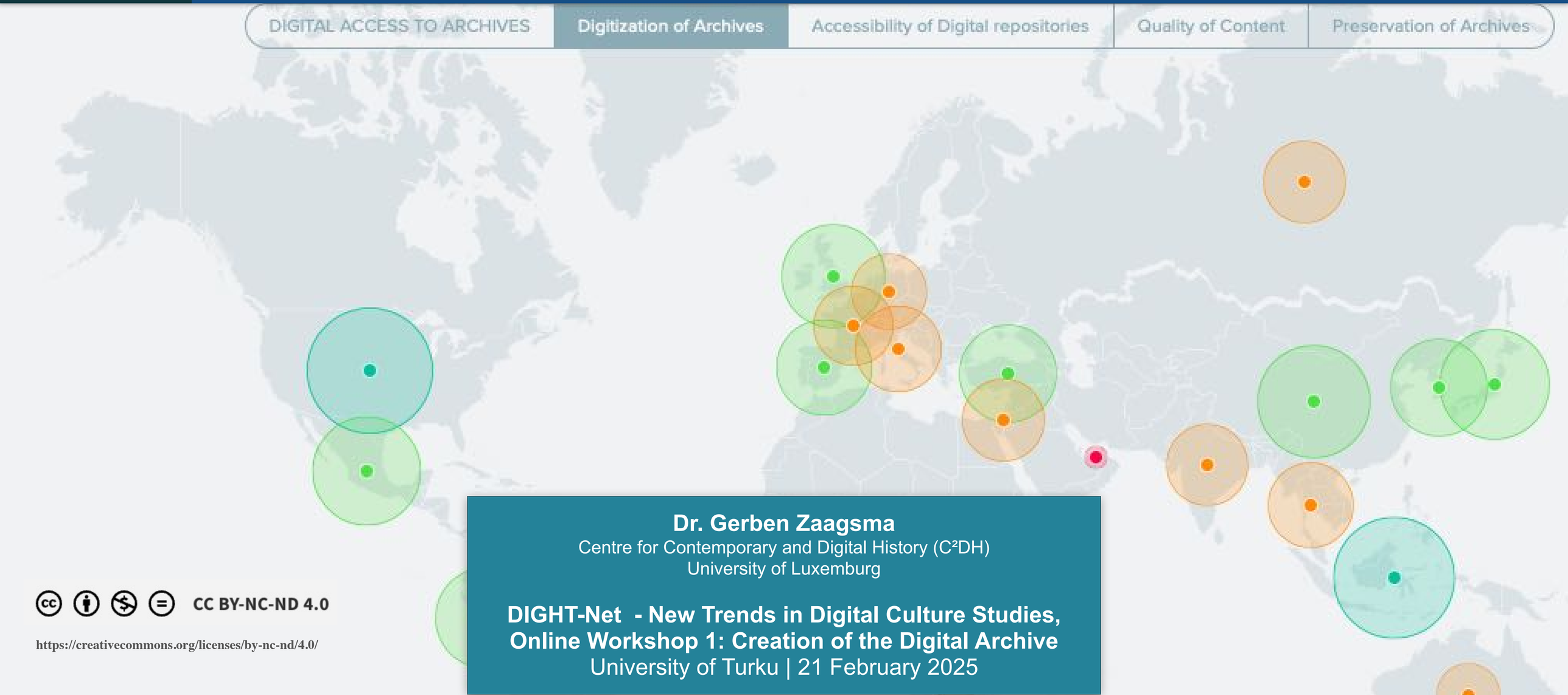
DIGHT-Net

The Digital Archive and the Politics of Digitisation



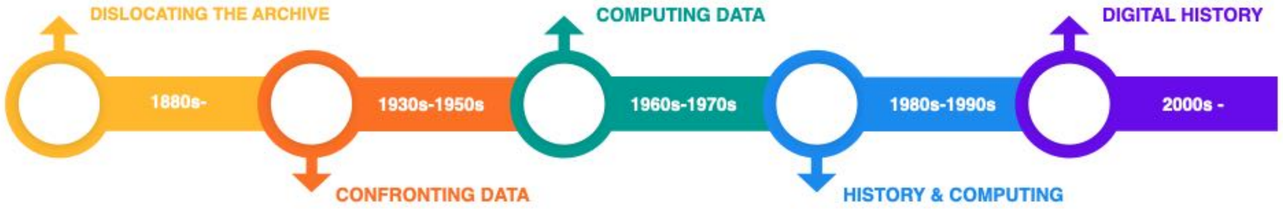
Dr. Gerben Zaagsma
Centre for Contemporary and Digital History (C²DH)
University of Luxemburg

DIGHT-Net - New Trends in Digital Culture Studies, Online Workshop 1: Creation of the Digital Archive
University of Turku | 21 February 2025

CC BY-NC-ND 4.0

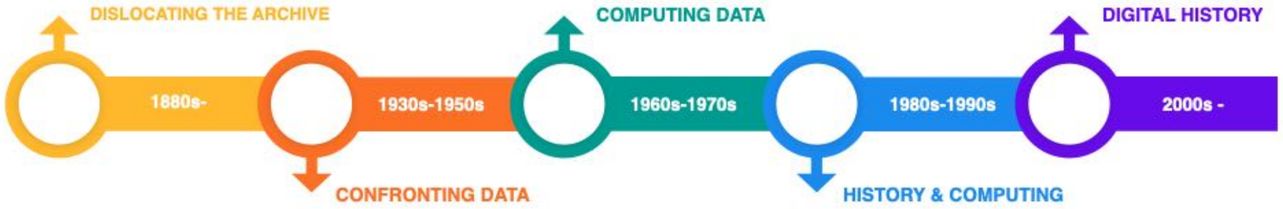
https://creativecommons.org/licenses/by-nc-nd/4.0/

Technology and historical knowledge production



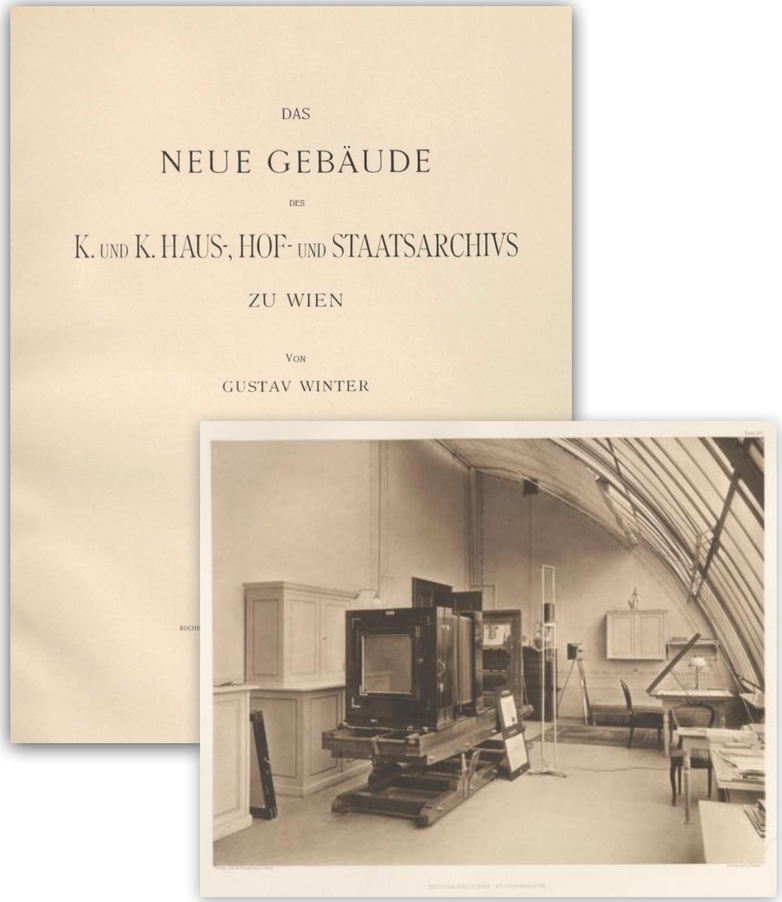
Gathering	Transcribing, excerpting, note taking sources in physical archive > acquiring photos > photostats > microfilm > retro-digitisation + born-digital data
Processing	Manual organising, listing, ordering, tabulating > mechanical aids > computers.
Analysis	Manual interpretative > computer-aided methodological approaches: quantitative and qualitative practices of text, mapping, network and image analysis, etc.
Dissemination	Print publication > cd-roms> project websites, blogs, social media, visualizations and 3D reconstructions

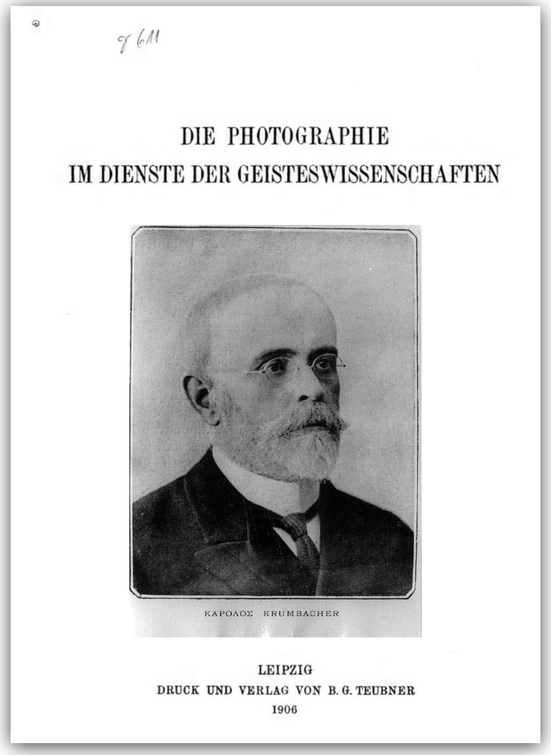
Technology and historical knowledge production

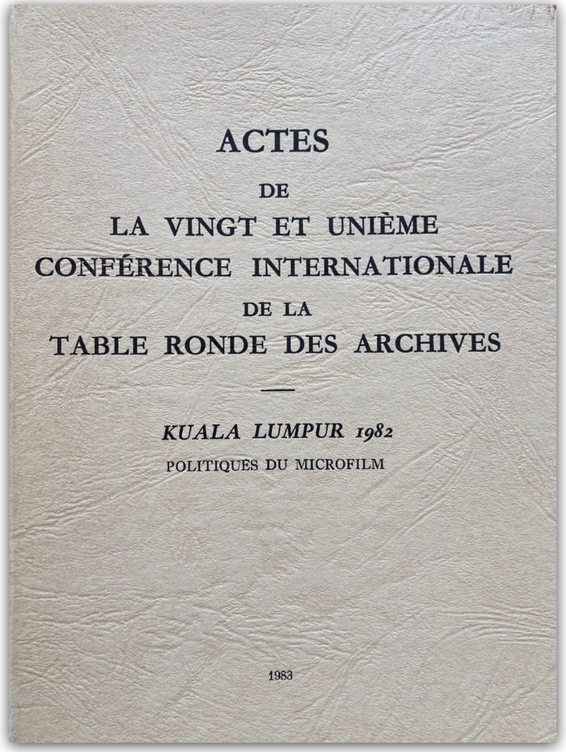


Gathering	Transcribing, excerpting, note taking sources in physical archive > acquiring photos > photostats > microfilm > retro-digitisation + born-digital data
Processing	Manual organising, listing, ordering, tabulating > mechanical aids > computers.
Analysis	Manual interpretative > computer-aided methodological approaches: quantitative and qualitative practices of text, mapping, network and image analysis, etc.
Dissemination	Print publication > cd-roms> project websites, blogs, social media, visualizations and 3D reconstructions

Digitisation in historical perspective

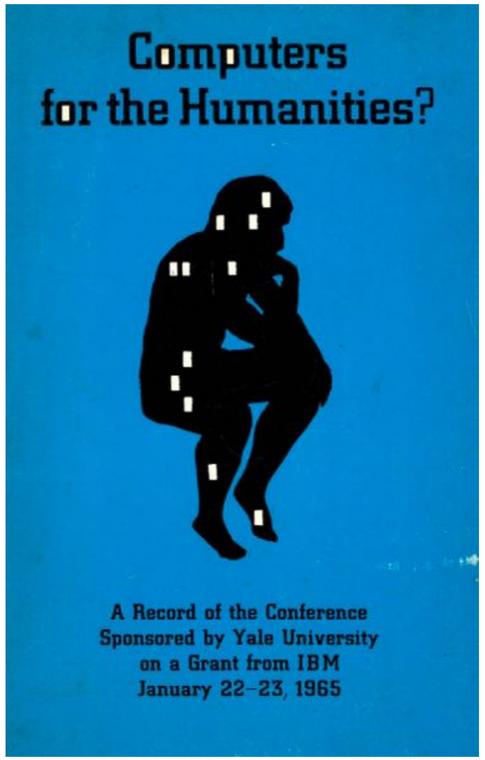






Gustav Winter, *Das neue Gebäude des k. und k. Haus-, Hof- und Staatsarchivs zu Wien* (Wien: Carl Gerold's Soh, 1906).

Abundance in historical perspective



Problems of Abundance

First I shall mention several historical projects facing what might be called the problem of abundance. It seems to me that the computer has had a significant role in eliciting and making easily available large bodies of information about the past. One extremely ambitious project is being carried on by the Ad Hoc Committee of the American Historical Association on the Collection of the Basic Quantitative Data of American Political History under the Chairmanship of Lee Benson.

the past, and how we find and define historical evidence.

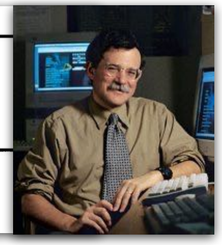
Historians, in fact, may be facing a fundamental paradigm shift from a culture of scarcity to a culture of abundance. Not so long ago, we worried about the small numbers of people we could reach, pages of scholarship we could publish, primary sources we could introduce to our students, and documents that had survived from the past. At least potentially, digital technology has removed many of these limits: over the Internet, it costs no more to deliver the *AHR* to 15 million people than 15,000 people; it costs less for our students to have access to literally millions of primary sources than a handful in a published anthology. And we may be able to both save and quickly search through all of the products of our culture. But will abundance bring better or more thoughtful history?[6]

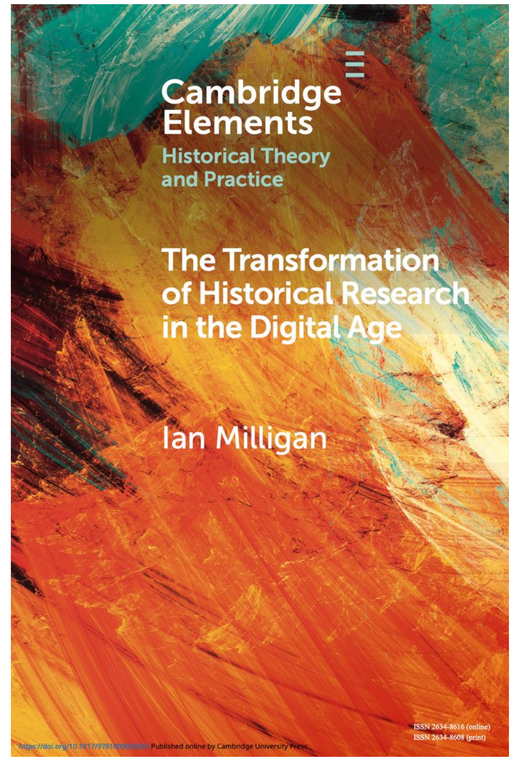
Historians are not unaware of these challenges to the ways that we work. Yet,

Forum Essay

Scarcity or Abundance?
Preserving the Past in a Digital Era

ROY ROSENZWEIG





Outline

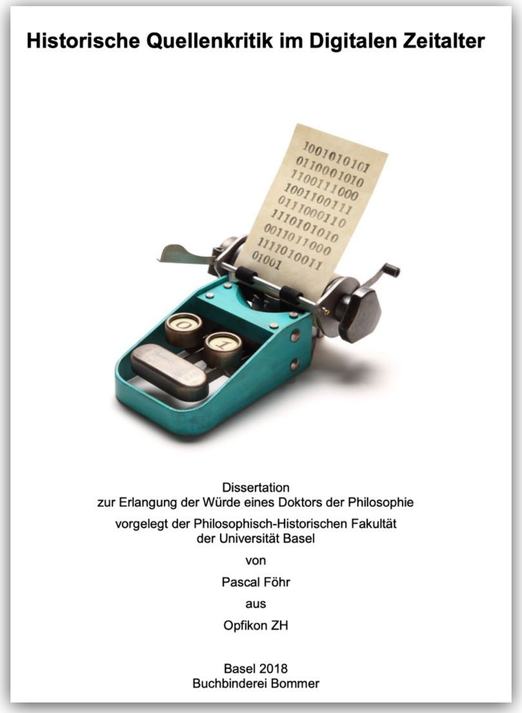
- Technology and historical knowledge production.
- A critical turn in digital history.
- The politics of (digital) cultural heritage.
- The state of digitisation.
- Political dimensions of online archival and curatorial work.
- Concluding remarks.

Top: Staatsarchiv Wolfenbüttel. Source.
Left: Internet Archive book scanner. Source.

Critical turn in digital history & humanities

The critical turn in digital history: from **technological determinism** to **methodological and epistemological reflection**:

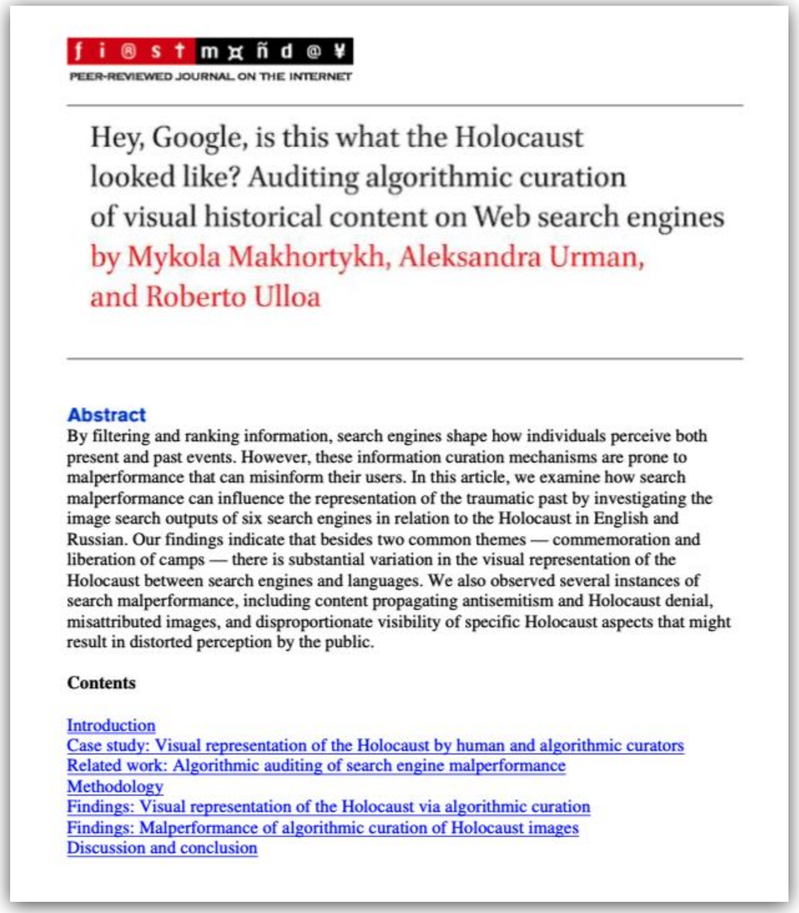
- **Tool** criticism
- **Data** criticism
- **Algorithmic** criticism
- **Code** criticism
- **Interface** criticism
- **Digital source** criticism



Historische Quellenkritik im Digitalen Zeitalter

Dissertation
zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Historischen Fakultät
der Universität Basel
von
Pascal Föhr
aus
Opfikon ZH

Basel 2018
Buchbinderei Bommer



PEER-REVIEWED JOURNAL ON THE INTERNET

Hey, Google, is this what the Holocaust looked like? Auditing algorithmic curation of visual historical content on Web search engines

by Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa

Abstract

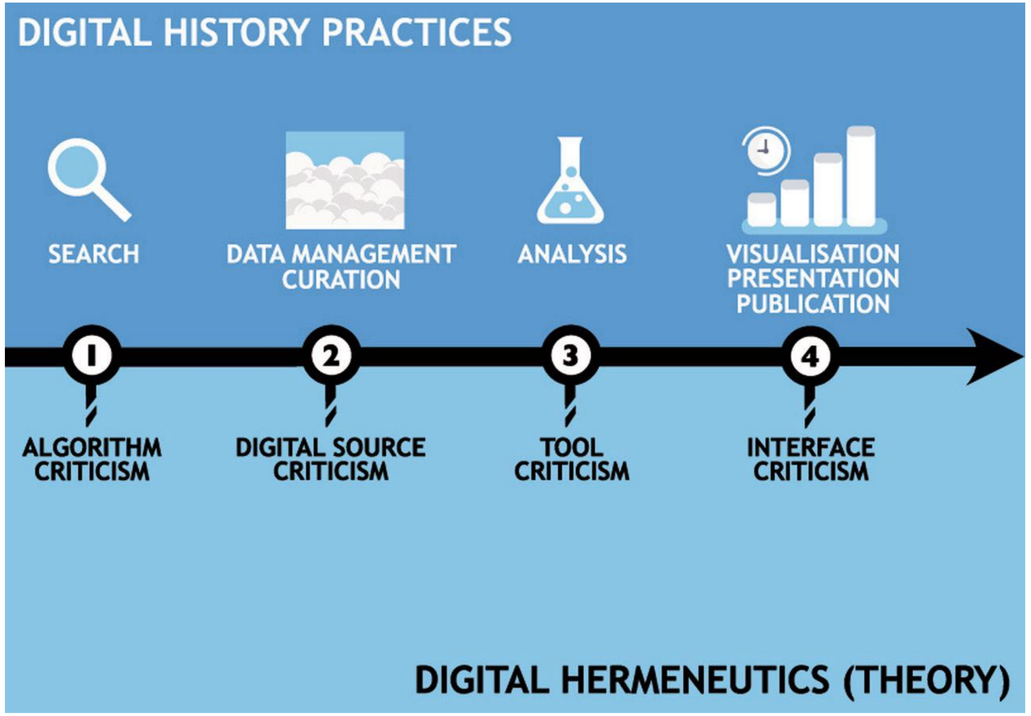
By filtering and ranking information, search engines shape how individuals perceive both present and past events. However, these information curation mechanisms are prone to malperformance that can misinform their users. In this article, we examine how search malperformance can influence the representation of the traumatic past by investigating the image search outputs of six search engines in relation to the Holocaust in English and Russian. Our findings indicate that besides two common themes — commemoration and liberation of camps — there is substantial variation in the visual representation of the Holocaust between search engines and languages. We also observed several instances of search malperformance, including content propagating antisemitism and Holocaust denial, misattributed images, and disproportionate visibility of specific Holocaust aspects that might result in distorted perception by the public.

Contents

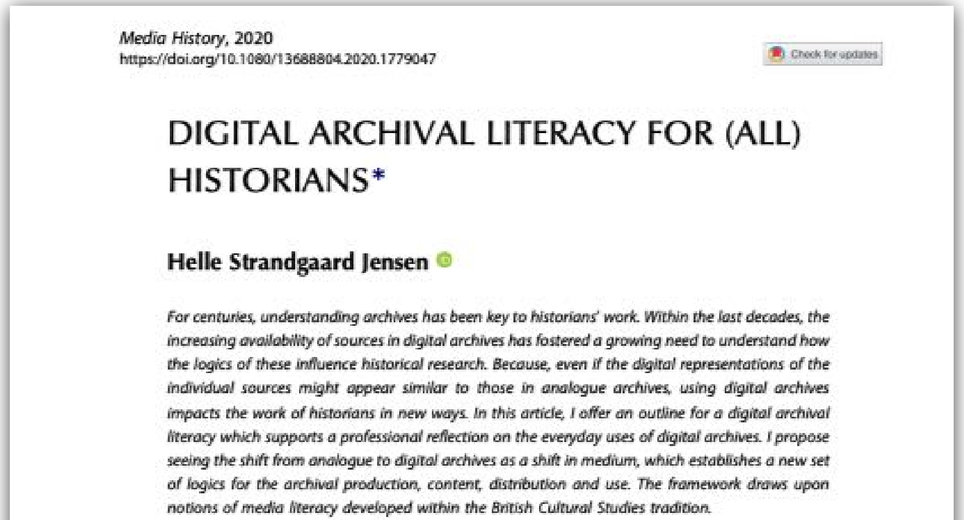
Introduction
Case study: Visual representation of the Holocaust by human and algorithmic curators
Related work: Algorithmic auditing of search engine malperformance
Methodology
Findings: Visual representation of the Holocaust via algorithmic curation
Findings: Malperformance of algorithmic curation of Holocaust images
Discussion and conclusion

Critical turn in digital history & humanities

"We use "**digital hermeneutics**" as a concept that enables historians to critically reflect on the various interventions of digital research infrastructures, tools, databases, and dissemination platforms in the process of thinking, doing and narrating history."



Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, 'Digital history and hermeneutics – between theory and practice: An introduction' in: Andreas Fickers and Juliane Tatarinov eds., *Digital History and Hermeneutics: Between Theory and Practice* (Berlin: De Gruyter Oldenbourg, 2021) 1-23, 6-7.


Media History, 2020
https://doi.org/10.1080/13688804.2020.1779047

DIGITAL ARCHIVAL LITERACY FOR (ALL) HISTORIANS*

Helle Strandgaard Jensen

For centuries, understanding archives has been key to historians' work. Within the last decades, the increasing availability of sources in digital archives has fostered a growing need to understand how the logics of these influence historical research. Because, even if the digital representations of the individual sources might appear similar to those in analogue archives, using digital archives impacts the work of historians in new ways. In this article, I offer an outline for a digital archival literacy which supports a professional reflection on the everyday uses of digital archives. I propose seeing the shift from analogue to digital archives as a shift in medium, which establishes a new set of logics for the archival production, content, distribution and use. The framework draws upon notions of media literacy developed within the British Cultural Studies tradition.

Critical turn in digital history & humanities

Article

JoLIS

Journal of Librarianship and Information Science
1–13
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0961000620918647
journals.sagepub.com/home/lis
SAGE

Defending the digital: Awareness of digital selectivity in historical research practice

Jon Coburn
School of History and Heritage, University of Lincoln, UK

Abstract

Existing research offers fearful conclusions on the use of online archival collections, finding that historians ignore and overlook the limitations of digital sources. However, an attitudinal case study at Newcastle University contradicts this consensus. This article discusses this study's detailed findings, determining that historians and library professionals demonstrate abundant and nuanced awareness of issues relating to 'digital selectivity'. Nevertheless, the interviewees suggested that this does not radically undermine their practice. The study also revealed compelling aspects of digital selectivity not currently observed in existing research, including the effects of career stage and the importance of cooperation with library professionals. While the existing literature provides appropriate and prudent concerns, it lacks tangible evidence pointing to a widespread phenomenon of poor digital historical practice. The article closes by recommending further research into historians' digital information-seeking behaviour, but argues that standardized metadata practices must come first. Limitations, including the absence of rigorous quantitative data, must be resolved before a fuller appreciation for digital historical practice is developed.

Keywords

Digitization, historians, archives, academia, libraries, digital selectivity, awareness, information literacy, research practice, information-seeking behaviour

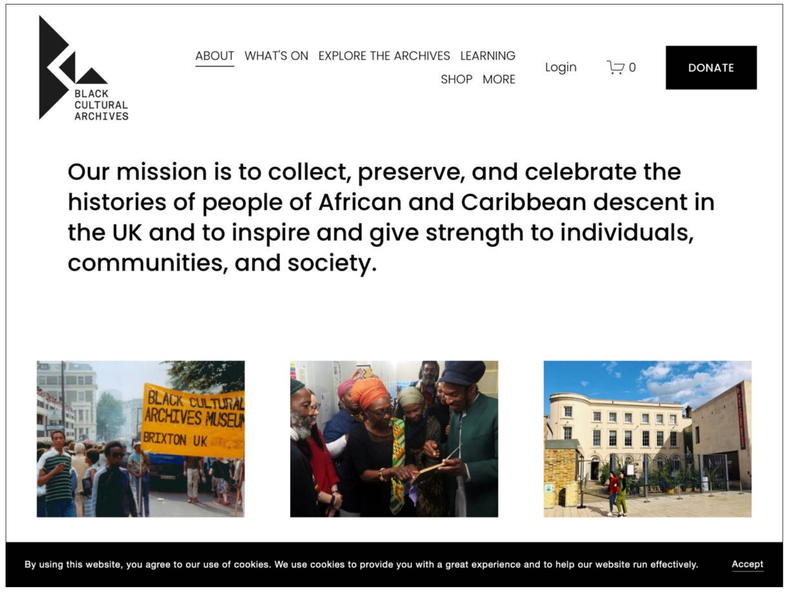
This rebuff to the more surgical form of data retrieval that occurs with digital archives is grounded in the participant's long experience of performing research prior to the widespread adoption of digital archival collections. They continued by explaining:

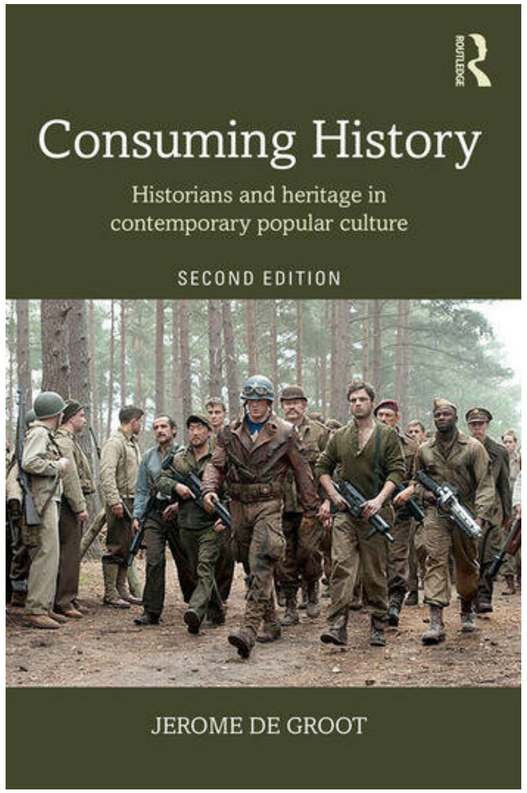
> Quite often when you're going through loads of stuff that you think 'This is boring me rigid', you know, 'When am I going to come across something relevant?', something relevant comes up and it's not one of those keywords you used if you were in a digital archive situation. (Interviewee 2)

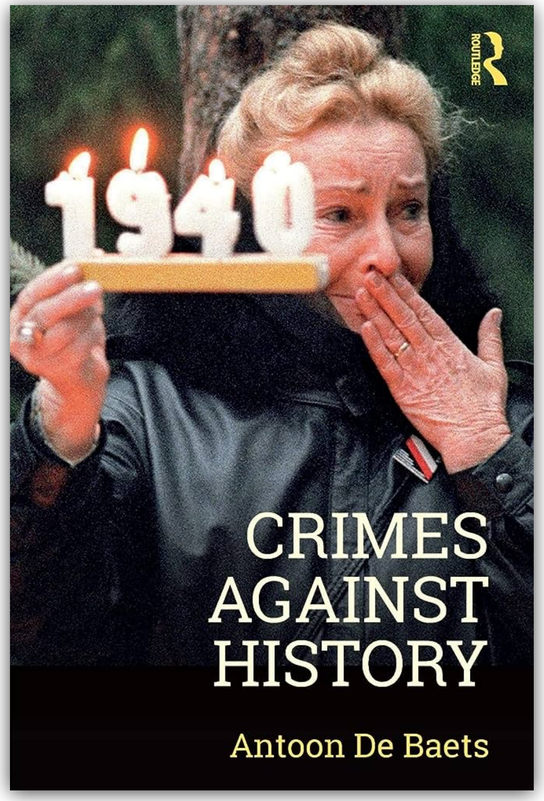
Once more, the participant made a distinct separation between the conduct in a physical archive and that in a digital archive. Early career researchers (ECRs), on the other hand, were more inclined to talk about the benefits of keyword searches and surgical discovery. Concerns over digital selectivity are, therefore, not consistent across the profession. Different academics prioritize different aspects.

Critical turn in digital history & humanities

- Which history is consumed online?
- How is history consumed online?
- How is history co-created online?
- How is history censored online?

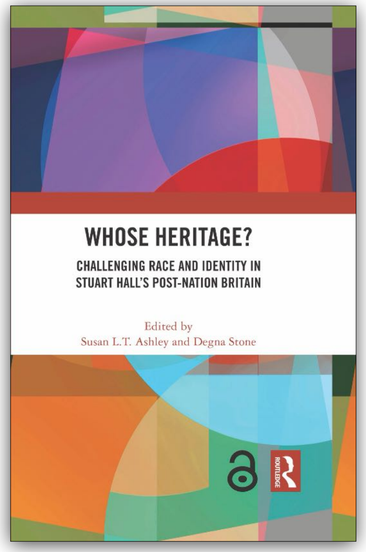


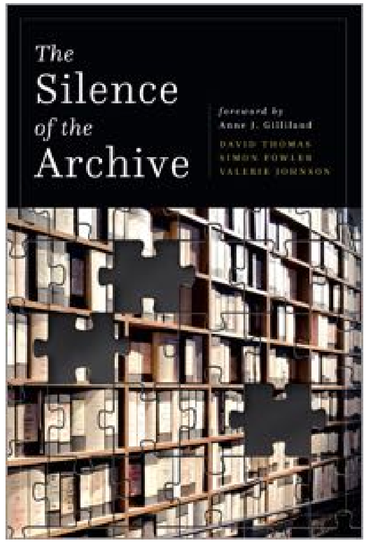


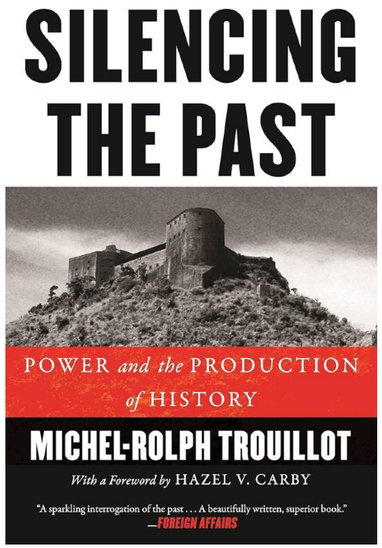


The politics of (digital) cultural heritage

- What counts as heritage has always been political and contested.
- Archives are not neutral but actors, and potential gatekeepers, in the production of historical knowledge.
- Silences in traditional archives and the digital age
- Digitisation ≠ neutral but a selection of a selection.
- What kind of histories can (and cannot) be told with digital resources, and by whom?







INTERNATIONAL JOURNAL OF HERITAGE STUDIES
2019, VOL. 25, NO. 11, 1148–1168
https://doi.org/10.1080/13527258.2019.1570964

Routledge
Taylor & Francis Group
Check for updates

Critical heritage studies and the legacies of the late-twentieth century heritage canon

Kynan Gentry[a] and Laurajane Smith[b]

[a]School of Social Sciences, University of Western Australia, Perth, Australia; [b]Centre for Heritage and Museum Studies, The Australian National University, Canberra, Australia

ABSTRACT
In recent years an interest in 'critical heritage studies' (CHS) has grown significantly – its differentiation from 'heritage studies' rests on its

ARTICLE HISTORY
Received 8 August 2018
Accepted 12 January 2019

Digital approaches to historical data absences:

- Network analysis
- Visualization tools
- Digitization and databasing of historical sources
- Topic modelling
- Textmining
- Bibliometrics

JOURNAL ARTICLE

Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive[1]

Alexandra Ortolja-Baird, Julianne Nyhan

Digital Scholarship in the Humanities, Volume 37, Issue 3, September 2022, Pages 844–867, https://doi-org.proxy.bnl.lu/10.1093/llc/fqab065

Published: 19 October 2021

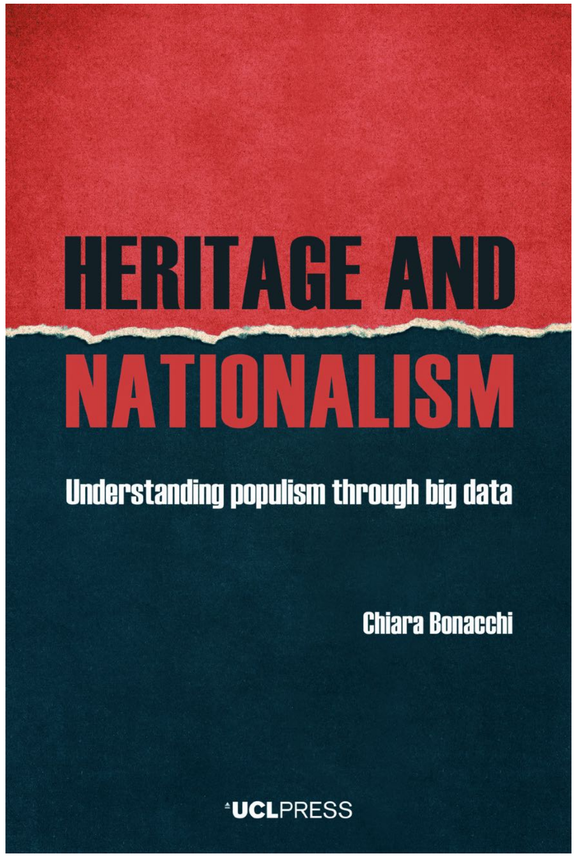
PDF Split View Cite Permissions Share

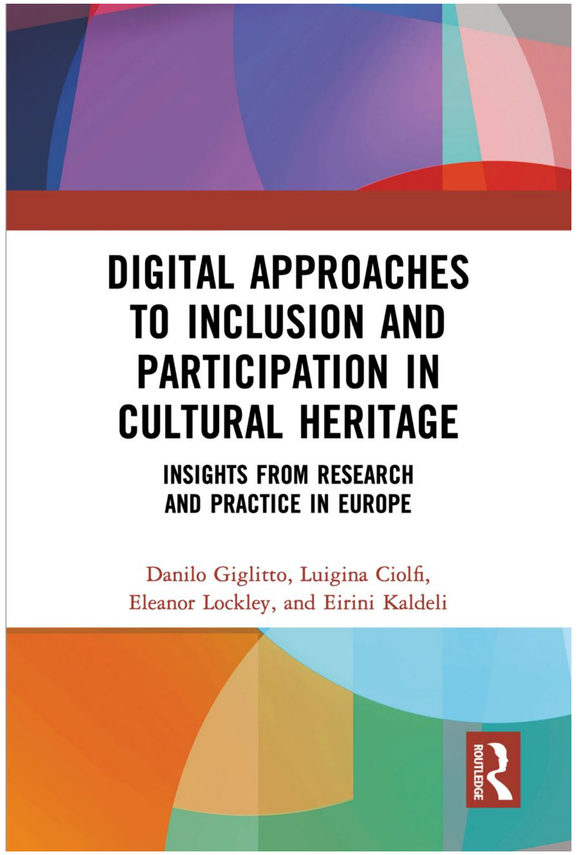
Abstract

The subjectivities that shape data collection and management have received extensive criticism, especially with regards to the digitization projects and digital archives of galleries, libraries, archives and museums (GLAM institutions). The role of digital methods for recovering data absences is increasingly receiving attention too. Conceptualizing the absence of non-hegemonic individuals from the catalogues of Sir Hans Sloane as an instance of textual haunting, this article will ask: to what extent do data-driven approaches further entrench archival absences and silences? Can digital approaches be used to highlight or recover absent data? This article will give a decisive overview of relevant literature and projects so as to examine how digital tools are being realigned to recover, or more modestly acknowledge, the vast, undocumented network of individuals who have been omitted from canonical histories. Drawing on the example of Sloane, this article will reiterate the importance of a

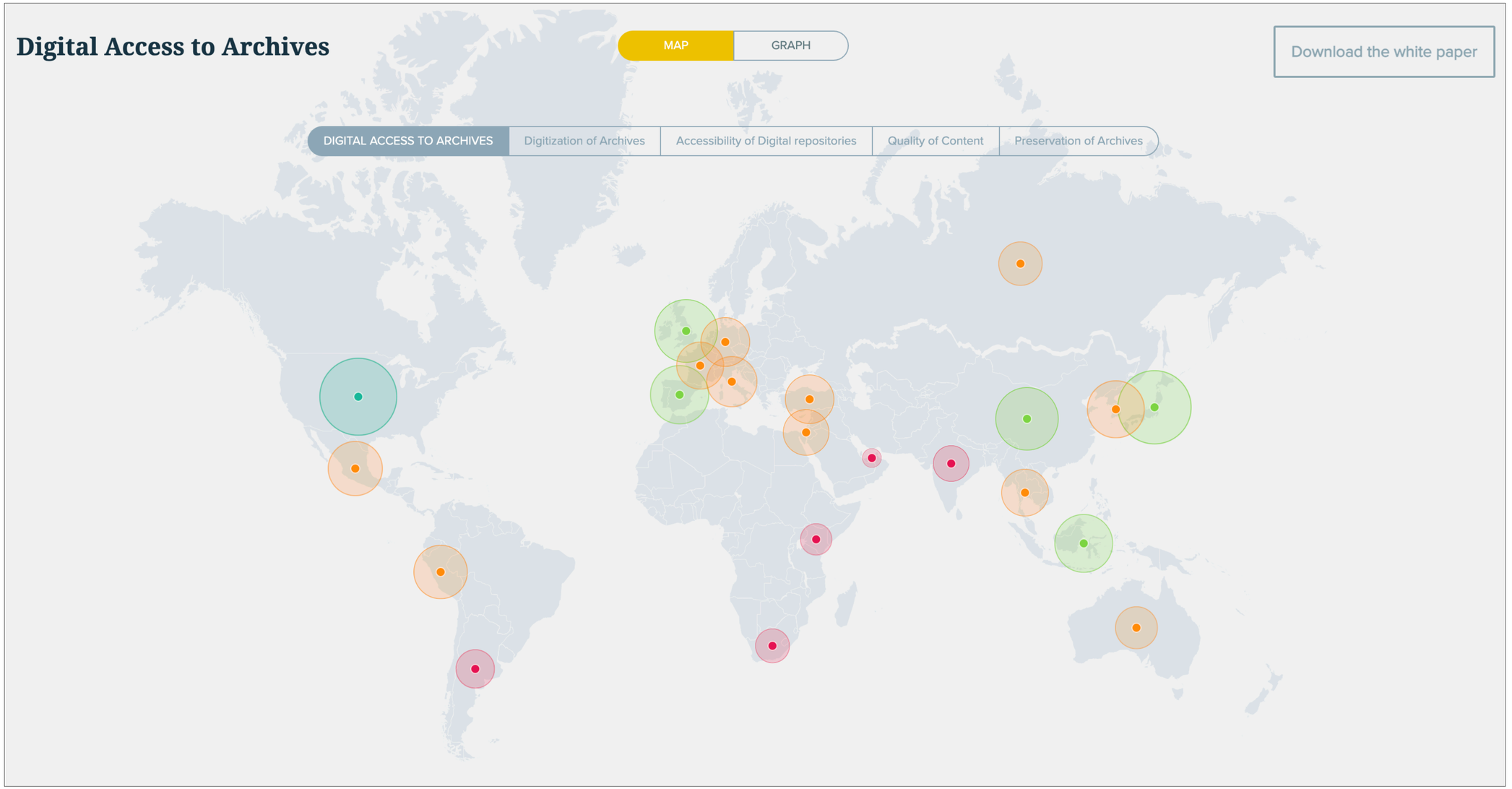
The politics of (digital) cultural heritage

- Digitisation: transcending the nation?
- The nation still looms large in the background.
- The construction of the (supra-)nation increasingly takes place online and so does the struggle for the past.
- How does this play out globally: Global North & South and beyond.
- What does this mean for the heritage of minorities and indigenous peoples.

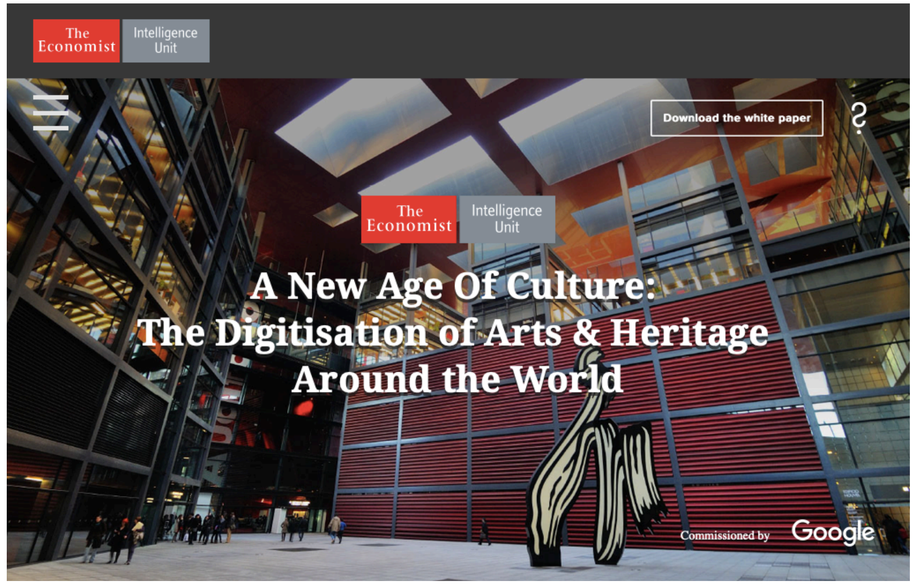




The state of digitisation: globally



The State of Digitisation: Globally



Source: Economist Impact.

Cultural Digitisation Scorecard

	1. Website assessment	2. Social media presence	3.Interactive experience capacity	4. Digital access to archives	5. Digital education initiativeas
Argentina	56.7	33.5	31	27.9	15.5
Australia	71.7	64.2	28.8	32.1	19.5
China	62.7	35.9	35.3	56.4	26.2
France	80.8	64.2	49.2	38	5.2
Germany	76.2	44.4	42.7	40.2	30
India	51.2	28.2	32.3	24.7	21.4
Indonesia	42.7	36.4	17.5	50.8	9.5
Israel	61	37.3	35.9	36.5	30
Italy	68.6	54.2	60.3	41.8	16.7
Japan	80.1	27.5	42.6	69.1	36.4
Kenya	51.8	20.4	11.3	19.5	15.6
Mexico	52.4	46.4	48	46.5	42.9
Peru	61.4	44.4	38.1	45.7	39.7
Russia	60.5	52.3	38.6	33.9	19
South Africa	59.2	45	36.1	22.6	35.1
South Korea	58.2	32	50	49.9	21.4
Spain	68	64.1	60.6	51.4	45.5
Thailand	54.7	21.6	31.8	37.8	14.3
Turkey	59.5	23	29	39.9	12.9
United Arab Emirates	67.6	58	27.9	4.7	7.9
United Kingdom	79.1	85.7	51.3	56.6	62.3
United States of America	78.6	82.7	53.7	73.4	49

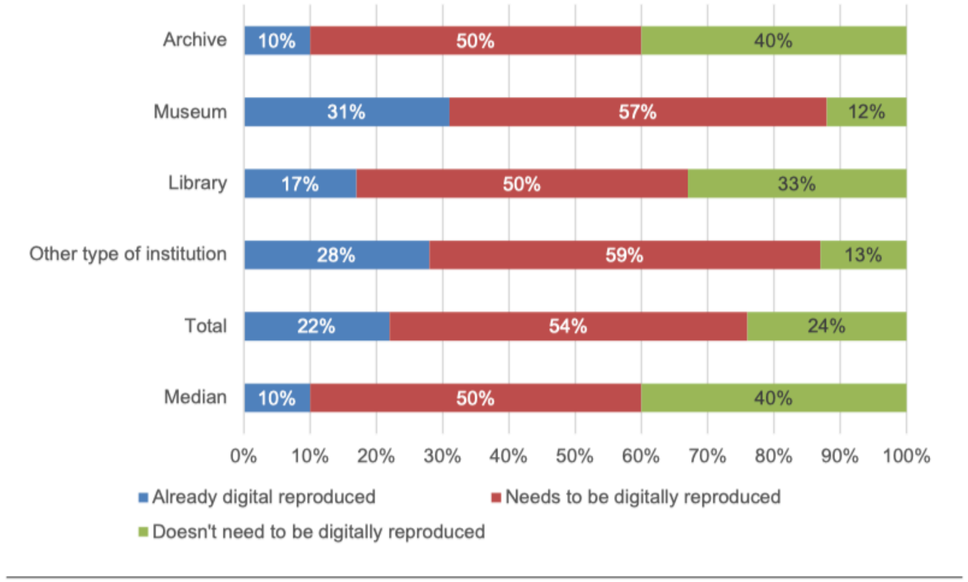
Note: all scores are on a 0-100 scale. Includes mobile.

Legend: Advanced (71-100) Progressing (51-70) Emerging (31-50) Nascent (0-30)

The State of Digitisation: Europe

- Numeric and eNumerate surveys conducted 2007 - 2017
- Note: much of our cultural record remains **offline**.
- Note: a **lack of online cataloguing** (only 58% of the heritage collections has been catalogued in a collection database).
- Access to cultural heritage in the digital age: not just about digitisation but starts with cataloguing and access to descriptive metadata.

Figure 3.7: Estimated percentage of analogue collection that has been digitally reproduced (n=757) or still needs to be reproduced (n=765)



Source: Gerhard Jan Nauta, Wietske van den Heuvel and Stephanie Teunisse, *Europeana DSI 2– Access to Digital Resources of European Heritage* (DEN Foundation: The Hague, 31.08.2017). See: https://pro.europeana.eu/page/enumerate.

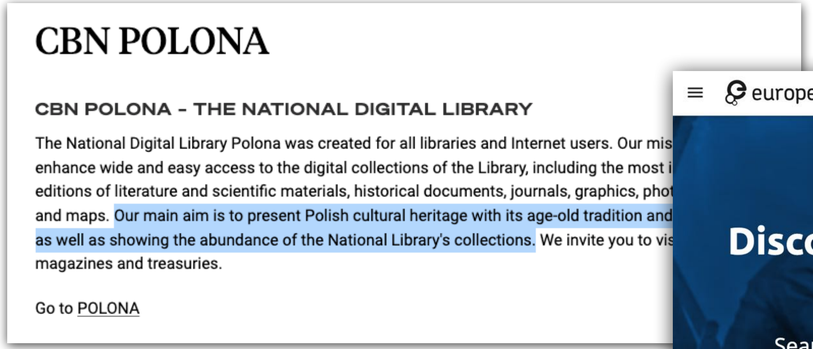
Political dimensions of online archival and curatorial work

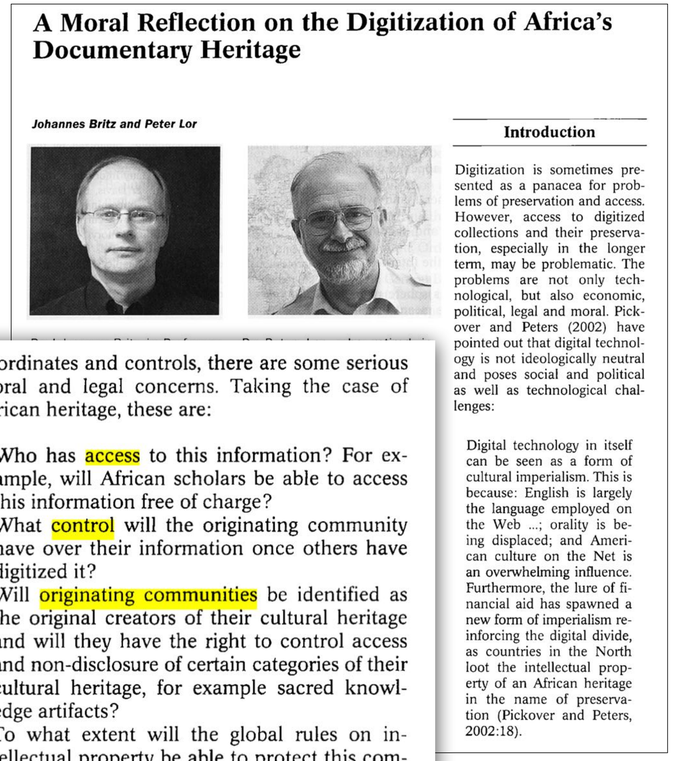
Analog dimensions [Brown & Brown 1998]	Digital equivalents
Collections are allocated to different depositories, libraries, or archives in the name of efficiency in avoiding redundancy = allocation of control.	**Which institutions digitise and control digital collections?**
Collection development refers to decisions concerning what is and what is not collected, what is merely stored but not catalogued (and hence made intellectually accessible), and what is thrown.	**What is digitised and why? What is metadated and/or OCR/HTR'ed?**
Cataloging and classification refer to the organizational and intellectual description of what is held. Whose schema will be used?	**How is it classified and how is it metadated?**
Circulation and access refers to decisions about who gets to see what, and this is shaped in part by the classification system or categorical order.	**How is access provided and mediated?**

Richard Brown and Beth Davis-Brown, "The Making of Memory: The Politics of Archives, Libraries and Museums in the Construction of National Consciousness", *History of the Human Sciences* 11, no. 4 (1998): 17–32.

Political Dimensions: Allocation

Which institutions digitise and control digital collections?




A Moral Reflection on the Digitization of Africa's Documentary Heritage

Johannes Britz and Peter Lor

Introduction

Digitization is sometimes presented as a panacea for problems of preservation and access. However, access to digitized collections and their preservation, especially in the longer term, may be problematic. The problems are not only technological, but also economic, political, legal and moral. Pickover and Peters (2002) have pointed out that digital technology is not ideologically neutral and poses social and political as well as technological challenges:

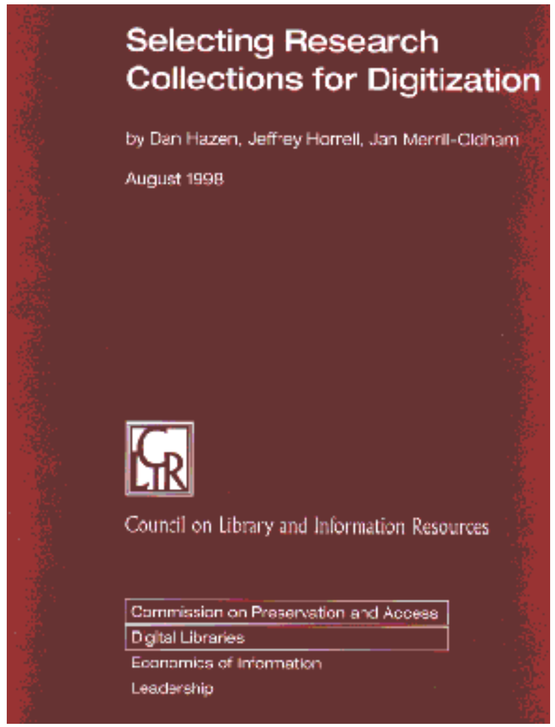
> Digital technology in itself can be seen as a form of cultural imperialism. This is because: English is largely the language employed on the Web ...; orality is being displaced; and American culture on the Net is an overwhelming influence. Furthermore, the lure of financial aid has spawned a new form of imperialism reinforcing the digital divide, as countries in the North loot the intellectual property of an African heritage in the name of preservation (Pickover and Peters, 2002:18).

coordinates and controls, there are some serious moral and legal concerns. Taking the case of African heritage, these are:

- Who has access to this information? For example, will African scholars be able to access this information free of charge?
- What control will the originating community have over their information once others have digitized it?
- Will originating communities be identified as the original creators of their cultural heritage and will they have the right to control access and non-disclosure of certain categories of their cultural heritage, for example sacred knowledge artifacts?
- To what extent will the global rules on intellectual property be able to protect this common heritage of Africa and prevent it from becoming exclusive, private property? Will the international intellectual property regimes be able to maintain the balance between private ownership and common heritage of the people of Africa?
- Will the people of Africa be fairly compensated for the use of their knowledge by others and what incentives will there be for them to make their body of knowledge available to the rest of the world?

Political Dimensions: Selection

What is digitised and why? What is metadated and/or OCR/HTR'ed?



Political Dimensions: Classification

How is it classified and how is it metadated?

Journal of the Society of Archivists
Vol. 32, No. 2, October 2011, 173–189

Routledge
Taylor & Francis Group

‘We Are What We Keep; We Keep What We Are’: Archival Appraisal Past, Present and Future

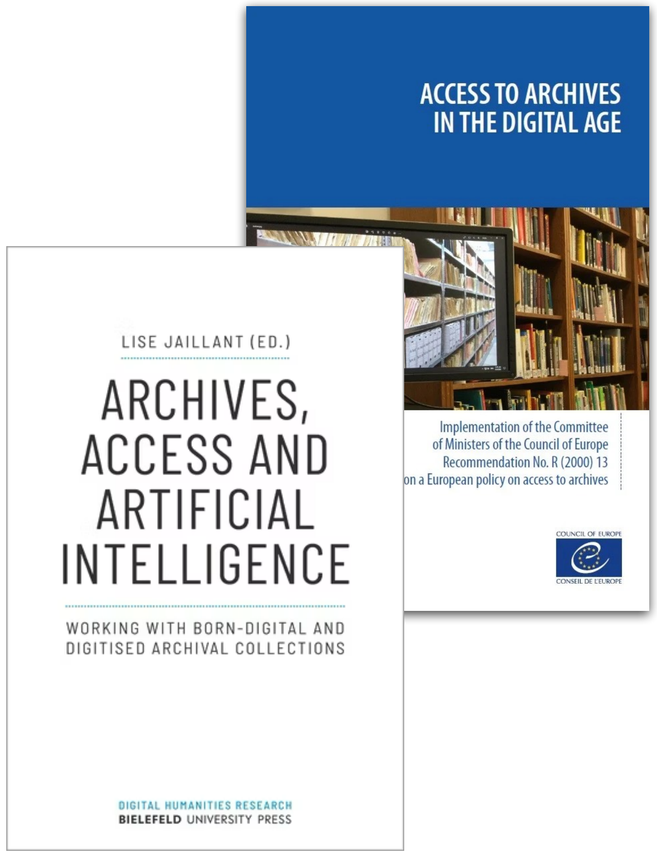
Terry Cook

Archival appraisal has its own history and is highly contested ground within the profession, and increasingly with our external communities. This article analyses the evolution of appraisal thinking through three well-established phases: the curatorial guardian assigning appraisal responsibility to the creator or administrator of records; the historian-archivist making appraisal decisions indirectly through the filter of trends in academic History; and the archivist as expert directly assessing contexts of function and activity to discern appraisal value. A fourth phase is now beckoning: participatory appraisal with various communities of citizens so that silences long haunting our archives may at last be heard.

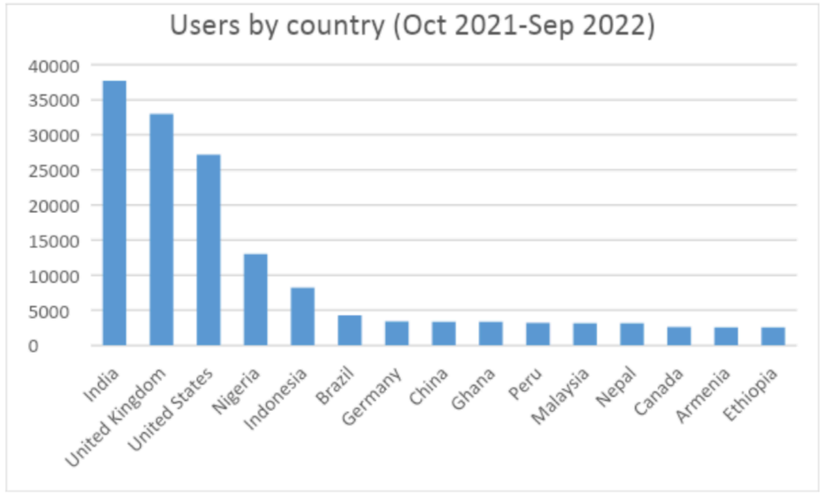
Kā Huru Manu https://www.kahurumanu.co.nz/

Political Dimensions: Access

How is access provided and mediated?



MUKURTU



Endangered Archives Programme

BRITISH LIBRARY

Annual Report

Phase 2, Year 4

1 October 2021 to 30 September 2022

Political Dimensions: Access

What is FAIR DATA?

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE

Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

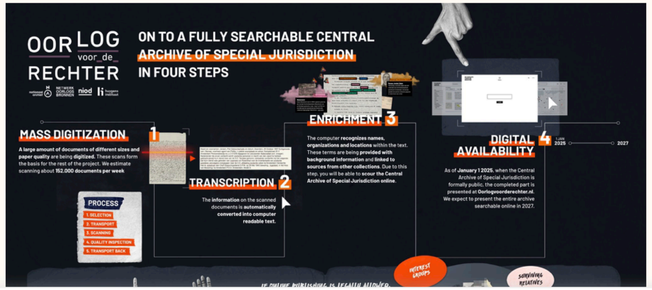
CARE Principles for Indigenous Data Governance

Political Dimensions: Access

ABOUT OORLOG VOOR DE RECHTER

NL

The Central Archives of the Special Jurisdiction (CABR) is one of the most important archives on WWII. Because this archive contains numerous special personal data, it is open to a limited extent on the basis of the Dutch Archives Act. This means that it can only be viewed under certain conditions. Nevertheless, the CABR is the most consulted war archives in the Netherlands. This public access restriction expires i
the Second World War. The youngest defendants wi
that very few of them will still be alive in 2025.



In an
Archi
Holoc
Cultu
Juris

We ar
conte
with
same

Wij gebruiken cookies

Wij plaatsen functionele cookies om deze website naar behoren te laten functioneren en
Meer informatie vind je in onze **cookie-** en **privacyverklaring**.

Political dimensions of online archival and curatorial work

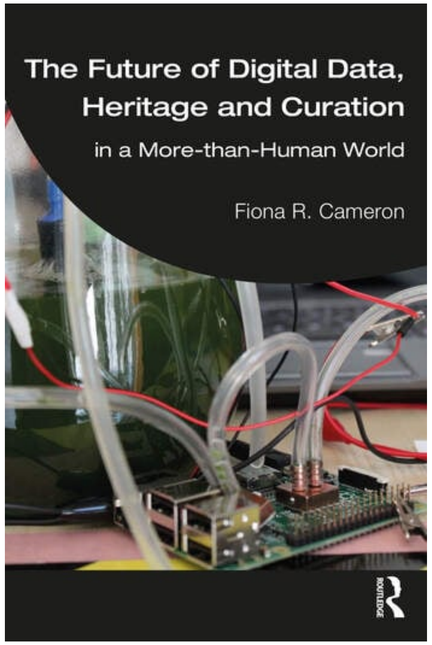
Analog dimensions [Brown & Brown 1998]	Digital equivalents
Collections are allocated to different depositories, libraries, or archives in the name of efficiency in avoiding redundancy = allocation of control.	**Which institutions digitise and control digital collections?**
Collection development refers to decisions concerning what is and what is not collected, what is merely stored but not catalogued (and hence made intellectually accessible), and what is thrown.	**What is digitised and why? What is metadated and/or OCR/HTR'ed?**
Cataloging and classification refer to the organizational and intellectual description of what is held. Whose schema will be used?	**How is it classified and how is it metadated?**
Circulation and access refers to decisions about who gets to see what, and this is shaped in part by the classification system or categorical order.	**How is access provided and mediated?**

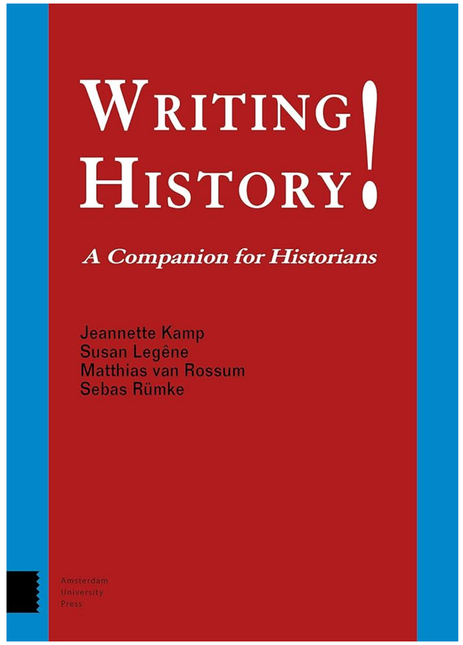
Richard Brown and Beth Davis-Brown, "The Making of Memory: The Politics of Archives, Libraries and Museums in the Construction of National Consciousness", *History of the Human Sciences* 11, no. 4 (1998): 17–32.

Concluding remarks

Some considerations for digital archives:

- **Engage** a broad range of stakeholders in development: heritage professionals: archivists, librarians, information specialists, historians/humanists, publics.
- Address **biases** through co-design of interfaces, co-development of classifications & metadata, crowdsourcing, selection for digitisation.
- **Transparancy**: include information on funding, selection criteria, metadata, access, interface and search options, links to relevant “offline” sources.
- **Education**: promoting digital literacy.
- Rethink and reevaluate purposes and uses of digitisation, and the **value of the non-digital** in digital humanities practices> think complementary not either/or.





Thank you for your attention!

M: gerben.zaagsma@uni.lu

W: gerbenzaagsma.org

Article (Scientific journals)

Facing the History Machine: Towards Histories of Digital History

ZAAGSMA, Gerben

In press • In *History of Humanities*

Peer Reviewed verified by ORBi

Permalink
https://hdl.handle.net/10993/61381

Top: https://www.journals.uchicago.edu/doi/10.1086/731827.
Right: https://doi.org/10.1093/llc/fqac050.

JOURNAL ARTICLE

Digital History and the Politics of Digitization

Gerben Zaagsma

Digital Scholarship in the Humanities, Volume 38, Issue 2, June 2023, Pages 830–851,
https://doi.org/10.1093/llc/fqac050
Published: 16 September 2022

PDF Split View Cite Permissions Share ▾

Abstract

Much has been made in recent years of the transformative potential of digital resources and historical data for historical research. Historians seem to be flooded with retro-digitized and born-digital materials and tend to take these for granted, grateful for the opportunities they afford. In a research environment that increasingly privileges what is available online, the questions of why, where, and how we can access what we can access, and how it affects historical research have become ever more urgent. This article proposes a framework through which to contextualize the politics of (digital) heritage preservation, and a model to analyse its most important political dimensions, drawing upon literature from the digital humanities and history as well as archival, library, and information science. The first part will outline the global dimensions of the politics of digital cultural heritage, focusing on developments between and within the Global North and South, framed within the broader context of the politics of heritage and its preservation. The second part surveys the history and current state of digitization and offers a structured analysis of the process of digitization and its political dimensions. Choices and decisions about selection for digitization, how to catalogue, classify, and what metadata to add are all political in nature and have political consequences, and the same is true for access. The article concludes with several recommendations and a plea to acknowledge the importance of digital cataloguing in enabling access to the global human record.

Issue Section: Article