assessments, leading in the worst case to misclassification of data. Thus it is desirable to explicitly distinguish auto-generated content from contributions by real users, and provide a means to edit or delete such false assessments. This not only provides orientation to the user, but also gives them a way to exercise quality control over a semi-automated RDM.

Furthermore, by incorporating contextualized ontologies, they may shape a system in a way not intended by content creators, e.g. by over-emphasizing users' metaviews rather than following the path suggested by the creator. It seems as though especially community-based recommender systems have the potential to distort the experience in favour of prominent content, to the detriment of research located at the fringe. We believe that when incorporating these features into a system, it should be done sparingly and considered an addendum rather than the only or best way by which to experience the knowledge system.

Lastly, contextual ontologies pose the risk of losing the reference-character and orientation a fixed ontology provides. A fixed ontology communicates structure and substance of a knowledge space, it provides a system of references and referents that helps navigate a potentially vast space of information. Contextualized ontologies, however, are rarely clearly formulated; rather, they should always be considered "in flux" and evolving, adapting to the participants interacting with them, and the data they are grounded upon. Therefore, care should be taken to consider contextualized ontologies as means to an end, but not the end itself.

4.4 Exploiting Contextualized Ontologies in an Integrated RDM Environment

Within the Cluster of Excellence *Africa Multiple*, and specifically within *WissKI*, we wish to explore two key ideas in greater detail. We believe that both the automated and manual creation of metaviews provide interesting approaches for researchers to explore the rich repository of data containing contributions from a variety of different branches of research. Understanding one's own field consists not only of knowing what elements are similar, but also where they share borders with other fields. In this sense, visualizing the space of similarities and differences can empower researchers to gain insight into how their field relates to others, giving hints to potentially fruitful topics for interdisciplinary collaboration. Using the semantic graph that underlies *WissKI* may be a rewarding first route into this exploration, as it allows us to establish relationships between similar items, and determine difference perhaps in the simple absence of interrelations between concepts. By implementing crawlers that collect these relationships, we can ensure an adaptive and evolving system that updates existing views as and when new connections are formed, and avoid disconnecting users' views from the updated state of the repository.

A second idea we wish to follow is that of personalization. Previously, we noted that user-context can provide rich interaction histories, and data-context allows us to aggregate latent metadata inherent in the repository. We believe that when combined, these approaches have potential to open up the vast knowledge system to its users, be they creators or consumers. Creators may benefit from personalization when the system utilizes its knowledge of the content usually created by them to suggest categories from the ontology, or popular connections to related items. Consumers, on the other hand, may gain a richer experience of the research space when their

interaction histories are used to suggest items that are either similar in nature, as in the case of pieces of art that are visually similar; or similar in exhibition, or placement, as in the case of literary texts and musical compositions sharing a common underlying theme, e.g. "conflict".

In conclusion, we wish to augment the experience of a heterogeneous and broad knowledge system such as *WissKI* by incorporating elements commonly found in contextualized ontologies, specifically those of metaviews and personalization, for both content creators that contribute new research, as well as consumers that wish to utilize the system for their own research by browsing and interacting with the repository and its users.

5 Conclusion

In this paper, we outlined our efforts in building an integrated and contextualized environment for Research Data Management (RDM) which serves as a central hub for research data occurring at the various institutions from around the world being part of the Cluster of Excellence *Africa Multiple*.

At the heart of our platform lies a semantic graph serving as an ontology among data objects. In order to establish a novel way of accessing data collections, we will enrich the graph structure with context information building a system of recommendations for research data. Such context information will originate from the data itself as well as from interactions with the data and the collection.

This enables our system to interconnect researchers from various fields, thus paving the way to formulating novel research questions which transcend borders of scientific disciplines.

6 References

Andraschke, Udo, and Sarah Wagner. 2021. "Objekte im Netz." In *Objekte Im Netz*, 115–28. transcript-Verlag.

Candela, Leonardo, Donatella Castelli, and Pasquale Pagano. 2013. "Virtual research environments: an overview and a research agenda." *Data Science Journal*, GRDI–013.

Chaudhari, Sneha, Amos Azaria, and Tom Mitchell. 2017. "An entity graph based recommender system." *AI Communications* 30 (2): 141–49.

Deutsche Forschungsgemeinschaft. 2019. "Guidelines for Safeguarding Good Research Practice. Code of Conduct." Deutsche Forschungsgemeinschaft. https://doi.org/10.5281/zenodo.3923602.

Deutsche Forschungsgemeinschaft. 2021a. "WissKI- wissenschaftliche KommunikationsInfrastruktur." *GEPRIS – Geförderte Projekte Der Deutschen Forschungsgemeinschaft*. https://gepris.dfg.de/gepris/projekt/64181911 (last accessed: 9 May 2022).

Deutsche Forschungsgemeinschaft. 2021b. "EXC 2052: Africa Multiple: Reconfiguring African Studies." *GEPRIS – Geförderte Projekte Der Deutschen Forschungsgemeinschaft*. https://gepris.dfg.de/gepris/projekt/390713894?language=en (last accessed: 9 May 2022).

Drupal Association. 2021. "Drupal." https://www.drupal.org/ (last accessed: 9 May 2022).

Fichtner, Mark, and Vincent Ribaud. 2012. "Paths and Shortcuts in an Event-Oriented Ontology." *Metadata and Semantics Research* 343: 214–26. https://doi.org/10.1007/978-3-642-35233-1_22.

GFISCO. 2021. "GO FAIR." https://www.go-fair.org/ (last accessed: 9 May 2022).

Görz, Günther. 2011. "WissKI: Semantische Annotation, Wissensverarbeitung und Wissenschaftskommunikation in einer virtuellen Forschungsumgebung." https://www.academia.edu/9817244/ (last accessed: 9 May 2022).

Gruber, Thomas R. 1993. "A translation approach to portable ontology specifications." *Knowledge Acquisition* 5 (2): 199–220.

Heer, Jeffrey, and Maneesh Agrawala. 2008. "Design Considerations for Collaborative Visual Analytics." *Information Visualization* 7 (1): 49–62. https://doi.org/10.1057/palgrave.ivs.9500167.

Hohmann, Georg, and Bernhard Schiemann. 2009. "An ontology-based communication system for cultural heritage: Approach and progress of the WissKI project." *Scientific Computing and Cultural Heritage. Contributions in Mathematical and Computational Sciences* 3: 127–35. https://doi.org/10.1007/978-3-642-28021-4_14.

Huang, Zan, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. 2002. "A Graph-Based Recommender System for Digital Library." In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, 65–73. JCDL '02. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/544220.544231.

ontotext. 2021. "What is an RDF Triplestore?" https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/ (last accessed: 9 May 2022).

Rossi, Ryan A., and Nesreen K. Ahmed. 2016. "An Interactive Data Repository with Visual Analytics." *SIGKDD Explor. Newsl.* 17 (2): 37–41. https://doi.org/10.1145/2897350.2897355.

Scholz, Martin, and Günther Görz. 2012. "WissKI: A Virtual Research Environment for Cultural Heritage." In *Proceedings of the 20th European Conference on Artificial Intelligence*, 1017–18. ECAI'12. NLD: IOS Press.

Srinivasan, Ramesh, and Jeffrey Huang. 2005. "Fluid ontologies for digital museums." *International Journal on Digital Libraries* 5 (3): 193–204. https://doi.org/10.1007/s00799-004-0105-9.

Studer, Rudi, V Richard Benjamins, and Dieter Fensel. 1998. "Knowledge engineering: Principles and methods." *Data & Knowledge Engineering* 25 (1-2): 161–97.

SUB Göttingen. 2021. "DARIAH-DE." https://de.dariah.eu/ (last accessed: 9 May 2022).

Van Andel, Pek. 1994. "Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability." *The British Journal for the Philosophy of Science* 45 (2): 631–48. http://www.jstor.org/stable/687687.

Wang, Wei Min, Tobias Göpfert, and Rainer Stark. 2016. "Data management in collaborative interdisciplinary research projects–Conclusions from the digitalization of research in sustainable manufacturing." *ISPRS International Journal of Geo-Information* 5 (4): 41.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (1): 1–9.

Linguistic Diversity and Bias in Online Dictionaries

Gábor Bella (University of Trento)[95]
Khuyagbaatar Batsuren (National University of Mongolia)
Temuulen Khishigsuren (National University of Mongolia)
Fausto Giunchiglia (University of Trento)

Abstract

Traditional bilingual dictionaries, once pivotal translation tools, have been superseded on the Web by *multilingual lexical databases* that interconnect the lexicons of hundreds of languages, built for both human and computational uses. A close look at the structure of such databases reveals, however, a form of *linguistic bias*, namely an inbuilt preference towards the semantic space of dominant cultures, preventing the definition of locally specific words or failing to provide the means to connect them meaningfully to other languages. This paper presents our efforts in building a "diversity-aware" multilingual lexical database designed to integrate locally specific vocabularies for any language. We demonstrate our results so far and outline the human-centred methodology by which locally relevant linguistic content can be formalised and inserted into the global semantic space of multilingual lexicons.

1 Introduction

While the Web has been a successful catalyst of cross-cultural exchange, the push for unity also led to greater uniformity and did not benefit all communities to the same degree. The fading of local cultures has accelerated, especially of those that lack a dominant online presence. Within the context of language use, this two-faced phenomenon is known as the *digital language divide*: the

[95] Contact: gabor.bella@unitn.it https://orcid.org/0000-0002-3868-1740. Khuyagbaatar Batsuren https://orcid.org/0000-0002-6819-5444. Temuulen Khishigsuren https://orcid.org/0000-0001-5251-5733. Fausto Giunchiglia https://orcid.org/0000-0002-5903-6150.

vast majority of online content is expressed in a handful of dominant languages, while thousands of smaller—minority or "low-resource"—languages enter into a vicious circle of decreasing use and, in turn, decreasing utility perceived by its speakers or learners, ultimately ending in *digital language death* (Kornai 2013).

The preference of online tools and resources for the dominant languages has deeper and more pervasive effects than one might think. Beyond the presence or absence of written material in a given language, the language used to express the information exerts bias on the content itself (Graham and Zook 2013). If a certain community (for example, from central Africa) is devoid of the means of self-description in their own language—e.g. through Wikipedia pages about themselves—and consequently is described in the language of a more influent (Anglo-Saxon, French-speaking, etc.) community, the information will likely reflect the cultural or geo-political views of the latter. Likewise, it has been demonstrated that the same web search expressed in different languages, such as a search for restaurants in the West Bank expressed in Arabic, Hebrew, or English, will typically provide markedly different results: the content served depends on the language spoken (Graham et al. 2014).

This paper focuses on the online support of languages from the perspective of language technologies and their underlying resources. More specifically, we turn our attention to multilingual lexical databases (MLDB), the modern equivalents of bilingual dictionaries. These resources are widely used by humans and automated language technologies alike, and consequently have an influence over what can and cannot be expressed online, whether directly by people or by automated tools such as a translator website.

As we show, MLDBs are indeed prone to bias against certain languages due to both *incorrect* and *incomplete* coverage of their vocabularies. While missing content is a reason for bias that is trivial to understand (even if far from trivial to address), a more insidious reason for bias is the *structural inability* of MLDBs to express what we call *lexical diversity,* that is, lexical meaning that is specific to a given language or culture. If an electronic dictionary is structurally unable to lexicalise certain words, the consequence is that those words will not be proposed to human users, may not appear in machine translation results, which eventually leads to a loss of linguistically and culturally specific vocabulary. Based on the idea from (Ricœur 2004) and also (Hartmann 1985), according to which equivalence of meaning is *produced* by translation rather than presupposed by it, it is important to understand that, in practice, translation resources and their authors *prescribe* cross-lingual correspondences, rather than merely describing them. The resources created today are, therefore, eminently responsible for the quality of inter-lingual communication and understanding in the future.

In Section 2 of this paper, we provide a general presentation and illustrations of lexical diversity and bias. Section 3 presents how standard bilingual dictionaries have traditionally addressed lexical diversity. Section 4 provides a more detailed look into the structural reasons of diversity and bias within recent multilingual lexical databases. Section 5 presents our research on the building of a diversity-aware and bias-free database, the Universal Knowledge Core. Section 0 illustrates the development of the lexical database on two case studies. Finally, Section 7 provides perspectives for our work.

2 Lexical Diversity and Bias

Translation has been a widely studied topic in linguistics, and is understood to be a complex problem, with pervasive phenomena of *untranslatability* appearing on multiple linguistic levels. On the lexical level, on which we are focusing, untranslatability manifests itself in the clearest way through the phenomenon of *lexical gap*, when a word in a source language does not have a precise and concise translation into a given target language. Table 2 below illustrates this phenomenon for seven, more or less specific word meanings around the concept of *rice* (cooked/uncooked, white/brown, polished/in the husk, etc.), in five languages.

One can observe that none of the five languages has concise lexicalisations for all of the meanings given, yet each meaning is lexicalised in at least one language. In a context of automated or human translation, depending on the degree of formality or precision required—everyday conversation, a cookbook, or a scientific article—choosing an approximate correspondence may or may not be acceptable. Substituting a term with a broader one, such as using *rice* to translate the Swahili *mchele* (which means *raw rice*), is logically correct but results in information loss, while the same substitution in the other direction may result in injecting unintended meaning, such as *"I am eating raw rice"*.

Table 2: Lexicalisations of seven word meanings around the concept of "rice", in five languages.

Definition	English	French	Swahili	Hindi	Japanese
grains used as food, polished or not	**rice**	**riz**	*GAP*	*GAP*	*GAP*
grains still in the husk	**paddy**	*GAP*	**mpunga**	**धान**	**籾**
uncooked rice, brown or white	*GAP*	*GAP*	**mchele**	**चावल**	米
brown rice, cooked or uncooked	**brown rice**	**riz complet**	*GAP*	*GAP*	*GAP*
cooked rice	*GAP*	*GAP*	**wali**	**भात**	ご飯
uncooked brown rice	*GAP*	*GAP*	*GAP*	*GAP*	玄米
uncooked white rice	*GAP*	*GAP*	*GAP*	*GAP*	白米

It is not generally possible to explain or predict lexical gaps by abstract rules or recurrent patterns (Lehrer 1970). Still, gaps are often the linguistic manifestation of spatially or culturally defined specificities of a community of language speakers, as in the example above: in societies where rice constitutes a staple of the diet and is a culturally important concept, the related vocabulary is richer and finer-grained, while in other (e.g. Northwestern) societies the general lexicon includes only a few broader terms, and the rest of the meanings are either expressed as free combinations of words, such as in 'uncooked white rice', or are considered as specialised terms used only by domain experts and unknown to the general public.

From the example it should be clear that such cases of *lexical diversity* should not be considered as exceptions that are safe to ignore in everyday language use. On the contrary, linguistically or culturally specific words often bear a high local importance, as rice does in East Asia, and the

recurrent failure of a resource or tool—such as a dictionary or a translation device—to represent them results in what we call *lexical bias* against the language in question.

3 Traditional Bilingual Dictionaries

The bilingual dictionary has been the principal tool supporting foreign-language communication. Dictionaries have existed ever since the dawn of writing: we know of a Sumerian–Akkadian dictionary written in cuneiform (Landsberger et al. 1970). While their appearance and structure have varied across languages and cultures, they have always been considered as pragmatic tools for solving real-world translation problems, as opposed to their monolingual counterparts that have traditionally been more grounded in linguistic theory (ten Hacken 2016). Modern bilingual paper dictionaries cover general language separately from specialised languages, among others due to size limitations, and they provide a limited number of translations to choose from. The most important lexico-semantic phenomena they tackle are polysemy and homography (the same word having multiple meanings), by providing numbered translation entries, while synonymy within the target language is addressed by providing a list of target words.

Dictionaries sometimes are also able to represent cross-lingual diversity, such as untranslatability, approximations, or meaning shifts. High-quality dictionaries contain lexical gap entries, where the translation is given as an approximate description of the meaning using a free combination of words. Bilingual dictionaries being asymmetric by design—the list of headwords from language A to language B and those from B to A are not mirrors of each other and are, to some extent, authored independently—they are structurally capable of covering language and culture-specific words in both directions: meanings in A that are gaps in B, and vice versa. In the case of existing lexicalisations, however, the "faithfulness" of the translation is not explicitly indicated: each headword provides one or more translations, usually not providing the precise mapping type (such as towards an equivalent, a broader, or a narrower term). The practical task of translation is always apprehended as the best possible, or at least a "good enough" approximation of the original meaning. In order to help in the translation process and provide clues for choosing the target word, dictionaries sometimes annotate words by definitions, examples of use, or indications of style and register.

Of course, the mapping ability of the dictionary in terms of structure and form is merely a prerequisite, and not a sufficient condition, for unbiased coverage. Whether the actual dictionary takes advantage of these structural abilities depends on the lexicographers and their familiarity with either language and their respective cultural contexts. Lexicographers who are native speakers of the local language would necessarily provide markedly different coverage from, say, European missionaries operating within communities of lower degrees of literacy, leading to differing levels and forms of bias in lexical coverage.

4 Online Dictionaries

Digitisation has done away with the size limitations of paper dictionaries, allowing them to grow both in terms of the number of headwords and in the richness of information provided (etymology, derivations, morphological properties, etc.). The traditionally top-down, expert-driven data collection methods were complemented by bottom-up approaches based on the input

of a wider pool of native speakers, as in the case of Wiktionary.[96] Another novelty was the formal, machine-interpretable representation of certain resources, such as in the case of monolingual wordnets (Miller 1995). Machine-exploitability has become a catalyst of a wide variety of language technologies that exploit monolingual or multilingual lexicons, from sentiment analysis to machine translation.

A major paradigm shift on how word and their meanings are mapped across languages was introduced by multilingual lexical databases. Examples of such multilingual lexicons are PanLex (Kamholz et al. 2014), EuroWordNet (Vossen 1998) and its successor the Open Multilingual Wordnet (Bond and Foster 2013), BabelNet (Navigli 2010), or the Universal Knowledge Core (Giunchiglia et al. 2018). The shift consisted of moving from language pairs to the simultaneous coverage and mapping of tens, hundreds, and even thousands of languages within a single resource. It is easy to see that the traditional pairwise approach is not scalable in such settings: full mappings among n languages would require $n(n\text{-}1)$ separate dictionaries, which for a high number n of languages is beyond what could be considered as reasonable effort.

MLDBs thus had to offer a more scalable internal structure, which they did through the trick of introducing a *pivot* representation of lexical meaning, to which all monolingual lexicons are mapped. The pivot is sometimes an abstract meaning space, and sometimes that of one of the languages covered, as we will see later. Mapping languages through a pivot requires only n mappings as opposed to $n(n\text{-}1)$. Using a pivot, however, has major theoretical side-effects on translation ability. Let w_A be a word in language A, w_B a word in language B. Let us suppose that the respective meanings m_A and m_B of these two words are mapped together through a pivot meaning p: $m_A \leftrightarrow p \leftrightarrow m_B$. This mapping assumes the symmetry of the translation operation, which is known by both linguists and translators not to be the case in general (Adamska-Sałaciak 2010). To take an example from Table 2 above, translating the Swahili *mchele* into *rice* may be a reasonable approximation, but translating *rice* into *mchele* may be wrong in certain contexts (e.g. in *"I am eating uncooked rice"*). The assumption of symmetry is correct if and only if the mapping relation across word meanings is truly a symmetric operation, such as in the case of equivalence: however, the equivalence of meanings across languages is generally an overly strong assumption.

Another premise encoded into the mapping structure $m_A \leftrightarrow p \leftrightarrow m_B$ is that translation can be *transitive*: if p is a faithful translation of m_A, and m_B is a faithful translation of p, then m_B is, in turn, a faithful translation of m_A. Again, this is not necessarily true in practice, as cross-lingual mappings of meaning are rarely perfect, and multiple cross-lingual "hops" lead to more and more important shifts in meaning. Needless to say, the choice of the pivot meaning space greatly influences the overall representation ability of the entire multilingual resource. As the pivot connects all languages together, it ultimately determines the meaning space of the entire database. It is thus crucial that the pivot should be as rich and fine-grained as possible if it is to embrace the lexical diversity of hundreds, if not thousands, of languages.

[96] http://www.wiktionary.org (last accessed 6 May 2022); Wiktionary, however, should rather be considered as a collection of monolingual dictionaries organised around word forms, as it does not systematically provide cross-lingual translations.

Let us examine a few of the most widely used MLDBs today, in terms of their use of pivots and their representation ability of lexical diversity. In the case of Open Multilingual Wordnet (OMW), the pivot representation is the meaning space of the English language, as defined by the Princeton WordNet (Miller 1995). On the one hand, this choice makes practical sense, as among all similar resources the Princeton WordNet offers by far the widest coverage of word meanings. On the other hand, its meaning space is clearly oriented towards Anglo-Saxon cultures (e.g. it lexicalises the constituents of the school system with such culture-specific entries as *preschool*, *grammar school*, or *Eton College*). The OMW would map the Swahili and Japanese words for rice through English as follows:

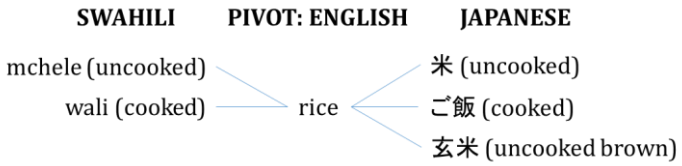


Figure 15: Mapping of word meanings in the Open Multilingual Wordnet.

The degree of information loss is flagrant: while both Swahili and Japanese provide fine-grained lexicalisations about the various forms of rice, the many-to-many mapping that results from passing through English masks all fine-grained differences, resulting in *both* a loss of detail and potentially incorrect mappings. The OMW thus presents a clear bias towards the English language and Anglo-Saxon cultures.

In recent work (Bond et al. 2020), the authors of the OMW promise to address this representational weakness through replacing the Princeton WordNet as pivot by a synthetic meaning space called the Collaborative Interlingual Index (CILI). The idea is that the CILI should be the union of word meanings from all languages, edited collaboratively. They also introduce the explicit representation of lexical gaps as a further means to encode linguistic diversity. Thus, the various meanings of *rice* are mapped as follows:

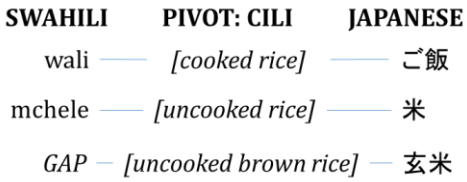


Figure 16: Mapping of word meanings in the Open Multilingual Wordnet 2.0.

Clearly, one-to-one mapping through the CILI instead of English offers a much finer-grained and more precise translation ability, and the indication of gaps allows the distinction between a lexical gap and a missing word (i.e. lexicon incompleteness). Note that the CILI is a flat list of lexical concepts: it does not indicate that the Japanese 玄米 (meaning uncooked brown rice) is a narrower concept than the Swahili *mchele* (meaning uncooked rice), and it only supports equivalence mappings. Thus, while the addition of gaps and the CILI eliminate the explicit English bias of the OMW, the mapping model limits the representation ability of the MLDB: cross-lingual

reasoning over word meanings—such as inferring that *wali* is a narrower term for *rice*—can only be done based on language-specific lexico-semantic relations.

BabelNet is another large-scale MLDB, geared towards applications in natural language processing. It was gathered from multiple sources including Wikipedia, which explains its partly-lexical, partly-encyclopædic coverage. BabelNet has its own meaning space constituted of BabelSynsets. The main difference between OMW and BabelNet is that the latter encodes lexico-semantic relations, such as hypernymy or meronymy, within the pivot, while the former deliberately refrains from doing so, letting monolingual dictionaries (wordnets) bear the brunt of organising the meaning space. The centralised approach gives BabelNet an added coherence, but also an increased responsibility over cross-lingual word meanings. Furthermore, BabelNet lacks an explicit representation of lexical gaps, and thus cannot indicate the lack of Swahili lexicalisation for *uncooked brown rice*. There is no way to know whether the word is missing because BabelNet is incomplete, or because it is truly a gap.

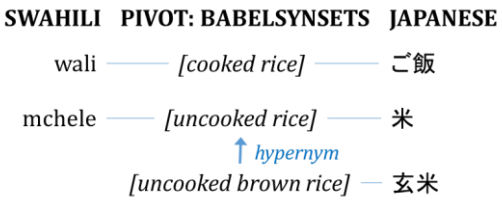


Figure 17: Mapping of word meanings in BabelNet.

5 Building a Diversity-Aware Lexical Database

Our efforts in building the Universal Knowledge Core[97] (UKC) as a diversity-aware MLDB, while designed independently, combines the representational strength of both the OMW and BabelNet (Giunchiglia et al. 2018), (Bella et al. 2022). The *concept core* of the UKC provides a pivot supra-lingual meaning space that can include lexical concepts from all languages. Relations across concepts are encoded within the pivot, as in BabelNet. However, realising that a full unification of the lexico-semantic structure of all languages into a single graph is a never-ending effort, the UKC also allows lexicalisations and their relations to coexist on the individual language level. The figure below shows such a transitional state of cross-lingual integration, where the graph of broader/narrower relationships is shared between the pivot and the Japanese lexicon. Another feature of the UKC lexical model is its explicit support for lexical gaps.

[97] Universal Knowledge Core: http://ukc.datascientia.eu (last accessed: 6 May 2022).

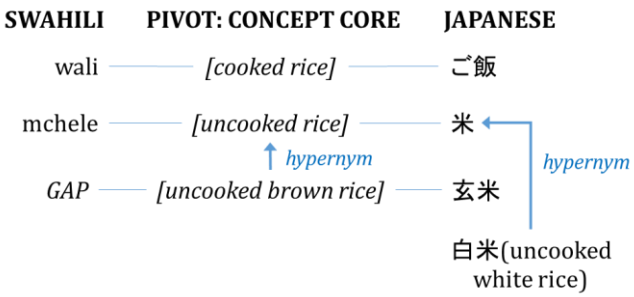


Figure 18: Mapping of word meanings in the UKC.

Among the resources presented in this paper, the UKC offers the most powerful representation of lexical diversity. However, one thing is to provide a model, and another is to fill it with diversity-aware content on language-specific concepts, their relatedness to other concepts, lexicalisations, and gaps. To populate the UKC in a controlled manner and ensure its high quality, we employ methods that always involve humans, preferably native speakers, to some extent. By "involvement" we understand either the production of content or its validation: it is thus possible that we integrate content produced automatically, such as in (Giunchiglia et al. 2017), provided that human validation is applied at some point. This philosophy goes against current trends in computational linguistics that are dominated by fully automatic approaches. While the cost of human work is often evoked as motivation for full automation, in our experience, the short but concerted effort of a few human language contributors can go a long way and may not cost more than paying a computer scientist to work on an algorithm for months or years.

We employ three kinds of agents in the roles of content provider or validator: algorithms, human language experts, and human crowds. Each kind of agent has its strengths and weaknesses: algorithms can process huge amounts of information, but their quality is highly dependent on their input and is bound by the logic implemented by the programmer. Experts provide the highest-quality results, but they are typically expensive and the volume of their output is limited. Crowds usually are inaccurate, but their work can be scaled up, and they may bring in a diversity of social, cultural, and economic viewpoints that would be difficult to obtain otherwise. Due to these differences, we prefer combining different kinds of agents within the same task. Table 3 shows actual projects for extending the UKC involving various types of actors, and the corresponding results in terms of the nature and quantity of linguistic knowledge collected and integrated into the UKC: the *Unified Scottish Gaelic Wordnet* (Bella et al. 2020), the *Mongolian Wordnet* (Ganbold et al. 2018), the *MorphyNet* morphological database (Batsuren et al. 2021b), the *IndoUKC* as a diversity-aware resource for Indian languages (Chandran Nair et al. 2022), the *CogNet* large cognate database (Batsuren et al. 2019, 2021a), the *KinDiv* database on kinship terms and lexical gaps for 699 languages (Khishigsuren et al. 2022a), and the *Universal Metonymy* database on metonymy relationships for 189 languages (Khishigsuren et al. 2022b).

The current version of the UKC includes about 110 thousand supra-lingual concepts, 2 million words, 40 thousand lexical gaps, and 8 million relations in over 2,000 languages (see Figure 19), collected from individual wordnets, Wiktionary, as well as from projects such as those in Table 3.

The lexicons and the diversity data contained in the UKC are freely downloadable from our LiveLanguage Data Catalogue.[98]

Table 3: Example language resource generation projects by the DataScientia Foundation, extending the UKC database.

Data generated	Producer	Validator	Example Project	Results
words and meanings	expert, crowd	expert	Unified Scottish Gaelic Wordnet	~14,000 meanings, ~10,000 words, ~660 lexical gaps
words and meanings	algorithm	crowd	Mongolian Wordnet	~1,000 meanings, ~2,600 words
intra-lingual relations	crowd, algorithm	expert	The MorphyNet morphological database	>10 million derivations and inflections
cross-lingual relations	expert	crowd, algorithm	IndoUKC: a diversity-aware lexical database for Indian languages	~35,000 meanings analysed, ~11,000 mapped
cross-lingual relations	algorithm	expert	The CogNet large cognate database	~8 million cross-lingual cognate pairs
lexical gaps	expert, algorithm	expert	The KinDiv multilingual database of kinship terms and lexical gaps	699 languages, ~37,000 lexical gaps
intra-lingual relations	algorithm	expert	The Universal Metonymy database of metonymy relations	189 languages, xxx metonymy relations

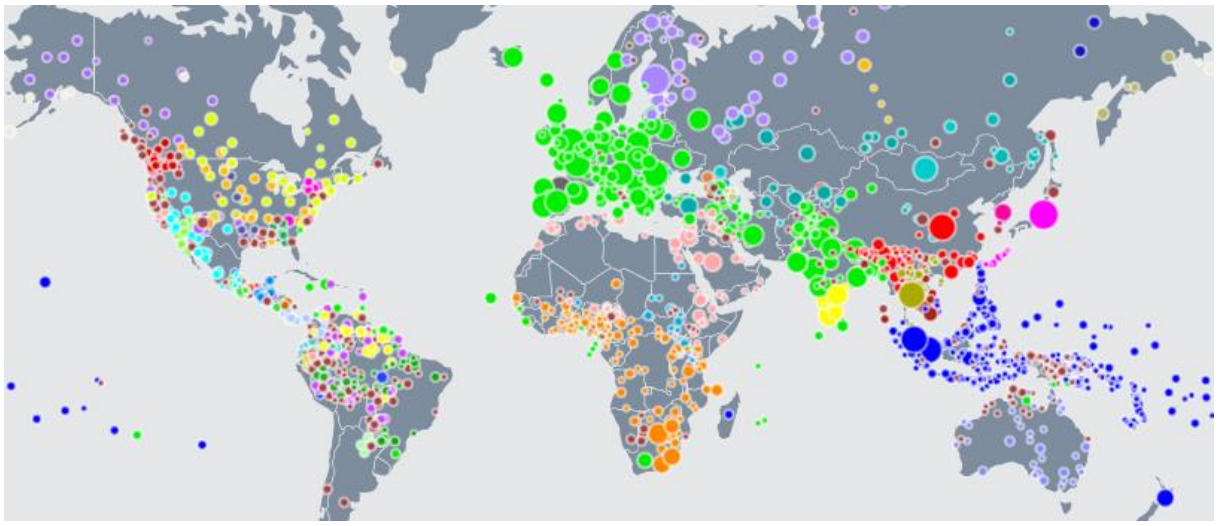


Figure 19: Languages currently supported by the UKC, colour and circle sizes representing language families and lexicon sizes, respectively.

[98] The LiveLanguage Data Catalogue. http://www.livelanguage.eu (last accessed: 6 May 2022).

It is clear that a large-scale resource such as the UKC can only be produced and maintained through an international collaborative effort, with language experts and native speakers contributing and overseeing language-specific work. In order to carry out such collaboration in an institutional framework, we have created the not-for-profit DataScientia Foundation[99], involving university partners (currently from seven countries and four continents). Partners take advantage from shared language development tools and infrastructure, carry out research collaboratively, and exchange best practices through a workshop organised yearly.

6 Case Studies on Lexical Diversity

In this section we present two example projects in which we have developed language resources that explicitly address language- and culture-specific aspects of lexical diversity. The first is the Unified Scottish Gaelic Wordnet[100] (Bella et al. 2020), the largest wordnet resource for Scottish Gaelic so far. The second is the KinDiv project[101] that specifically focuses on kinship terminology, known to be extremely diverse across languages and cultures.

6.1 The Lexicon of Scottish Gaelic

The Unified Scottish Gaelic Wordnet is a human-driven language development project destined to provide a new wordnet resource and simultaneously to extend the UKC with new content. This project employed the so-called expansion approach to extend the coverage of a small existing wordnet (itself converted from the crowdsourced Wiktionary) by about 6,500 new words through lexical translation from English. Proceeding concept by concept, translation and subsequent validation were carried out by two paid language experts, based on the corresponding English lexemes, definition, and example sentences. During translation, about 2,400 neologisms were coined by the translator, a practice that is not uncommon in the context of revival efforts for endangered languages that suffer from a general lack of modern vocabulary.

In an attempt to address diversity and bias, about 600 Gaelic lexical gaps were also identified by the translator, i.e. English words that are not translatable to Gaelic and for which the translator did not feel the need to create neologisms (e.g. the English "*barnburner*" meaning *an impressively successful event*). While methodologically mostly straightforward and efficient, the main disadvantage of the translation-based approach is that it ignores words that are specific to the target language, i.e. that are gaps in the source language, in our case English. In other terms, translation from a dominant language creates bias towards it, as it does not reveal the linguistically or culturally specific vocabulary of the target language. Within the Gaelic project, we partially addressed this issue by a separate expert effort of collecting and defining Gaelic-specific words that are English gaps. As a first effort, about 60 such words were identified and described. These words, such as "*onfhadh*" for *the raging sound of the sea* or "*turadh*" for *when the rain stops*, have become new concepts inside the UKC. We believe that other approaches, such as

[99] The DataScientia Foundation. http://datascientia.eu (last accessed: 6 May 2022).

[100] The Unified Scottish Gaelic Wordnet. http://ukc.disi.unitn.it/index.php/gaelic/ (last accessed: 6 May 2021).

[101] The KinDiv project. http://ukc.disi.unitn.it/index.php/kinship/ (last accessed: 6 May 2022).

crowdsourcing, can be more efficient in identifying lexical gaps and should be investigated as future work.

6.2 Diversity in Kinship Terminologies

Kinship terms designating family relationships—grandparents, aunts and uncles, nephews and nieces, cousins, etc.—are known to show great variety across languages and cultures (Murdock, 1970). For example, the relationship expressed by the single English term *uncle* is designated by three different terms in Northern Sami: *eahki* (father's elder brother), *čeahci* (father's younger brother), and *eanu* (mother's brother). While translating between these two languages, whether by a human or a machine, the differences in meaning need to be fully clear to the translator in order to avoid injecting unintended content.

In order to support this understanding, the Universal Knowledge Core needs to contain the following information:

- the English *uncle* is a lexical gap in Northern Sami;
- the Northern Sami *eakhi*, *čeahci*, and *eanu* are lexical gaps in English;
- *eakhi*, *čeahci*, and *eanu* are all more specific than *uncle*.

The *KinDiv* project (Khishigsuren et al. 2022a) has generated about 250 kinship-related supra-lingual concepts, 1,900 terms, and 37 thousand lexical gaps over 699 languages. The project mostly relied on native speaker input collected by field experts (Murdock, 1970), but also on automated commonsensical inference rules. To our knowledge, KinDiv is so far the largest resource that provides lexical gaps in a formal, computer-interpretable manner.

7 Perspectives

The UKC continues to grow with lexical data produced by ongoing projects: word derivations, importing of words from other MLDBs, lexical gaps. The collection of gaps is a challenging research problem, as there is no systematic way to find them. While lexicon translation, i.e. providing translations for words/meanings from language A into language B, can uncover gaps within B, this approach presupposes the existence of an extensive lexicon available for A, containing language and culture-specific entries that are potential gaps in other languages. As this is not generally the case for under-resourced languages, alternative methods need to be sought, such as corpus-based algorithms or crowdsourcing.

Besides extending the UKC, we are working on a more formal—quantitative—characterisation of language diversity through the notions of lexicon similarity (Bella et al. 2021) and bias. We wish to extend the scope of computing bias from lexico-semantic databases to corpus-based resources used in computational linguistics, such as word embeddings and cross-lingual transfer matrices. Such resources are very widely used in contemporary research on under-resourced languages, and the impact of linguistic bias on results has not yet been studied.

Finally, we are actively seeking partners with profiles in linguistics, computational linguistics, computer science, or from other disciplines, to collaborate on building language resources especially for under-resourced languages. We are interested in approaches under both human-

driven and computational modalities, with the DataScientia Foundation being able to provide the necessary software tools and institutional framework for collaboration.

8 References

Adamska-Sałaciak, Arleta. 2010. Examining Equivalence. *International Journal of Lexicography* 23(4): 387-409.

Batsuren, Khuyagbaatar, Gábor Bella, and Fausto Giunchiglia. 2019. CogNet: A Large-Scale Cognate Database. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3136–3145. Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1302.

Batsuren, Khuyagbaatar, Gábor Bella, and Fausto Giunchiglia. 2021 (a). A Large and Evolving Cognate Database. *Language Resources and Evaluation* 1–25. https://doi.org/10.1007/s10579-021-09544-6.

Batsuren, Khuyagbaatar, Gábor Bella, and Fausto Giunchiglia. 2021 (b). MorphyNet: A Large Multilingual Database of Derivational and Inflectional Morphology. *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* 39–48. https://doi.org/10.18653/v1/2021.sigmorphon-1.5.

Bella, Gábor, Fiona McNeill, Rody Gorman, Caoimhín Ó Donnaíle, Kirsty MacDonald, Yamini Chandrashekar, Abed Alhakim Freihat, and Fausto Giunchiglia. 2020. A Major Wordnet for a Minority Language: Scottish Gaelic. In *12th International Conference on Language Resources and Evaluation*, ed. Nicoletta Calzolari, 2812–2818. Paris: The European Language Resources Association (ELRA).

Bella, Gábor, Khuyagbaatar Batsuren, and Fausto Giunchiglia. 2021. A Database and Visualization of the Similarity of Contemporary Lexicons. In *Text, Speech, and Dialogue: 24th International Conference, TSD 2021,* ed. Kamil Ekštein, František Pártl and Miloslav Konopík, 95–104. Cham: Springer International Publishing; Imprint Springer. https://doi.org/10.1007/978-3-030-83527-9_8.

Bella, Gábor, Erdenebileg Byambadorj, Yamini Chandrashekar, Khuyagbaatar Batsuren, Danish Ashgar Cheema, and Fausto Giunchiglia. 2022. Language Diversity: Visible to Humans, Exploitable by Machines. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, Dublin, Ireland.

Bond, Francis, and Ryan Foster. 2013. Linking and Extending an Open Multilingual Wordnet. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* 1352–1362. Sofia.

Bond, Francis, Luis M. Da Costa, Michael W. Goodman, John P. Mc Crae, and Ahti Lohk. 2020. Some Issues with Building a Multilingual Wordnet. In *12th International Conference on Language Resources and Evaluation*, ed. Nicoletta Calzolari, 3189–3197. Paris: The European Language Resources Association (ELRA).

Chandran Nair, Nandu, Rajendran S. Velayuthan, Yamini Chandrashekar, Gábor Bella, and Fausto Giunchiglia. 2022. IndoUKC: a Concept-Centered Indian Multilingual Lexical Resource. In: *13th International Conference on Language Resources and Evaluation*, Marseille, France.

Ganbold, Amarsanaa, Altangerel Chagnaa, and Gábor Bella. 2018. Using Crowd Agreement for Wordnet Localization LREC (2018).: In: *Proceedings of the 11th International Conference on Language Resources and Evaluation*. Miyazaki, Japan.

Giunchiglia, Fausto, Khuyagbaatar Batsuren, and Gábor Bella. 2017. Understanding and Exploiting Language Diversity. In *International Joint Conferences on Artificial Intelligence,* 4009–4017. Red Hook, NY: Curran Associates Inc. https://doi.org/10.24963/ijcai.2017/560.

Giunchiglia, Fausto, Khuyagbaatar Batsuren, and Abed Alhakim Freihat. 2018. One World—Seven Thousand Languages. In: *Proceedings of the 19th International Conference on Computational Linguistics and Intelligent Text Processing*.

Graham, Mark, and Matthew Zook. 2013. Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web. *Environment and Planning A: Economy and Space* 45 (1): 77–99. https://doi.org/10.1068/a44674.

Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers* 104 (4): 746–764. https://doi.org/10.1080/00045608.2014.910087.

Hacken, Pius ten. 2016. Bilingual Dictionaries and Theories of Word Meaning: In Proceedings of the XVII EURALEX International Congress 61–76.

Hartmann, Reinhard. 1985 [2007]. Contrastive Text Analysis and the Search for Equivalence in the bilingual Dictionary. In *Symposium on Lexicography II.* ed. Karl Hyldgaard-Jensen and Arne Zettersten, 121–132. Tübingen: Max Niemeyer.

Kamholz, David, Jonathan Pool, and Susan Colowick. 2014. PanLex. Building a Resource for Panlingual Lexical Translation: Proceedings of the 9th International Conference on Language Resources and Evaluation. *European Language Resources Association* 3145–3150.

Khishigsuren, Temuulen, Gábor Bella, Khuyagbaatar Batsuren, Abed Alhakim Freihat, Nandu Chandran Nair, Amarsanaa Ganbold, Hadi Khalilia, Yamini Chandrashekar, and Fausto Giunchiglia. 2022 (a). Using Linguistic Typology to Enrich Multilingual Lexicons: the Case of Lexical Gaps in Kinship. In: *13th International Conference on Language Resources and Evaluation*, Marseille, France.

Khishigsuren, Temuulen, Gábor Bella, Khuyagbaatar Batsuren, Daariimaa Marav, and Fausto Giunchiglia. 2022 (b). Metonymy as a Universal Cognitive Phenomenon: Evidence from Multilingual Lexicons. In: *Proceedings of CogSci 2022*, Toronto, Canada.

Kornai, András. 2013. Digital Language Death. *PloS One* 8 (10): e77056. https://doi.org/10.1371/journal.pone.0077056.

Landsberger, Benno, Erica Reiner, and Miguel Civil. 1970. The Series HAR-ra = hubullu: Tablets XVI, XVII, XIX and Related Texts. *Rome: Pontificium Institutum Biblicum*.

Lehrer, Adrienne. 1970. Notes on Lexical Gaps. *Journal of Linguistics* 6 (2): 257–261. https://doi.org/10.1017/S0022226700002656.

Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38 (11): 39–41. https://doi.org/10.1145/219717.219748.

Murdock, George Peter. 1970. Kin Term Patterns and their Distribution. *Ethnology*, 9(2): 165-208.

Navigli, Roberto, and Simone Paolo Ponzetto. 2010. BabelNet. Building a Very Large Multilingual Semantic Network: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. *Association for Computational Linguistics* 216–225.

Ricœur, Paul. 2004. Sur la traduction. *Paris: Bayard.*

Vossen, Piek (ed.). 1998. EuroWordNet: A multilingual database with lexical semantic networks. *Dordrecht: Kluwer.*

Connecting Knowledge

Exploration of interdisciplinary research networks with VIVO

Matthias Fuchs, Stefan Wolff (Saxon State and University Library Dresden)[102]

Extended Abstract

Interdisciplinary cooperation is becoming increasingly important and is sometimes a decisive prerequisite for applying for research funding. With the search for adequate research partners, however, also come numerous challenges. For example, finding the right person also depends on entering suitable search terms. However, the specialist terms for certain research topics differ in some cases in the individual disciplines. The search effort is also increased by the large number of possible platforms on which researchers can present their expertise. At the same time, the active maintenance of these profiles also implies extra work for the researchers. Therefore, new approaches and tools are needed to support the interdisciplinary exchanges of researchers in the best possible way.

Against this background, the Saxon State and University Library Dresden (SLUB) has developed a new and innovative tool within the framework of the project "Specialised Information Service

[102] Contact: Matthias.Fuchs@slub-dresden.de, https://orcid.org/0000-0003-3365-3158.
Stefan Wolff https://orcid.org/0000-0003-0015-9671.

Mobility and Transport Research" (FID move[103]): the Mobility Compass[104]. The aim of the Mobility Compass is to make existing research networks in this subject area visible and accessible, thus enabling users to identify interesting researchers as cooperation partners.

The basic concept of the Mobility Compass is to connect and link various open data sources with the research outputs of scientists under the principles of Linked Open Data. These include, among others, data from the German National Library, Global Research Identifier Database and Springer Nature. The linked data are indexed using a multilingual specialist vocabulary supplemented with synonyms. The search for suitable persons is then carried out on the basis of, for example, the researched topics or the researchers' places of work. A central result of this selection is a list of corresponding researchers, whereby a detailed view can be retrieved for each person (see Figure 20).

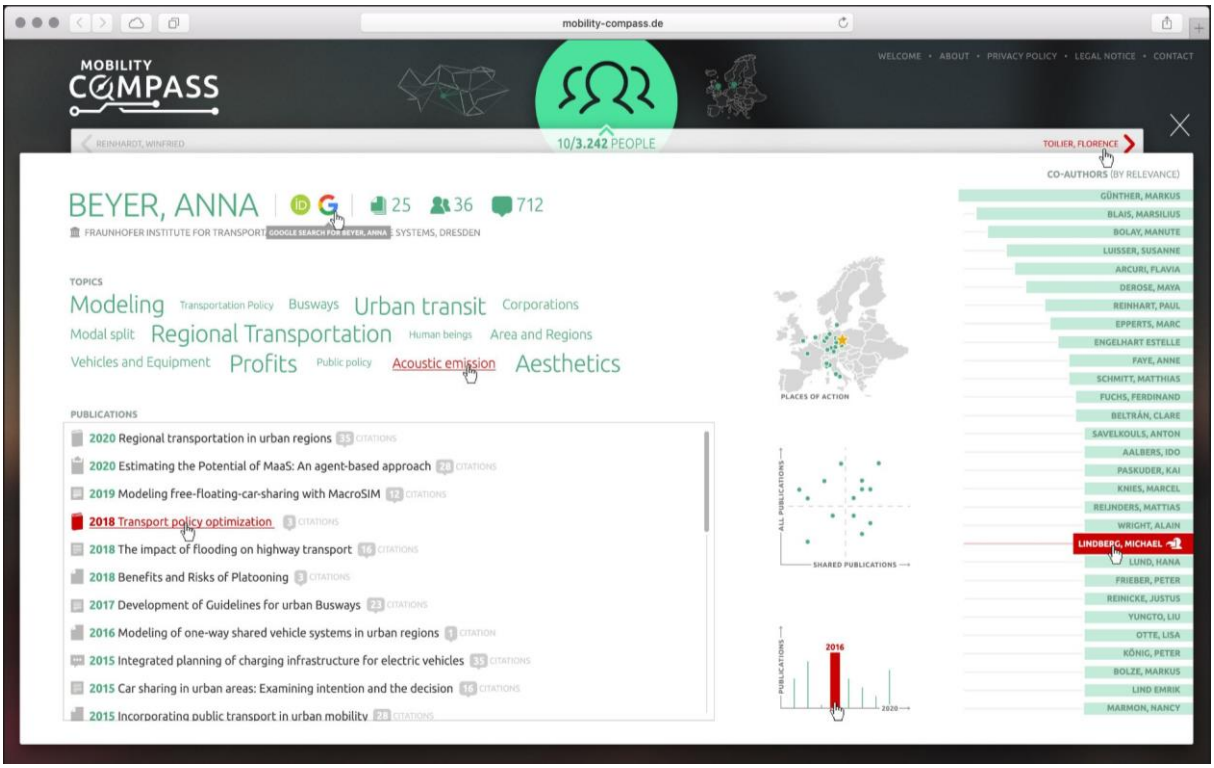


Figure 20: Detailed view for a single researcher. Copyright (CC BY-NC 4.0) by the Saxon State and University Library Dresden.

[103] FID move is a DFG-funded cooperation project between the Saxon State and University Library Dresden (SLUB; https://ror.org/03wf51b65, last accessed 22 December 2021) and the German National Library of Science and Technology – Leibniz Information Centre for Science and Technology and University Library Hannover (TIB; https://ror.org/04aj4c181, last accessed 22 December 2021).

[104] The previous version of the tool, the *Forschungskompass*, was shown at the conference.

The technical base of the Mobility Compass is the already existing and established tool VIVO. This is an open-source current research information system (CRIS). Due to the open conception and development of the tool, the Compass approach can be transferred to other subject areas and its focus can be expanded.

A deeper insight into the technical specifications and functionalities of the tool is provided in the article The Mobility Compass: A VIVO-based approach for exploring interdisciplinary research networks (Wolff et al. 2021). The Mobility Compass can be accessed at www.mobility-compass.eu (last accessed 22 December 2021).

References

Wolff, Stefan, Maria Rutschke and Matthias Fuchs. 2021. The Mobility Compass: A VIVO-based approach for exploring interdisciplinary research networks. *O-Bib. Das Offene Bibliotheksjournal 8*(1): 1–9 https://doi.org/10.5282/o-bib/5642.

Digital Repositories in African Heritage and African Studies

Section 4

Digital Return as a Question of Ethics

Transformation at the International Library of African Music (ILAM) in South Africa

Lee Watkins

1 Introduction

The primary question addressed in this paper is concerned with the following: In view of the International Library of African Music's (ILAM) colonial legacy how could ILAM transform the social value of its holdings in keeping with the general call for transformation in the academic sector as well as in society at large? A secondary question is: how may the ethical be determined and practiced in prevailing circumstances of poverty, a lack of social development, and the protection and sustainability of the intangible heritage of the continent?

Hugh Tracey (1903–1977), the founder of ILAM, holds a legacy that has been regularly critiqued by outsiders to the discipline of African music studies as aiding the goals of British colonialism.[105] His recordings and collection of music instruments are deemed by many as the ill-gained pickings of colonial and racial privilege. Coetzee, for example, regards Tracey as a "part of a mid-twentieth century movement which sought to marshal positive representations of traditional African culture in the interest of maintaining and strengthening colonial rule" (2014: 152). Further, she claims that his work "functioned to promote racist exclusion in the manner of its production, distribution and claims to expertise." In similar vein, Edwards (2016) considers Tracey an agent

[105] During the course of his lifetime, Tracey recorded the traditional and popular musics of sub-Sahara Africa far and wide. His contribution to the conservation of African music places him at the forefront of ideas about conservation in general, and the study of African music in particular. Towards the end of his career in the 1970s, his role in the study of traditional African music was augmented by others such as John Blacking, JH Kwabena Nketia, and Hugh Tracey's son, Andrew Tracey, among others, with each making a substantial contribution to the canon of African music studies in various publications. Their contribution is further evident in a range of collections of recordings currently held at ILAM.

of colonialism and apartheid in later years in as far as his stereotyping and essentialising of African music are concerned. However, despite their academic merit, I argue that the critiques offered by Coetzee, Edwards, and others such as Richards (1993), whose pioneering study on empires and archives is of critical importance, may be crucial in claiming moral and historical ground, but they do not necessarily offer a way forward for archival practices. Instead, Stoler's (2010) view on the colonial archive is far more helpful. To her credit she refrains from prescribing a specific future for the former colonial repository. Seeing the archive as "the supreme technology of the late nineteenth-century imperial state, as a repository of codified beliefs that clustered and (bore witness to) connections between secrecy, law, and power" (2010: 97), she argues that "scholars need to move from the archive as source to the archive as subject, and to view archives as sites of knowledge production rather than knowledge retrieval." In her discussion on the Gallo archive in Johannesburg, Allen similarly observes that the archive represents a configuration of power as it offers a selection which serves particular, institutional interests rather than seeking the means to becoming more inclusive (2007: 266). Echevvaria (1990) further situates the archive as both relic and ruin, a repository of codified beliefs, genres for bearing witness, clustered connections between secrecy, power, and the law.

In view of the above, much could be argued around the stigma attached to the origin of a music archive such as ILAM which is regarded as a testament to colonial, hegemonic power. However, I argue that ILAM's future lies in reading the archive against the grain because, if anything, its current realities project the archive as a fluid repository of artefacts that need to be interpreted in view of its possible biases (Allen 2006: 266). These are reasons why ILAM decided to follow an ethical approach which would challenge its colonial past while at the same time adapting to a transforming world. Following consultations with members of the community, students and academic peers, ILAM adopted the following objectives in its everyday and research activities:

- Ethical research practices.
- Negotiated rights to access, and fair usage.
- The protection of intellectual property rights.
- Commercial exploitation conducted in collaborative and participatory manner for developmental purposes.
- Outreach as in community-based archives and education.
- Repatriation – the digital return of collections to communities of origin.
- Archives undertake research/restudies that can repatriate/return existing collections in ways that promote education, conservation and cultural revitalisation.
- Maintain an ethic of reciprocity as far as possible.
- Responsible archiving of field recordings.
- A use agreement negotiated with the performers that covers possibilities for reproduction, sale and/or repatriation (digital return).

Among others, the ethical lies in removing the secrecy behind which the colonial archive thrived. Transformation is evident in the participatory manner by which new knowledge emerges from a colonial repository. Digitised content goes a very long way towards realising these objectives above as it provides for greater mobility and easier accessibility. Specifically, ILAM's past is being redeemed by several projects undertaken in collaboration with stakeholders who share our vision for sustainable practices emanating from the archive. These projects deal with digital return as an exercise where interest in the recordings is revived with a view to providing young African artists with the means to economic independence. ILAM is involved with the establishment of satellite archives where local stakeholders become involved in developing an interest in the musical heritage of local communities, thereby creating an income for these stakeholders. The archive is a source of materials related to music education and it has become a key partner in efforts at combining heritage with social and economic development. The demands for a decolonised music archive such as ILAM have been vociferous over the past decade.

ILAM's current mission must be seen within the context of its founder, and how his legacy compels one to interrogate the role and function of the archive in the twenty-first century. How, for instance, does one transform this legacy into a mission that produces knowledge in keeping with the demands of a transforming society, and in particular, the challenges endured by many rural communities? ILAM is based in the rural parts of the Eastern Cape Province. Indeed, with these associations, reimagining the archive has become a matter of urgency. This vision is aimed at investing the hollow sounding rhetoric of "transformation" and "decolonialism" with efforts that are ethically inspired. Moving away from a simplistic binary in which Hugh Tracey's work has been either vilified or revered, events at ILAM bring together several components aimed at transforming ILAM such that it becomes more appropriate to the demands for ethical research and appropriate archival practices.

I argue that ILAM is not merely a colonial repository of sounds and objects. It is a community organisation critically engaged with both grassroots and globally, transnational institutions. Along with colleagues, students and members of the community in Grahamstown (Makhanda), the archive is being transformed into an agent for developing new networks of relationships across disciplines. While many ethnomusicologists deem themselves as curators of information, the future of archives resides not only in innovation but also in new methodologies which are necessarily participatory and ethical in a decolonial understanding of the term. Members of the community, particularly those who had previously been denied access to the resources of an archive, have to be included in discussions on the future of the archive, and must be engaged in the everyday prospects of maintaining such an archive.

2 Digitisation and Collaborations for the Benefit of Sustainable Development

ILAM has been actively seeking partnerships with other private and semi-institutional collectors and archivists. Since 2002 ILAM has been digitising collections of reel tapes, cassettes, shellac records, documents and photographs.

To date, ILAM has digitised the collections of Hugh Tracey, Andrew Tracey and the African National Congress, among others, and is in the process of digitising and cataloguing at least 13 other collections. The digitisation of Hugh Tracey's collections followed as a result of renewed archival practices around the globe, an activity aimed at adapting to new technologies for preservation and dissemination. The issue of digital return also followed the practices at numerous archives and museums around the globe, to return recordings to communities of origin. Thus digitising at ILAM has transpired in keeping with the global emphasis on the digitisation of materials, and as a matter of conserving music recorded on very fragile technology. Over the course of this process staff at ILAM, especially its sound engineer, Elijah Madiba, have acquired a wealth of experience. This wealth of experience, in addition to the digitised collections, have enabled ILAM to provide open accessibility, to encourage digital return, and in turn, to develop new knowledge through the production and dissemination of digitised holdings. For instance, digital return has allowed for the correction of much of the information captured by Hugh Tracey on his field cards. Digitisation has facilitated transformation at the archive as well as the pursuit of relations which are ethically sound in as far as participatory archiving is concerned.

Currently, ILAM has several agreements in place which provide an ongoing source of revenue. While the archive is recognised by the South African government as a valuable asset, this recognition does not yield much support in real and financial terms. As a result, ILAM is compelled to seek and generate funds from other sources. These funds are used to initiate and maintain various projects such as conducting research on the jazz heritage of the Eastern Cape and to provide funding for a local children's arts programme called Sakhuluntu. Sakhuluntu is a children's arts project based in Joza, a local township in Makhanda. Among ILAM collaborators is an organisation called Singing Wells, whose owner, James Allen, records traditional music in central Africa. With the help of ILAM's recordings he visited numerous countries in central and east Africa to which he has returned recordings in the hope of reviving and maintaining interest in traditional local musics and instrument making. He is providing local participants with funds to build instruments and he brings together traditional and popular musicians. These new renditions are returned to ILAM and catalogued.

A second collaborator is Beating Heart, an emerging music distribution company in the UK which has been revitalising interest in the recordings of various African countries through remixes of music which are sold on the market and currently being marketed in a number of forums including the BBC. Income generated from this source is used to establish and support various organisations including community-based music organisations in Makhanda. Due to the intervention of ILAM, and in the spirit of the decolonial such that the powers are as equitable as possible, Beating Heart has shifted from a "charity model" to one that is more appropriate to our circumstances, namely, that of "development." ILAM's tentacles spread from the cosmopolitan world right down to its immediate circumstances. Up to now, ILAM staff and students have also been assisting Sakhuluntu. Another project gaining from ILAM's digitisation efforts is Access Music Project! (AMP!), a township-based music education project. ILAM staff assist AMP! in numerous ways from fund-raising to serving on its board. Through the relationship with Beating Heart, ILAM secured much needed equipment for the production and administrative requirements of AMP!.

ILAM is situated in a deeply rural part of the province and operates within numerous constraints. The lack of capital, human and intellectual resources in rural Eastern Cape means that community-based organisations rely extensively on the few individuals at universities who have an interest in their programmes. This is the reason why my colleague and I find ourselves constantly in the service of organisations such as those mentioned above. While we may have begun with a solely academic understanding of the archive, these engagements with the community have compelled us to interrogate the functions of the archive in both practical and theoretical terms. We ask ourselves about the nature of the institution, about how best to make the resources in the archive available to those who are in need, and how, in turn, recipients could make a contribution to the everyday functioning of the archive? We engage in this kind of questioning to ensure that the top-down hierarchy associated with colonialism is not perpetuated.

3 Digital Return Projects

The digital return of recordings was initiated by the former Director of ILAM, Diane Thram. Since 2014 she has been returning music to Zanzibar, Tanzania, Malawi and Durban. My colleague, Elijah Madiba, returned recordings to artists in the local, amaXhosa community in Makhanda. His concern was not merely with digital return but to evaluate how these recordings could be used to generate incomes for local slam poets and hip hoppers. Previously, mostly foreign interests had taken the music, repackaged it, sold it and made a profit with sales. Our concern was to allow local musicians similar opportunities for taking the music to generate and sustain incomes for themselves. As part of his doctoral research at Oxford University, Noel Lobley conducted an ethnographic analysis of The Sound of Africa field recordings made and published by Hugh Tracey between 1933 and 1973 in Makhanda. During 2007-2007 he sought to develop a new way to circulate recordings among a source community that has never before been reached through institutional archival practice. A former Masters degree student from Botswana, Gomolemo Mojaki, repatriated recordings from ILAM's Hugh Tracey collection to the Bangwaketse community in 2014 and 2015. She learnt much about the state of indigenous musics at schools and in the community. Robbie Campbell, a doctoral student at SOAS at the University of London, returned audio Mp3s and three of the films of Chopi music on flash drives to local musicians in Mozambique. His research demonstrates critical parallels between Specific Learning Difficulties such as dyslexia, language and music making. A former doctoral student, Luis Amoros Gimenez, conducted digital return in Zimbabwe. His experiences there are recorded in his book entitled '*Tracing the Mbira Sound Archive in Zimbabwe'*, which was published by Routledge in 2018. Jocelyn Moon, a doctoral student at the University of Washington in Seattle, returned recordings of matepe music to Zimbabwe. She has created social networking sites which offer recordings of matepe music and contribute to the awareness of existing archival resources from ILAM. Relying on digital platforms, her efforts at repatriation and revitalisation are particularly innovative and have far-reaching influences.

Through digitisation and repatriation, these projects are providing new knowledge about archival practices in as far as digital return allows for opportunities that are participatory, economically empowering for young musicians, and a greater access to resources.

The next challenge is to determine the possibilities of the digital return of objects such as music instruments. While the return of music instruments has invited debates as to direct ownership

and the welfare of the instruments which had for a long time been preserved in the archive, there is now the possibility of recreating the instruments digitally. In this manner, a three dimensional, digitised version of the object will open up the way to enhancing the accessibility initiated by the recording. This is a possibility in the near future.

4 Conclusion: Whither the "Ethical"

Above I have provided examples of how the ethical may be realised in the collaborative practices of a music archive. The term "ethical" is not mere rhetoric but hopefully animated by a response which questions the dominance of a Euro-American time and space and the exclusivity of a tertiary environment.

For a continent on which many of its music cultures are imperilled, ILAM has become a beacon of hope for scholars and musicians. It attracts students and scholars from just about everywhere, but more crucially, it has an influence on developing closer relationships between both overseas institutions (such as Bayreuth University in Germany) and local grassroots communities and organisations in this part of the Eastern Cape. These are important connections and relationships that could evolve as a result of performance studies filtering into other pressing concerns such as social cohesion and justice in South Africa.

ILAM, of course, is not the only organisation that is trying to explore new paths in decolonial and applied work. There exists a network of contacts in other parts of South Africa and the continent who make themselves available for setting similar agenda for transformation. Through collaborative interventions, the relationship between digital return and revitalisation is remade. Through supporting artists in residence, music education, cataloguing and ongoing research on African music, ILAM has made critical steps towards transforming the value of its holdings. In so doing, it challenges the dominance of western art music and its associated value system. As I have shown, a key underlying theme is the reliance on partnerships, and collaboration with sister and community organisations. This development and sufficient funding are the only means through which ILAM will confidently proceed into the future.

Current practices at ILAM raise into prominence several questions related to decolonial ethics. While processes and relationships at ILAM are explored and developed within the context of how the "decolonial" may be imagined in the South, from the outside it may be seen that the decolonial hinges on what seems ethically "good" in the here and now, as opposed to the abstract or on philosophical grounds. The concept of decolonialism may provide a pretext for exploring the sound object in more critical ways. How does the "decolonial" speak to issues of race, class, access, aesthetics, and gender? But existing discourses on decolonialism also tends to valorise the institution as if it is the only agent with a right to a voice in the conversation. Further, the decolonial overlaps with the colonial in as far as many current practices such as collecting, conservation, and so on, were indeed very much part of the colonial enterprise. Is it then possible to perceive of the decolonial as a byword for maintaining the status quo at institutions which are inherently conservative and resistant to change? What kind of future lies there for the ethical in the immanence of the disciplines we practice and in transforming the archive?

5 References

Allen, Lara. 2007. Preserving a Nation's Heritage: The Gallo music archive and South African popular music. Fontes Artis Musicae 54 (3): 263–279.

Berger, Peter L. 2006. Some reflections on culture and development. In Developing cultures: Case studies, ed. Lawrence E. Harrison and Peter L. Berger, xvii-1. New York: Routledge.

Coetzee, Paulette. 2014. *Performing whiteness, representing otherness: Hugh Tracey and African music*. Rhodes University, Grahamstown, South Africa. Ph D thesis.

Echevvario, Roberto. 1990. *Myth and archive: A theory of Latin-American narrative*. Cambridge: Cambridge University Press.

Edwards, Brent. 2016. The sound of anticolonialism. In *Audible empire: Music, global politics, critique*, ed. Ronald Michael Radano and Tejumola Olaniyan, 269–291. Durham, London: Duke University Press.

Richards, Thomas. 1993. *The imperial archive: Knowledge and the fantasy of empire*. London: Verso.

Stoler, Ann Laura. 2002. Colonial archives and the arts of governance. Archival Science (2): 87–109.

Is the Endangered Archives Programme easily accessible to researchers in Africa?

Actions toward a level playing field for digitised cultural heritage

Jody Butterworth (British Library, London)[106]

Abstract

Since 2005, the Endangered Archives Programme (EAP) has been offering grants to digitise documentary heritage. In attempting to understand whether EAP is user-friendly to researchers within Africa, there are two main points to consider - the application process and the accessibility of the digitised content on the website. Over the past fifteen years, there has been little opportunity to analyse the success rate of applications to the Programme. We have begun to look into this and the results pertaining to applications for projects within Africa will be scrutinised by looking at relevant data visualisations. After considering the discrepancies between successful and unsuccessful applications, there will be a discussion of possible solutions that may rectify these imbalances. EAP has also done little to share its findings on the locations of people accessing the website and its content; this will be addressed in the remaining part of the paper by looking at Google Analytics and will hopefully reassure readers that the digitised content is not exclusively destined for researchers based in academic institutions within Europe and North America. As

[106] Contact: Jody.Butterworth@bl.uk https://orcid.org/0000-0001-7251-7269.

such, the paper stems from an internal practical approach rather than a theoretical and methodological study. Although there are several organisations that fund the preservation of cultural material, similar programmes – unless funded by Arcadia – do not, necessarily, make it a priority to make digitised content available online and it is therefore difficult to compare EAP with similar programmes that both support the digitisation of documentary heritage while also making the material available for research online. This paper will hopefully stimulate such a discussion.

1 The Principles of the Endangered Archives Programme

The Endangered Archives Programme (EAP) seeks to preserve documentary cultural heritage through digitisation. The Programme offers grants on an annual basis, with the aim of enhancing local capacity in managing and preserving archival collections. Therefore, all applications must involve at least one partnering institution from the country where the material is located. The archival partner agrees to look after the master copy of the digitised material and to make it available locally for researchers. A further copy is deposited with the British Library; this copy is more widely accessible as the images are made freely available on the EAP website.

EAP is administered by the British Library and supported by Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin. Since its first round of grants in 2005, EAP has supported more than 400 projects in over 90 countries and the Programme currently has over 9 million images and 25 thousand sound recordings online for research, inspiration and enjoyment[i]. Although EAP supports projects wherever resources are limited, this paper will be focussing on those based in Africa.

2 Introduction

The first time I was asked to speak on EAP projects based in Africa was at the SCOLMA conference (the UK Libraries and Archives Group on Africa) in 2017. The theme of the conference was ”Document to Digital: How does digitisation aid African Research?“ This was a celebration of the 60th anniversary of the organisation and the 10th anniversary of their publication “African Studies in the Digital Age: DisConnects?”. What struck me was that there had only been three fleeting references to EAP in the publication. Knowing how many projects there had been within the continent, this was a surprise and shock, until it dawned on me that material from our projects was only published on our website from 2010 and although EAP had been funding digitisation for much longer, it had not been made available. Therefore, rather naturally, no one was writing about it.

I would like to believe we have come a long way since then. I was very pleasantly surprised when a British Library colleague recently mentioned that she had been sitting on a red London double-decker bus when she had overheard a conversation behind her - one woman telling another that the material digitised through EAP in Sierra Leone had made her research possible. I just wished the colleague had interrupted their conversation and had asked a few more questions because one of the drawbacks of making the material freely available online is that we do not necessarily know anything about the individuals who are accessing the material or why.

3 Data Visualisations for Applications to EAP

The SCOLMA talk had raised a few questions. Although EAP was beginning to receive more applications from within African institutions than western universities, it did not mean the same proportion of applications from Africa were successfully funded. It also appeared as if the Programme was popular in certain areas but not others. These questions enabled EAP along with colleagues within the Digital Research team to bid for a 3-month PhD placement to carry out data visualisations on our applications and projects relating to Africa. We were extremely lucky to have had Sarah FitzGerald, a PhD candidate at Sussex University[107], on our team. She wrote three blogs highlighting her work and internally for EAP, she produced a more detailed report analysing our applications and projects within Africa. Additionally, because she is a linguist, she included a very helpful corpus analysis to account for the discrepancy between the number of applications coming from within Africa as compared to the home countries of all other successful applications that had been awarded grants.

FitzGerald analysed the first 14 years of the Programme, which covers the first phase of EAP funding from Arcadia. With the exception of the first three years, most applications to preserve African archives have come from people living in Africa. Applications from North America and Europe on average seemed to be fairly equal. Applications from elsewhere were almost non-existent. There were three applications from Australia, and one western academic based at an institution in Asia. That being said, it is disappointing to see in the graph (Figure 21) that the increase of applications from within Africa was not more consistent and continuous.

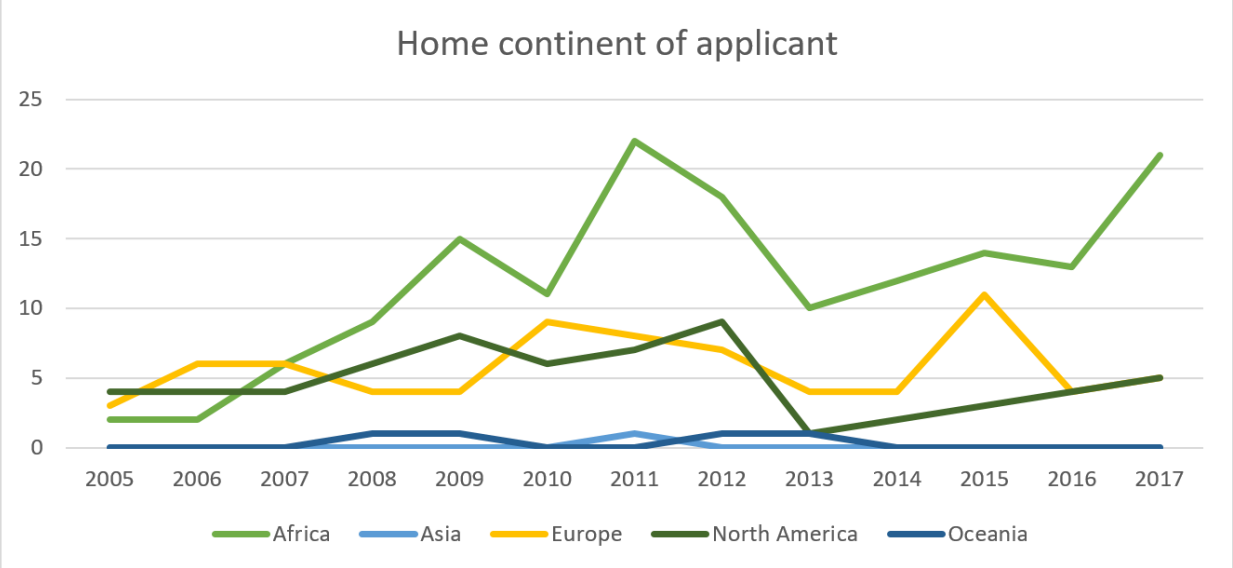


Figure 21: The locations of projects within Africa.

At the start of lockdown, just a few months after the conference in Bayreuth, I began a deeper analysis by using the tools OpenRefine and PowerBI to understand the entire EAP application

[107] https://www.bl.uk/case-studies/sarah-fitzgerald (last accessed 05 March 2021).

process during its first phase of funding (2005-2018). This enabled me to take a closer look at the number of successful applications in relation to the number of submissions for the entire region. For the first time, it also allowed us to compare the types of institutions and the likelihood of funding from within these institutions, the decision process during the annual grant cycle, whether projects were able to complete during their grant period, the various budgets per country, as well as carrying out more detailed analysis by looking at material type (manuscripts, books, newspapers, photographs, sound etc. – though this will not be discussed within this paper).

Looking at the dashboard below (Figure 22), it was rather sobering to see that fewer than half of the funded projects came from a main applicant within Africa and made up only 16% of the total applications received for the region - 40 grants. Although FitzGerald had correctly assessed that more applications were being received from within Africa than elsewhere, the graph in the dashboard below (Figure 22) depicting the successful applications showed that there was no improvement in the success rate. In the final round of phase one in 2018, the Programme funded six applications from within Africa and six from outside.

Consolidating all this information also made it apparent where the Programme needs to be more proactive in providing training for writing grant applications. There has not been a single successful submission from a private archive or library; however, we have had, for example, a project to digitise three private libraries in Timbuktu, where the project leader was an independent researcher from outside of Mali. Of the ten funded projects where the applicant was based at a national archive, seven of these were from the National Archives of Malawi and only three came from elsewhere - despite having received 25 in total.

When examining the annual grant cycle, it became clear that most of the failed applications were turned down at the preliminary stage. Several applicants also chose not to continue and withdrew before the detailed application deadline - 42 between 2005 and 2018. Nevertheless, it was very encouraging to learn, that when the Panel agreed on 65 conditional offers, 54 were funded. When the Panel were unable to make a judgement and deferred their decision - meaning the applicant had 12 months to reapply - 27 of 41 were funded. A deferred decision is no longer an option during EAP Phase 2.

Another optimistic insight was the successful conclusion of the projects. Only one grant was terminated and fewer than 3% of projects were still to submit their digitised outputs at the time of creating the data visualisations. Although completed, only two projects failed to submit a final report. It was therefore difficult to analyse whether they had submitted the digital content on time. The grant agreement stipulates that the outputs must be submitted within three months of the official end date to the grant. Fortunately, 40% of the grantees achieved this. A further 13% submitted their outputs within a three to six-month period, and the remaining 40% submitted the outputs after six months. The remaining 7%, a reassuringly small figure, consisted of the terminated project, those that had not submitted a final report or were long-overdue.

Figure 22: PowerBI Dashboard representing budget per region for successful applications.

4 Attempts to tackle training issues

Helping applicants who have not had previous grant-writing experience is something EAP is keen to address. One way of solving this issue is by putting applicants in touch with successful grant holders. This has been helpful, as often there may be language barriers, regional problems or concerns that the EAP team based in London is unable to fully address.

Another solution to help applicants was the publication of *Remote Capture: Digitising Documentary Heritage in Challenging Locations* (Butterworth et al. 2018), a *how-to* guide for people considering a scholarly digitisation project. It is probably too soon to say whether the book has helped improve the quality of applications. It is freely available as a PDF through Open Book Publishers, and one can see in which countries the book has been downloaded. Interestingly, it has been downloaded 86 times in Nigeria, 36 times in Ethiopia, 122 times in Botswana, 85 time in South Africa, 45 times in Zimbabwe, and 34 times in Kenya[108]. It may be too simplistic to relate the number of downloads to the number of applications, but it is positive to note that we had previously only funded one project in Zimbabwe, and in the current round, we had three applications. Previously, only two projects from Botswana were funded, and this round alone we had two applications. Similarly, we received two applications from both Kenya and Ethiopia. From this round of applications, 25 of the 92 were project proposals within Africa. Of the 17 that have been invited to submit a detailed application, 12 are from applicants based within Africa – it will

[108] https://reports.openbookpublishers.com/public/report/10.11647/obp.0138 (last accessed 23 December 2021).

be interesting to see how many of these actually get funded[109]. The Panel noted that the quality of applications this year had improved. This is, hopefully, due to our continuing work to refine the EAP application process, improved clarity in the application forms, and being able to consult *Remote Capture*. EAP has also held webinars to help applicants apply, including webinars in languages other than English. We have historically received fewer applications from Francophone Africa, so last summer, we held a webinar in French. We tried to promote the webinar as widely as possible. Despite our efforts, only five people attended, and only one of whom was based in Francophone Africa - Côte d'Ivoire. Conversely, the webinar in Spanish had nearly 50 attendees, and the Arabic webinar had over 20 participants.

5 The EAP website and location analysis of researchers accessing content

Since the SCOLMA conference in 2017, EAP upgraded its website as part of the renewal agreement with Arcadia. This enabled mobile devices to now access EAP content, which had not been possible previously. The images are now all IIIF compliant, allowing for the sharing and embedding of images through the manifests, again this was impossible with the earlier website. The British Library, like several libraries, uses the free tool Google Analytics to understand and improve user experiences. EAP uses GA to try to get a better understanding of the location of researchers and a sense of the most popular items online. '[T]racking crude effectiveness measures, usually in terms of hits or visits'[110] is currently the main objective for EAP at present. Although the 'crude' statistics are only able to focus on the number of clicks on a particular item, or the location of a researcher, it has truly transformed how we understand the impact of making digitised material freely available online but it does not provide any insight about the individual users to the site.

The analytics discussed below only refer to our images online. I do want to mention, in passing, that our sound projects are not accessed through the EAP website. We are aware that the Guinean project to digitise the Syliphone Label has always been EAP's most popular. When the nearly 8,000 sound recordings were made available through BL Sounds, there were 30,000 hits during the first month, making the landing page, one of the most successful British Library web pages at the time.

Looking at the last quarter of 2019 and the EAP website, there are some fascinating statistics. Examining the cities and the number of page views for October 2019, the highest African city was Nador in Morocco with 3,103 page views (ranking it at number 8). Addis Ababa was 12th with 1,651 page views, Lagos was 14th with 1,369 and Sohag in Egypt was 16th with 1,097. Surprisingly, the highest scoring American city was Boston coming in at 18th with 964 page views.

[109] The Panel was postponed due to the Covid-19 pandemic and took place in April 2021. Only four projects were funded within Africa: Ghana (EAP1425), Kenya (EAP1357), Nigeria (EAP1418) and Sierra Leone (EAP1366)

[110] Steven J. Turner M.L.I.S. (2010)

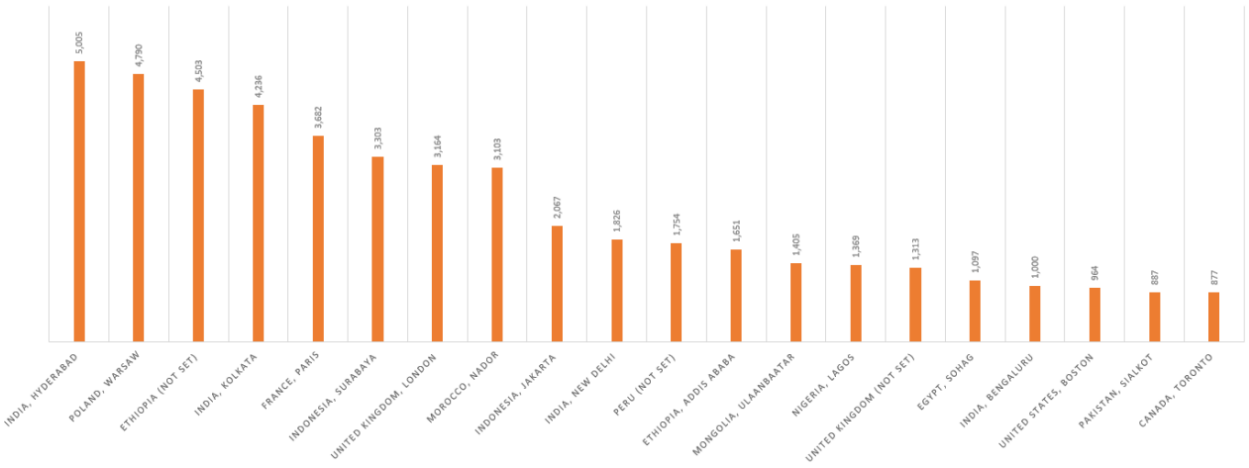


Figure 23: Web statistics for October 2019.

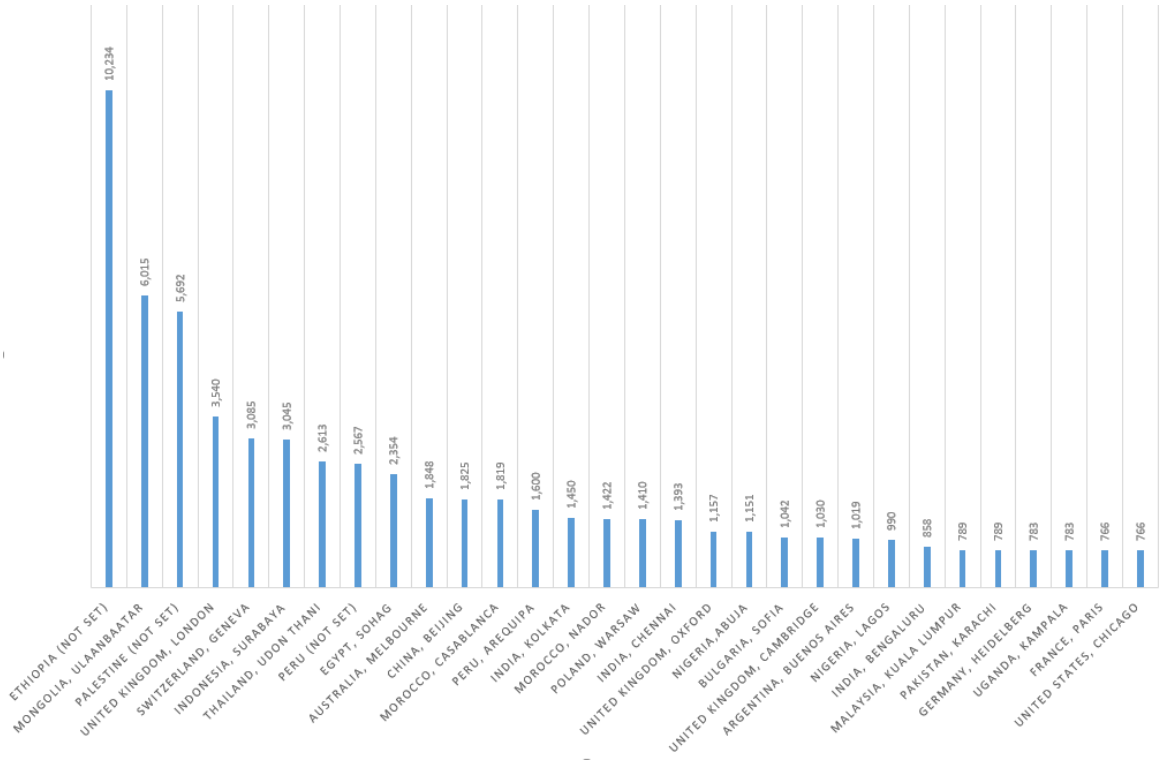


Figure 24: Web statistics for November 2019.

In the following month, Ethiopia as a whole came first with just under 10% of the total page views (10,234 out of 107,476), London came fourth with 3,540 page views. Morocco had two cities within the top 15 with high page views, Abuja was 19th with 1,151, just slightly in front was Oxford at 18th and Cambridge came in at 21st. Chicago was the first American city to be sited, and

came in at number 30 with 766 page views, Lagos, and Kampala both ranked higher. The assumption has been that the digitised material is only viewed by academics based in the Global North. Looking at all of the locations where people are accessing the website – this is clearly not the case. Of course, I am aware of the benefits of VPN and it would be a mistake to take these figures as sacrosanct, but I cannot think of any reason why users of the EAP website would deliberately use such software to mask their location when accessing these archives. Even if some of the IP addresses used in Google analytics are somewhat misleading, it cannot detract from the clear overall picture that researchers are accessing EAP content from all parts of the world.

It is also interesting to explore the actual pages on the EAP website that have the highest number of views. During 2019, one of the most popular items on the EAP website was a printed history of the Hausa people from the National Archives of Nigeria (EAP535/1/2/25/2). Perhaps this is not surprising as Nigeria has a high number of people accessing the website. What is somewhat more intriguing is that the most popular item in November with 82 sessions was a letter written in Bamum script regarding Forgiveness of Sins during the month of Ramadan (EAP051/1/1/12/373). Cameroon does not make it into the 100 places accessing EAP and so this is clearly a researcher or researchers from outside of the country. 82 sessions seems rather high for a single researcher of a two-sided letter and so perhaps it has been used within an academic teaching session outside of Cameroon – but of course, it is hard to be sure.

December 2019 was a quieter month. Though again, African collections were within the top 20 pages accessed by researchers (which also included general hits to the home pages, information for applicants, grants, resources, training and guidelines). EAP387 a project focussing on Fulfulde Ajami manuscripts was at number 8 with 182 page views. Each researcher spending an average of more than two minutes on the project home page). Second was EAP286, one of several EAP projects that have digitised Ethiopian manuscripts. During the month, this project had 114 viewers, over 85% of people landing on this page then delved further within the collections.

6 Encouraging a positive perception of EAP's objectives

I asked Paul Lihoma, Director of the National Archives of Malawi and previous member of EAP's International Advisory Board his views on the fact that a prerequisite to being awarded an EAP grant is that all the digitised material is made freely available online. His reply was:

> In promoting EAP among the ESARBICA Board members in Namibia in 2014, I was surprised that a majority of the members expressed disapproval of submitting digital copies to the BL for free publication on the BL's website. The Board's majority view was that assisting countries to undertake digitisation and requiring them to deposit copies of any digitised material was similar to externalisation of Africa's unique archival heritage to Britain. When Professor Mnjama[111] made a presentation on EAP to the ESARBICA general conference, similar sentiments were expressed by some

[111] Professor Mnjama is a current member on EAP's International Advisory Panel More information can be found at https://eap.bl.uk/whos-who (last accessed 23 December 2021).

> commentators. This is why, perhaps, not many applications are submitted by heads of national archival institutions in the Eastern and Southern Africa region.
>
> Personally, I hold a different view on this matter. Any archival institution exists to preserve and provide access to historical records and the joy of any archivist is to see that acquired material is catalogued and preserved professionally so that the material is accessed at all time. Resource-constrained archival institutions such as those in this part of the world, where precious archival material is being lost due to, among other reasons, precarious conditions and as a result, this material remains inaccessible and obscure. Which is better, seeing precious archival heritage disintegrating thereby denying both present and future generations access to the material or rescuing the material through digitisation to ensure that it is preserved permanently and digital copies of the material shared for wider publicity and easy access?

What is clear from the perception of many, and the International Council of Archives Board for Eastern and Southern Africa in point, is that having the material made available on the EAP's website suggests that it is meant for a British audience, and perhaps to put it more forcefully, is a source of neo-colonial appropriation – our statistics clearly show otherwise. This also helps explain why Dr Lihoma's positive attitude to EAP and the sharing of archival material has resulted in seven projects based at the National Archives of Malawi, whereas the Programme has only funded three applications from other national archives. There has to be internal support and trust at national level for collaboration with funding partners such as EAP – It is clear there is still a degree of wariness and this can only begin to be addressed by publicising our web statistics. In addition, EAP follows the British Library's digital preservation guidelines by making sure that copies of the content are stored on separate servers and accompanied by checksums so that the integrity of the digital files can be checked. The local archival partner looks after the master digital copy but may not have the resources to store the content at various locations, it would be interesting to know how much of the content remains fully intact. Naturally, one of EAP's roles would be to supply to archival partners any digital files that have become corrupt.

Recently, EAP has done more to promote the Programme through displays and public talks. In September 2018, the Library held a free photographic display marking the 8,400 manuscripts from Djenné and surrounding local areas within Mali that went online for the first time. This was the first display at the British Library to have bilingual labelling because the display was being shown concurrently at the National Archives in Bamako. It was part of the official celebrations to mark the digitised manuscripts being received by the National Archives, which acted as the local archival partner for the projects in Djenné.

When I was asked to speak at the Bayreuth conference, I used it as an excuse to ask some of our most recent grant holders questions so that they could reflect on their projects. Their answers certainly helped me to understand some of the practicalities of having an EAP project, but it has also clarified what it means to receive the grant. Of course, it is the grant holders and EAP teams within the countries that are the soul of the Programme and I would like to end my paper by sharing some of their comments.

The first question I asked was how being awarded a grant had helped with their own research. Fallou Ngom, Professor of Anthropology and Director of the African Studies Center at Boston University has had two EAP projects (EAP334 and EAP1042) that focussed on Wolof and Mandinka Ajami manuscripts, and he commented:

> They have allowed me to work with local scholars who have been made invisible by the Eurocentric definition of literacy as the ability to read and write in European languages.

Anne Bang, Professor of History at the University of Bergen has also been the recipient of two EAP grants (EAP466 and EAP1114) and she pointed out that:

> Working with the staff who do the actual photography is also research in its own right. These are trained teachers in Islamic studies/madrasa teachers who have their own clear perceptions of what the material is. Its value in the past and today, its usage in the past and today. The cultural situating of the material comes out very clearly in the photo room, as the material is being discussed, page by page, item by item. Sometimes it is also recited or sung, which in itself is a method for identification and a way of understanding.

It was clear from all of the feedback, that the grant holder gains unfettered access to material that had not been widely seen or scrutinised before. This access often resulted in journal articles, or books, and in some cases award-winning publications. It was also evident that being involved with the digitisation helped with teaching in the classroom and sharing interesting and relevant material with their students. Being a grant holder has also led to applications and support for fully-funded PhD students and research assistants.

Another of my questions was - What aspect of your EAP grant has made you the most proud? I would like to finish with the response of two current grant holders. Mohamed Mwamzandi at University of North Carolina at Chapel Hill, who has a project digitising Pulaar Islamic texts (EAP1245).

> By allowing us to work closely with very senior scholars in Senegal and Mali, our own people in those respective countries could see a great relevance in the work we do in the United States. It helped us contribute back home in a very meaningful way. At UNC, being awarded the grant increased our scholarly visibility in the university, both at the departmental and college levels.

And finally Henny Ziai, a PhD candidate at Columbia University, who was awarded a grant last year to digitise the Abdallah Bey Hamza Collection in Sudan (EAP1284).

> [This project] has been helpful in allowing me to build some of the skills required as a historian of Sudan, namely, providing an opportunity to learn about and acquire an overview of the range of archival sources available in the nineteenth century, their strengths and limitations for historical research...The ability to include 9-10 people, based in Sudan, in a project that involves the preservation of their archival heritage.

However, more so than this, I am proud that this does not involve exploiting their labour, since the grant permits a decent level of pay for their work.

7 Conclusion

Digitised collections and any training material that the Programme makes freely available online have a truly global reach. It is clear from our analytics that the content is not just for academics based at western universities. As a Programme, we are reactive and do not target particular collections, but we are very aware that we need to do more in terms of outreach, promoting our grants and publicising our web statistics. Hopefully attending the 'Frontiers in African Digital Research Conference' and the subsequent publication of this paper is one step towards addressing this. If other funding bodies were able to share their results, the next stage, perhaps, would be to make comparisons to see whether we have had similar results and experiences.

8 Acknowledgements

I would like to thank all the project holders who replied to my questionnaire. I would also like to thank Sarah FitzGerald, for the insightful research she carried out during her placement with EAP and to Pomerol Partners who helped with creating the PowerBI Dashboard.

9 References

Barringer, Terry and Wallace, Marion, eds. 2014. *African Studies in the Digital Age: DisConnects?* Leiden: Brill

Butterworth, Jody, Andrew Pearson, Patrick Sutherland, and Adam Farquhar, eds. 2018. *Remote Capture: Digitising Documentary Heritage in Challenging Locations.* Cambridge: Open Book Publishers.

FitzGerald, Sarah "Visualising the Endangered Archives Programme project data on Africa, Part 1. The project" https://blogs.bl.uk/digital-scholarship/2018/08/visualising-the-endangered-archives-programme-project-data-on-africa-part-1-the-project.html. (last accessed 5 March 2021).

FitzGerald, Sarah "Visualising the Endangered Archives Programme project data on Africa, Part 2. Data visualisation tools" https://blogs.bl.uk/digital-scholarship/2018/08/visualising-the-endangered-archives-programme-project-data-on-africa-part-2-data-visualisation-tools.html. (last accessed 5 March 2021).

FitzGerald, Sarah "Visualising the Endangered Archives Programme project data on Africa, Part 3. Finishing up" https://blogs.bl.uk/digital-scholarship/2018/09/visualising-the-endangered-archives-programme-project-data-on-africa-part-3-finishing-up.html. (last accessed 5 March 2021).

Steven J. Turner M.L.I.S. (2010) "Website Statistics 2.0: Using Google Analytics to Measure Library Website Effectiveness", *Technical Services Quarterly*, 27:3, 261278, https://doi.org/10.1080/07317131003765910. (last accessed 5 March 2021).

Setting Up a Digital Archive for Ethnographic Data

Challenges, Strategies, Experiences

Wolfgang Kraus (University of Vienna)[112]

1 Introduction

Can material from ethnographic research and related qualitative methodologies be made useful and meaningful beyond its original research context? Many ethnographers are deeply sceptical about the possibility of others reusing their data, and they have good arguments to support their position. A major one among them is that submitting to the increasing demand for open data might have a damaging impact on relations between researchers and research subjects because it risks to undermine the relation of trust that is fundamental to most ethnographic research. Giving away data means giving up control about what is done with them, while many researchers feel they can never hand over to others the responsibility for the ways in which their data are used (cf. Imeri 2017; 2019: 49 f.; Pels et al. 2018).

These arguments and others against the open data optimism prevalent in many other fields are well founded and must be taken very seriously. Conversely, there are equally good reasons for archiving ethnographic data in order to preserve and make them available for various kinds and scenarios of reuse – provided that it is done right.

In what follows I will report on what might be described as a grassroots and researcher-centred initiative that I have been leading at the University of Vienna, the establishment of an

[112] Contact: Wolfgang.Kraus@univie.ac.at

ethnographic data archive since 2017. I will discuss the main motivations, strategic considerations and experiences made in a project that is still in its early stages, even though it is permanently established by now. Most of my argument is framed in terms of the specificity of ethnographic research; however, a large part of the challenges we faced and the strategies we developed is relevant for many kinds of qualitative data in the social sciences and humanities.

I have been considering questions of data archiving at various points in my career as a researcher, not least because of my interest in ethnographic photography and audio documentation, which tend to produce *data objects* in a more obvious sense than other forms of ethnographic communication. I also had the good fortune of cooperating with and being supported by an audio archive right from the start of my own field research – indeed, the oldest institution of its kind world-wide, the Vienna-based Phonogrammarchiv of the Austrian Academy of Sciences – where I was able to archive a selection from my field recordings made in Morocco in the mid-1990s (see Kraus 2007).

The specific idea of setting up an ethnographic data archive at the Vienna Department of Social and Cultural Anthropology first came up in 2013–14, while I was head of department, in conversations with my deputy and friend, Gertraud Seiser. We were aware that the department was heading towards a major generational transition, with several colleagues bound to retire over the next years, and discussed the possibility of creating a local archival infrastructure to preserve the ethnographic data they had collected during their research careers, involving them in the process while they were still available. In addition to such considerations, our ideas about why and how this should be done were based on our own long-term experiences of conducting and teaching ethnographic research (e.g., Seiser 2012; Seiser and Schweitzer 2010; Kraus 1998, 2004).

For both of us, the increasing call for data management and open access to research data was not yet relevant when we began to reflect on meaningful ways of preserving ethnographic material for secondary reuse. However, starting in the natural sciences, this has become a major concern in all scientific fields over the last decade (see Allianz 2010). The debate on researchers' responsibilities in producing and handling their data, under such headings as Open Research Data, FAIR data (FORCE11 n.d.) or data management, now provides an important context for all considerations of data preservation and archiving (see, e.g., Imeri 2017, 2019; Mosconi et al. 2019; Pels et al. 2018).

2 Managed and Open Research Data

Therefore, I will begin by looking at this wider context of demands to manage, share and open research data. It is evident that these demands are predicated on an understanding of what research data are, itself based on certain assumptions about the scientific process of producing knowledge. Let us see how appropriate these assumptions are for ethnographic research.

The notion of research data is often treated as self-evident or only descriptively specified in listings of what may constitute research data[113]. Trying to turn this into a definition, as is

[113] "Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden.

sometimes done[114], risks becoming a circular argument: research data are defined by their role in the research process, while research itself can be defined through the systematic use of data.

A definition in the strict sense of the term must in some way also specify how data relate to the object of research. A much-quoted formulation originating from the United States Office of Management and Budget reads, "Research data is defined as the *recorded factual material* commonly accepted in the scientific community as necessary to validate research findings…" (OMB 2006, my emphasis). A slightly expanded variation has been picked up by several UK institutions (see, e.g., EPSRC n.d., itself quoted by other institutions). Another definition states, "Research Data [are defined as] Data that are *descriptive* of the research object, or are the object itself" (University of Bath 2011, my emphasis).

The understanding of data entrenched in such definitions seems to me to be based on two sets of assumptions. In epistemological terms, it assumes that research data document aspects of the real world in a factual or, at least, descriptive sense and in a way that is largely independent of their specific research context. Although due allowance is often made for disciplinary specificities and differences, the basic model is often that of the natural sciences. If data reflect the real world independently of their research context, then they are unproblematic to reuse in a different context. Another important assumption is that access to research data serves "to reproduce and verify the results" of research, as the FWF, Austria's main funding institution for basic research, states in the context of its Open Access policy (FWF n.d.).

A second, related set of assumptions goes like this: once public money is invested to fund research, the results (publications) should be made openly accessible. "Taxpayer-funded research" is an important buzzword here, and one that raises interesting questions concerning the role of national boundaries with regard to access to knowledge (see, e.g., the US Alliance for Taxpayer Access[115]; for a British example, see the Foreword in Concordat 2016: 2). By the same logic, the data collected during research become assets that cannot be owned by the researcher. Rather, they must be shared and the public – including other researchers – has a right to access and re-use them.

Even though I can partially agree with some of the ideas involved in the second set, while remaining sceptical of their neoliberal thrust, I want to underline that taken together these assumptions imply three things that sit rather uncomfortably with the practice and self-understanding of ethnographic research:

- Ideals of objectivity and replicability of research that I consider epistemologically mistaken;

Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden" (DFG 2015).

[114] "Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen" (Allianz n.d.).

[115] https://www.taxpayeraccess.org/. Last accessed 1 June 2022.

- Principles of cost efficiency as well as efficiency in generating knowledge;
- A sharp discontinuity between everyday knowledge/experience and research-based knowledge.

3 Characteristics of Ethnography

Most readers are probably aware of the basic principles underlying ethnographic research. Nevertheless, for the sake of my argument and in contrast to the assumptions outlined above, let me quickly sum up the main characteristics and assumptions of ethnography. Since the methodological approach of ethnography emerged in the field of social and cultural anthropology, I use *ethnographic* and *anthropological* as overlapping and almost synonymous terms, notwithstanding the fact that ethnographic methods are also practiced in many other fields in the social sciences and humanities.

Research is typically done over extended periods of time in close collaboration and exchange with research subjects. It is based on communicative relations with those being researched and sees them as active participants rather than passive objects of observation. Much ethnographic research deals with personal lifeworlds, a fact that implies important issues of trust and responsibility.

Ethnography is not a method but rather a methodology that relies on an open and flexible combination of tools. Hence, ethnographers tend to produce varied and multiple forms and formats of data. Whether in analogue or in digital formats, ethnographic data tend to be technically diverse, and different kinds of data are interpreted in relation to each other and to the overall research context and experience. In recent years and with the establishment of new digital tools, media and ways of disseminating and sharing knowledge based on ethnographic research, this characteristic has come to be discussed as "multimodal anthropology" (e.g., Collins and Durington 2018).

In typical suggestive but vague fashion, Clifford Geertz wrote of the fact "that what we call our data are really our own constructions of other people's constructions of what they and their compatriots are up to" (Geertz 1973: 9). Even if the understanding of the task of anthropology, and the category of data in ethnography, may have expanded a lot since Geertz (1973) proposed his notion of "thick description", most anthropologists today would subscribe to a vision of ethnographic research as based on a relation of dialogue between researchers and the people they work with. From such an understanding, several fundamental assumptions of ethnographic research follow:

- Data are not simply "found" but co-constructed in a process of dialogue between the researcher and the research subjects. They do not merely document facts "out there" but are representations that contain the voices, or perspectives, of both sides involved: the researchers and the researched.
- Therefore, ethnographic data cannot simply belong to the researchers (and even less to their institutions). They also belong to the research subjects and their communities, who have their own interest in the data.

- There are no "raw", uninterpreted data in ethnography. The dialogic process of "making" data is by necessity a process of interpretation.
- Being products of relation and dialogue, ethnographic data are neither objective nor subjective. Both of these notions presume a clear distinction between the observer and the observed object, a distinction that is neither meaningful nor possible in ethnographic research. Therefore, ethnographic data are never simply "descriptive of the research object" (University of Bath 2011), but first of all of the ethnographic relation between researchers and research subjects.
- Both sides involved have their social and cultural context which they bring into the ethnographic encounter. Therefore, there is a gradual difference but no discontinuity between ethnographic knowledge and everyday knowledge and experience.
- Taking into account the historical context in which the ethnographic methodology emerged and, even more importantly, the responsibility that comes with its practice, ethical considerations clearly must take precedence over considerations of efficiency.

On a side note, I should make it clear that I am not claiming that there should be a separate epistemology for anthropological and ethnographic research in contrast to, say, the natural sciences. Rather, I am convinced that the epistemological basis of the assumptions outlined above holds for all forms of science, but that is not a point to be elaborated here.

A second clarification concerns the established distinction between "raw" and "processed" or interpreted data, a distinction that makes no sense with regard to ethnographic material[116]. In what appears to be an opposite view, the anthropologist Peter Pels states, "Anthropologists should… insist on making an epistemological distinction between 'raw' and 'processed data'" (Pels et al. 2018: 394), but it turns out that he means something different by these terms. Simplifying his complex argument, the main point is that ethnographic data typically contain deeply personal information, thus, "Extensive processing of raw materials (beyond mere anonymisation) becomes inevitable if others are to reuse them" (Pels et al. 2018: 394). Yet, that risks to render them useless for further research. I agree; however, "processing" here is not used in the predominant sense of making data usable for analysis; rather, it refers to the task of preparing them for reuse by others. Nevertheless, I do not agree with Pels' argument – if I am reading him correctly –that sharing data in principle cannot be reconciled with the research subject's "rightful claims to knowledge shared with researchers" (Pels et al. 2018: 394). It can – provided that research subjects are explicitly conceptualised as forming part of those for whose access and reuse data should be prepared and archived, and that necessary precautions are taken. These topics are discussed at length in later sections.

4 Towards an Ethnographic Data Archive

Most anthropologists will probably tend to agree with my characterisation of principles of ethnographic research. However, the project of setting up an ethnographic data archive and the

[116] The distinction is also problematic with other kinds of data. From a statistician's point of view, Barrowman (2018) argues convincingly that quantitative data are never raw nor context-free.

archival strategies we devised in the course of our activities relied on additional assumptions which may hold more potential for controversy.

Our basic premise is that, generally speaking, ethnographic data have an intrinsic interest beyond the primary research context, for two main reasons. First, they are complex and rich in ways which are hardly ever fully exploited in the original analysis. This is often a matter not only of complexity but also of sheer quantity.[117]

Second, being situated in time and space, ethnographic data are historical by nature. With the passage of time and as a consequence of transformation and change, they may become interesting and relevant in unforeseeable ways.

Thus, there are good reasons to preserve ethnographic data in order to make them accessible and reusable beyond the original research context – reasons that are entirely unrelated to the rationale of the Open Research Data discourse. However, it would be mistaken to treat them as independent of their original research context. Ethnographic knowledge is embedded in social relations and in complex corporeal experience. Specific data sets can represent the research process only incompletely. As stated above, ethnographic data tend to be diverse, and different kinds of data must be interpreted in relation to each other and to the overall research context. A guiding principle for our archival activities is therefore that the link between specific data and their research context must be retained as far as possible.

I proposed the idea of a pilot project for setting up digital ethnographic data archive in early 2015 via the Faculty of Social Sciences and again a year later. It took the University of Vienna almost a year to evaluate how this idea fit into the overall digitisation and emerging data management strategy before the decision was taken to finance a two-years pilot project that would be attached to the Vienna University Library. The data were to be archived in PHAIDRA, the "repository for the permanent secure storage of digital assets at the University of Vienna"[118]; our cooperation with the PHAIDRA team has proved to be a highly productive and most pleasant experience.[119]

The main people involved were Igor Eberhard, who was hired for a half-time position as archive manager, Birgit Kramreither, the head of the Social and Cultural Anthropology Library, as coordinator, and myself as scientific leader. One of our more obvious insights during the pilot phase was that work capacity was notoriously short. When several months before its end we pondered how to make the best use of the remaining funds, we decided to employ a student assistant, Jasmin Hilbert, to support us.

[117] In my own case, a significant proportion of the material from my second field research phase in central Morocco between 1995 and 2005 remains unused. I doubt that this is a singular experience.

[118] See the PHAIDRA website, https://phaidra.univie.ac.at, last accessed 1. June 2022.

[119] Several people have substantially contributed to establishing and developing the *Ethnographic Data Archive*; most noteworthy among them are Maria Seissl, Susanne Blumesberger, Raman Ganguly, Rastislav Hudak and Claudia Feigl. I and the *eda* team are deeply indebted to them for their ongoing support.

After the pilot project ended in February 2019, the Vienna University Library decided to continue the activities as part of its regular operations. Eberhard's employment was extended and has now been made permanent. Hilbert also continues working with us but on a terminable basis. With the permanent establishment of the archive, we changed its name from the earlier *Projekt Ethnographische Datenarchivierung* to *Ethnographisches Datenarchiv*, or *Ethnographic Data Archive* in English; the acronym *eda* remains the same.[120]

5 Objectives and Strategic Considerations

As mentioned above, the point of departure for *eda* was the generational change at the Department of Social and Cultural Anthropology. In the case of colleagues bound to retire, as in Seiser's and mine, the primary focus was on the digital preservation of historical data from earlier researches, mostly in analogue form. The insight that it was also necessary to support ongoing research and provide data management expertise followed logically from our experience of working with the historical material. It is important to note that our initiative came out of our and our colleagues' research practice. Thus, it is not representative of the "'top down' policy push" that Mosconi et al. identify as a characteristic of Open Science, but rather of what they refer to as the "collegial desire to share data" (Mosconi et al. 2019: 756).

The *eda* team set its main objectives as follows:

- Defining archival and metadata strategies and standards adapted to the specificities of ethnographic research.
- Testing and defining best practice digitisation workflows.
- Networking and exchange with other data management and archival initiatives in related fields.
- Identifying the ethical and legal issues involved and proposing solutions.

In the medium to long term, we aim at developing a comprehensive research data management strategy for social and cultural anthropology and related fields.

Among the guiding strategic considerations, one is that cooperation with researchers makes more sense than the administration of legacies. This reflects experiences we made working with materials existing in the departmental archive, with often insufficient metadata and context information. More importantly, this conviction is based on our holistic understanding of ethnographic data as representing an interactive research process rather than separate aspects of an independent reality. This means for us that all data must be linked to the research setting and the researchers' biographies or careers and should be made accessible in a way that enables the user to take this context into account. Only the researchers themselves can accomplish this task of contextualising and interlinking the data objects in a comprehensive manner.

[120] The *eda* website (https://eda.univie.ac.at/) gives a short overview of the team, activities and cooperations; the *eda* team can be contacted at: eda.ksa@univie.ac.at.

Another guiding principle is the respect and support for the legitimate interest of research subjects and source communities in the data, and their protection from harm. Here again, only the researchers can make informed decisions about the interests and risks involved with specific data sets. As Pels notes, it is their "ethical duty to control how research materials 'go public'" (Pels et al. 2018: 395).

On a more pragmatic level, we aim at sustainability through optimised workflows, appropriate file formats, standardised procedures and metadata, and ongoing quality control. Devising best practice digitisation workflows requires a balancing of contradictory demands. The amount of work and cost involved and the required storage space should be kept low, while the technical quality of a digital copy should be such that it can be expected to be taken as an adequate representation of the analogue original even several decades from now.

Note that this also involves a judgement about the relevant aspects of an object. Imagine the case of a faded photograph. Are we mainly interested in its current appearance – that is, in the effects of history on it – or in the original information that can be restored by proper illumination and by digital editing, or both? In each case, the optimal digital copy or copies will be different. Once again, the researchers can help to make such decisions.

A further non-technical aspect is an ideal of autonomy in the digitising processes. We try to avoid having to hand over objects to third parties for digitising purposes, and have so far been successful in doing so. Finally, our digitising workflows also rely on considerations of the relation between analogue objects as potential carriers of knowledge and their digital representations, and the transformation from one state to the other. The question of the faded image raised previously shows how a conception of what constitutes the data object in relation to the research context must guide the digital representation. Another example is a Compact Cassette, which can contain several field recordings. Is the single recording the object we are interested in, or is it the cassette as an entity representing a specific moment or time span in the field? We have opted for the second as our predominant perspective. Finally, in line with the considerations above, we leave the selection of what to archive to the researchers, while offering our advice when being asked for it.

6 Challenges

Archiving ethnographic data for reuse clearly is a case of contradictory demands. On the one hand, we wish to make the data accessible while preserving their richness and complexity and the connections between them in order to keep them meaningful. On the other hand, precisely these aspects may expose the research subjects or their communities, but it is our duty to respect their interests and keep them safe from harm as best we can. This may force us to restrict or postpone access to the data and/or to process data so as to avoid identification. Finally, researchers themselves may also have an interest in protecting their privacy.[121]

[121] These and other related issues have extensively been discussed in the last few years (see, e.g., Cliggett 2016; Eberhard and Kraus 2018; Imeri 2017, 2018, 2019; Lederman 2016; Pels et al. 2018; Sterzer and Kretzer 2019; Zeitlyn 2012).

Following from this dilemma of principles, there are many practical problems that must be addressed during the process of archiving data. There are no abstract general solutions for these problems, nor is there a single recipe for balancing the underlying contradictions and tensions. Viable compromises must be found for each single case. Therefore, rather than proposing dubious answers, I phrase some of these issues – the list is not exhaustive – as necessary and pressing questions of ongoing relevance:

- How should we adequately represent the wider research context of data objects, their interlinked nature, and the researcher's positionality?
- How can we best protect the research subjects' privacy, interests and security?
- How should we deal with data from historical research where research subjects were not, and can no longer be, asked to make informed decisions about the material?
- How can we pseudonymise without risking to make the data next to useless due to loss of context? (Anonymising in the strict sense is hardly an option with ethnographic material.) The question is further compounded in the case of visual material.
- How do we make sure that the value we attach to personal non-identifiability does not harm interests that research subjects in a given field may value higher, such as social embeddedness, visibility and recognition?[122]
- What could be the possible consequences for research subjects if we decide not to pseudonymise them?
- How can we best make data accessible and meaningful for the research subjects and source communities?
- At least in the case of data that pose a potential threat to personal interests or security, it is also necessary to break up the simplifying notion of "source communities". How can we identify who is entitled to represent communities in claiming access to data?

7 Data, Digitising and File Formats

The forms and formats of data we typically deal with include:

- Text (on paper and digital, the latter often in obsolete file formats): e.g., notebooks, field diaries, transcripts.
- Images (on film, paper and digital): e.g., photos, documents, drawings, maps.
- Audio recordings (analogue and digital on various kinds of data carriers): e.g., interviews, narrations, recitations, rituals, music. With some digital material, again, the problem of file formats comes up, as with some files from closed system dictation devices in highly

[122] See Zeitlyn 2012: 470 f.

compressed proprietary formats. Today, most high-quality digital recorders produce files that can be archived without any conversion, depending on the settings.

Film and video material have not yet been part of our activities. This is due to the specific complexities of digital video such as the multiplicity of formats and the need for data compression; unlike image and audio, storing video uncompressed and in highest resolution is not a practical option. Furthermore, we do not have a lot of personal experience in handling video data. This does not however mean that we exclude video as a matter of principle.

Text or images on paper and similar support materials are scanned or photographically copied. Depending on the size of the original, a photographic copy with a high-resolution digital camera and a high-quality macro lens can provide better real resolution and image quality than most flatbed scanners and always allows better control of lighting characteristics. Digital copies are stored either as image or PDF files. The former option is preferred if it might make sense to edit the image during use, for instance to improve the readability of manuscripts; the latter provides easier access, especially with multi-page documents.

For digitising photos on film (negatives or diapositives) we employ a Hasselblad Flextight X1 scanner, one of the highest-quality film scanners on the market today. Alternatively, and more often, we copy film material with a high-resolution digital camera (Pentax K-1) and extremely high-resolution industrial (line-scan) lenses.[123] Extensive testing proved that with a carefully optimised workflow this is a much faster and in most respects technically superior solution, at least for 35mm film.[124] Image files are archived as TIFF files, usually in Adobe RGB (1998) or Gray Gamma 2.2 space with a resolution of 16 bits per channel (or rarely as lossy JPEG 2000 files with some images reproducing text information).[125]

We use an RME Babyface USB interface as analogue-to-digital converter to transfer analogue audio material – typically on Compact Cassette tape – to LPCM encoded audio with a sampling rate of 96 kHz and 24 bits resolution. A challenge during playback is to determine whether the source material has been recorded using Dolby noise reduction, a fact that is often undocumented. Audio

[123] With its Pixel Shift Resolution system, the 36-megapixel Pentax K-1 DSLR provides full colour information for each pixel, unlike competing full-frame cameras with the same nominal but less real resolution. This is an essential feature to achieve best image quality and avoid the typical imaging artifacts of conventional Bayer sensor cameras. It has a very visible advantage for copywork too where it helps to avoid aliasing issues, such as moiré.

[124] The Flextight X1 has a maximum nominal resolution of 6300 ppi while the Pentax K-1 images provide about 5000 ppi for 35mm film. However, testing showed that this does not translate into a visible advantage of the X1 scans. With medium format film the X1 is able to provide visibly better resolution: 3200 ppi versus about 2200 ppi with the Pentax K-1 for 6x6 negatives or slides.

[125] 16-bit resolution is mandatory for all image editing steps. It is also preferrable for archival files because there may be a need for further editing during reuse. Note that the lossless LZW compression that may be used with TIFF files does *not* reduce files sizes with 16-bit images.

files are archived as WAVE or as losslessly compressed FLAC (with the advantage of smaller file sizes).[126]

Born-digital materials – those that originate in digital form – are best archived in their original format, but this may be problematic in a long-term perspective in the case of proprietary and/or obsolete formats. In these cases, we try to convert them to suitable, preferably open archival formats and archive these as well. Moreover, they may be stored on obsolete data carriers and require special hardware to convert them to archivable files. So far, we have been able to gain experiences with obsolete text files which we managed to read using the open-source LibreOffice software. The preferred archival format for text files is PDF/A. We also processed a large number of digital audio recordings on MiniDisc and from a proprietary and closed dictation system, both with their own specific difficulties. These recordings are stored in their original sampling rate with 16 bits resolution.

We have not yet started archiving born-digital photographs. Here again, image files are best stored in their original format. The lossy JPEG format commonly used by digital cameras is acceptable for digital field photographs but should be avoided for digitisation purposes. Whenever raw file formats – which provide better image quality and allow users to profit from future developments in raw converting technology – are used, it makes sense for archival purposes to convert proprietary raw formats to the open DNG format using the Adobe DNG Converter or other equivalent software. However, it is wise to archive processed TIFF files too. There is disagreement as to whether DNG is a suitable format for long-term archiving,[127] but it should be understood that raw files hold information which is lost when converting them to conventional image formats. Regardless of how this may be judged from the archival perspective, it must be kept in mind that converting raw files into images is not just a technical process but involves of necessity a creative interpretation. With all relevant digital image formats the options of embedding metadata, e.g. the IPTC and EXIF standards, can be extensively used in order to provide technical, descriptive and context information.

For some of the routine processes, we have written detailed step by step guidelines based on our equipment and preferred workflows. For others, this remains to be done. We have also developed a file naming convention and guideline.[128] These guidelines are meant to ensure consistency in processes and compliance with best practice workflows.

8 Metadata

In the process of establishing *eda* we gave a lot of thought to metadata-related questions, for several reasons. Seemingly innocent descriptive information – such as names attached to collectivities – could transport and cement problematic perspectives or ethnocentric

[126] See https://xiph.org/flac/index.html. Last accessed 1 June 2022. The downside is that FLAC files are somewhat less accessible than WAVE, but the open-source VLC media player, for instance, can be used for playback.

[127] See, e.g., DFG 2016: 19 for a negative view.

[128] File names also serve as call numbers for digital objects, but are separate from the object numbers and handles assigned by the repository.

assumptions. Attention to the research context was a guiding principle for our approach to metadata. Another important question was how to facilitate access to data objects for research subjects and source communities. The PHAIDRA team proved extremely flexible and supportive in helping us to develop an *eda*-specific metadata scheme and submit form.[129]

For metadata the PHAIDRA repository relies on a linked-data approach combining several established metadata standards, including Dublin Core, BIBFRAME, SKOS and others. We suggested several additional metadata fields that we considered useful, which were then translated into existing categories.[130]

While access to data objects in PHAIDRA can be restricted or blocked – an indispensable feature for *eda* for obvious reasons –, a basic given of metadata in PHAIDRA is that they are always public and open. They cannot be restricted, and there can be no authorship and no copyright (or *Urheberrecht* in the German and Austrian sense, which is based on a different reasoning) in them.

Our interest in making the research context transparent and tangible turned out to be in potential conflict with this policy. What about personal information that helped to contextualise the data but was not fit for public access? What about background information about historical materials that relied on a research effort and personal interpretation and opinion, in other words, an authorship that should be credited? Our solution is the notion of *context data* that we devised as a pragmatic workaround for these issues. Context data are data objects – generally in text format (PDF/A) – providing information on other objects that is not suited for metadata because it is too complex and extensive, needs being protected, or requires authorial responsibility and copyright. As separate data objects, they can have access restrictions; they are referenced in the metadata of the related objects. However, in the present form, they are not searchable. A significant category of context data is information on research projects and researchers' biographies and research careers, as in the biographical interviews that Igor Eberhard conducted in 2017 with Elke Mader, then professor at the Vienna Department of Social and Cultural Anthropology.[131]

Another consideration was that, given the potential complexity of data objects, we wanted to clearly distinguish between several object categories in metadata, some of which might coincide in the case of a given object. We referred to these categories as:

- born digital-object,
- digital copy,
- first-order analogue object, and
- second-order analogue object.

[129] Rastislav Hudak from the Vienna University Computer Center's Software Design & Development team deserves special credit for his invaluable support in this context.

[130] For an example of an object with fairly complete metadata, see https://phaidra.univie.ac.at/o:1069269.

[131] See https://phaidra.univie.ac.at/o:1146526. Sadly, Elke Mader passed away in August 2021.

While the first two are obvious, the others require an expanded definition. Imagine the case of a musical recording. A straight WAV file from a digital recorder is a born digital-object. A recording on a Compact Cassette, when digitised, is the digital copy of the audio content of a first order analogue object, the cassette. When the recording, beyond the immediate musical event, is considered a representation of a specific instrument that is being played, the instrument is a second-order analogue object. Each one of these categories, as far as they exist in a given object, requires its own metadata. An analogue object in the second sense may also exist in the case of a born-digital object. The *eda* metadata scheme lets us follow this approach, even if an appropriate terminology – not necessarily using our working terms, which might be improved – has not yet been implemented because it is difficult to translate into existing metadata standards.

The submit form that the PHAIDRA team designed for *eda* is relatively transparent and self-explanatory. This will allow us to provide researchers with personalised or project-specific accounts in order to upload their own data and metadata (something we have not yet tried). This option is further facilitated by the possibility to create project-specific metadata templates.

A major metadata-related issue that remains is the lack of appropriate controlled vocabularies. The need for such vocabularies is now widely perceived as pressing, and activities are taking place in various contexts, e.g., in the project GND for Cultural Data (GND4C) (DNB 2019). Our aim is to stay abreast of these developments and, if possible, to participate in them in the future. Some other problems with the standardisation of metadata, concerning for instance the local context of data objects, are also not yet satisfactorily resolved. For these and other reasons, the whole area of index terms is still very underdeveloped in *eda*.

9 Archived Material

We started working with material from a few selected persons who were prepared to participate in the process of finding our way through the intricate practicalities of archiving ethnographic data. The first were Elke Mader with text, audio and photographic material from her two research phases in South America (1979–81 in Peruvian Amazonia and 1990–99 in Ecuador), and Gertraud Seiser with interviews from a field school with MA anthropology students in rural Austria which she taught in the context of the KASS project in 2005.[132] In addition to this material, we worked with images and documents from the ethnographic collection and archive of the Department of Social and Cultural Anthropology. Finally, I brought in field notes, photographs and audio recordings from my research in Morocco (1983–85 and 1995–2005). Simultaneously working with concrete data objects and developing our strategies, workflows and metadata approach proved very productive because different items tended to raise specific issues and forced us to come up with workable solutions.

In addition to the data that have already been archived, we prepared and processed a large number of objects which have not yet been ingested in PHAIDRA for various reasons – lack of work capacity, incomplete metadata, open questions concerning access and so on. By January 28,

[132] http://web.eth.mpg.de/kass/, last accessed 1 June 2022; https://phaidra.univie.ac.at/o:1071491; Seiser and Schweitzer 2010.

2021, *eda* has archived 561 objects. More than half of them are photographs,[133] several are text documents (text- or image-based PDF/A files)[134] and single audio files[135]. A few are collections which can be used to bring together related objects for easier access.[136] The rest are container objects.

The container object is a data object that consists of of several files representing the same analogue object. We conceived this notion based on the idea that the original analogue object is an entity capable of better elucidating the research context than the detached information it might hold. Consider, again, the example of a Compact Cassette containing several recordings made over a time span. Rather than archiving separate tracks, we opted for the format of the container object allowing us to retain the original connection between the recordings. In addition, a Compact Cassette often comes with notes written on it, referring to its content, or on the cardboard insert inside the box. Most of the container objects archived so far represent Compact Cassettes and consist of audio files and photos of the two cassette faces and inserts.[137] Other container objects represent photographs from the ethnographic collection of the Department with extensive captions on the verso.[138]

In addition to these objects, a large number of items from the archive and ethnographic collection of the Department has been digitised and ingested using the *eda* metadata scheme and submit form, but with the Ethnographic Collection as owner rather than *eda*. Most of these are documents such as correspondence and educational slides.

Other than the persons named above, several colleagues have voiced their interest or agreed to work with *eda* in order to archive material from their researches, but we have not yet started cooperating. In principle we are also open to working with researchers from other institutions than the University of Vienna. We have made contacts with some researchers whose material seems to us to promise synergies with data sets we already have but, again, this has not yet resulted in concrete cooperations.

10 Conclusion and Outlook

The main assumption underlying *eda*'s activities is that ethnographic data constitute historically situated representations of a world in flux. As such they do have a value beyond the primary research context and should in principle be preserved and shared. However, the dialogic character

[133] E.g., https://phaidra.univie.ac.at/o:1096509.

[134] E.g., https://phaidra.univie.ac.at/o:947378.

[135] E.g., https://phaidra.univie.ac.at/o:1048725. This object and some other audio files are short excerpts from longer field recordings containing oral traditions in *Tamazight* Berber language that I published and analysed (Kraus 2019: 79). Being cross-referenced with the publications, they exemplify one way of using the repository to connect data and publications.

[136] E.g., https://phaidra.univie.ac.at/o:1146527.

[137] E.g., https://phaidra.univie.ac.at/o:957555.

[138] E.g., https://phaidra.univie.ac.at/o:936768.

of ethnography and the access ethnographers gain to personal life-worlds raise important issues of confidentiality, privacy and reciprocity.

As discussed above, there is a basic contradiction between the aim of making data accessible to others outside the primary ethnographic relation, and the ethical obligations that follow from that relation. This contradiction cannot be resolved on the level of principles. Nevertheless, despite the difficulties and challenges discussed above, it makes sense to find pragmatic but ethically sound compromises that allow us to preserve and share – with various possible forms of restrictions – ethnographic data without harming our research subjects. After all, the situation is not so unlike the challenges that come along in the process of ethnographic writing and publishing – challenges for which we routinely find pragmatic single-case solutions.

However, these inherent tensions require a number of additional measures that *eda* will have to deal with in the future, and before archiving more sensitive data than we have had so far. They have to do with its implementation within the PHAIDRA repository but also with matters of principle such as the relation between research subjects and the data they helped to generate. The categories of ownership of, control of, access to and licence to use data need to be distinguished and clarified. Since for ethical and legal reasons in many cases opening data is not an option, flexible models of access management are required, but complex and difficult to implement. If we want to avoid the necessity of defensively precluding access to all sensitive material – an option that would defeat the short to mid-term purpose of the archive – then in the future development of PHAIDRA several graded levels of access to data must be defined and implemented.[139] This will be an important step, even if making decisions about appropriate levels then will add another layer of complexity to the archiving process.

On the surface, these are ethical and legal issues that arise after the fact of data production, unless they have been clarified through informed consent, which should be the case with more recent research. Nevertheless, I hope to have made it clear that, given the collaborative nature and other characteristics of ethnography, it is impossible to draw a strict line between research and subsequent forms of organising knowledge. It is both consistent and productive to extend the understanding of ethnography as a co-production of knowledge to the archiving process. We have not yet made serious steps in that direction, but consider it necessary not only in making decisions about access to data but also with regard to producing metadata and context data. In recent years, various approaches to organising knowledge have emerged that build on communities and collaboration and could provide models of how to proceed.[140]

If I read him correctly, Peter Pels suggests there is more than one "reason to consider social science data as indigenous or global heritage" (Pels et al. 2018: 393) – a perspective that makes both the preservation of ethnographic data and the inclusion of research subjects and their communities in the process an ethical obligation. Only such an inclusive approach will be able to

[139] For an elaborate model of how this can be done, see Sterzer et al. 2018.

[140] For an example of a community-centred approach concentrating on scholarly communities, see Herrmann and Kurzawe 2020; for a museum-based approach of collaboration with source communities, see Scholz 2017a, 2017b.

safeguard data that are findable, accessible, interoperable and re-usable (FORCE11 n.d.) in the long term for all interested parties.

11 References

Allianz der deutschen Wissenschaftsorganisationen. n.d. Forschungsdaten. https://www.allianzinitiative.de/archiv/forschungsdaten/. (last accessed 31 May 2022).

Allianz der deutschen Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten. https://doi.org/10.2312/ALLIANZOA.019. (last accessed 31 May 2022).

Borrowman, Nick. 2018. Why Data Is Never Raw. *The New Atlantis* (56): 129–135. https://www.thenewatlantis.com/publications/why-data-is-never-raw. (last accessed 31 May 2022).

Cliggett, Lisa. 2016. Preservation, Sharing, and Technological Challenges of Longitudinal Research in the Digital Age. In *eFieldnotes: The Makings of Anthropology in the Digital World*, ed. Roger Sanjek and Susan W. Tratner, 231–250. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812292213-014.

Collins, Samuel Gerald, and Matthew Slover Durington. 2018. Multimodality. In *The international encyclopedia of anthropology*, ed. Hilary Callan, 1–7. Hoboken, NJ, Chichester, West Sussex: Wiley Blackwell. https://doi.org/10.1002/9781118924396.wbiea2304.

Concordat. 2016. Concordat on Open Research Data. https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf. (last accessed 31 May 2022).

DFG. 2015. Leitlinien zum Umgang mit Forschungsdaten. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf. (last accessed 31 May 2022).

DFG. 2016. DFG Practical Guidelines on Digitisation: DFG form 12.151 – 12/16. https://www.dfg.de/formulare/12_151/12_151_en.pdf. (last accessed 31 May 2022).

DNB. 2019. GND für Kulturdaten (GND4C). https://www.dnb.de/DE/Professionell/ProjekteKooperationen/Projekte/GND4C/gnd4c.html. (last accessed 31 May 2022).

Eberhard, Igor, and Wolfgang Kraus. 2018. Der Elefant im Raum: Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 41–52. https://doi.org/10.31263/voebm.v71i1.2018.

Engineering and Physical Sciences Research Council. n.d. EPSRC policy framework on research data: Scope and benefits. https://epsrc.ukri.org/about/standards/researchdata/scope/. (last accessed 31 May 2022).

FORCE11. n.d. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing Version b1.0. https://www.force11.org/fairprinciples. (last accessed 31 May 2022).

FWF. n.d. Open Access to Research Data. https://www.fwf.ac.at/en/research-funding/open-access-policy/open-access-to-research-data/. (last accessed 31 May 2022).

Geertz, Clifford. 1973. Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures*, ed. Clifford Geertz, 3–30. New York, NY: Basic Books.

Herrmann, and Felix Kurzawe. 2020. Discuss Data: Community-zentrierter Ansatz für das Forschungsdatenmanagement in den Geistes- und Sozialwissenschaften (2): 56–64. https://doi.org/10.17192/BFDM.2020.2.8277.

Imeri, Sabine. 2017. Open Data? Zum Umgang mit Forschungsdaten in den ethnologischen Fächern. In *E-Science-Tage 2017*, ed. Jonas Kratzke and Vincent Heuveline, 167–178. University Library Heidelberg: heiBOOKS, 167–178. http://doi.org/10.11588/heibooks.285.377.

Imeri, Sabine. 2018. Archivierung und Verantwortung: Zum Stand der Debatte über den Umgang mit Forschungsdaten in den ethnologischen Fächern. In *Archivierung und Zugang zu Qualitativen Daten*, ed. Betina Hollstein and Jörg Strübing, 69–79. Berlin: Rat für Sozial- und Wirtschaftsdaten. https://doi.org/10.17620/02671.35.

Imeri, Sabine. 2019. Open Data' in den ethnologischen Fächern: Möglichkeiten und Grenzen eines Konzepts. In *Forschungsdesign 4.0: Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive*, ed. Jens Klingner and Merve Lühr, 45–59. Dresden: Institut für Sächsische Geschichte und Volkskunde. https://doi.org/10.25366/2019.07.

Kraus, Wolfgang. 1998. Contestable Identities: Tribal Structures in the Moroccan High Atlas. *Journal of the Royal Anthropological Institute* 4 (1): 1–22.

Kraus, Wolfgang. 2004. *Islamische Stammesgesellschaften: Tribale Identitäten im Vorderen Orient in sozialanthropologischer Perspektive.* Wien/Köln/Weimar: Böhlau. https://doi.org/10.26530/oapen_574825.

Kraus, Wolfgang. 2007. Als Anthropologe in Zentralmarokko: Methodologische Reminiszenzen. In *Um-Feld-Forschung: Erfahrungen - Erlebnisse - Ergebnisse*, ed. Julia Ahamer and Gerda Lechleitner, 165–179. Wien: Verl. der Österreichischen Akad. der Wiss. https://austriaca.at/0xc1aa5576%200x00152aec.pdf.

Kraus, Wolfgang. 2020. Overlooking the past from the High Atlas mountains: Ayt Ḥdiddu oral traditions. *Études et Documents Berbères* 41 (1): 71–86. https://doi.org/10.3917/edb.041.0071.

Lederman, Rena. 2016. Archiving Fieldnotes?: Placing "Anthropological Records" among plural Digital Worlds. In *eFieldnotes: The Makings of Anthropology in the Digital World*, ed. Roger

Sanjek and Susan W. Tratner, 251–271. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812292213-015.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three Gaps in Opening Science. *Computer Supported Cooperative Work (CSCW)* 28 (3-4): 749–789. https://doi.org/10.1007/s10606-019-09354-z.

OMB. 2006. OMB Circular A-110. Uniform Administrative Requirements for Grants and Other Agreements with Institutions of Higher Education, Hospitals and Other Non-Profit Organizations (11/19/1993) (further amended 09/30/1999). *Office of Management and Budget Circulars and Guidance, Washington D.C.: The White House* 39–68. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A110/2cfr215-0.pdf. (last accessed 25 May 2021).

Pels, Peter, Igor Boog, J. Henrike Florusbosch, Zane Kripe, Tessa Minter, Metje Postma, Margaret Sleeboom-Faulkner, Bob Simpson, Hansjörg Dilger, Michael Schönhuth, Anita von Poser, Rosa Cordillera A. Castillo, Rena Lederman, and Heather Richards-Rissetto. 2018. Data management in anthropology: The next phase in ethics governance? *Social Anthropology* 26 (3): 391–413. https://doi.org/10.1111/1469-8676.12526.

Scholz, Andrea. 2017a. Sharing Knowledge as a Step toward an Epistemological Pluralization of the Museum. *Museum Worlds: Advances in Research* 5 (1): 133–148. https://doi.org/10.3167/armw.2017.050112.

Scholz, Andrea. 2017b. „Wissen teilen" als postkoloniale Museumspraxis: Ein Kooperationsprojekt zwischen der Universidad Nacional Experimental Indígena del Tauca (Venezuela) und dem Ethnologischen Museum Berlin. *Sociologus* 67 (1): 59–81. https://www.jstor.org/stable/45174713.

Seiser, Gertraud. 2012. Kooperationen in der Landwirtschaft kontextualisiert: Ethnografische Fallstudie von der Mühlviertler Alm. *Österreichische Zeitschrift für Soziologie* 37 (4): 385–402. https://doi.org/10.1007/s11614-012-0063-1.

Seiser, Gertraud, and Peter P. Schweitzer. 2010. Houses, families, and the making of community: The Austrian rural case. In *The view from below: Nineteen localities*, ed. Patrick Heady and Peter Schweitzer, 109–123. Frankfurt, New York: Campus Verlag.

Sterzer, Wjatscheslaw, and Susanne Kretzer. 2019. Archivierungsstrategien anpassen: Herausforderungen und Lösungen für die Archivierung und Sekundärnutzung von ethnologischen Forschungsdaten. *Bibliothek Forschung und Praxis* 43 (1): 110–117. https://doi.org/10.1515/bfp-2019-2011.

Sterzer, Wjatscheslaw, Sabine Imeri, and Matthias Harbeck. 2018. Zugriff auf ethnologische Forschungsdaten: Anforderungen und Lösungen. Poster for the 107. Deutscher Bibliothekartag in Berlin 2018. https://nbn-resolving.org/urn:nbn:de:0290-opus4-157193.

University of Bath. 2011. ERIM Terminology V4: Terminology for Research Data Management. https://wiki.bath.ac.uk/display/ERIMterminology/ERIM%20Terminology%20V4. (last accessed 31 May 2022).

Zeitlyn, David. 2012. Anthropology in and of the Archives: Possible Futures and Contingent Pasts. Archives as Anthropological Surrogates. *Annual Review of Anthropology* 41 (1): 461–480. https://doi.org/10.1146/annurev-anthro-092611-145721.

Portals and Platforms

Making the Indiana University Sub-Saharan African Collections Digitally Accessible

Mireille Djenno (Indiana University Bloomington)[141]

1 Introduction

This essay differs slightly from the presentation I gave at the Frontiers in African Digital Research Conference in early 2020. This is in part to take into consideration the extraordinary global events of 2020 and beyond, as well as to use the opportunity to broaden the discussion and problematize some of the issues raised by the case study at the heart of the presentation. The focus of the present essay is the modes by which Indiana University (Bloomington) currently creates digital access to its holdings of African archival materials, and the complex set of factors, including some ethical considerations, that influence the provision of access to these collections. This essay is also written within the context of a renewed interest in centering the work of African scholars of Africa in the wake of Jean Allman's presidential lecture at the African Studies Association (ASA) 2018 conference. The caveat that accompanied the Bayreuth presentation also applies here: while the case study presented aims to be instructive, it should not be taken to be exemplary of comparable institutions and/or their means or methods of creating digital access to their African archival collections.

1.1 Case Study

Indiana University (Bloomington), where I am the African Studies Librarian, holds an extraordinary collection of materials from and related to sub-Saharan Africa, including its renowned collections of material related to Liberia, the H.K. Banda Archive, the Archives of

[141] Contact: mdjenno@indiana.edu, https://orcid.org/0000-0003-0790-2915.

Traditional Music, the personal papers of world figures Ngũgĩ wa Thiong'o, Nadine Gordimer, Ousmane Sembène, among other notable collections. These materials have been acquired, through purchase and donation, for close to 80 years at the time of writing. In the last 30 years it has been possible, because of technological advancements, to digitize some of this material, and to make some of it accessible remotely. In the discussion that follows, when I refer to digitization, I am using the widely adopted UNESCO definition of digitization, i.e. "the creation of digital objects from physical, analogue originals by means of a scanner, camera or other electronic device" (UNESCO, n.d.). When I refer to "creating access", I mean these digitization processes as well as the use of various portals and platforms to link directly to digitized objects or to make archival material known (discoverable) to the user.

The essay takes a deductive approach by starting with the broad context of access provision and from there, drilling down to as many of the specifics as the essay parameters allow. This approach highlights the hindrances and affordances of the systems currently in place. I begin with some information about Indiana University itself.

1.2 Context

Indiana University (IU), located in Bloomington, Indiana, is a large (43,000 students), public, Research One, predominantly white institution (PWI). A few of these characteristics are particularly notable for contextualizing the digital access currently provided by IU to its African archival holdings.

In the American context, a Carnegie Classification of "Research One" or "R1", refers to "institutions that awarded at least 20 research/scholarship doctoral degrees and had at least $5 million in total research expenditures" (The Carnegie Classification of Institutions of Higher Education, n.d.). More colloquially, and better suited to our purposes, the designation denotes a well-resourced institution.

The significance of "PWI" goes beyond the ratio of racial groups that is implied. A PWI designation indicates that neither the mission nor the compositional diversity of the institution reflects the interests of racial minority groups. The term is often contrasted with "minority-serving institution" (Bourke 2016). IU's status as a PWI is in keeping with its location in southern Indiana, in a racially homogenous, mid-sized city of 84,000, not easily connected by mass transit to domestic or international metropolitan centers. Because of their location, IU's archival African collections remain out of easy physical reach of most scholars of Africa, regardless of their origins and/or resources. This is even more the case for African scholars of Africa, many of whom lack the resources to consult archives in person and therefore rely upon digital access for their research.

Without an implicit or explicit institutional mandate that collections be made accessible to anyone (scholar or not) beyond the campus community, there has hitherto been no programmatic urgency to digitize the IU African archival collections en masse and the collections are, by default, most accessible to those with no proximal African heritage. These two factors arguably contribute to IU's peculiar instance of what Jean Allman calls the "racial landscape of knowledge production about Africa" (Allman 2019). This context is consequential if for no other reason than that put forth by Michele Pickover: "In this globalized world, knowledge and information and, as a

consequence heritage, are seen as strategic resources and tools. The manner in which information is used and who controls it is therefore pivotal" (Pickover 2009).

2 National Level

2.1 Cooperative African Materials Project / Center for Research Libraries

One of the longest standing collaborations for provision of remote access in which IU has been involved at the national level is the Cooperative African Materials Project (CAMP). CAMP is administered by the Center for Research Libraries (CRL),"an international consortium of university, college, and independent research libraries" based in Chicago, Illinois. Originally created in 1963 to share microfilm of sub-Saharan newspapers among participating institutions, CAMP has evolved to include the preservation of other types of documentary resources, many of which are accessible to member institutions via a web portal maintained by CRL. IU's most significant contribution to CRL has been content from the Indiana University Liberian Collections, a set of collections first associated with the IU Department of Ethnomusicology, and later, transferred to the IU Libraries.

While Indiana University is involved with national organizations and entities such as the Council on Library and Information Resources (CLIR), the Digital Library Federation (DLF), and the Digital Public Library (DPLA), we do not currently contribute any of our Africa-related content. These entities represent opportunities for creating wider access to our African archival collections, but care will need to be taken in providing such wide access. As Zinaida Manžuch notes, digitization projects undertaken with an eye toward creating worldwide access often create new inequities and conflicts in the process (Manžuch 2017). Or to put it in Pickover's terms "the digitization of heritage material for publication on the worldwide web is a site of struggle and the real challenges are not technological or technical but social and political" (Pickover 2009).

IU's location in the West as well as the campus' lack of cultural connection to the substrate material make the digitization of its African archival holdings susceptible to biased decision-making with regard to the means and extent of access of the digitized material (Manžuch 2017). Moreover, while privacy concerns of persons implicated in archival material are heightened when information is made available in an easily shared form, these concerns are even more acute when the material was never intended for wide consumption and is used without the consent of the communities in which it originates. Such would be the case with information regarding the Poro, an indigenous West African secret society whose beliefs and practices are represented in some of IU's archival collections, to take but one example.

3 Regional Level

The most important academic cooperative arrangement in the Midwestern region is the Big Ten Academic Alliance (BTAA), a consortium of 14 member institutions. The dominant library discussion currently happening within the BTAA consortium revolves around shared print collections but there are also ongoing discussions about how member institutions negotiate digital access to published content, and how we might collectively provide digital access to rare and unique content. As of this writing, African Studies librarians at BTAA institutions are forming

their own group within which to discuss the implications of these consortial decisions on access to their collections, inclusive of archives.

4 Campus Level

4.1 Media Digitization and Preservation Initiative

IU also undertakes digitization initiatives at the campus level, the most renowned of which is the Media Digitization and Preservation Initiative (MDPI), which began in 2013. MDPI, which sought to make the audiovisual holdings of Indiana University digitally accessible, is an example of an initiative that was executed as a coordinated effort by various campus repositories, including campus libraries[142].

	IULC Inventory for MDPI (minus film)		
	2009	**2017**	**Total by Format**
Audio Cassette	51	101	152
Dictaphone Tapes		2	2
Audio Open Reel	75	104	179
CD-R		6	6
DAT		8	8
DVD		11	11
MiniDV		149	149
LP Disks (Commercial)	92		92
Video-Umatic	24		24
VHS	9	54	63
Totals	**251**	**435**	**686**

Phase I: In the first phase of the project, $15 million dollars was allocated for digitization, which did not include funding for storage or staff. In this phase, time-based media was selected to be digitized by a custom/purpose-built system created by Memnon (later acquired by Sony). Items whose copyright was not clearly established or were commercially available were excluded from

[142] https://mdpi.iu.edu/ (last accessed: 6 May 2022).

this otherwise ecumenical process. Approximately 700 time-based items related to Liberia were sent to MDPI for digitization during this phase. This number excludes any 16mm, 8mm and Super8mm films.

Phase II: The second phase of digitization involved film. No items from the African collections were added during this phase.

Phase III: In total, over 350,000 items were digitized over the course of the MDPI project. Very importantly, all of the items were digitized to the same standard. Collection managers are currently creating access to the digitized material, facilitated by this standardization.

Although creating access involves sorting through various rights issues, donor stipulations, and other considerations, and then determining the appropriate amount of access to a given item or collection, individual campus units and/or collection managers do not have to devise systems for making this content available. The content digitized by MDPI is made available (if it is made available at all) via Media Collections Online (MCO), an IU Libraries platform, accessible on the web. Some collections are available strictly to the IU community (current students, faculty and staff) and subsets thereof; others still are available to everyone in the world via the web. Within the broad legal framework set by the IU Office of the Vice President & General Counsel, collection managers, in consultation with the Head of the IU Libraries' Copyright Program, determine if access restrictions are needed. In keeping with the current ethos of openness in archives, and the Indiana University Libraries Digital Preservation Policy of providing access to the digitized objects within its purview "as broadly and openly as possible", worldwide access is granted whenever possible[143].

5 Libraries Level

Below the campus-level, smaller scale digitization projects are often initiated by individual library collection managers who obtain funding by various means (including external funding), in order to complete the digitization of material held by the Libraries. In these cases, it is left to the initiating managers to put together ad hoc teams of campus-based technical experts to support these projects. Over the years, portions of the IU African Collections have been digitized in this manner, including our collection of posters from Somalia, a project which represented "the unified efforts of several departments of Indiana University's Library system" (Gibbs 1997). The posters were digitized with external funding from the United States Department of Education in the form of a Title II-C grant, awarded in 1994. The purpose of Title II is to "provide grants to State educational agencies and subgrants to local educational agencies to—increase student achievement...; improve the quality and effectiveness of teachers, principals, and other school leaders; increase the number of teachers, principals, and other school leaders who are effective in improving student academic achievement in schools…" (US Department of Education 2015). The broad rationale therefore, for federal funding of the poster digitization project under Title II-C was that digital access to these resources held by IU would be of benefit not just to the IU community,

[143] https://media.dlib.indiana.edu/ (last accessed: 9 May 2022).

but to educators everywhere. It is often the case that grant funders determine the type and extent of access, with federal agencies typically stipulating broad access as a condition of funding.

5.1 IU Libraries African archival collections

In an article published in 2010, my predecessor Dr. Marion Frank-Wilson described the digitization efforts related to collections of Africana personal papers held by IU Libraries, including the H.K. Banda Archive, and the Nuer Field Notes collection, whose digitization she oversaw (Frank-Wilson 2010). In the years since, additional archival collections of major significance to African Studies, have come to various campus repositories, including IU Libraries.

- Lilly Library[144]

Archival material related to Africa has largely come to the university via donation but significant collections have been purchased by the IU special collections Lilly Library. Administratively distinct from the main (Wells) library collections I oversee, the Lilly holds the papers of Kenyan Nobel Prize winner Ngũgĩ wa Thiong'o, Senegalese filmmaker Ousmane Sembène, South African writer Nadine Gordimer, South African playwright Athol Fugard, and Senegalese novelist Boubakar Boris Diop. To date, these collections have not been digitized and plans for digitizing them are unknown to the author. It should be noted, in keeping with the present discussion, that the (current) lack of digital proxies to these collections of monumental importance to their respective countries of origin does nothing to re-center African scholarship in the production of knowledge about Africa.

- Archives of Traditional Music (ATM)[145]

A memorandum of understanding (MOU) was signed in 2013 transferring the non-musical portions of Indiana University Liberian Collections to the IU Libraries main (Wells) library collection. The traditional Liberian music material which formed the nucleus of this varied collection remains part of the 421 collections of field recordings from the African continent held by the IU Archives of Traditional Music. These collections (which exclude commercial recordings) represent close to 7800 individual recordings, and a steadily increasing number of them are accessible through MCO.

- Wells Library[146]
- Liberian Collections

The non-musical Liberian collections that were transferred to the main library collections are vast and varied. They constitute, by far, the largest portion of the archival material related to Africa held by Indiana University, and include government documents, field research notes, and personal correspondence, among many other types of material. In addition to brokering many of the donations to the collection before it became part of the IU Libraries, Dr. Verlon Stone, the

[144] https://libraries.indiana.edu/lilly-library (last accessed: 9 May 2022).

[145] https://libraries.indiana.edu/archives-traditional-music (last accessed: 9 May 2022).

[146] https://libraries.indiana.edu/herman-b-wells-library (last accessed: 9 May 2022).

coordinator of the Indiana University Liberian Collections (IULC), initiated several digitization projects, the most notable of which is the digital publication of the Liberian Government Archives. In 2012 and 2013, with grants from CAMP, along with contributions from libraries across the country receiving U.S. Department of Education Title VI grants, Dr. Stone coordinated the digitization of Liberian National Archives documents that had been collected by Dr. Svend E. Holsoe. Other parts of the collections were partially processed by Dr. Stone, with the assistance of a project archivist that was in place for two years, as well as student workers.

- External partnerships

Yet other parts of the Liberian collections were digitized as part of other projects, notably with external partnerships with the British Library Endangered Archives Programme (EAP) and the National Science Foundation. The IULC received EAP grants from in 2005 and 2007, with which it microfilmed the papers of Liberia's long-serving President William V.S. Tubman. The Tubman papers and photographs were later digitized and are now accessible via the British Library website as well as on IU platforms.[147]

Funded by the National Science Foundation, and developed by the University of Wisconsin-Madison, George Mason University, the IULC, and the Liberian Center for National Documents and Records Agency (CNDRA), the multimedia project. *A Liberian Journey* digitized film originally shot in the 1930s related to the advent of the Firestone plantations. The original film was converted to 16mm in the 1970s. The project was managed by Dr. Gregg Mittman, of the University of Wisconsin-Madison, where the materials were digitized. IU scanned some photos, film, and diaries, which are stored in our high-capacity/high-security Auxiliary Library Facility (ALF).[148]

The US-based community of Liberians and Liberianists, some of whom are active scholars, others of whom have a strong heritage interest in these materials, is very close-knit. In his work in building these collections prior to their ownership transfer to the IU Libraries in 2013, Dr. Verlon Stone identified this community as the primary beneficiaries of access (Manžuch 2017) but curricular and other institutional priorities of the university, both within and outside of the Libraries, have limited resources for the processing and digitization of these collections. This has meant that both the scholarly and heritage communities do not enjoy ready access to these materials.

Anecdotally, the collections of materials related to Liberia at Indiana University are among the most important of their kind in the world, and the lack of digital proxies to much of this material means that important raw materials of scholarship about Liberia remain easiest to access for non-Liberian scholars, and materials crucial to the understanding of Liberian history remain out of reach for most Liberians. As Samantha Winn notes in her discussion of displaced archives, while "general users may be inconvenienced when displaced archives are stored remotely or in

[147] https://eap.bl.uk/project/EAP139 (last accessed: 9 May 2022).

[148] http://liberianhistory.org/ (last accessed: 9 May 2022).

undisclosed locations…the removal of archival materials from their country of origin creates a unique hardship for users who regard the materials as their cultural patrimony" (Winn 2015).

- H.K. Banda Archive

This archive, dating mostly from the 1950s to the 1990s, was given to the IU Libraries by Dr. Don Brody, the official biographer of Prime Minister and later President H.K. Banda of Malawi. "It includes published and unpublished correspondence, speeches, manuscripts, diaries, and extensive background information about Southern and Central Africa)" (Frank-Wilson 2010). A subset of the collection was digitized a decade ago, and is accessible via Archives Online, described in the following section.

- Nuer Field Notes

Former missionary to South Sudan, Eleanor Vandevort, donated this collection comprised of "[unpublished] linguistic field notes on the Nuer language, correspondence, slides, as well as a scanned copy of her ethnographic monograph, A Leopard Tamed: the Story of an African Pastor, his People, and his Problems." It is a simple HTML site hosted on library servers and is discoverable via the African Studies Research Guide, described further down.

- Digital Somali Library

The Digital Somali Library which was recently migrated to an IU platform called Sitehosting from the previous Linux-based central web server called Webserve, is the portal to the IU Libraries collections of Somaliana which include digitized language books, and posters (discussed above).

6 Creating Access at the Library Level

Developed by what is now the IU Libraries Digital Collections Services unit, two platforms host the finding aids and digitized images of various archival collections (called "repositories" in the systems). Although separate, these two platforms are linked, with items listed the finding aids of some collections linking directly to digitized images. At the heart of these platforms is a practice of standardized ways of digitization that facilitate reuse by end users.

- Archives Online[149]

Archives Online "is a portal for accessing descriptions of Special Collections and Archives - ones chiefly containing materials other than books - from libraries, archives, and other units at Indiana University Bloomington and from other institutions around the state of Indiana."[149] Finding aids related to the various Liberian collections, as well as the H.K. Banda Archive are hosted on Archives Online for the time being but are being shortly migrated to ArchiveSpace, an open source platform that has been adopted by libraries of comparable size (Calahan 2016).

[149] https://libraries.indiana.edu/databases/archivesonline (last accessed: 9 May 2022).

- Image Collections Online (ICO)[150]

The IU Libraries are also in the process of overhauling our image curation infrastructure. For now, the purpose-built Image Collections Online (ICO) database, launched in 2011, remains the repository for freely available images that can be used for non-commercial educational or research purposes. The images in the database that relate to Africa come principally (but not exclusively) from a sub-collection titled Liberian Photograph Collections, itself comprising four constitutive collections, including the William V.S. Tubman Collection. Tubman was President of Liberia, and the collection of photos having to do with his personal and professional life are part of the over 30,000 images related to Liberia from the 1940s to the present, held by IU.

Given that these are sub-collections of much larger repositories on two different and disparately administered platforms, download and access statistics are hard to come by for these digitally accessible collections. This makes it difficult to know how frequently the collections are used, how this has changed over time, and what that says about the demand for remote access. It is hoped that the new infrastructure will allow us to gauge the use of these collection more accurately.

- African Studies Research Guide[151]

Conversely, the usage statistics for the African Studies Research Guide are very easy to determine. The research guide, which I began building a few months after I got to IU (toward the end of 2016), is an attempt to gather all of the existing web-accessible IU resources related to Africa in one interface and to serve as a platform for resources to come.

The software used to build the guide is LibGuides, a near ubiquitous commercial library web service, and its capabilities include providing usage statistics and reusable content from other guides. While the proprietary nature of LibGuides means that users are limited by the strictures Springshare (the owner of LibGuides) chooses to impose, LibGuides has been widely adopted by libraries in North America and is therefore familiar to many library users, and easy for librarians without technical expertise to use to create portals to primary and secondary sources.

The guide itself is organized by collection type and in addition to links to the platforms created and supported by the Libraries, i.e. Archives Online and Image Collections Online, the Archival Collections section contains links to the following platforms created by my predecessors, and maintained by me and sometimes a half-time staff assistant, namely The Digital Somali Library and the Nuer Field Notes collections.

Elsewhere in the guide, there are links to other archival resources created by librarians, including the Africa Interactive Map Project (AIMP), an improvised open source discovery tool (Partlow et al. 2019), the digitized collection of Somali posters, and the Omeka-hosted Pamphlet Collection, a diverse collection of printed ephemera that includes brochures, flyers, pamphlets, booklets,

[150] https://libraries.indiana.edu/databases/imageonline (last accessed: 9 May 2022).

[151] https://guides.libraries.indiana.edu/africanstudies (last accessed: 9 May 2022).

reprints, conference papers, and other items written in a variety of languages, including African vernacular languages. (Hardesty 2014).

7 The Labor of Access

As the foregoing illustrates, access to IU archival collections related to Africa is created through the labor of many people. This distributed model of labor includes full-time library staff, as well as student workers, and any personnel hired ad hoc for grant-funded digitization projects.

The in-house technical expertise in the IU Libraries is mostly concentrated in the Digital Collections Services (DCS) department, which created the Archives Online and Image Collections Online platforms in the early 2000s. This department, which concentrates on project work, sometimes with the support of grant funding, executes many of its functions with the help of student workers –in ready supply because of the university's fee structure.

Other non-experts, including me, are often responsible for the day-to-day maintenance of these platforms/interfaces. For example, in addition to maintaining the African Studies Research Guide, I create and edit finding aids for the collections I oversee, as well as maintain the collections of pamphlets and posters hosted on Omeka. While there are benefits to being thus empowered, and cross-training and other forms of support are available, there are serious limitations to such an approach. Balancing this work with myriad other duties means that the maintenance of these systems (often subject to infrastructure changes at the university level and/or policy changes at the Libraries' level) is uneven, resulting in varied user experiences.

8 Access Challenges

8.1 Siloing

The various means by which our collections are made accessible are reflective of the various institutional "silos" that exist. As is the case in many cultural heritage institutions, the prevailing ethos of access to archival materials runs up against institutional constraints, siloing chief among them. While university-level initiatives such as MDPI are vast in their scope (as are the infrastructural changes at the library-level, e.g. the migration from Webserve to IU Sitehosting), they cannot meet every access need that arises with growing and evolving collections.

Additionally, silos are not easily scalable. Again, the standardization achieved by MDPI points the way to a scalable model but it would mean collection managers relinquishing control of how they make archival material available, and would inevitably involve the kinds of delays associated with unit-level or campus-level initiatives.

8.2 Funding

While IU is a well-resourced institution, funding that supports the maintenance of existing resources as well as the creation of new digital access/collections is often severely limited, requiring collection managers to seek external funding to support initiatives. This requires time and effort rarely available to collection managers whose job duties involve more than collection management.

8.3 Rights & Permissions

As mentioned, most of the African archival collections held by IU came to us by donation. Older gift agreements did not have digitization provisions whereas the deeds of gift currently in force often make distribution via digitization a condition of donation. This has meant that some agreements have had to be revisited with donors (or their executors) in order to bring the gifts in alignment with current practices.

8.4 Visibility

During its first three years of existence, the African Studies Research Guide received 3300 visits. Changes by our User Experience unit to support LibGuides rather than Drupal for discovery layers managed by collection managers, and indexing by Google of the IU Libraries' LibGuides (beginning in January 2020) has meant a significant increase in visits – 13,200 visits as of March, 2022. The increased visibility provided by adopting Libguides as a platform underscores the fact that digitally accessible collections can remain "hidden" even to local researchers, depending on portal infrastructure and indexing mechanisms.

9 The Path ahead

The foregoing is meant to illustrate the complex environment in which the African Studies Collections of Indiana University Libraries tries to preserve and create access to archival collections, and to go some distance toward explaining the hindrances and affordances of our current configuration. Our goal is consistent, predictable access via persistent URLs and digital preservation and, like everyone else, we are not facing easy or straightforward solutions to these challenges. We are nonetheless proceeding in imaginative ways, keeping focused on what we want for end users, and continuing to experiment with the tools and approaches that make the most sense for those objectives. We are well aware, as Michael Levine-Clark puts it "…the services and tools that libraries have built to provide access to content have evolved to the point that they may become just as important to the institutional mission as the collection itself" (Levine-Clark 2014).

At the organizational level (IU Libraries) we are hoping to build capacity on the foundations of a service model that emphasizes unmediated access. Ideally, users will be able to access/download digital objects for which we have cleared the rights (our MDPI project is in some ways a local model for this) or which have been placed into the public domain by their rights holder(s). At the collection level, we have different digitization priorities/objectives; emphasizing digital humanities capacity for some collections; basic access for others; ("big data" is not an issue for African materials); and are open to contributing content (not necessarily master files but certainly derivatives and metadata) to DPLA and others.

10 Conclusion

Digital access to collections is usually constrained by available resources. In the case of generally resource-rich institutions, the constraints are different, but they exist nonetheless. In the case of the IU Libraries, where we are fortunate to have good storage conditions, there is less urgency to digitize for the sake of preservation than from the desire to make these collections more easily accessible first and foremost to our university community and then, to the wider world.

Within that broad effort to make collections more accessible, our efforts to digitize our African archival collections has never privileged accessibility for African scholars of Africa. As I stated close to the outset, only those with the wherewithal to make their way to Bloomington have had access to the substrate, which reinscribes other asymmetries of access and reinforces by default, the current conditions of knowledge production about Africa. I think it is therefore incumbent upon us to make further digitization and digital enhancement of our collections part of an explicit mission to increase access to non-published resources originating from Africa, more accessible to African scholars of Africa and thereby seize the opportunity to meaningfully intervene in the existing modes of knowledge production about Africa.

11 References

Allman, Jean M. 2019. #HerskovitsMustFall? A Meditation on Whiteness, African Studies, and the Unfinished Business of 1968. *African Studies Review* 62 (3): 6–39. https://doi.org/10.1017/asr.2019.40.

Bourke, Brian. 2016. Meaning and Implications of being labelled a predominantly white Institution. *College and University* 91 (3): 12–18.

Calahan, Lisa, and Kate Dietrick. 2016. Setting the Stage and Keeping Sane: Implementing ArchivesSpace at the University of Minnesota. *Journal of Archival Organization* 13 (3-4): 114–126. https://doi.org/10.1080/15332748.2018.1443502.

Frank-Wilson, Marion. 2010. Africana personal papers at Indiana University: issues and questions. *African Research and Documentation* (112): 15–24.

Gibbs, Ann. 1997. Indiana University Libraries' Somali Poster Preservation Project. *Microform & Imaging Review* 26 (1): 34–36. https://doi.org/10.1515/mfir.1997.26.1.34.

Hardesty, Juliet, L. 2014. Exhibiting library collections online: Omeka in context. *New Library World* 115 (3/4): 75–86. https://doi.org/10.1108/NLW-01-2014-0013.

Indiana University Libraries. (n.d.). Digital Preservation Policy. https://libraries.indiana.edu/digital-preservation-policy (last accessed 9 Mai 2022).

Levine-Clark, Michael. 2014. Access to Everything: Building the Future Academic Library Collection. *Libraries and the Academy* 14 (3): 425–437. https://doi.org/10.1353/pla.2014.0015.

Manžuch, Zinaida. 2017. Ethical Issues In Digitization Of Cultural Heritage. *Journal of Contemporary Archival Studies* 4 (2): 1–17.

Mnjama, Nathan. 2015. Migrated archives: The African perspectives. *1Journal of the South African Society of Archivists* 48:45–54. https://doi.org/10.4314/jsasa.v48i0.

Partlow, Mia, Theresa Quill, and Mireille Djenno. 2019. An open source solution to a common collection problem: The Africa image and map portal. *Library Hi Tech News* 36 (4): 7–9. https://doi.org/10.1108/LHTN-01-2019-0004.

Pickover, Michele. 2009. *Contestations, ownership, access and ideology: Policy development challenges for the digitization of African heritage and liberation archives*. Conference Proceedings. First International Conference on African Digital Libraries and Archives (ICADLA-1), Addis Ababa, Ethiopia, 1st-3rd July 2009. https://core.ac.uk/download/pdf/39667845.pdf (last accessed 6 Mai 2022).

The Carnegie Classification of Institutions of Higher Education. (n.d.). About Carnegie Classification. https://carnegieclassifications.iu.edu/ (last accessed 9 Mai 2022).

United Nations Educational, Scientific and Cultural Organization (n.d.). Fundamental Principles of Digitization of Documentary Heritage. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/digitization_guidelines_for_web.pdf (last accessed 9 Mai 2022).

United States Department of Education. 2015. *TITLE II - Part C: Preparing, Training, and Recruiting High-Quality Teachers, Principals, or other School Leaders* https://www2.ed.gov/policy/elsec/leg/essa/legislation/title-ii.pdf (last accessed 9 Mai 2022).

Winn, Samantha. 2015. Ethics of Access in Displaced Archives. *Provenance, Journal of the Society of Georgia Archivists* 33 (1): 5–13.

Dreiser, Anja R. and Cyrus Samimi, eds. 2022. Frontiers in African Digital Research. *University of Bayreuth African Studies Online* 9. Bayreuth: Institute of African Studies.

ISSN 2628-1791
DOI: 10.15495/EPub_UBT_00005720

UNIVERSITÄT BAYREUTH

INSTITUTE OF AFRICAN STUDIES

africamultipl CLUSTER OF EXCELLENCE