

The digital cliff of death and the digital dark decades: the challenges for teaching and researching contemporary history

Eva Pfanzer

Department of Contemporary History
University of Innsbruck

Vienna, 24 November 2021

Digital Humanities & interdisciplinarity

- different from traditional Humanities, digitization produces DATA:
 - different approaches (eg., what, who, how accessible)
 - different tools (eg., algorithms, geo-visualizations)
 - different methods (eg., “Big Data”-analyses, text mining)
 - ▶ different questions?
 - ▶ different disciplinary skills?
- DATA usually = interdisciplinary collaboration:
 - humanities & ...
 - different disciplinary approaches:
 - ▶ libraries: standards, specifications, material, users/audience
 - ▶ humanities: workflow from selection to interpretation of specific / individual material
 - ▶ computer science: automate ‘distinctive small steps’

DATA from a modern/contemporary historian's perspective

- **eg., newspapers in modern/contemporary history**
 - large number of discoveries and accelerated industrialization
 - ~1800 to today
 - newspapers develop from chronicles to opinion-forming press organs
 - distribution on a large scale all over Europe simultaneously
 - ▶ ideal for digitization efforts
 - ▶ copyright-free data older than 100/70 years (moving walls)

contemporary history				modern history			
2020	2000	1990	1950	1900	1850	1800	
digital born / digitally created / digitally archived		not digitized / not free of copyright / paywalls		partially digitized / accessible depending on nationality		digitized / copyright-free	

1950 as the Digital Cliff of Death?

Digital Dark Decades

- **serious challenges = Digital Dark Decades**
 - no historical (newspaper) data
 - copyright & privacy issues
 - ▶ digitization efforts not feasible for libraries / archives
 - ▶ archives of digital data ≠ archives of digitized data

contemporary history				modern history		
2020	2010	1990	1950	1900	1850	1800
digital born / digitally created / digitally archived	/	not digitized / not free of copyright / paywalls	partially digitized / not possible depending on nationality	digitized / copyright-free		

Digital Dark Decades

Terry Kuny, "A Digital Dark Ages?
Challenges in the Preservation of
Electronic Prevalent Information, in:
63RD IFLA, Sep 1997

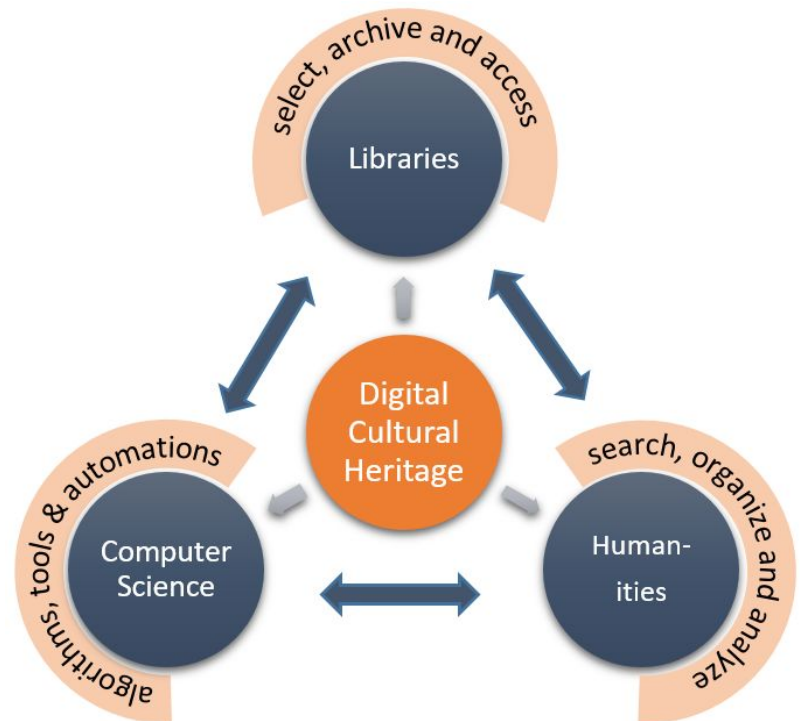
Importance for research and teaching?

- **Time-period in question, some examples**
 - 1945 - 1989: Cold War
 - 1957 - 1993/- today: European Communities / European Union
 - Social Movements:
 - ▶ “1968”
 - ▶ Civil Rights Movement
 - ▶ Second Feminist Movement
 - ▶ Labour & Cooperative Movement
 - ▶ Anti-nuclear Movement
 - ▶ Environmental Movement
 - ▶ Human Rights Movement
 - currently topical: mandatory vaccinations in eur. countries (eg., smallpox ~ 1948; polio late 1970s)
 - **Solutions??**

1. Interdisciplinarity

from NewsEye project:

- libraries: critically question digitization processes, communicate transparently (metadata, contexts, educational material)
- computer science: discover automation opportunities for a broad, non-standardizable user group (“diverse use cases and needs”)
- humanities: critical reflection on results but also on algorithms, interfaces, tools, methoden
 - ≠ not multidisciplinary approach: several disciplines work on 1 problem
 - = “something” between the disciplines that defines AND bridges disciplinary boundaries
 - collaboration: concept of integrated interdisciplinary workflows

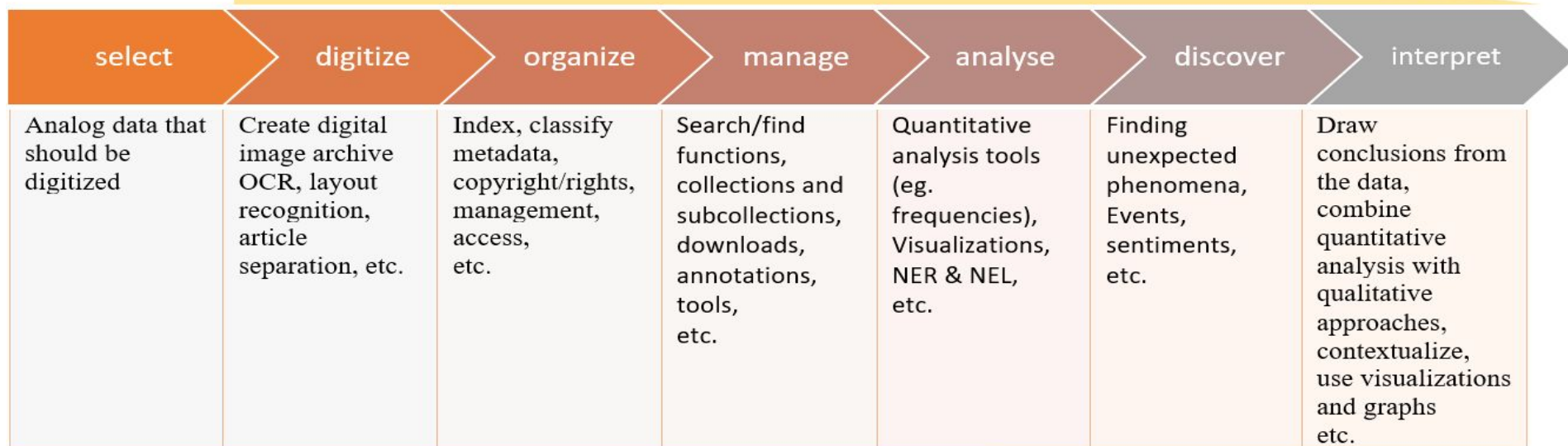


© <https://doi.org/10.1002/asi.24565>

Interdisciplinary Collaboration

- from the NewsEye project:
 - humanities (research scenarios)
 - computer science (automation)
 - libraries (data)

BUT: needs and interests of DH scholars and users vary
→ define workflows



Digital Cultural Heritage
Curators/Libraries

(Digital) Humanities

2. Collaboration on the meso-level

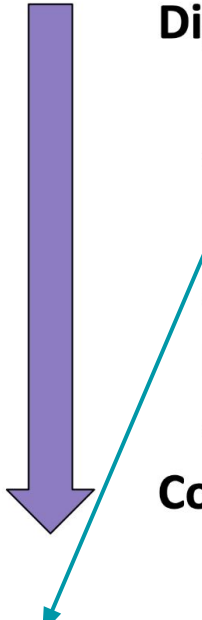
Digital History

Micro: individual research, smaller-scale, close reading, traditional digital library-based platform, e.g. newspaper portal

Meso: collaborative research, medium-scale, mixed methods, corpus-based, some-level of computational analysis, e.g. notebooks

Macro: collaborative research, larger-scale, focus on computational analysis, use of High Performance Computing

Computational History

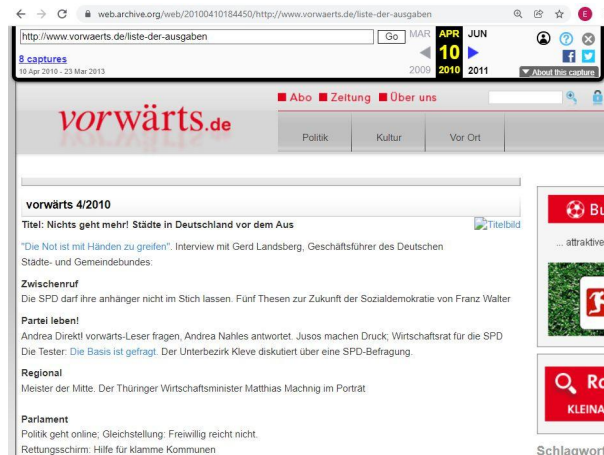
- 
- research tailored corpora
 - data / tools / methods that bridge the gap between 1950 and 1990
 - collaborate with (external) data providers
 - align digitized and digital data

© Sally Chambers

Collaboration on the meso-level

Example “Labour Movements in Austria”

- Arbeiter-Zeitung 1889-1985
 - official newspaper of Austrian Social Democracy (incl. exile press)
 - ANNO: 1889-1936 (including some exile press)
- “private” initiative: <http://www.arbeiter-zeitung.at/> 1945-1989 (offline since 2021, flash & html-access, no db-access via wayback)
- regional newspaper: Burgenländische Freiheit 1922-2007 (registered users)
- Austrian press agency news: “APA historisch” (offline since 2019 will not be revived)
- webarchives: Vorwärts



© left: <https://web.archive.org/web/20190205084208/http://www.arbeiter-zeitung.at/>
center: <https://web.archive.org/web/20100410184450/http://www.vorwaerts.de/liste-der-ausgaben>
right: <http://www.bf-archiv.at/>

3. Open Science

Humanities:

- individually tailored corpora
- size: extremely different (several dozen articles to millions of pages “big data”)

Libraries:

- as dataproviders demands difficult to meet
- innovative approaches needed
- Labs environment as areas of trial & error

Computer Science:

- challenges remain high error rate in OCR, layout etc.
- automation is difficult with tailored meso-level corpora

→ **Open Science?** → **Open research infrastructures?** → **Open data?**

- data management plans on various levels (up to EU-level)
- collaborate with (external) data providers (eg., APA, publishers)
- legal & ethical issues; copyright / privacy right expertise
- win-win-solutions (eg., publishers offer data, libraries offer data storage, data curation)

Literatur

Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Prevention Information", in: 63RD IFLA (International Federation of Library Associations and Institutions) Council and General Conference, September 1997.

Maev Kennedy, Maev "National Archive project to avert digital dark age". *The Guardian*, 4 July 2007).

Toni Weller, *History in the Digital Age*, London 2013.

Oberbichler, Sarah/Boroş, Emanuela/Doucet, Antoine/Marjanen, Jani/Pfanzelter, Eva/Rautiainen, Juha/Toivonen, Hannu/Tolonen, Mikko, Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians, in: *J Assoc Inf Sci Technol* (2021), S. 1–15, <https://doi.org/10.1002/asi.24565>, eingesehen 19.8.2021.

Oberbichler, Sarah/Pfanzelter, Eva, Topic-specific corpus building: A step towards a representative newspaper corpus on the topic of return migration using text mining methods. (issue 1: Scalable historiography: Storytelling, Hermeneutics, Data), in: *Journal of Digital History* jd001 (2021), 1, <https://journalofdigitalhistory.org/en/article/4yxHGiqXYRbX>, eingesehen 8.11.2021.

Danke! Thank you!

Questions?