Data papers – An introduction

Alexandra Schneider and Malte Hagener

Abstract
In order to diversify the scope of scholarly formats within NECSUS, the new section
Data Papers offers a curated platform for publishing commented datasets from film
and media studies research projects. It invites researchers to share insights into the
often invisible collaborative work of data preparation and dataset collection.

Keywords
data paper, digital methods, digital humanities, critical data studies

Data has become a key term in recent years – from the neoliberal 'data is the new gold' to the more critical 'raw data is an oxymoron'.[1] The significance of data seems beyond doubt, but the debate around data has only just begun. What counts as data in our field? What do we need data for? Where, and how, do we gather data? How 'neutral' is data? Who owns data? What is a responsible and sustainable way of dealing with data?

This new section in NECSUS is aiming to extend and pursue this debate by way of concrete engagement with specific data sets. We need theoretical reflections and discussions of data, of the way it is used and abused, but we also need to look closely and concretely at data and what we can learn from it. This is what data papers are supposed to accomplish.

As the formats of scholarly publications have been diversified, NECSUS has been an early adopter and a key journal in facilitating this development – for example, by including a curated section dedicated to audiovisual essays, and by integrating scholarly perspectives on the GLAM sector with reviews of exhibitions and festivals. In this spirit of generative novelty, NECSUS is broadening the formats of scholarly publications again with the inclusion

of data papers in a section specifically dedicated to them. A data paper presents research in film and media studies through data sets that have been carefully selected and curated. The data paper not only makes the data available, but also adds a critical discussion and reflection on the specificities of the data set, on the rationales behind the gathering, on the methods and implicit assumptions.

In the field of digital methods, digital humanities, and digital history the publication of data papers has been established over the recent past. In the last years a number of specialised journals for data papers in the humanities and social science have been founded: two prominent examples are the *Journal of Open Humanities Data* (https://openhumanitiesdata.metajnl.com/about/) and the *Research Data Journal for the Humanities and Social Sciences* (https://brill.com/view/journals/rdj/rdj-overview.xml). The recently established *Journal of Digital History* (https://journalofdigitalhistory.org/en) adds data and layers of visualisations to its articles, thus combining traditional research with the idea of data papers. In order to reach a disciplinary audience in our field, NECSUS now offers a platform for data papers from film and media studies scholars.

A data paper is usually understood as a publication of a documented data set (this can also be a digital collection of sorts) on a specific topic with an associated documentation. This documentation describes how, when, and why the data was collected and what the data set contains. The documentation might also contain a critical reflection on the selection of data, the questions one could raise in conjunction to the specific data set, and suggestions for future research. The documentation gives the context for the data itself, which is an integral part of the publication. The data should be published under an open license (usually CC-BY) in order to also facilitate and encourage the reuse and expansion of the data sets.

In the humanities, data is challenging; it is often incomplete and sometimes contradictory; it might be ambiguous and it sometimes resists classification. Data also often requires context to be properly understood. Yet again, if we use digital methods or if we are aiming to build a larger data pool for research, we need data of a certain quality. Constructing these data sets is a laborious and complex task that remains mostly invisible. It also requires specific skills and knowledge that is not yet fully established in the field, so sharing such information is also a way of anchoring these practices within our field.

The new section Data Papers contributes to the visibility of data work in our field. This section thereby honors and makes visible work that has been often below the radar and thus invisible for academic recognition – especially for emerging, early career scholars. Too often distinctions and honours in media studies – be it on the job market, regarding prizes, or in procedures of promotion and tenure – are based on traditional markers, most often publication in specific outlets. Moreover, data papers are often multi-authored collaborative publications. By including data papers in NECSUS, collaborative scholarly research will also

gain additional momentum, which we see as a welcome challenge in our fields. It allows staff that often gets no credit in the humanities (data and IT specialists, student assistants, early career researchers, infrastructural staff) to be credited as part of the work in a larger team.

The data paper section, like the audiovisual essay section, is curated and can consist of one or more papers for each issue of NECSUS. In the future, we hope to highlight a variety of approaches in terms of the methods used, but also regarding the topics and the sources. Data papers can also be a way of getting to know a new field. We are happy and proud to kick off the new section with the publication of a data paper by Skadi Loist and Zhenya Samoilova that is the culmination of a multi-year research project on the circulation of films on the festival network. It has become something of a truism to claim that film festivals provide an alternative circuit in which specific films (so called 'festival films') find their distribution. This data paper provides the groundwork for further discussion of this question, and it allows us to pose and answer many others.

The data set also discloses the way that the data has been constructed, which was a long and complicated process. To anyone who thinks that one only needs a good plan and then to stick to it almost religiously, this is a must-read because it discloses how complicated such a process is. The data paper – and the data sets that are part and parcel of the paper – allows us to understand and retrace the decisions and ambiguities that are necessarily part of such collections. Transparency, one could argue, is one of the main goals that data papers strive for, so researchers can re-use the data but also understand and amend, if necessary, for other purposes. Data papers guarantee reproducibility of results, but they also facilitate other uses of the data, some of which the original makers did not think about.

The new section understands itself as an ongoing experiment to explore our fields of expertise and the formats we critically reflect and communicate in and about our research. As our research (environment) becomes increasingly digital in manifold ways, from the way we access information to the channels of publication, we need to take the shape of data more strongly into account.

Authors

Alexandra Schneider is Professor of Film and Media Studies at Johannes Gutenberg-University Mainz. Her fields of expertise include film historiography with a focus on amateur media and format studies. She is one of the co-founders of NECS and is currently working on two film historiographical projects with computational components.

Malte Hagener is Professor of Film and Media Studies at Philipps-University Marburg and the managing director of the Marburg Center for Digital Culture and Infrastructure (MCDCI). His research interests concern film historiography, digital methods, and the new infrastructure of academic knowledge production. He is a co-founder of NECS and a founding editor of NECSUS.

References

Couldry, N. and Meijas, U.A. *The costs of connections: How data is colonizing human life and appropriating it for capitalism.* Stanford: Stanford University Press, 2019.

D'Ignazio, C. and Klein, L., *Data Feminism* (Cambridge, MA: MIT Press, 2020).

Es, K. van and Verhoeff, N.(eds.), *Situating Data: Inquiries in Algorithmic Culture.* Amsterdam: Amsterdam University Press, 2023.

Gitelman, L (ed.). *'Raw data' is an oxymoron.* Cambridge: MIT Press, 2013.

Kitchin, R. *The data revolution: Big data, open data, data infrastructures & their consequences.* London: Sage, 2014.

Zuboff, S. *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* London: Profile Books, 2019.

Notes

[1] A few seminal publications on the question of data and its uses include Gitelman 2013, Kitchin 2014, Couldry & Meijas 2019, Zuboff 2019, D'Ignazio, C. and Klein, L. 2020, Van Es & Verhoeff 2023.