

Hansson, Karin; Cerratto Pargman, Teresa; Dahlgren, Anna (2020): Datafication and cultural heritage: provocations, threats, and design opportunities. In: Proceedings of the 18th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing and the Design of Cooperation Technologies - Workshop Proposals, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.18420/ecscw2020_ws05

Datafication and cultural heritage: provocations, threats, and design opportunities.

Karin Hansson*, Teresa Cerratto Pargman, Anna Dahlgren

Stockholm University

khansson@dsv.su.se, tetsy@dsv.su.se, anna.dahlgren@arthistory.su.se

Abstract. Increasing digitization and the emergence of new data sharing practices are likely to change how our understanding of history is negotiated. The curation of data is always culturally and ideologically inflected. Accordingly, archiving practices are not only fundamental for our understanding of the past but vital in navigating the present. We have to pay particular attention to the consequences of the interfaces that curate history, especially in relation to big data.

Crowdsourcing, social media, linked open data, and other participatory and open science practices challenge the archiving practices in cultural heritage institutions due to the character of the networked publics involved and the established structures between and within institutions. However, they also open up new opportunities and practices when it comes to understanding and defining our shared culture.

In this workshop we will bring together researchers who have studied these issues or are working to develop critical perspectives on archiving practices.

Copyright 2020 held by Authors, DOI: 10.18420/ecscw2020_ws05

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, contact the Authors.

Background

The digitization of cultural heritage, increased datafication in all areas of society, and new infrastructures for sharing data within research and with the wider public are likely to change our views of the production of history. Participatory practices such as citizen science are, for example, enabled and further developed in projects that includes a broad public in crowdsourcing projects (Borowiecki, Forbes, & Fresa, 2016; Cameron, Kenderdine, 2007; Cerratto Pargman, Joshi and Wehn, 2019; Manzo, Kaufman, Punjasthitkul, & Flanagan, 2015). Digitization entails an increased datafication and development of data-driven practices in diverse societal sectors, which results in more and more human activities being monitored, analyzed, quantified, archived, classified, and linked to other data (Boyd & Crawford, 2012; Lyon, 2014).

Easier access to large-scale and linked datasets have opened up new possibilities to explore and create value from data and have enabled the development of more evidence-based quantitative scientific practices. This quantitative turn has influenced new research fields such as digital humanities, digital social science, e-learning, e-business, learning analytics, etc., where access, storage, and analysis of data drives the development of new infrastructures and methodologies. This creates changes not only at the level of data techniques and methods but also for researchers' practices. For instance, some of these changes include increased standardization, development of shared infrastructures, and direct publication of research data, which changes the preconditions for research not only in the natural sciences but also in the humanities (Borgman et al., 2012; Carusi & Jirotko, 2009; De La Flor et al., 2010; Ribes & Lee, 2010; Schroeder, 2008).

Increasing datafication of research practices and their infrastructure create opportunities to nurture an open research culture, enabling researchers to share their results through open access (Kidwell et al., 2016; Nosek et al., 2015; Roche et al., 2014).

Datafication and increased measurement practices are also reflected in areas such as education (Biesta, 2015), journalism (Coddington, 2015), and politics (Milan & Velden, 2016). Such practices are not only possible through the access, storage, and analysis of people's data but via the tools for gathering and analyzing data. The proliferation of results is democratized through easily accessible infrastructures of people, survey tools, opinion polls and petitions, visualization, and dissemination to the public. This scientification, in which research methods are increasingly developed and integrated with everyday working practices, also creates expanded demands for a digital literacy. Not only is it necessary to understand how information is created and disseminated (boyd, 2011), but it is also necessary to foster deeper insights into research methodologies and archiving processes.

The critics of datafication claim that the belief in quantification poses dangers. For example, transparency decreases (it is often unclear on what grounds the data have been curated) (Andrejevic, 2014; Boyd, 2012; Bunnik et al., 2016; Driscoll & Walker, 2014), important values are lost if they are not easy to compile as numbers, and threats to personal integrity increase as data collections are disseminated, linked, and combined (Schradié, 2011).

Efforts to generate, collect, identify, and classify data and data collections risk obscuring the multifaced work practices around history production, including reward structures, authority structures, formalization of knowledge, interdependencies among groups, trust mechanisms, and the transitional quality of data collections (Borgman, 2012). Datafication can also be seen as an increased commodification of various aspects of human practices, especially due to the datafication of our life-worlds. People's opportunities to express themselves and organize themselves through the use of social media also contributes to new forms of surveillance and sources for consumer research (Hansson et al., 2018).

In the context of the humanities, there are critics who think that the qualitatively founded criticality that is at the core of contemporary research in the humanities is threatened and downplayed by, for example, politically controlled funding (Belfiore & Upchurch, 2013). Also, the majority of crowdsourcing and open science projects are within the realm of the natural sciences and areas where data is easy to quantify (Burke, 2012; De León, 2015). In the humanities, scientific processes are often different, and they demand other considerations. Research in the humanities is more about creating heterogeneity and differences than collaborating around one shared common goal (Anderson & Blanke, 2015).

Furthermore, critics have pointed out that when archiving practices are distributed and maintained broadly over diverse sectors and groups in society, enabling a multifaced and fragmented notion of history, cultural heritage institutions might need to reevaluated their role in society (Fredheim, 2018).

However, the developers of infrastructure and the critics rarely meet. Few academic studies or commercial design processes take criticism seriously by developing practices and tools that combine qualitative and quantitative approaches in a self-reflexive way.

Furthermore, while areas such as e-research, cyberinfrastructure, and crowdsourcing, are generally well covered by the ECSCW community, the intersection between these areas and the increasing digitalization processes and datafication in the humanities is less explored.

In this workshop we bring together researchers from HCI and CSCW who have studied these issues or are working to develop critical perspectives on technology, design, and research practices. We particularly welcome empirically-based

research that looks into digitizing and digitalization processes in cultural heritage institutions. We also welcome theoretical contributions that put research practice into a philosophical and historical context while also questioning established norms of what constitutes cultural heritage.

Suggested topics and inquiries for the workshop

In this workshop we are inviting 10–15 researchers to discuss common issues. Topics and inquiries for the workshop can for example be:

The transformation of cultural heritage institutions in the age of datafication:

- How do we make sense of the complex network of systems, information, people, values, theories, histories, ideologies, and aesthetics underlying various types of archiving practices? What are the design challenges?
- What happens when data structures become central for how a cultural heritage institution operates? What are the unintended consequences?
- How are critical archiving practices supported in cultural heritage institutions?
 - What are the cultural and ideological aspects of data curation?
 - How do we adopt intersectional perspectives in classification systems?

Quantification of research practices:

- What are the crucial design decisions when developing sharing platforms for research in the humanities? What are the opportunities and new practices in relation to understanding and sharing culture?
- How do we navigate infrastructures, rewards structures, and social structures when designing systems that help preserve and share the cultural heritage?
- What are the consequences and opportunities when using crowdsourcing, usually developed for micro tasks, in more qualitative research practices?

Participation in archiving practices:

- How is participation constructed and enacted in citizen science and crowdsourcing practices? How is participation constrained, for example, by infrastructural arrangements, technological affordances and social norms?
- What are the implications for transparency, surveillance, and trust when designing for participation in the development of the cultural heritage sector?
- What are the coping strategies and resistance to or appropriation of datafication?

- What are the tactics, structures and normative foundations necessary for supporting participatory metadata practices? What are the challenges? How do we negotiate standardization versus complexity when developing metadata practices?
- What are the implications of the (lack of) transparency and accountability of data practices in different sectors? What are the challenges this poses for users' data literacy?
- What are the new asymmetries and power relations that data practices may bring between memory institutions and audiences, or between different segments of audiences?

We are especially interested in bringing together researchers and practitioners working with digitizing and digitalization processes.

Description of the workshop activities

This one-day workshop will explore the topics through prototyping and brainstorming sessions. The workshop is divided into two sessions. The first half consists of participant presentations of their research topics. The second half will be a brainstorming session where the topics of the workshop are further explored through collaborative prototyping.

In human-computer interaction (HCI) design we are used to co-design methods such as sketches, prototypes, cases and scenarios to achieve a more informed design, grounded in the reality of potential users. Artistic techniques are also used to involve participants as informants and co-designers such as probes, scenarios, and role-playing. However, unlike most problem-focused design research, the aim with this workshop is not to use these methods to achieve a more informed design. Instead, we use the design process as a method to collaboratively materialize our own understanding of our research.

Information on the workshop will be disseminated through our website (<https://dataficationandculturalheritage.blogs.dsv.su.se>) and via emailing lists relevant for the ECSCW community as well as a broader interdisciplinary research community. Accepted papers will be circulated beforehand to prepare attendees for discussions at the workshop. Beyond the themes highlighted here by the workshop organizers, other themes for the workshop that emerge from the position papers will be posted on the website. A key discussant, identified among the workshop attendees, will be assigned to each position paper to facilitate interaction and engagement in the workshop. The participants will prepare a 5-minute presentation to be delivered in the introduction of the workshop, but focus the will be on developing our ideas through collaborative prototyping.

We will take the workshop as an opportunity to explore future collaboration (e.g., a mailing list and/or collaborative research projects). The results from the workshop may be developed further for a special issue or anthology.

Organization

The workshop is organized by an interdisciplinary group of researchers from the fields of Human Computer Interaction and Digital Humanities.

Karin Hansson, Associate Professor in Computer and Systems Sciences at Stockholm University, has written extensively about technology-based participation from a design perspective. She is currently part of a research project on the development of #MeToo activism in Sweden, and part of the “Metadata culture” research group at Stockholm University.

She has previously organized workshops on CSCW themes such as: The Morphing Organization – Rethinking Groupwork Systems in the Era of Crowdwork at ACM GROUP 2014, Sanibel Island, USA; Examining the Essence of the Crowds: Motivations, Roles and Identities at ECSCW 2015, Oslo, Norway; Toward a Typology of Participation in Crowdwork at ACM CSCW 2016, San Francisco, USA; Crowd Dynamics: Exploring Conflicts and Contradictions in Crowdsourcing at ACM CHI 2016, San Jose, CA, USA; Ting: Making Publics Through Provocation, Conflict and Appropriation, The 14th Participatory Design Conference 2016, Aarhus, Denmark.

Together with Thomas Ludwig, she recently edited the Special issue: Crowd Dynamics: Conflicts, Contradictions, and Cooperation, The Journal of Collaborative Computing and Work Practices (JCSCW), Volume 28, September 2019. This SI started in a previous workshop at CHI. Right now she is co-editor together with Shaowen Bardzell, Malin Sveningsson, and Teresa Cerratto Pargman on another SI of JCSCW called “Materializing Activism”, which is a result of a workshop with the same title at ECSCW 2019.

Teresa Cerratto Pargman, Associate Professor in Human-Computer Interaction (HCI) at the Department of Computer and Systems Sciences at Stockholm University. Her work contributes to the study of how digital technologies and applications reflect and configure socio-material practices and how emerging practices shape the development and design of digital technologies. In particular, she has published on design and appropriation of emerging technologies in the educational sector and on the epistemic, value-laden and social infrastructures, such technologies make possible but also disrupt. She coordinates the research area of Technology-enhanced learning and leads the Critical Computing group at the Department of Computer and Systems Sciences at Stockholm University.

Anna Dahlgren, professor of Art History at the Department of Culture and Aesthetics at Stockholm University, has published on different aspects of photography and vernacular visual culture including the digital turn, print culture and archives and museum practices. Right now she is managing the research project Metadata Culture (<http://metadataculture.se>) at Stockholm University, an interdisciplinary research project that investigates and develop methods for obtaining qualified and extensive metadata for images in digitalized cultural heritage collections.

References

- Anderson, Sheila, and Tobias Blanke. (2015). Infrastructure as intermeditation – from archives to research infrastructures. *Journal of Documentation*, vol. 71, no. 6, pp. 1183–1202. doi:10.1108/JD-07-2014-0095
- Andrejevic, Mark. (2014). The big data divide. *International Journal of Communication*, vol. 8, no. 1, pp. 1673–1689.
- Belfiore, Eleonora, and Anna Upchurch. (2013). *Humanities in the twenty-first century : beyond utility and markets*.
- Biesta, Gert J. J. (2015). *Good Education in an Age of Measurement*. Routledge. doi:10.4324/9781315634319
- Borgman, Christine L., Jillian C. Wallis, and Matthew S. Mayernik. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work*, vol. 21, no. 6, pp. 485–523. doi:10.1007/s10606-012-9169-z
- Borowiecki, Karol Jan, Neil Forbes, and Antonella Fresa (Eds.). (2016). *Cultural Heritage in a Changing World*. Cham: Springer International Publishing. doi:10.1007/978-3-319-29544-2
- boyd, danah. (2011). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *A Networked Self* (pp. 39–58). Routledge.
- Boyd, Danah, and Kate Crawford. (2012). Critical Questions for Big Data. *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679. doi:10.1080/1369118X.2012.678878
- Bunnik, Anne, Anthony Cawley, Michael Mulqueen, and Andrej Zwitter (Eds.). (2016). *Big Data Challenges. Society, Security, Innovation and Ethics*. Palgrave MacMillan.
- Burke, Katie L. (2012). Interface Facts. *American Scientist*, vol. 100, no. 6, pp. 463–465.
- Cameron, Fiona. Kenderdine, Sarah. (2007). *Theorizing Digital Cultural Heritage*. Cambridge, Massachusetts: MIT Press.
- Carusi, Annamaria, and Marina Jirotko. (2009). From data archive to ethical labyrinth. *Qualitative Research*, vol. 9, no. 3, pp. 285–298. doi:10.1177/1468794109105032
- Coddington, Mark. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism*, vol. 3, no. 3, pp. 331–348. doi:10.1080/21670811.2014.976400
- De La Flor, Grace, Marina Jirotko, Paul Luff, John Pybus, and Ruth Kirkham. (2010). Transforming scholarly practice: Embedding technological interventions to support the collaborative analysis of ancient texts. *Computer Supported Cooperative Work*, vol. 19, no. 3–4, pp. 309–334.

- doi:10.1007/s10606-010-9111-1
- De León, Mireia Alcalá Ponce. (2015). Crowdsourcing in the memory institutions: Mass transcriptions. *BiD*, vol. 35, no. 35. doi:<https://dx.doi.org/10.1344/BiD2015.35.26#sthash.DdLXRxNe.dpuf>
- Driscoll, Kevin, and Shawn Walker. (2014). Big Data, Big Questions| Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. *International Journal of Communication*, vol. 8, no. 1243170, pp. 20. Retrieved from <http://ijoc.org/index.php/ijoc/article/view/2171>
- Fredheim, L. Harald. (2018). Endangerment-driven heritage volunteering: democratisation or 'Changeless Change.' *International Journal of Heritage Studies*, vol. 24, no. 6, pp. 619–633. doi:10.1080/13527258.2017.1399285
- Hansson, Karin, Thomas Ludwig, and Tanja Aitamurto. (2019). Capitalizing Relationships: Modes of Participation in Crowdsourcing. *Computer Supported Cooperative Work: CSCW: An International Journal*, vol. 28, no. 5, pp. 977–1000. doi:10.1007/s10606-018-9341-1
- Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, ... Brian A. Nosek. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology*, vol. 14, no. 5, pp. e1002456. doi:10.1371/journal.pbio.1002456
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, vol. 1, no. 2, pp. 2053951714541861. doi:10.1177/2053951714541861
- Manzo, Christina, Geoff Kaufman, Sukdith Punjasthitkul, and Mary Flanagan. (2015). "By the People , For the People ": Assessing the Value of Crowdsourced , User - Generated Metadata. *Digital Humanities Quarterly*, vol. 9, no. 1, pp. 1–21.
- Milan, Stefania, and Lonneke van der Velden. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, vol. 2, no. 2, pp. 1–11. doi:10.14361/dcs-2016-0205
- Nosek, B A, G Alter, G C Banks, D Borsboom, S D Bowman, S J Breckler, ... T Yarkoni. (2015). Scientific standards. Promoting an open research culture. *Science (New York, N.Y.)*, vol. 348, no. 6242, pp. 1422–5. doi:10.1126/science.aab2374
- Ribes, David, and Charlotte P. Lee. (2010). Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories. *Computer Supported Cooperative Work*, vol. 19, no. 3–4, pp. 231–244. doi:10.1007/s10606-010-9120-0
- Roche, Dominique G., Robert Lanfear, Sandra A. Binning, Tonya M. Haff, Lisa E. Schwanz, Kristal E. Cain, ... Loeske E. B. Kruuk. (2014). Troubleshooting Public Data Archiving: Suggestions to Increase Participation. *PLoS Biology*, vol. 12, no. 1, pp. e1001779. doi:10.1371/journal.pbio.1001779
- Schradie, Jen. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, vol. 39, no. 2, pp. 145–168. doi:10.1016/j.poetic.2011.02.003
- Schroeder, Ralph. (2008). e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. *Social Science Information*, vol. 47, no. 2, pp. 131–157. doi:10.1177/0539018408089075