Open Research Data and Innovative Scholarly Writing: OPERAS highlights

Elisa Nury
Digital Humanities +
SIB Swiss Institute of Bioinformatics
Lausanne, Switzerland
elisa.nury@sib.swiss

Claire Clivaz
Digital Humanities +
SIB Swiss Institute of Bioinformatics
Lausanne, Switzerland
claire.clivaz@sib.swiss

Marta Błaszczyńska
Institute of Literary Research of the
Polish Academy of Sciences
Warsaw, Poland
marta.blaszczynska@ibl.waw.pl

Michael Kaiser
Max Weber Foundation
Bonn, Germany
Kaiser@MaxWeberStiftung.de

Agata Morka
European Coordinating Office
Open Access Books
Berlin, Germany
agata@openbookpublishers.com

Valérie Schafer
C²DH
University of Luxembourg
Luxembourg, Luxembourg
valerie.schafer@uni.lu

Jadranka Stojanovski
Department of Information Sciences
University of Zadar
Zadar, Croatia
Centre for Scientific Information
Ruđer Bošković Institute
Zagreb, Croatia
jadranka.stojanovski@irb.hr

Erzsébet Tóth-Czifra
Open Science
DARIAH-EU
Berlin, Germany
erzsebet.toth-czifra@dariah.eu

Abstract—We present here highlights from an enquiry on the innovations in scholarly writing in the Humanities and Social Sciences in the H2020 project OPERAS-P. This article explores the theme of Open Research Data and its role in the emergence of new models of scholarly writing. We examine more closely the obstacles and fostering conditions to the publication of research data, both from a social and a technical perspective.

Keywords—Open Research Data, Scholarly writing, Academic publishing, Innovation, Social Sciences, Humanities

I. INTRODUCTION

Since the last decade, new models of scholarly writing have emerged alongside the practice of sharing Open Research Data. The transformation has manifested itself in different ways in Social Sciences and Humanities (SSH) and STEM disciplines according to their respective epistemic culture. The SSH have focused more on encoding standards such as the Text Encoding Initiative (TEI) or the integration of multimedia. In STEM, publishing descriptions and providing links to datasets and databases has become a prominent topic, in journals like *Earth System Science Data*[65], or on platforms featuring datasets with observations like *ScienceMatters* founded in 2016[66]. Certain fields such as Neurosciences, Astronomy or the Life Sciences have already been engaged in data sharing and open science practices for several years, while the transition also impacted SSH, that goes progressively in the direction to sharing Open Data (Vanholsbeeck et al., 2015).

In this paper, we present first results demonstrating that the production and publication of research data may deeply transform the creation of knowledge SSH. These results have been produced in the H2020 project OPERAS-P (Open Scholarly Communication in the European Research Area for Social Sciences and Humanities – Preparation). OPERAS-P wishes to help prepare "a long-term, evidence-based strategy for the development of [OPERAS] infrastructure and its services", one of the four purposes of the project[67]. OPERAS is an emerging European research infrastructure[68] that aims to address the scholarly communication needs of European SSH researchers with the

---

[65] See https://www.gbif.org/data-papers.

[66] https://www.sciencematters.io/. About data and publication process, see Parsons and Fox (2013).

[67] See https://www.operas.unito.it/projects/operas-p/. The H2020 project OPERAS-P is the preparatory phase of the building of OPERAS.

[68] Since the end of 2019, OPERAS has been founded as an AISBL, *Association internationale belge de droit public*, and counts today 56 organizations from 17 countries; website: https://operas-eu.org.

appropriate infrastructure. It supports the successful implementation of the emerging global, European, and national data policies, along with the European Research Infrastructures (ERICs) in SSH[69].

Within the OPERAS-P framework, an enquiry on the transformation of scholarly writing is led by the Institute of Literary Research of the Polish Academy of Sciences (IBL PAN), in collaboration with five other institutions[70]. In the context of this paper our research questions are: What are the necessary conditions for research data to become new scholarly models of writing? What is propelling us in that direction? What is preventing one from accommodating the novel means of scientific writing to one's own practices? After conducting an overview of the existing literature on innovative scholarly writing practices, the team has led about forty qualitative interviews with diverse stakeholders: scholars, editors, publishers, librarians. The interviews have been transcribed and translated into English when needed. We present here the first highlights of this enquiry.

II. OVERVIEW OF THE CURRENT LANDSCAPE

The publication of research data comes within a larger context. One aspect is the adoption of the digital medium for both reading and writing, an important milestone in the history of written communication (Vandendorpe, 2013). However, there is a gap between scholarship creation practices, which have adopted the digital medium, or even create complex digital scholarly objects as outcomes, and publication mechanisms which are still very close to the digital surrogates following paper-based publishing paradigm (Tóth-Czifra, 2019). Still, new possibilities offered by digital medium have encouraged the growth of innovative models of writing and publishing. New trends are emerging, such as demands for publishing multimodal content alongside the text, or a growing interest in research data among SSH scholars as well as funders who now often require different forms of data sharing and Data Management Plans (DMPs) from applicants. Changes in conducting research, managing data and reporting on the research results are influenced by the pressure of "publish or perish", the biases and flaws in the peer review process (Lee et al., 2013), or an "insupportable economic model" (Fitzpatrick, 2011).

Publishing research data provides researchers with the benefits of reusing data, being able to reproduce results, and assessing the value of a publication (Pettifer et al. 2011). In SSH, reproducibility is a contested notion, but sharing research data facilitates scholarly transparency to understand where the source ends and where the interpretation starts. Tracking the provenance of data and sources as well as the layers of interpretation that have been added to it is central to SSH research[71]. Scholars in the field of the Life Sciences also pushed the data paper as a new form of scholarly publication following academic standards which describes datasets and the circumstances of their collection, and provide a link to their repository (Chavan & Penev, 2011). The following years have seen the infrastructure being developed for the purpose of publishing datasets and data papers and preserving them for the long term: the creation of open data repositories, such as Nakala in France, DANS in Netherlands or DARIAH-DE Repository in Germany, and data journals published by prestigious academic publishers[72].

The SSH have followed the same direction shortly thereafter. In 2015 the *Journal of Open Humanities Data* started publishing data papers, and the first volume of Brill *Research Data Journal for SSH* came out in 2016. Both journals are Open Access and have collaborations with dedicated data repositories, for instance DANS and *Dataverse Network of Harvard University*, and more general repositories like Figshare and Zenodo. More recently, De Gruyter, with the C²DH at the University of Luxembourg, launched the *Journal of Digital History*[73], an innovative publication platform for "multi-layered" articles that include data, methodology and a narrative layer. Therefore, the data paper is becoming an established form of scholarly writing. The purpose of a data paper is, among other things, to give credit for the effort required to prepare, curate, and contextualize data with the proper metadata. As such, data and its description indeed form a new model of scholarly writing in SSH.

---

[69] DARIAH (Digital Research Infrastructure for the Arts and Humanities), https://www.dariah.eu/; CESSDA (Consortium of European Social Science Data Archives), https://www.cessda.eu/; CLARIN (Common Language Resources and Technology Infrastructure), https://www.clarin.eu/; SHARE (Survey of Health, Ageing and Retirement in Europe), http://www.share-project.org.

[70] DARIAH-EU with SIB (CH) as partner, the Max Weber Foundation (DE), Open Book Publishers, the University of Luxembourg and the University of Zadar (Croatia).

[71] See for example the *Research Data Journal for the Humanities and Social Sciences*; the *Journal of Open Humanities Data Dataverse*; the *Journal of Open Archaeology Data*.

[72] Brill has its own data repository with Figshare: https://web.archive.org/web/20201020000158/https://brill.figshare.com/. Elsevier and Nature have both launched journals to publish research data and data papers in 2014: see http://web.archive.org/web/20200325153944/https://www.elsevier.com/authors/author-resources/research-data, and http://web.archive.org/web/20200325154111/https://www.nature.com/sdata/about. De Gruyter also integrates a widget to publish code alongside articles since 2018: http://web.archive.org/web/20200325154206/https://www.eurekalert.org/pub_releases/2018-07/co-dgp071118.php

[73] See SIBDARIAH03 and 04; all interviews are referred in the final bibliography. See also: https://journalofdigitalhistory.org/en/about

SSH data often depend on artefacts owned by various Cultural Heritage institutions that impose their own policies and copyrights restrictions, or on qualitative interviews that may contain sensitive personal information, which affect the culture of data publication (Tasovac et al., 2020). It should also be highlighted that scholars receive no credit for data publication as such, a point related to national or/and institutional open science policy. Researchers need to present the data into a journal article format. The reasons for this are complex, but one explanation has to do with the lack of scholarly information management systems that are inclusive with digital scholarly objects rather than publication texts. However, the emergence of search engines like Google Datasets[74], discovery systems like the OpenAIRE Research Graph[75], and TRIPLE[76] platform for SSH data discovery and reuse will improve the situation.

III. CONDITIONS, OBSTACLES, AND FOSTERING ELEMENTS

What are the opinions of SSH researchers regarding the publication of research data? During our interviews, the respondents expressed a variety of opinions when asked "What is your opinion about publishing the entire material from a given study in SSH: whole interviews, annotated texts, research protocols, data collected in the research process etc.?" There were doubts related to the time-consuming aspect of academic life that already requires researchers to read, write and peer review articles:

> *"It doesn't make any sense. I already don't have time to read all the articles I want to read. I understand it intellectually, but given the time I have, I don't think I would take the time to get into an underground area below the article." (UniLux01, 2020)*
>
> *"In History, we are already happy if one person takes time to read what we write! Who is going to read research notes on a subject for which the final monograph or publication will already be read by too few scholars?" (SIBDARIAH02, 2020)*
>
> *"I'm very much in favour of there actually being digital data repositories that allow as much data as possible to be accessed by people who are interested. I think that [...] the accessibility part of the data should be increased online [...]. The problem is that the research data is only relevant to a very small portion of the readers. That is to say that, in fact, it's like footnotes, footnotes are very important for the epistemological and ethical guarantee of the work." (SIBDARIAH08, 2020)*

As summarized by the last quotation, research data are very important as a guarantee of valid research processes, but they may be used only by a minority of scholars, which means that it represents a large investment in time for what might seem like little return. However, the availability of research data is crucial for the accuracy and reliability of peer review. On the other hand, interviewees also stressed the importance of transparency, with caveats about privacy, copyrights, and reuse of data:

> *"Yes, I would. It is even necessary, or it is becoming more and more mandatory in certain cases. Today, transparency is very important." (SIBDARIAH01, 2020)*
>
> *"I'm really in favour of that. [...] I think it depends on the field a lot but in my field having to publish alongside your manuscript which should really be your reflections on the data that you've collected. Publishing the data and publishing what you did with that data so publishing some form of code that you used to get from data to conclusions, and to create visualizations, and tables, and stuff like that I think that would be very beneficial [...]. Also, it would make the whole process much more transparent and it would not eliminate, it would reduce the margin for foul play." (UNIZD01, 2020)*
>
> *"One of the things that we come up against is that, culturally, people expect transparency. That becomes dangerous because then you can violate things like privacy. [...] But if I put that stuff out there, scholar X is going to take that data and write that next book that I'm not going to right now. Because the incentives of scholarship are what they are, you still have to be careful about what full publication would look like." (SIBDARIAH07, 2020)*

In some cases, the publication of research data was a necessity in the context of reporting mathematical and statistical experimentations where a traditional article was not sufficient: the sources needed to be made available as the software used and the raw data from the experiment. Ideally, the research data would be published with the same standards of rigor as traditional academic publications, however the peer review of data would raise an enormous

---

[74] https://datasetsearch.research.google.com/
[75] https://www.openaire.eu/
[76] https://operas.hypotheses.org/projects/triple

challenge in terms of the workload it would impose on reviewers that may already be short on time. Even researchers who agree that research data should be peer-reviewed admit that realistically, we will never be able to do that at scale:

> *"And really the labour involved in evaluating these things just goes through the roof. And I just don't think people are going to have time to do that kind of evaluation for every piece of digital scholarship that emerges in the next few years. So, I think there's a looming crisis for the labor of peer review." (DAE03, 2020)*

The interview samples above highlight an interest in the need to publish research data in SSH, although there are limitations and obstacles. The question for us was to better understand those obstacles and identify the conditions that could foster the publication of data. From the preliminary analysis of our interviews, as well as the scientific literature on the subject, we identified two broad topics, a social one and a technical one. The social one covers research challenges of the relationship between the various stakeholders involved in research data management – Galleries, Libraires, Archives, and Museums (GLAM) professionals, data processors such as Digital Humanities labs, research teams, repository managers, data stewards and publishers – and the current misalignments between data sharing policies and academic rewarding criteria. The second area encompasses technical problems of data curation and storage.

A. Research Challenges

One of the biggest obstacles to innovative scholarly publications, such as research data, is the "reward structure", that is, how research is assessed and credited within academic institutions (Moore & Adema, 2020). In fact, scholarly publications are not only about disseminating knowledge, but they also play an important role in assessing academic success, in evaluating and promoting researchers. However, the currency of academic credit is not money, it is "reputation" (Andrews, 2017), largely measured by a series of analytics: the number of publications, the number of citations received by those publications, the impact factor of journals where they are published, the prestige of the publisher, or the publication type (in certain cases SSH scholars receive more credit for a book than an article, for example).

As a result, there is a strong incentive for researchers to publish traditional scholarship in prestigious venues for their field of study, to receive the credit, they need to get funding, a stable position, and to advance their career. In this context, and within the time constraints of research projects, this situation creates a tension for scholars to balance the need for traditional publications and desire for innovative practices, as highlighted during our interviews: publishing data is time-consuming, which is a disadvantage (SIBDARIAH10); the work often has to be done twice, once for preparing and depositing the digital output, and once for a more traditional publication (SIBDARIAH01). These problems are not new and have been repeatedly highlighted in the field of Digital Humanities: "Digital humanists find, time and time again, that they are expected to perform twice the labor of traditional scholars; once for the work itself and once again for its evaluation" (Eve, 2020; see also Baillot, 2016 and Fitzpatrick, 2011).

Closely related to the concept of academic credit is "data hugging" [77], the opposite of data sharing. Scholars are often reluctant to release data, as they must cope with a culture of perfection, and they dare only present data of utmost perfection. Since good publications bring credit, this is a valid fear, especially considering that SSH scholars may take years to gather their data, to fully analyze them, interpret and write monographs. On the other hand, the emerging practice of data publications carry the potential to immediately claim early attribution and credit the authors which deconstruct the dynamics that fuel the current data hugging phenomena. The publication of research data would also improve the recognition that collecting data is already doing valid academic research (Truan, 2019).

One solution to encourage data sharing would be for scholars to get acknowledgement early in the process, and not after, for instance, the final monograph is published. In STEM, there is a culture of sharing preprints, and it can be adopted for SSH articles as well, but for monographs we need more innovative writing models: for instance, web books where chapters can be published consecutively (SIBDARIAH01)[78]. Promoting a data citation culture would bring research data into the spotlight: scholars cited for their data would receive academic credit, which would in turn be considered by funders and by committees for promotions. But to say it in short with Tóth-Czifra, "the information systems measuring the (re)use and impact of digital tools and scholarly data are still in their infancy" (2020).

However, it is often not the reservations of researchers that are the main obstacle to developing data sharing practices. The nature of SSH data, already discussed in the previous section, has implications in the areas of copyright and privacy. If most of the research data used in a research project consists of existing artefacts, such as objects owned by a third party (artworks, texts, audiovisual materials), the opportunities for publishing the dataset will largely be

---

[77] The term has been coined by Dr. Hans Rosling in a quite famous talk about the "Data-Hugging Disorder" given in May 2009 (see Frydman, 2009).

[78] Web books are books presented in the format of a website, which are regularly updated (Fauchié, 2016).

constrained. While such information should be explicitly provided (Angelaki et al., 2020), sometimes it is even difficult to verify the ownership of an object or to check on what license it was originally shared. Furthermore, social scientists conducting surveys or interviews will have to obtain an explicit consent from respondents to reshare the data (raw answers, transcriptions or recordings). In many cases time-consuming – and sometimes complex – anonymization and pseudonymization procedures will be necessary.

This also raises the question of times and temporalities. More and more funders are requiring provision of a DMP from the very beginning of the project, but this raises complementary questions about the maintenance of data as well as derived research data and scientific outputs at the end of the project. Willingness is often not sufficient when researchers at the same time must deal with long-term needs of data preservation, mid-term funding and the paradoxically short life cycles of the data, formats, devices, tools and platforms. As demonstrated in Barats et al. (2020), there are "a number of different temporalities […] and multi-stakeholder issues that require collective reflection to clearly identify the actors and locations that are best adapted to implement and support the challenges of data sustainability."

In summary, the challenges described above therefore require for the academic community to rethink research assessment practices, to change the metrics of academic credit, and to take into account the time and amount of labor necessary for the publication of quality research data. There are initiatives in that sense, for instance the SF Declaration on Research assessment (DORA), that has been signed by many institutions in Europe, or the HuMetricsHSS initiative[79]. This claims for new peer review practices, as developed in Digital Humanities for evaluating scholarship (Baillot, 2016). Peer reviewed data papers can be part of the solution, but we need to develop criteria and procedures that certify the quality of research data.

B. Data Storage and Curation

On the side of the more technical problems, there is a need for infrastructures to access and store data, along with the relevant metadata that is necessary to interpret and reuse data. A first concern is about access to data. As noted by Rieder & Hofmann (2020), "the concept of observability starts from the recognition of a growing information asymmetry between platform companies, a few data brokers, and everyone else. The resulting data monopoly deprives society of a crucial resource for producing knowledge about itself." Some datasets may be stored behind a paywall and thus only accessible for researchers with funding. This may increase the gap between those who can afford paying to access data and those who cannot.

One challenge that came up during the OPERAS-P interviews is also the fragmented nature of data repositories (Mostern et al., 2016) and the need for a single point for discovery, such as a European wide search engine, e.g., Isidore. As noted by (Gregory et al., 2020), "[b]efore data can be reused, they must first be discovered", and data finding can be hampered by the technical infrastructure (researchers use Google with mixed success) and is also dependent on the researchers' social context. One may add the fragmented nature of data and sources that may be divided up in several repositories and the issue of hosting of complex corpora. In addition, one researcher dealing, for example, with born digital heritage may share derived data, some metadata and permalinks to web archives that are preserved in national institutions, but will not be able to share more, because authors' rights apply. This also highlights the need for interoperability, but the large number of metadata standards can make it complicated, as there is a variety of standards in SSH: general standards (DublinCore), standards for text (TEI), images (IIIF), archival materials (EAD), cultural heritage (CIDOC-CRM), and so on.

Another concern is how to link various outputs of a project, the data, the articles, the code, the source materials etc. The common practice now is to use persistent identifiers such as DOI. This also has implications for publishers and libraries: publishers will have to deal with projects that have multiple outputs (SIBDARIAH07) – how do those outputs hold together as a unified, complex entity? How do librarians' catalogue and provide access to a publication made of multiple parts? Are PIDs sufficient? Can they be used to keep track of citations for the data? There seems to be no accurate information management system in place for that[80]. Moreover, while certain writing tools allow for a greater integration of data into the scholarly text, often only the minority of researchers use it. One of the interviewees feels that in their field:

> *"the relation between data and writing is still a bit conflictual because people write in Word and*
> *there's no way to integrate nicely your statistics or your lines of code and to have good*
> *synchronization between the data and the text you're writing or to provide interaction between the*
> *text and the reader." (IBL08, 2020)*

FAIR data and current trends in Open Science underline the possibilities and opportunities of use and reuse of data but this also raises other challenges. First, there is a need for new models of peer review, as scholars who are

---

[79] https://humetricshss.org/. See also https://www.dariah.eu/activities/working-groups/impact-factors-and-success-criteria/.

[80] See the *Journal of Open Humanities Data*: https://web.archive.org/web/20201027160132/https://openhumanitiesdata.metajnl.com/about/.

really able to evaluate these data and review them may be rare. One may also push for an interdisciplinary peer review, mixing several levels of digital, engineering, and scientific skills. Innovative scholarly outputs including datasets may be challenging for the readers who are more used to traditional publications. In addition, some interviewees underlined their lack of time to read all papers related to their field of research, therefore being very skeptical about their availability to go deeper in the reading and discovery of data. Regarding reuse, the point is not just about sharing data, but also about contextualizing them to allow a genuine reuse. Finally, there is also a challenge of maintenance - and eventually repair. This is often a part which is forgotten in the process and may create plenty of data lakes that are unexploitable because they lack transparency, contextualization, updates, etc. Morselli & Edmond (2020) note that work is lost due to resource and technical challenges, but they also illustrate how the sustainability of the results of digital research projects can be thought of as a process instead of an end product that involves more than ensuring a long-term hosting data infrastructure.

IV. CONCLUSION

A complex environment is at stake, consisting of data brokers, engineers, researchers, publishers, funders, several kinds of data, as well as several legislative environments in an internationalized world. There is a need for incentives at all levels and for understanding that this investment has a cost (may it be in terms of funding, maintenance, engineering, time, etc.), but this may cost less than losing vast amounts of data and research. Capitalizing on the existing infrastructures may in a mid-term perspective create a strong reward. Parallel to these systemic changes we also need a cultural shift to view the publishing of data as a valuable scholarly output.

Conceptual models may be needed to help design intelligent and efficient solutions. One example is the data scope concept by Hoekstra et al., which, for example, suggests "classifying data" to group them "to reduce complexity". This adds a level of abstraction to the data" (2018). How do we shrink the gap between those who are able and those who are not to share data, and how might we direct a whole generation of researchers to this transition? Are there new skills that all researchers should develop, and will these new skills create new research profiles and kinds of support? Data stewardship (Mons, 2018) is developing and may in the future be more and more pertinent, becoming a new fundamental position.

REFERENCES

[1] Andrews, T. (2017, video). After the Spring: Digital forms of scholarship and the publication ecosystem [Conference Keynote] (26.09.17). *20th International Conference on Electronic Publishing Positioning and Power in Academic Publishing: Players, Agents and Agendas. 7–9 June 2016 in Göttingen, Germany, Göttingen*. https://www.youtube.com/watch?v=44FaPFjYxMk

[2] Angelaki, G., Badzmierowska, K., Brown, D., Chiquet, V., Colla, J., Finlay-McAlester, J., Grabowska, K. et al. "How to Facilitate Cooperation between Humanities Researchers and Cultural Heritage Institutions. Guidelines." *Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences, 10 March 2019*. https://doi.org/10.5281/zenodo.2587481.

[3] Baillot, A. (2016, conference communication). A certification model for digital scholarly editions: Towards peer review-based data journals in the humanities. Conference communication. *Digital Scholarly Editing: Theory, Practice, Methods*, Université d'Anvers, Oct 2016, Anvers, Belgium, https://halshs.archives-ouvertes.fr/halshs-01392880

[4] Barats, C., Fickers, A., & Schafer, V. (2020), Fading Away... The challenge of sustainability in digital studies. *Digital Humanities Quarterly* 14 (03), http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html

[5] Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, *12*(S15), S2. https://doi.org/10.1186/1471-2105-12-S15-S2

[6] DAE03: Tóth-Czifra, E. (2020). Transcript Interview DAE03 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/10.34847/nkl.3e9bjm89

[7] Eve, M. P. (2020). Violins in the Subway: Scarcity Correlations, Evaluative Cultures, and Disciplinary Authority in the Digital Humanities. In: Edmond, Jennifer (ed.) *Digital Technology and the Practices of Humanities Research.* Cambridge: Open Book Publishers, 105-22.

[8] Fauchié, A. (2016, blog post). Le livre web, une autre forme du livre numérique (24.10.16). *Quaternum.Net,* https://www.quaternum.net/2016/10/24/le-livre-web-une-autre-forme-du-livre-numerique/

[9] Fitzpatrick, K. (2011). *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. NYU Press; JSTOR. www.jstor.org/stable/j.ctt9qg9mh

[10] Frydman, G. (2009), "e-Patients Demand: Put An End To Data-Hugging Disorder" (04.06.09), *Society for Participatory Medicine. Transforming the Culture of Patient Care Blog.*

https://participatorymedicine.org/epatients/2009/06/e-patients-do-not-suffer-from-database-hugging-disorder.html

[11] Gregory, K. M., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). Understanding data search as a socio-technical practice. *Journal of Information Science*, *46*(4), 459–475. https://doi.org/10.1177/0165551519837182

[12] Hoekstra, F. G., Koolen, M., Burgers, J. W. J., van Faassen, M., Nijenhuis, I. J. A., & Derks, S. (2018, report). Inleiding Data Scopes. *Huygens ING*. https://doi.org/DOI: 10.13140/RG.2.2.23812.01928

[13] IBL08: Błaszczyńska, M. (2020). Transcript Interview IBL08 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/10.34847/nkl.e87827w3

[14] Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. https://doi.org/10.1002/asi.22784

[15] Mons, B. (2018). *Data Stewardship for Open Science. Implementing FAIR Principles*. Routledge Press.

[16] Moore, S., & Adema, J. (2020, blog post). COPIM Experimental Publishing Workshop—Part 1: Inhibitions Towards Experimental Book Publishing (19.10.20). *COPIM*. https://copim.pubpub.org/pub/experimental-publishing-workshop-part-1/release/2

[17] Morselli, F., & Edmond, J. (2020). Sustainability of digital humanities projects as a publication and documentation challenge. *Journal of Documentation*. https://doi.org/10.1108/JD-12-2019-0232

[18] Mostern, R., and Arksey, M. Don't Just Build It, They Probably Won't Come: Data Sharing and the Social Life of Data in the Historical Quantitative Social Sciences. *International Journal of Humanities and Arts Computing* 10 (2016), 205–24. https://doi.org/10.3366/ijhac.2016.0170

[19] Parsons and Fox (2013), Is data publication the right metaphor? *Data Science Journal* 12, https://www.jstage.jst.go.jp/article/dsj/12/0/12_WDS-042/_pdf/-char/en

[20] Pettifer, S., McDermott, P., Marsh, J., Thorne, D., Villeger, A., & Attwood, T. K. (2011). Ceci n'est pas un hamburger: Modelling and representing the scholarly article. *Learned Publishing*, *24*(3), 207–220. https://doi.org/10.1087/20110309

[21] Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1535

[22] SIBDARIAH01: Nury, E. (2020). Transcript Interview SIBDARIAH 01 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/11280/d4aafc66

[23] SIBDARIAH02: Nury, E. (2020). Transcript Interview SIBDARIAH 02 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/11280/133a7853

[24] SIBDARIAH03 and 04: Nury, E. (2020). Transcript Interview SIBDARIAH 03 and 04 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/11280/d2905a1b

[25] SIBDARIAH07: Nury, E. (2020). Transcript Interview SIBDARIAH 07 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/11280/a7f68f73

[26] SIBDARIAH08: Nury, E. (2020). Transcript Interview SIBDARIAH 08 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/10.34847/nkl.64b9ratx

[27] SIBDARIAH10: Chiquet, V. (2020). Summary Interview SIBDARIAH 10 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/11280/ac0ffef5

[28] Tasovac, T., Chambers, S., & Tóth-Czifra, E. (2020, position paper). *Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper*. https://hal.archives-ouvertes.fr/hal-02961317

[29] Tóth-Czifra, E. (2019, blog post). Laying the Pavement Where People Actually Walk: Thoughts on Our Chances of Bringing Scholarship Back to the Heart of Scholarly (23.10.19). DARIAH Open. https://dariahopen.hypotheses.org/645

[30] Tóth-Czifra, E. (2020, blog post). 10 practical tips to fight against the culture of non-citation in the humanities (03.03.2020). *DARIAH OPEN blog*. https://dariahopen.hypotheses.org/747

[31] Truan, N. (2019, blog post). "How to make the most of your publications in the humanities?". (22.01.19), *Ici et là*. https://icietla.hypotheses.org/994

[32] UniLux01: Schafer, V. (2020). Transcript Interview UniLux01 (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/10.34847/nkl.66b98590

[33] UNIZD01: Zauder, K. (2020). Transcript Interview UNIZD01. (H2020 OPERAS-P). *[Text] Nakala*. https://doi.org/10.34847/nkl.6aaf8jl9

[34] Vandendorpe, C. (2013). Reading on Screen: The New Media Sphere. In S. Schreibman & R. Siemens (Eds.), *A Companion to Digital Literary Studies*. Wiley Online Library. https://doi.org/10.1002/9781405177504.ch10

[35] Vanholsbeeck, M., Engels, T., & Istenič Starčič, A. (2015). Guidelines for Data Sharing and Data Citation in Social Sciences and Humanities Journals Perspectives and Insights from the Cost Action Enressh. *Archives et Bibliothèques de Belgique - Archief- En Bibliotheekwezen in België* 106, 83–92.