JANA KECK

How Meaningful are Digital Humanities Projects When it Comes to Training Early-Career Scholars in Digital Literacy?

Newspapers spread dramatically throughout the nineteenth century in terms of sheer numbers of papers, their size, and their influence. This created a global culture of rapidly circulating information. Fuelled by falling printing costs, new editorial exchanges, steam-driven transport, and telegraphy, newspapers linked cities and small towns to a global network. These periodicals became the first big data for a mass audience. Large-scale projects digitising historical newspapers run by national and local libraries and commercial companies have grown exponentially in scope, scale and ambition since the 1990s (Terras 2010). In recent years, the mass digitisation of these sources has led to new efforts in periodical studies. Technical innovations in both hardware and software have inspired scholars to seek collaborations with institutions that are digitising material and researchers from computer science who have experience in working with, linking and analysing large amounts of data. From 2017 to 2019, I was working as a doctoral researcher in a digital humanities project titled *Oceanic Exchanges: Tracing Global Information Networks in Historic Newspaper Repositories, 1840–1914* (Oceanic Exchanges 2017) that brought together scholars in computational periodicals research from the US, Mexico, Germany, the Netherlands, Finland, and the UK. Large-scale digitised newspaper repositories from Europe to Australia, from public to private institutions, were collected to study how information spread across national and linguistic borders in the nineteenth century.

'Information flow' has historical resonances and contemporary saliency to networked collections of digital materials. The objective of *Oceanic Exchanges* was to figure out how to compare digitised newspapers from different national and local repositories that were created using different digitisation software and that follow distinct metadata guidelines. In order to detect and study media events in the nineteenth century that had a global impact, we needed to develop innovative *data modelling*, *linking* and *text mining* methods. Once we identified these events, we used and merged diverse methods from computer science and the humanities, including *interactive visualisation systems*, *network analysis,* and *close reading* to examine how these events were textualised in different nations and how information that was political, scientific, literary, or religious in nature changed through its travel in time and space. To find solutions to these research questions, scholars from different academic, public, and commercial institutions, and diverse disciplinary, national, and linguistic backgrounds, collaborated to develop innovative computational methods

© 2024 Jana Keck (CC BY-SA 4.0). Published in: Carsten Wergin and Stefanie Affeldt (Eds.),
Digitising Heritage. Transoceanic Connections between Australia and Europe. Heidelberg 2024, 175–185.
DOI: https://doi.org/10.17885/heiup.1305.c18426

for the analysis of big humanities data. Interdisciplinary cooperation fostered the development of novel methodological, theoretical, and practical approaches relevant for scholarship in *archival studies*, *transnational history*, *computational linguistics*, and *information visualisation*. The collaborative efforts between experienced and newer-to-their-fields researchers resulted in various project outcomes such as datasets, ontologies, algorithms and implementations, documentation, and articles that focus on questions about media history, as well as technical, political, and ethical challenges of digitisation.[1]

For early-career scholars, doing historical research in the digital age requires training in digital literacy. In this article, I want to illustrate how *Oceanic Exchanges* offered an international and inter- or transdisciplinary training environment for doctoral researchers to prepare them for career paths in and outside of academia in the digital age. This article can be seen as a report, in which I share my own experiences as a doctoral researcher with a background in linguistics and American Studies. I will share examples in producing some of the above-listed project outcomes: from evaluating the interoperability of digital archives to developing an *interactive visualisation system* to analyse the geographical, temporal, and spatial dissemination of news. *Oceanic Exchanges* provided us – and by us, I am referring to doctoral researchers from the humanities and computer sciences – with the opportunity to acquire skills in digital literacy and knowledge about the legal and ethical status of digitisation, and to build networks across disciplines and research institutions. By witnessing the operations of large-scale digital research and digitisation projects, I not only gained experience in project management, but also increased my consciousness of intercultural communication by learning about the different political, economic, and social factors that shape digital humanities work in different locations. By providing examples of my own research tasks, I want to show how *Oceanic Exchanges* turned into a fruitful learning platform for global perspectives on data, code, and tool criticism. These skills seem increasingly relevant in times of the digital transformation of society, the disciplines needed to be able to study large amounts of digitised or born-digital material, and the influence of algorithms in finding information online. Additionally, these collaborations and learned skills have encouraged me to consider other career options in research management, communication, or environments outside of academia. Being mentored on different job opportunities is important in the early-career stage because not every doctoral student will receive a professorship in the future, and not everyone aspires

[1] For a current state of the various project outcomes, see: https://oceanicexchanges.org/outcomes/ (accessed 24/8/2023). Some articles that focus on media history are currently under review or will be published soon. One of the reasons for this “delay” is that we had to make sure that the novel methodological approaches or innovative tools are first published in computational journals in order to support the careers of young scholars, before we can use them for the publication in journal of the humanities disciplines.

to such a position.[2] To make sure that early-career scholars get the necessary education to prepare them for job markets in the digital age, we need to change evaluation and review systems, accept and reward diverse forms of publication within disciplines, and normalise inter- and transdisciplinary scholarship. If these measures do not find wider acceptance and implementation, it will be detrimental especially to the education of up-and-coming scholars (KÖNIG 2020).

The digital universe is growing very fast, and the amount of information contained in digital representations of historic newspapers as data is exploding. How can early-career scholars receive training in data literacy in the humanities? Working in *Oceanic Exchanges* has shown me that one way of learning how to collect, manage, evaluate, and apply data in a critical manner (RISDALE et al. 2015) is to make sure that early-career scholars get directly in touch with institutions that are digitising sources. The project gave me the opportunity to work with institutions from the following public, private, and public-private sectors: Australia's Trove, the British Newspaper Archive, Chronicling America, Europeana Newspapers, the Hemeroteca Nacional Digital de México, the National Library of Finland, the National Library of the Netherlands, the National Library of Wales, Aotearoa New Zealand's Papers Past, and Cengage Publishing, one of the leading commercial custodians of digitised newspapers. Bringing together the digitised newspaper repositories of these different institutions resulted in a collection of more than one hundred million newspaper pages. I received information about which technologies have been used in different nations to digitise heritage; they told me about their own technical, political, or ethical challenges, and what they envision for the future of archival studies in a digital age.

To investigate the different national corpora, to explore both historical content and modern data storage and linkages, we first had to examine the data structure and develop "more nuanced understandings of (digital) archives as both sources and subjects of history" (HAUSDEWELL et al. 2020, 146). Digital humanities projects do not provide quick results by simply pushing on a button. What I had to learn to start with was that one cannot simply use, link, and mine these distinct data repositories to ask historical questions about the press system in the nineteenth century. When these institutions turned analogue sources into digital representations, they used very different digitisation programs and different metadata guidelines. For this purpose, we had to examine which digitisation software each had used, and for which reasons. Additionally, we found out that selection criteria differ among institutions, and range from prioritising the digitisation of the most vulnerable objects to providing a more diverse selection of languages or publication locations and offering content from urban and rural places. We documented these

[2] For a similar advocacy of hands-on training in academic collections, see Polly LOHMANN: *Digitising from Scratch*, pp. 103–121 in this volume.

findings in teams.[3] Apart from gathering this information for our own research, the purpose was to formulate recommendations for the institutions. Even though this process was extensive and time-consuming, the development of a shared and source-specific ontology for describing the form and content of nineteenth-century newspapers is paramount to the future integration and linking of distinct collections. Bringing together these different findings resulted in an open access guide titled *The Atlas of Digitised Newspapers and Metadata* (Beals et al. 2020). The guide brings together the technical and political histories of the individual databases that will allow other researchers and institutions insights into their digitisation choices, with a deeper look at the language of the digitised newspaper, the variety of newspaper terminologies, and the metadata inherent in these collections. *Oceanic Exchanges* did not aim to create a totalising research infrastructure, but rather to expose the conditions by which researchers can work across collections, helping guide similar projects in the future seeking to bridge national collections.

Creating this guide required intensive dialogue between researchers to discuss the current state of digital archives and above all their future as democratic knowledge infrastructures. As it turns out, only few digitisers have integrated scholarly expertise into their decision-making processes of digitising heritage so far. It seems that they only bring in academic researchers when they themselves have identified them as core users. The Koninklijke Bibliotheck (the National Library of the Netherlands) turned out to be the only institution that has established a committee of experts from periodical research (Koninklijke Bibliotheck 2008). The German Newspaper Portal (*Deutsches Zeitungsportal)* that is currently being developed as a “sub-portal” of the German Digital Library follows the example of developing an infrastructure that will meet the demands of scholars. “The initial impulse – and funding – to develop a national [German] newspaper portal was given by the scholarly community, especially the (Digital) Humanities” (Dinger/Landes 2019). My collaboration in *Oceanic Exchanges* resulted in me being asked to become a member of the German Newspaper Portal’s advisory board to provide perspectives as an early-career scholar. As members, we give feedback on questions related to the quantity and quality of the sources: for example, as a scholar, would you rather have access to more newspapers in the digital archive or would you recommend that we prioritise steps in improving the quality of digitised sources such as *Optical Character Recognition (OCR) Post-Correction.* Apart from questions about selection and quality, they ask questions about the interoperability of the interface: if we allow scholars to download sources, which data types should we make available? Should we integrate tools for advances searches and analyses? Such efforts not only allow me to critically reflect on digital practices, but also give me as an early-career

[3] I am especially grateful to Clemens Neudecker from the State Library of Berlin, who taught me about the diverse digitisation programs in Europe and helped me in writing data reports of the German-language sources in *Oceanic Exchanges*.

scholar an active role in shaping how other scholars collaborate, make discoveries, create, communicate, and publish knowledge.

Technical challenges regarding the interoperability of digital archives are not only influenced by digital technologies but often shaped by the demands of the end users. When I examined the findings of the other digital archives, I gained a different picture. According to the majority of the ten data partners, academic researchers are not the primary audience for whom these platforms are being created; "the general public" is (Hausdewell et al. 2020, 151). Some platforms, for instance, trace user behaviour and develop and adapt strategies for information search and retrieval accordingly. For scholars, this implies that the way they can access and search through digital archives is shaped by the work of software developers and archivists as much as it is influenced by the user behaviour of the general public. What happens if scholars do not get in touch with such institutions? Key principles of sustainability go beyond storage and access strategies. A continuous and mutually informing dialogue between institutions that are digitising sources and scholars from different disciplines and positions in their career is of critical importance to discussing how we can and will "move around" in digital archives. Conversations with the other data partners revealed that using digital archives for advanced research remains limited. "Whether in terms of information retrieval possibilities or interface design, providers tended to aim for simplicity" (Hausdewell et al. 2020, 154), and simplicity means that the traditional method of keyword searching "is here to stay for the foreseeable future, since the majority of users have been habituated to this mode of search" (Hausdewell et al. 2020, 155). As it is difficult, time-consuming, and cumbersome, if not impossible, to predict in advance which search terms will best identify documents, queries are typically refined through testing and sampling. Keyword searching requires the scholar to have prior knowledge and possesses the seemingly contradictory weaknesses of finding too few and too many documents. Consequently, there remains no other option to evaluate the digitised material other than simply reading it. While these other digital archives provide big humanities data, they offer no innovative methods of analysing them. "It remains to be seen whether digital archives will continue to support relatively prescribed and limited modes of searching, browsing, and viewing newspapers alongside more advanced functionalities and what divisions will arise out of different funding models" (Hausdewell et al. 2020, 156).

Working on *The Atlas* and being a member of the advisory board of the German Newspaper Portal has shown me that we need to get in touch with these institutions to have a voice in what is being digitised and how we can access and study it online. *The Atlas* aims to form the foundation of a wider mapping of collections beyond its current North Atlantic and Anglophone-Pacific focus. This continuous process is influenced by motivations and decisions about the future inclusion and exclusion of digitised newspapers in digital archives. However, to foster such efforts in the future, other forms of publication, such as guides, need to find their place in publication lists. Sustainability means meeting our own needs without compromising the ability of future generations to meet their needs. Conversations about digital project sustainability, however, usually address technical aspects such as storage, or data formats and their long-term access. What these conversations often leave out is the social resources needed to guarantee their long-term access and use. A documentation, dataset, a tool, or an interface will only be sustainable when we make sure that we develop it in a way that means it can be used for further projects and by other scholars. The digital humanities are not just a service provider for humanities scholars who want to have a dataset or analyse a digital archive.

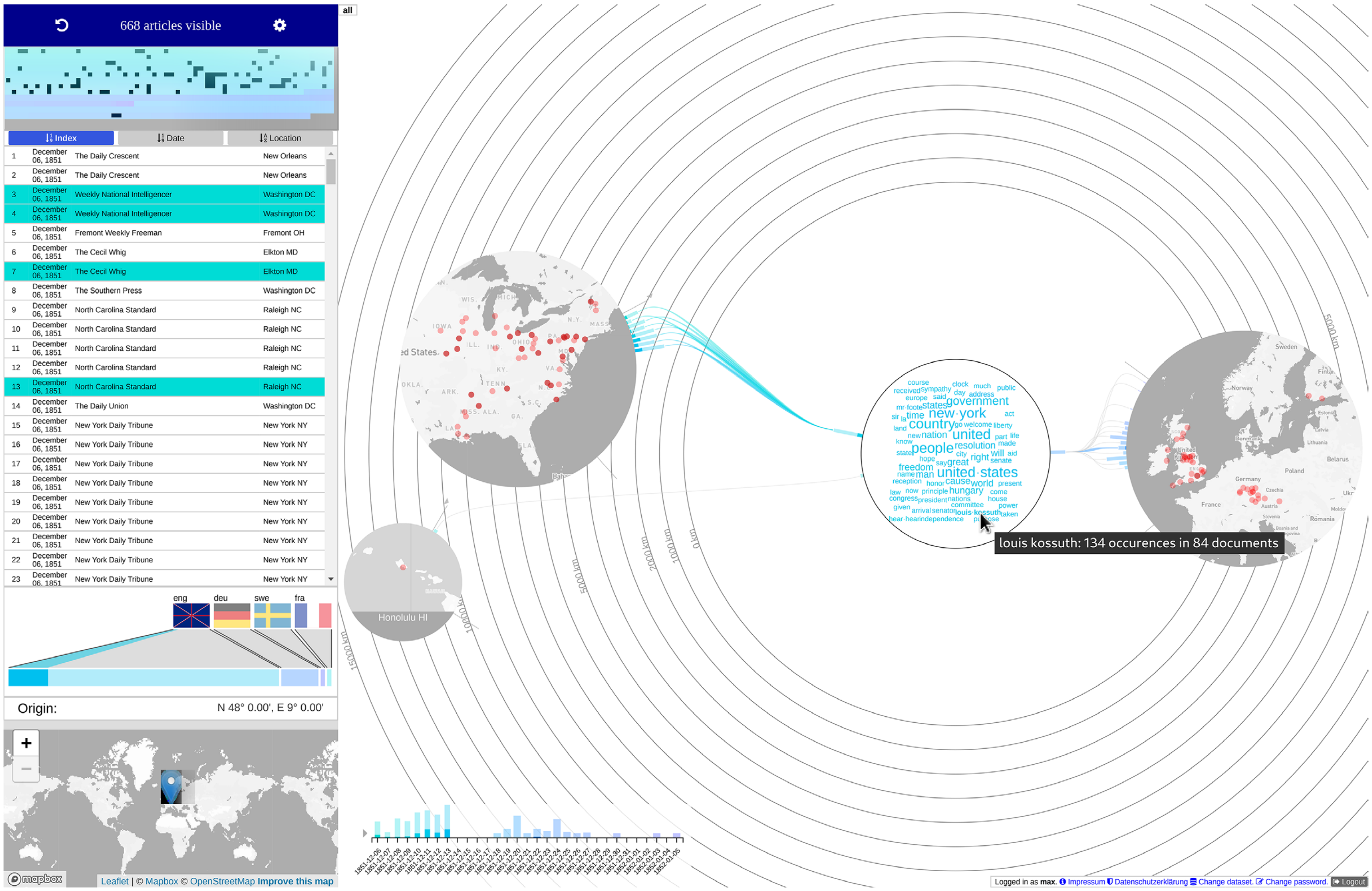
The application of digital methods and tools is not only time-consuming, it requires an adaptation of the way we think about scholarly work, project outcomes, and publications in the humanities in the digital age. The media change in the digital age presupposes that we rethink not only how we interpret sources as data, how we search and analyse, but also how we present, publish, and evaluate other forms of research output. These activities also presuppose that we offer early-career scholars training in digital literacy, and not only in digital humanities departments. Such undertakings could range from motivating early-career scholars to contact libraries or archives that are digitising material, promote other forms of research outputs than publication of peer-reviewed articles in humanities journals, such as documentations, datasets or code, articles in computer science journals, and even PhD projects that focus primarily on the histories of digitised archives. Using – that is, learning to understand – existing software is a form of training. Collectively designing and producing new technology is a challenging, but more productive and sustainable process for both the humanities and computer science.

To illustrate why such inter- or transdisciplinary research activities could be productive and meaningful, but has so far become problematic for the careers of early-career scholars, I want to give an example of a tool that I developed with computer scientists from the field of *information visualisation*. Interdisciplinary collaboration brought to the fore the very dilemma of different practical and strategic assumptions about what is considered and valued as an important and relevant project outcome in the humanities versus computer science. The specific challenges we had when it came to tools was to develop a system that would allow scholars in the humanities to study global media events. At the same time, these tools had to be cutting-edge in computer science, because they were primarily being developed by early-career scholars. Developing a novel algorithm in computer science does not necessarily mean that that algorithm will be useful for the analysis of historical questions. One of the tools that I collaboratively developed with computer scientists to systematically explore datasets of media events and study what information spread, how, and how it differed among nations and languages is Lilypads (Franke et al. 2020). Interactive visualisation tools like Lilypads are based on an algorithmic transformation of existing data into an interactive visual representation. To construct this transformation process, we had to discuss together which data (both textual and numeric) should be visualised to answer such questions about information flow in the nineteenth century. Lilypads presents a novel, integrated interactive visualisation approach that supports brushing, linking, filtering, and drill-down that enable the exchange of ideas, hypotheses, and results with other researchers. The dataset used for Lilypads, as illustrated in Fig. 1, covered sources from different languages, places, and dates.

It included metadata (i.e. structured data such as newspaper title, place or date of publication) and data (i.e. newspaper texts as unstructured data in the form of plain text). It allowed distant and close reading approaches to studying data, with the aim not of replacing textual evidence with graphs, maps, or trees, but of providing numerical, textual, and visual representations of data to uncover and model new sets of evidence that is difficult to discern at the level of the individual newspaper. If users want to read the newspaper article, they can click on one of the news items (list on the left side) and open up a new tab to read the text, and also find a link to the digital archive. When users click on the link, they get redirected to the digital archive where they can examine, for instance, where the news item is embedded in the overall newspaper edition or examine what adjacent articles and images.[4] Linking such tools to other databases is sustainable because it guarantees an effective interplay between different platforms. It also provides scholars with the opportunity to examine data in different contexts and representations, from scanned images of the newspaper page to a data point in time.

Lilypads should be seen as a multifaceted product: it is a prototype, a tool, a method, a non-static dataset, a datasheet, and the result of several people attempting to advance transdisciplinary research. To develop Lilypads, both the early-career scholars from computer science and I received active mentoring by senior scholars from other disciplines. Lilypads is now being used to analyse global media events, with the findings being transformed into articles for journals in the humanities to show how dissemination, disinformation and censorship developed in the nineteenth century. Lilypads was presented at the 11th International Conference on Information Visualization Theory and Applications (2020), published in the conference proceedings, and even awarded "Best Student Paper". According to its reviewers, the integration of scholars from the humanities as co-authors is pivotal for the evaluation of the usefulness of such tools and its sustainability for future work. However, while quality management of cutting-edge research in computer science focuses primarily on algorithms, quality management in the humanities demands the publication of articles in prestigious peer-reviewed articles. Basically, neither disciplines either values or rewards the careful creation and curation of datasets. However, this form of research is a highly intellectual process that has to be reflected and documented on, and it plays a crucial role in determining what data we analyse. Non-printable products such as curated datasets and the well-documented sheets

[4] For a detailed explanation of the interoperability of Lilypads, see FRANKE et al. (2020).



1	December 06, 1851	The Daily Crescent	New Orleans
2	December 06, 1851	The Daily Crescent	New Orleans
3	December 06, 1851	Weekly National Intelligencer	Washington DC
4	December 06, 1851	Weekly National Intelligencer	Washington DC
5	December 06, 1851	Fremont Weekly Freeman	Fremont OH
6	December 06, 1851	The Cecil Whig	Elkton MD
7	December 06, 1851	The Cecil Whig	Elkton MD
8	December 06, 1851	The Southern Press	Washington DC
9	December 06, 1851	North Carolina Standard	Raleigh NC
10	December 06, 1851	North Carolina Standard	Raleigh NC
11	December 06, 1851	North Carolina Standard	Raleigh NC
12	December 06, 1851	North Carolina Standard	Raleigh NC
13	December 06, 1851	North Carolina Standard	Raleigh NC
14	December 06, 1851	The Daily Union	Washington DC
15	December 06, 1851	New York Daily Tribune	New York NY
16	December 06, 1851	New York Daily Tribune	New York NY
17	December 06, 1851	New York Daily Tribune	New York NY
18	December 06, 1851	New York Daily Tribune	New York NY
19	December 06, 1851	New York Daily Tribune	New York NY
20	December 06, 1851	New York Daily Tribune	New York NY
21	December 06, 1851	New York Daily Tribune	New York NY
22	December 06, 1851	New York Daily Tribune	New York NY
23	December 06, 1851	New York Daily Tribune	New York NY

Origin: N 48° 0.00', E 9° 0.00'

louis kossuth: 134 occurences in 84 documents

668 articles visible
all
Index
Date
Location
eng
deu
swe
fra
Leaflet | © Mapbox © OpenStreetMap Improve this map
mapbox
0 km
1000 km
2000 km
5000 km
10000 km
15000 km
Honolulu HI
Sweden
Norway
Finland
Belarus
Poland
Germany
France
Czechia
Austria
Romania
Lithuania
Denmark
course clock much public received sympathy day address europe said government mr foote states sir la time new york act land country welcome liberty new nation united part life know made state people resolution city will aid hope say great right senate freedom name man united states reception honor cause world present law now principle hungary come congress president nations house given committee power arrival senator louis kossuth taken hear independence
1851-12-06
1851-12-07
1851-12-08
1851-12-09
1851-12-10
1851-12-11
1851-12-12
1851-12-13
1851-12-14
1851-12-15
1851-12-16
1851-12-17
1851-12-18
1851-12-19
1851-12-20
1851-12-21
1851-12-22
1851-12-23
1851-12-24
1851-12-25
1851-12-26
1851-12-27
1851-12-28
1851-12-29
1851-12-30
1851-12-31
1852-01-01
1852-01-02
1852-01-03
1852-01-04
1852-01-05
Logged in as max. Impressum Datenschutzerklärung Change dataset. Change password. Logout

Fig. 1: The interface of Lilypads. The main view shows the total temporal distribution and a word cloud of frequent words from the visualised documents, alongside the spatial distribution of the documents. The approach breaks up the world map into map insets containing areas of interest, which can be shown with a higher level of detail. On the left side, a document mini-map, document list, and distribution of the documents' languages are shown.

that accompany these datasets have not yet become first-priority outcomes for scholarly communities in computer science or the humanities. Giving little recognition and attention to these forms of publications seems incompatible with the demands of funding institutions that are increasingly asking scholars to make their data publicly available by default, rather than on request. This will even become a mandated requirement from some funding agencies, such as the German Research Foundation (DFG 2020). Such efforts call for innovation, sustainable research data management, and sharing, and simultaneously for the acceptance of curated datasets and publications, models, and tools as genuine accomplishments by all scholarly communities.

There is a tension between innovation and motivation, not just a disinterest. In my opinion, if we learn to better appreciate the careful creation and curation of datasets, we will reduce the risk of creating highly biased ones. If we want to implement these measures, scholars have to be rewarded for creating computational models and publishing them in scientific journals or publishing datasets as much as they are for publishing their methodological use of the tool to study the past. As the data which play a role in digital humanities are themselves multiform and comprised of, among other types, visual, audio, geospatial, temporal, and statistical data, we can even say that by studying the digital humanities, early-career scholars become meta- or transliterate, and thus able to employ their knowledge, competencies, and skills in very diverse social and professional domains. The digitisation of cultural heritage data, along with the creation of born-digitals, the increasing impact of algorithms on the way how we present, analyse and publish knowledge, needs to be integrated into critical debates in the humanities, not only the digital humanities. Likewise, this presupposes an openness and enhanced understanding about digital data, code, and technological infrastructures. Digital source criticism demands that we understand how data have been encoded, indexed, and enriched with metadata. Working in *Oceanic Exchanges* has given me the opportunity to analyse data in teams and produce code, tools, datasets and -sheets. Generating these research products, like establishing guidelines on workflows, applying project management best practices, and evaluating the interoperability of new and existing digital archives, is indispensable in such large-scale projects, but also labour-intensive and time-consuming. While my involvement in these research outputs is being recognised and rewarded by scholarly communities in computer science or archival studies, there are no clear or updated structures for evaluating these non-traditional forms in the humanities. As Dorothea Salo notes, "[t]he present system of humanist scholarly communication relies on print monographs, mostly print journals" (SALO 2020, 221). Digital humanists "shoulder the doubled research burden of writing" monographs and articles next to their non-textual research products because of "books' and articles intelligibility as research products to tenure and promotion committees" (SALO 2020, 222). These arguments reflect wider debates in the digital humanities about considering non-traditional formats. In "Data Beyond Vision," Rebecca Sutton Koeser et al. call upon scholars to ask:

"What would it look like to consider non-textual research outputs as first-order scholarly work?" (Sutton Koeser et al. 2020).

Adapting to disciplinary changes requires the consideration of non-traditional formats. "Until the humanities consciously break the hegemony and path dependency of print," Salo concludes, "digital humanists will remain alienated from the rest of the humanities, preventing the humanities from adopting open processes such as data sharing and open-access publishing. In turn, this harms the reach and sustainability of the humanities as a whole" (Salo 2020, 215). If the current academic system does not adapt its evaluation criteria and educate its early-career scholars in digital literacy – data, tool, and interface criticism – it will foster the education of a generation that is, on the one hand, unable to critically reflect on and educate students about the massive amount of data that we produce on a daily basis, from mails to social media posts to the influence of algorithms in shaping public opinion. To illustrate where I consider *Oceanic Exchanges* to be sustainable or unsustainable by using humans as resources, I have shared some experiences that I gained through my own activities in the project alongside conceptualising and writing my doctoral thesis, planning workshops, conferences and meetings, and mentoring students. I have exemplified why I consider tasks such as cooperating with digitisers sustainable – because they show scholars work opportunities outside of academia and likewise guarantee that they are taking an active part in building digital archives to guarantee that their interoperability adheres to the needs of academic researchers.

Acknowledgements

Oceanic Exchanges (OcEx) is funded through the Transatlantic Partnership for Social Sciences and Humanities 2016 Digging Into Data Challenge. We are grateful to Deutsche Forschungsgemeinschaft for support (grants PR 1080/5-1, PA 1956/3-1, KO 5362/2-1).

Bibliography

Beals, Melodee/Bell, Emily (2020), *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges*, https://www.doi.org/10.6084/m9.figshare.11560059 (accessed 13/3/2021).

Deutsche Forschungsgemeinschaft (2020), "Digitaler Wandel in den Wissenschaften", https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/digitaler_wandel/index.html (accessed 27/2/2021).

Dinger, Patrick/Landes, Lisa (2019), "The German Newspaper Portal: A National Aggregator for Digitised Historical Newspapers", in: *europeana pro* (16: Newspapers), https://pro.europeana.eu/page/issue-16-newspapers#the-german-newspaper-portal-a-national-aggregator-for-digitised-historical-newspapers (accessed 1/3/2021).

Franke, Max/John, Markus/Knabben, Moritz/Keck, Jana/ Blascheck, Tanja/Koch, Steffen (2020), "LilyPads: Exploring the Spatiotemporal Dissemination of Historical Newspaper Articles", in: Andreas Kerren, Christophe Hurter and Jose Braz (eds.), *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020), vol. 3*, s.l., 17–28, https://doi.org/10.5220/0008871400170028 (accessed 2/3/2021).

Hauswedell, Tessa/Nyhan, Julianne/Beals, Melodee/Terras, Melissa/Bell, Emily (2020), "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria That Shape Digital Archives of Historical Newspapers", in: *Archival Science* 20, 139–165, https://doi.org/10.1007/s10502-020-09332-1 (accessed 15/2/2021).

König, Mareike (2020), "Geschichte digital – zehn Herausforderungen", in: Cord Arendes, Karoline Döring, Claudia Kemper, Mareike König, Thorsten Logge, Angela Siebold und Nina Verheyen (eds.), *Geschichtswissenschaft im 21. Jahrhundert,* https://blog.degruyter.com/geschichte-digital-zehn-herausforderungen/ (accessed 15/1/2021).

Koninklijke Bibliotheek (2008), "Selection Criteria", https://www.kb.nl/en/organisation/research-expertise/digitization-projects-in-the-kb/databank-of-digital-daily-newspapers/selected-titles-and-selection-procedure/selection-criteria (accessed 15/1/2019).

Oceanic Exchanges Project Team (2017), *Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840–1914*, https://doi.org/10.17605/OSF.IO/WA94S (accessed 2/3/2021).

Ridsdale, Chantel/Rothwell, James/Smit, Mike/Ali-Hassam, Hossam/Bliemel, Michael/Irvine, Dean/Kelley, Daniel/Matwin, Stan/Wuetherick, Brad (2015), *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*, https://doi.org/10.13140/RG.2.1.1922.5044.

Salo, Dorothea (2020), "Is There a Text in These Data? The Digital Humanities and Preserving the Evidence", in: Martin Paul Eve and Jonathan Gray (eds.), *Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access*, https://doi.org/10.7551/mitpress/11885.001.0001, (accessed 2/3/2021).

Sutton Koeser, Rebecca/Doroudian, Gissoo/Budak, Nick/Li, Xinyi (2020), "Data Beyond Vision," in: *Transformations* (1), https://doi.org/10.5281/zenodo.3713670, (accessed 2/3/2021).

Terras, Melissa (2011), "The Rise of Digitization: An Overview", in: Ruth Rikowski (ed.), *Digitization Perspectives*, Rotterdam, 3–20.

Figure Credits

Fig. 1 © Mapbox and OpenStreetMap. Reproduction from Franke et al.