# The Digital Transformation – Implications for the Social Sciences and the Humanities

*Harald Kümmerle*
*Franz Waldenberger*
*(Eds.)*

Deutsches Institut für Japanstudien

Miscellanea 21

**Deutsches Institut für Japanstudien**

Jochi Kioizaka Bldg. 2F
7-1, Kioichi
Chiyoda-ku, Tokyo 102-0094, Japan
Tel.: (03) 3222-5077
Fax: (03) 3222-5420
e-mail: dijtokyo@dijtokyo.org
homepage: http://www.dijtokyo.org

Direktor: Prof. Dr. Franz Waldenberger

Erscheinungsjahr: Oktober 2020
Erscheinungsort: Tokyo
Herstellung: IUDICIUM Verlag GmbH, München
ISSN 0941-1321

# Contents

Introduction
*Harald Kümmerle and Franz Waldenberger* . . . . . 7

*Part I: Technological developments and methodology*

Research Design in the Digital Age: Implementing Analysis of large Japanese Text Corpora using Topic Models on the web-platform TopicExplorer
*Christian Oberländer* . . . . . 13

E-Learning and Gender: A Proposal for Three Types of Collaboration
*Atsuko Sano* . . . . . 20

*Part II: Research data and infrastructure*

Sharing Social Science Data in Japan: JSPS's Plan for a Federated Data Catalog
*Yukio Maeda* . . . . . 27

How e-infrastructure could transform research activities
*Miho Funamori* . . . . . 32

*Part III: New research objects and perspectives through interconnectivity*

Insights into the Digital Transformation From the Fields of Medicine and Healthcare in Japan
*Susanne Brucksch* . . . . . 43

Politics in the Age of Connective Cynicism
*Fabian Schäfer* . . . . . 50

*Part IV: Shifting boundaries of the social sciences and humanities (SSH)*

Collaborative and constructive research design in the digital age
*Cornelius Schubert* . . . . . 61

Social Sciences and the Humanities in the age of Big Data. What is at stake?
*Martina Franzen* .......................................... 66

Knowledge production and the role of SSH in the digital age
*Franz Waldenberger* .......................................... 72

Postscript: Comparative and historical perspectives on the digital transformation
*Harald Kümmerle* .......................................... 81

List of authors .......................................... 85

## List of figures

Christian Oberländer: Figure 1 .......................................... 16

Christian Oberländer: Figure 2 .......................................... 17

Christian Oberländer: Figure 3 .......................................... 18

Fabian Schäfer: Figure 1 .......................................... 51

# Introduction

*Harald Kümmerle and Franz Waldenberger*

The term digital transformation (DT) captures the changes in our personal, social, economic, political, cultural and academic life brought about by the rapid diffusion of information and communication technologies. With the increasing digital connectivity, the rise of platform economies and the cost reductions in artificial neural networks, the collection, processing and analysis of large amounts of Big Data have become feasible and widespread. The combination of Big Data and artificial intelligence (BDAI) is exerting a great impact on the natural, life, physical and engineering sciences in which research has generally been conducted with a more positivist attitude. Here, BDAI reconfigures the scientific research process as it reduces the reliance on a priori theories: Making plausible, but not yet proven statements based on interpreting collected data (so-called "abductive reasoning") now plays a prominent role in the formulation of new hypotheses and theory-building. As has been argued convincingly, this represents a significant epistemological change (Kitchin, 2014, pp. 6–7). In contrast, for the social sciences and humanities, representing areas of scholarship that are highly diverse in their philosophical underpinnings, it is less clear what the consequences of the digital transformation will be.

The contributions gathered in this collection shed light on this question. They are the outcome of the workshop "The Digital Transformation – Implications for the Social Sciences and the Humanities" that took place at the German Institute for Japanese Studies (DIJ) on 24 and 25 September 2019. Co-organized by the Nippon Institute for Research Advancement (NIRA) and the German Centre for Research and Innovation Tokyo (DWIH), the workshop brought together scholars from Japan and from abroad, many of the latter being specialists on Japan. This is why the Japanese case is given special attention in many contributions. Even there, however, the insights are not limited to Japan, but address phenomena and implications of the DT of general relevance, even though they may manifest themselves elsewhere under different conditions.

It makes sense to group the contributions, which showcase ongoing discussions according to their focus of interest. The first pair of articles deals with methodological changes that follow the technological developments in data processing and networking. Christian Oberländer presents TopicExplorer, a tool that has been developed at the University of Halle-

Wittenberg. It provides a Big Data approach to analyze large text corpora in a selective, yet systematic way. While its theoretical foundation (the so-called topic models) has been laid almost twenty years ago, it has come to realize its potential in SSH only recently. Using English-language Wikipedia articles as an example, Oberländer outlines the approach TopicExplorer projects have taken with Japanese-language corpora, prominently blog posts. Atsuko Sano addresses several aspects in developing e-learning tools taking the example of adult education programs related to gender policy. To Sano, empirically assessing and quantifying the knowledge bias with regard to gender issues is an important first step in the design of curricula. However, she also stresses that the digital gender gap needs to be addressed when developing and designing digital tools.

The second pair of articles relates to the storage, accessibility and sharing of research data. Yukio Maeda describes the strategy pursued by the Japanese Society for the Promotion of Science (JSPS) for the social sciences. Because mobility of academic careers across institutions continues to be lower in Japan than in many other countries, data sets that have been gathered by one institution are generally difficult to access for researchers with other institutional affiliations. The Japanese initiative takes international standards explicitly into account. On the technical side, JSPS is cooperating with the National Institute of Informatics (NII), the most important point of reference for the DT of science in Japan. Miho Funamori discusses how establishing a digital research infrastructure is transforming all phases and aspects of the research process. Data sets are not limited to those that are gathered for research, but research outcomes themselves have to be seen as valuable data. Bibliometric methods that consider citations to calculate impact factors – all too familiar for scholars in the natural and social sciences – are part of the resulting infrastructure. A question that remains is how such methods can be applied to the humanities, where publishing in languages other than English tends to be more common or even the rule.

The third pair of articles discusses new research objects and perspectives that should be taken into account, resulting from the drastically increased interconnectedness of a digitalized society. Susanne Brucksch addresses the increasing importance of automatization in the Japanese health system. The deployment of tele-monitoring or telehealth in general allows for cost savings in health care facilities, generates synergies between healthcare providers in different locations and improves services in peripheries. However, the successful implementation of digital technologies crucially depends on how well their materiality and social embeddedness is taken into account. Fabian Schäfer investigates the transformation of the social sphere, which is prominently influenced by algorithms in social networks. During Japan's electoral campaign in 2014, bots active on Twitter engaged in verbal abuse and hate speech against female politicians in a fashion that imitated human users from the political right. Notably, the emerging algorithmic-connective sociality creates collective cynicism, which is inherently advantageous for populists. Schäfer's contribution thus vividly demonstrates how transformations in the digital space influence perceptions of the material space and vice versa.

The last three articles explore how the boundaries of the SSH are shifting in the context of the DT. Cornelius Schubert addresses the history and possibilities for cooperation between social scientists and computer scientists. These two groups first came into contact in the 1980s when software engineers tried to find out why products they developed were only hesitantly adopted. Since then, collaboration has suffered from the tendency of one party considering the other to merely be a service provider. While the DT may seem to give an inherent advantage to computer scientists, Schubert argues for a symmetrical approach in the design of collaborative projects between social scientists and computer scientists. Martina Franzen looks at how the growing relevance of Big Data is affecting the boundaries and cooperative as well as competitive relationships between science, business and civil society. While recent open data policies try to ensure access to research data, they do not address data collected by private platform companies. The asymmetric accessibility and the related issues of data sovereignty constrain academic research especially in the social sciences. The implied shifts in knowledge production brought about by the digital transformation and their implications are yet to be fully understood. Franz Waldenberger, too, looks at the impact of DT on the SSH, applying the conceptual framework of societal knowledge production. Taking a cross-disciplinary perspective, he points to the special role of SSH for a sound governance of data-driven knowledge production in the digital age. However, he also stresses that disciplinary boundaries are in many ways obstructing the SSH to fully perform their governance function.

In the postscript, Harald Kümmerle argues for a comparative and historical approach when studying the impact of the DT. Comparative research would be able to apply and contribute to conceptualizations of *digital spatiality.* Our understanding of the present "data revolution" would also greatly benefit from insights gained by studying *data acceleration regimes* found in historical research.

### *Reference*

Kitchin, R. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1: 1–12. doi:10.1177/2053951714528481

# Part I:
# Technological developments and methodology

# Research Design in the Digital Age

## Implementing Analysis of large Japanese Text Corpora using Topic Models on the web-platform TopicExplorer

*Christian* Oberländer

### 1. Large text corpora and topic models for a variety of research strategies

Text mining of Japanese texts using topic models is at the core of the TopicExplorer project that was established at the Department of Information Science and the Institute of Political Science and Japanese Studies of Halle University. The aim of this project is to make the application of topic models to large Japanese text corpora easily accessible for researchers in Japanese studies. Applying topic models in this way can support the implementation of different research strategies including content analysis, discourse analysis and web-ethnography. This contribution will briefly describe the motivation and first results of the TopicExplorer project.

### 2. Advantages of using topic models for social science research on Japan

Why should social scientists interested in Japan want to apply text mining using topic models? After all, computer programs that support quantitative content analysis have been around for many years now. However, for addressing almost any given research topic, researchers in the social sciences and the humanities today can access larger amounts of textual data than ever before. This digital data exceeds in volume and precision any sample of texts taken from analogue archives. These text corpora are often interesting precisely because they contain texts by numerous authors. They therefore provide a multitude of perspectives on a particular topic. However, frequently these corpora are so large that they cannot be processed comprehensively by humans. In addition, not all texts are equally interesting to researchers: while some documents may repeat mainstream positions, others may contain extreme or even abstruse views. Topic models can help in addressing these problems.

A second reason why social scientists should be interested in topic models is their ability to identify more complex structures of meaning in a given text corpus. Topic models can concentrate and structure large text corpora into often unexpected topics. They achieve this by analyzing the word distribution in the documents without the need for manual annotation. The topics identified in this way frequently offer social scientists focused access to content-saturated and interesting texts that are located between the mainstream and extreme statements. Explorative text mining with topic models can thereby make large text corpora interesting reading material for researchers. Therefore, topic models can be an important point of departure for formulating research questions, choosing an adequate research design and even implementing a particular research design.

Given the potential usefulness of topic models, it seems surprising that they are not used more widely in the social sciences. But for already more than ten years, projects have confirmed that two barriers exist that hamper cooperation between social scientists and computer scientists in the application of topic models. First, there is the problem of technical access to the results of topic models for social scientists. Second, there is the lack of confidence that social scientists have in the results of topic modelling.

Concerning the first barrier, i. e., that topic models do not appear easily accessible, the statistical and mathematical foundations of this method are obstacles that account for this. Even today, literature about the application of topic models in text mining still appears rather small compared to other methods of the qualitative social sciences. In addition, software applications needed for adjusting and implementing such methods often require technical knowledge. The second barrier is linked to the first one. Topic models calculate latent themes within texts not according to semantic aspects but follow only criteria based on information theory. There are few software solutions that facilitate the efficient transfer of such probabilistic inferences about word distributions in documents into hypotheses about their content. Therefore, we need analytic tools that allow users to draw meaningful conclusions from the calculation results of topic models.

These were the reasons to initiate the TopicExplorer project building a web platform with the aim to make the application and, in particular, interpretation of topic models readily available for social scientists working on Japan.

## 3. Developing the web-platform TopicExplorer to facilitate "blended reading"

In the process of developing the TopicExplorer software, it became clear that in the age of digitalization, the dualism of interpreter and text is muddled by the immense amount of digitally available data and the new electronic method of analysis. The reason for this is as follows. While social scientists often do not have background knowledge of text mining and probabilistic modelling, in order to apply these methods, they must at least generate from the original text data and the topic model calculated from it a plausibility structure for the interpretation of the topics suggested by the topic model. While this should ultimately lead to a better understanding of the content of the text corpus as a whole, at the same time, it must not suggest wrong conclusions.

From early on in the TopicExplorer project, we spent much time discussing how to address this problem. Finally, we decided that beyond the topic model-based analysis of the content of a text collection, the software must also support users in critically counterchecking their interpretations of the topic modelling results. Most importantly, the software was designed to allow users to track proposed topics back to the relevant original documents. Such backtracking has proven particularly useful not only because users can review their interpretation of the topics calculated probabilistically, but also because this process can lead them to particularly content-rich text passages that they can then integrate into their scientific arguments. The software also provides ready access to the information needed to correctly cite such text passages.

This approach led us to a research method that some call "blended reading". "Blended reading" is a procedure that integrates the advantages of semiautomatic text analysis with those of classical text reading, aiming to maximize the potential of analyzing large text corpora. The term "blended reading" is derived from Franco Moretti's differentiation between two modes of reading, that is "close reading" and "distant reading". The concept of "blended reading" tries to integrate both modes in order to combine human and computer-based abilities to achieve the best possible outcome. In our application, "blended reading" requires the iterative link of corpus analysis using topic models – "distant reading" – with the examination of individual texts – "close reading" – from the relevant topics in the corpus.

In addition to the analytic content, technical requirements were also difficult to meet. The interactivity of the analytic process posed great challenges to the efficiency of the software components that had to be newly developed. The pre-processing of texts, the modelling of topics and the linking for the calculated topics to the individual documents are complicated workflows that can take between several hours and several days to calculate for a given text corpus. All steps that consume relatively more time to complete were therefore organized into a separate software unit. Since the text corpus as well as the topic model calculated from it encompass large amounts of data, efficient algorithms for search and analysis had to be developed. These algorithms were then combined with specialized visualization techniques for topic models into one interactive, web-based user surface forming another separate software unit.

## 4. TopicExplorer application

The TopicExplorer system in its current design comprises several different functions. However, "blended reading" is at its core. This is realized through two different views on the text corpus. Figure 1 shows the user surface of TopicExplorer from the example of a collection of the 10.000 longest articles of the English-language Wikipedia. The topics are presented as word lists. Similar topics are placed close to each other. In the figure, for example, the four yellowish topics next to each other are all related to entertainment. Users can move the sliding bar to navigate through the entire spectrum of topics without encountering breaks. That is why the colors can be applied to the topics in a spectral manner from left to right. The color code supports quick visual referencing of topics in the other views of TopicExplorer. For a given

**Figure 1**

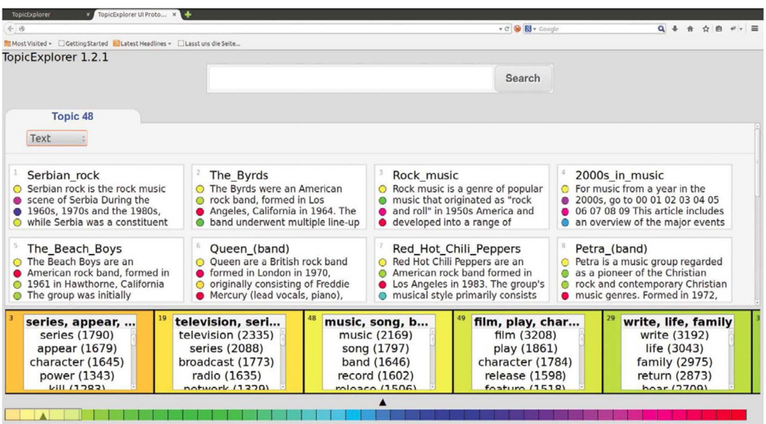
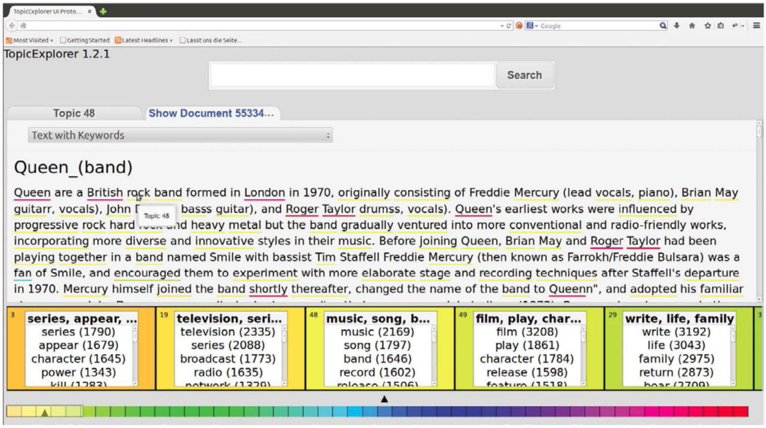


**Figure 2**

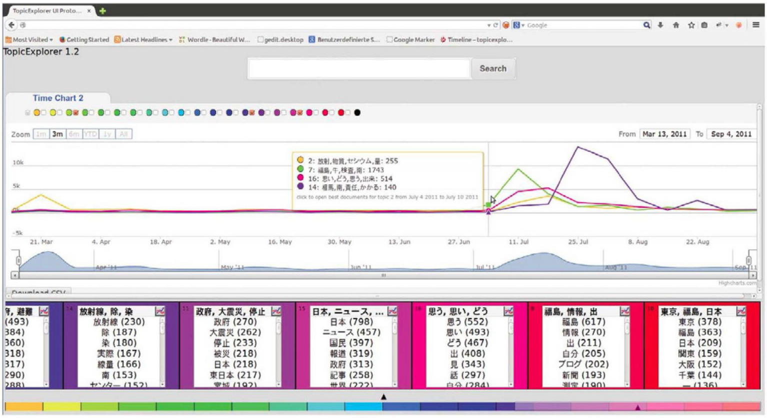



topic, the browser tab provides a document ranking exhibiting the most important documents for that topic: in Figure 1, topic 48 located in the middle concerns music. Each box represents a document. As a representation of the document, the box contains a title, the beginning of its text, and four colored dots that indicate the four most dominant topics in the text. The topic referenced by the color can be explored by moving the mouse over the dot. Clicking on the dot, the respective topic moves to the middle of the browser window.

When clicking on the title of an individual document, for example "Queen (Band)", a new tab will open (Figure 2). This view shows the full text of the document. Its words are assigned by colored underlining to the different topics that were generated by automated topic modelling. Through this view, the result of the topic analysis can be directly verified using the document content. As in the former view, here too, the colors assigned to topics can be checked by moving the mouse over them. Clicking on an underlined word moves the respective topic to the center of the browser view.

By clicking on the time-line icon of a topic (in the colored box of that topic on the top of the right side), the timeline view of TopicExplorer is opened for the respective topic in a new tab. Figure 3 presents an example from a corpus of Japanese blogs mentioning the word "nuclear power plant" (*genpatsu*). Timelines for other topics as well as the average of the timelines of all topics as a comparison can be interactively displayed via checkboxes. The legend displays the most important words of a given topic for that week to which the cursor has been moved. Through this, the

**Figure 3**




changes within a given topic can be followed interactively. Clicking on a week in the timeline opens a document ranking in a new tab that shows the most important documents for that week. Through this, timelines give insight into when and in which order topics reach a peak, for example.

## 5. Conclusion

"Blended reading" as introduced above is a possible best practice for using a digital tool – in this case: text mining for analyzing Japanese text data. By balancing the bird's eye view of the digital tool with the detailed analysis of conventional reading, it aims to facilitate a better judgement. In order to validate and control the results generated by computer-based analysis, the possibility of referencing the original texts in detail must be always secured. Referencing individual texts aims to ensure that there are no open contradictions between close and distant readings, which constitutes a kind of quality control. However, both perspectives do not just coexist but can be made to interact. For example, findings in close readings can motivate a recalculation of the topic model with changed settings. By creating a maximum of transparency, a process oscillating between conventional and digital research methods might be the most fitting way to integrate the new digital tools.

## Further Reading / Sources

Hinneburg, A. and C. Oberländer. 2014. Getting the Story from Big Data: Interaktive visuelle Inhaltsanalyse für die Sozialwissenschaften mit dem TopicExplorer am Beispiel Fukushima. In: Bubenhofer, N. and M. Kupietz (eds.). *Visualisierung sprachlicher Daten. Visual Linguistics – Praxis – Tools*. Heidelberg: Heidelberg University Publishing, 269–299.

Stulpe, A. and M. Lemke. 2016. Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung. In: Lemke, M. and G. Wiedemann (eds.). *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer VS, 17–62.

# E-Learning and Gender

## A Proposal for Three Types of Collaboration

*Atsuko Sano*

The expansion of e-learning, or the digitalization of education, may allow for the following three types of collaboration with gender-focused aspects. This paper addresses Collaboration 1, presented at workshop at the Deutsches Institut für Japanstudien (DIJ), and then goes on to propose Collaborations 2 and 3, which were formulated based on the DIJ workshop and following conference.

### Collaboration 1: Collaboration between e-learning and the analysis of its learning record data of local governmental staffers tasked with gender equality

At the workshop, I explained an e-learning project and its learning record data at the National Women's Education Center of Japan[1] (hereinafter called "NWEC"), where I work as an e-learning specialist. Special mention was made of the potential collaboration between e-learning and the analysis of the learning record data of local governmental staffers tasked with gender equality. Indeed, the NWEC is the sole national organization in Japan tasked with promoting gender equality through education, and has integrated e-learning into its five-year mid-term plan since 2016.

The result of this was the blending of e-learning into the NWEC's annual group training. The group training is intended for staffers of local governmental bodies tasked with gender equality. The level of background knowledge held by the training participants seemed to vary greatly, making it difficult for us to formulate a training plan. To address this challenge and account for the knowledge gap, we asked the participants to take a mandatory e-learning course before the group training. This established a common baseline upon which the group training could be organized. Using this baseline, we aimed to enhance the effects of the training. The e-learning course explained the basic terms, background, and history of gender equality, and was expected to take an hour or two.

[1] https://www.nwec.jp/en/index.html

Upon completion of the group training (including the prerequisite e-learning course), the NWEC obtained data from the course clarifying the challenges in rolling out further e-learning programs. Specifically, analyzing the learning record data and quantifying the background knowledge gap, as well as leveraging responses to the questionnaire at the end of the e-learning course, allowed us to consider possible explanations for the gap and how to respond to it.

Two specific outcomes were observed: general indifference toward international initiatives in the field (CEDAW[2], SDGs[3] and CSW[4]); and a clear gap/low awareness regarding knowledge about gender equality among the staffers of local governmental bodies. The former issue was addressed by updating the training content to add language and incorporate photos archived at the NWEC. Meanwhile, the results of the questionnaire determined that the latter issue was likely derived from the following three factors:

1. The majority of staffers assigned to gender equality tasks at local governmental bodies had not been exposed to gender equality matters previously;
2. In most cases, they learn from senior staff through on-the-job training, having few opportunities to participate in comprehensive training on gender equality; and
3. Due to periodic job rotation, there are very few veteran staffers who stay in the same capacity for more than five years.

Based on the above results, the NWEC plans to develop a more comprehensive e-learning course specifically targeting staffers of local government bodies being assigned to gender equality tasks for the first time. The course will furnish them with more thorough knowledge.

## Collaboration 2: Collaboration between e-learning and the digital gender divide

The second proposed collaboration is between e-learning and the digital gender divide. In order to roll out digital education effectively, it is desirable to formulate measures and policies based on information around how users interact with ICT (Information and Communications Technol-

[2] The UN's Convention on the Elimination of All Forms of Discrimination Against Women
[3] The UN's Sustainable Development Goals (Goal #5 pertains to gender equality)
[4] The UN's Commission on the Status of Women

ogy) such as digital tools and the internet. Indeed, the W20 (Women20), one of the advisory bodies for the G20 regarding women's digital inclusion, has issued the policy brief, "The Digital Gender Gap,"[5] where it outlines three components of the digital gender divide:

> (1) access [to] and use of digital technologies and the internet; (2) development of the skills needed to use digital technologies and to participate in their design and production; and (3) advancement of women to visible leadership and decision making roles in the digital sector. (p. 2)

Based on the above, I wonder whether it is possible to quantify—and hence clarify—the digital gender gap through data analysis. With this, it is important to be fully aware how women are digitally "disadvantaged" and implement measures accordingly. The digital gender gap can cause further gaps not only in information and skills, but also in income. To bridge the gap, it is advisable to quantify, and thereby clarify, the digital gender gap first of all. Such quantification will then enable us to come up with an appropriate response and thus provide more effective learning opportunities, including e-learning targeting women in particular.

What specific measures can be considered once quantification has taken place? One such measure is deploying e-learning tools on mobile devices. Indeed, such necessity is suggested in answers to the NWEC's e-learning questionnaire as touched upon in Collaboration 1. Some of the answers indicate that respondents spend more time on e-learning after work or at home, rather than during business hours. Further, these answers showed that a leading cause of this pattern was the stringent internet security measures within the workplace, which limit access to e-learning tools outside the network. If that is the case, the availability of e-learning tools on mobile devices may benefit learners by ensuring that the necessary tools are constantly available.

Still, a question arises: Is there a gender gap in terms of availability of mobile devices such as smartphones? Is how they are used considered when calculating the digital gender gap? Here, I would like to reiterate collaboration between e-learning and the digital gender divide. For example, if it is true that women primarily use mobile devices as a means of communication, it will be necessary to provide e-learning opportunities where they are exposed to information about gender equality in closed communication spaces. If the data suggests that more men are engaged in online games, it might be more appealing to men if the contents are gam-

[5] Retrieved, June 20, 2020 from https://w20japan.org/pdf/digital_equity%20_policy_biref_w20%20Japan_final.pdf

ified. As it stands, though, there is insufficient data at present on the deployment of e-learning tools on mobile devices regarding the digital gender gap. This challenge is also closely related to the lack of sufficient gender statistics in Japan.

Going further, rolling out digital applications in consideration of the digital gap among users has applications for promoting lifelong learning as well. In this case, we can consider what kind of accommodation or design is necessary to provide people from all walks of life with lifelong learning in a digital manner. Enabling each and every citizen to have equal access to the digital environment would serve as a prerequisite for developing citizen science, where citizens participate in and contribute to knowledge creation. This suggests that addressing the digital gap is an important challenge not only in promoting gender equality but also in guaranteeing the right to education in the digital era, as well as in developing civil societies.

## Collaboration 3: Collaboration between digitalization impacting all areas and the gender perspective border-crossing into multiple areas

Digitalization affects various areas and promotes interdisciplinary collaboration. As demonstrated by Collaborations 1 and 2, such an interdisciplinary characteristic of digitalization has an affinity with the promotion of gender equality.

According to the second German gender equality report, "Zweiter Gleichstellungsbericht,"[6] while stating that "digitalization offers huge potential (Die Digitalisierung bietet große Potenziale)," it also points to negative aspects of digitalization, including cyber harassment and the need to protect users from it. This part of the nature of digitalization, having both positive and negative aspects, requires further discussion in Japan, as well.

Digitalization has both positive and negative impacts on various areas and accelerates interdisciplinary collaboration, while the promotion of gender equality itself is interdisciplinary in nature. In order to ensure a win-win relationship between the two, I would like to encourage Japan and Germany to seek ways to further collaborate in this endeavor.

[6] Retrieved June 20, 2020 from https://www.gleichstellungsbericht.de/de/topic/2.zweiter-gleichstellungsbericht-der-bundesregierung.html

# PART II:
# RESEARCH DATA AND INFRASTRUCTURE

# Sharing Social Science Data in Japan: JSPS's Plan for a Federated Data Catalog

*Yukio Maeda*

Sharing research data requires infrastructures that curate, process, preserve, and disseminate machine readable data with enough information for secondary analysis. Some progress has been made in the past two decades in Japan, which were primarily initiated by a few research universities. However, neither the government nor the funding agencies have set any policy for data sharing in the humanities and social sciences in the past several decades. In this essay, I briefly review the current situation of social science data sharing in Japan, then introduce JSPS's project for strengthening data sharing infrastructures for the humanities and social sciences with a special emphasis on its data catalog project.

## Background

In Japan, the long-term management of social science research data has been mainly left to the responsibility of individual researchers. Lacking an institutional mechanism to share research data, valuable data tend to remain exclusively in the hands of the original principal investigators and their students. In the case of survey data with representative sampling, many datasets have been preserved and shared on an ad-hoc basis, depending on the practice of academic apprenticeship of the research groups that collected the original data. The situation surrounding data sharing for survey data in Japan has improved over the past 20 years, but many datasets are still maintained individually or simply lost.

Institutional support for sharing aggregate data is even weaker. The government publishes its data as a part of their policy or administrative activities, not for promoting research. Many researchers collect and compile published aggregate statistics from the government to make them usable for their specific research needs. Compared to the situation of survey data, currently no institutional mechanism exists to share the data that researchers have compiled and edited from the government sources. Furthermore, Japan lags in producing documentation for research data in

English, which makes collaboration between Japanese and foreign researchers difficult.

All these problems make academic research less efficient, which creates concerns about Japan's competitiveness in the humanities and social sciences. Policymakers and the academic community are aware of the weakness of research infrastructures in Japan. In April 2017, the Task Force to Strengthen Basic Science Capability hosted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) summarized its discussion as follows:

> It is necessary to build a data platform in the field of social science. … This data platform should be neutral and sustainable. It provides a framework through which research data in the social sciences are curated, preserved, shared and analyzed.[1]

Likewise, in June 2017, the Science Council of Japan issued "Recommendations from the fields of the humanities and social sciences," which state that:

> Building databases of historical and public documents, survey data, etc. is essential not only for the further development of the humanities and the social sciences, but also for establishing the foundation of international collaborative research.[2]

Against this backdrop, the Japan Society for the Promotion of Science (JSPS) initiated the Program for Constructing Data Infrastructure for the Humanities and Social Sciences starting in the Fiscal Year 2018. Its funding period is five years from FY 2018 to FY 2022, with an annual budget of 200 million Japanese yen (roughly 1.65 million euros and 1.83 million US dollars applying exchange rates of November 2019). While it is a large-scale project in the domain of the humanities and social sciences, it is not large enough to build an institution filled with professional staff members specialized in data sharing. Instead, JSPS tries to encourage, coordinate and institutionalize existing efforts made by research centers and institutions located at universities.

1 http://www.mext.go.jp/a_menu/kagaku/kihon/1384933.htm. Translation is mine.
2 http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t242–2e.pdf The excerpt is from the official English version of the report.

## Overall framework of the project

The project started in April 2018. In July 2018, JSPS called for a tender to choose a small number of data centers at academic research institutions to cooperate with JSPS for data preservation and sharing. After a rigorous evaluation and deliberations at the steering committee meeting, four data centers were selected in the fall of 2018. Then, around the same time, to spearhead the preservation and dissemination of research data, JSPS installed the Center for Constructing Data Infrastructures for the Humanities and Social Sciences as a headquarters of this project. JSPS also made an agreement with the National Institute of Informatics (NII) to secure NII's technical expertise necessary to achieve the goals of the project.

The headquarters itself does not deal with research data. Its expected role is 1) to promote data sharing among researchers, 2) to locate past studies that have not been preserved and shared, 3) to provide guidelines on data sharing and related issues to the academic community, 4) to coordinate the activities of the designated data centers and 5) to set up a federated data catalog to make research data available to researchers in Japan and abroad.

The data centers are responsible for curating research data from the original data collectors and for preserving them.[3] After the necessary processing and editing, the data centers make those data available to the academic community for secondary analysis. In order to increase the visibility of social science data collected in Japan, it is also important to prepare related documentation in English and other languages if deemed necessary. Furthermore, adding value to the existing data is also an important activity. Depending on the specific characteristics of data, it may be desirable to harmonize multiple datasets for comparative analysis or create cumulative data files for longitudinal analysis. In the case of aggregate data from official statistics, harmonization across time also adds value for secondary users.

In order to enhance the activities mentioned above, JSPS encourages the data centers to employ professional staff who work on research data. Professional staff can be data archivists, data librarians, data managers or IT specialists who develop tools for data sharing, or any other member who works exclusively on data management and sharing. The lack of funding for staff members is one of the reasons why data sharing has not been widely practiced in Japan. Through providing budgets to hire professional staff members, JSPS tries to promote data preservation and sharing among the academic community.

[3] The data centers can work on the datasets that they themselves collected for research in the early phase of the project.

## A plan for a federated data catalog

In order to make social science research data findable and accessible to the broader academic community, JSPS is developing a federated data catalog with the NII. The catalog is designed to enable users to search for research data in the social sciences across the four designated data centers as well as other institutions with a single interface.[4] For that purpose, we need to ask each data center to provide a description about their data in a uniform format. As of November 2019, the plan is as explained below.

For any catalog, it is necessary to set the rule about how the object (data) being catalogued is described. The description about data is called metadata. We adopted DDI 2.5 as a metadata standard. DDI is an abbreviation for the Data Documentation Initiative, which is a widely used metadata standard among the social science data repositories worldwide.[5] In order to design the federated data catalog, we compared the data catalogues of several well-known data archives such as ICPSR, GESIS, UKDS, etc., and selected typically used metadata elements. In order to make data search more efficient, some of the metadata elements need to be made consistent with the specific rules of descriptions determined in advance. The list of specific words for metadata elements is usually called "controlled vocabulary." The DDI controlled vocabularies for the unit of analysis, the sampling procedure and the mode of data collection have been translated into Japanese. As for keywords for research topics, the controlled vocabulary and its definition for CESSDA topic classification is now being translated.[6]

The federated data catalog ideally harvests metadata from each data center automatically to avoid mistakes arising from manual works and to keep the information up to date. Each of the four data centers has its own strength and weakness, and their existing data catalogs have been developed locally. JSPS is negotiating with each data center to make its local catalog compatible with OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), through which the metadata are mechanically harvested and then uploaded into the federated catalog. In the long run, we plan to offer an interface through which institutions other than the designated data center can upload their metadata into the federated data catalog as we expect that there is a vast number of research data that can be shared.

[4] In October 2019, one data center in the humanities was also selected following a call for tenders. The project now has five designated data centers.

[5] https://ddialliance.org/

[6] CESSDA stands for Consortium of European Social Science Data Archives. https://www.cessda.eu/

The catalog will be bilingual to make datasets findable and accessible to those who are not fluent in the Japanese language. For that purpose, metadata need to be prepared both in Japanese and in English. The descriptions in English may not be as rich and comprehensive as those in Japanese, but need to be detailed enough to allow users abroad to search the data effectively. We are negotiating with institutions abroad that maintain international data catalogues to confirm whether they can harvest the metadata from the JSPS catalog to increase the visibility of research data from Japan.

Currently, the test version of the data catalog is being developed internally. We hope to make it open during the third year of the project (FY 2020). It is the first attempt in Japan to gather information on research data across research institutions. Though the knowledge of the availability of research data is not a secret, that knowledge is usually held by domain specialists. Lacking an institution for data sharing across institutions and the boundaries of academic disciplines may have hindered researchers from finding data that could serve their research needs. The federated data catalog will make the existing research data more visible to those who are unaware of the data. We hope that the JSPS's data catalog will contribute to the further development of research by making social science data findable and accessible for researchers in Japan and abroad.

# How e-infrastructure could transform research activities

*Miho Funamori*

## 1. Introduction—E-infrastructures impacting peoples' lives

With the emergence of the World Wide Web around the turn of the millennium, the Internet has pervaded the world. People are living in a digitally connected world and the Internet, as well as the e-infrastructures established around it, have become indispensable for daily life. People are searching and extracting information, making purchases, using online analytical tools, enjoying music, games, and other forms of entertainment, and more importantly, connecting with other people through e-mails and SNS services.

For academe, too, the Internet and e-infrastructures have become indispensable. Various research tools have already been built along existing research flows. Future research processes can (at least ideally) probably be understood as follows: Each research step uses texts, numbers, images, or other kinds of digital content, and doing research can be seen as an activity that processes and transforms digital content over multiple steps. The research tools then will be built along the research flow as processors for these conversions. In this scenario, it can even be said that academe will be more reliant on—or an even heavier consumer of—e-infrastructures than the general public.

E-infrastructures for the public were in the first place structured to imitate certain parts of the real world. Letters and telephones became e-mails and SNS tools. Newspapers and magazines became information sites. Stores became e-commerce sites. Records and TVs became streaming sites, and so forth. As time passed, some of the e-infrastructures have disrupted businesses or workflows from the physical world that they, in part, once imitated. Newspapers, music industries, and other content-providing industries are still struggling to find a suitable business model in the digital age. Some new businesses have emerged, for instance, linking and analyzing data to enable personalized marketing services.

It can be assumed that digitization in academe will give rise to similar developments: First, physical research processes are replaced by e-infrastructures. For instance, printed academic journals have largely become available as or turned into e-journals. Calculations done manually are now generally carried out with computers. In the first place, this was done to enhance the efficiency of the workflow. However, after a certain period, some disruptions have set in. To name just one: Chasing journal impact factors, citations, h-index scores, and other bibliometrics is one of the most extreme changes in academe. It is altering the value system of academe regarding what kind of research output academe considers good research.

This short paper attempts to sketch the emerging digital transformations (DTs) seen throughout the research process by examining the functionalities and impact of e-infrastructures used in each step of research. The various kinds of impact for each functionality are analyzed to identify patterns across the research process. Finally, the possible DTs in research activities are discussed.

Note: This paper uses the terms "e-infrastructures" and "e-tools" interchangeably. The former gives a rather big and complex impression, whereas the latter is used for small and handy e-infrastructures.

## 2. E-infrastructures used throughout the research lifecycle and their impact

There are tremendous amounts of ever-evolving e-tools for research activities. Many researchers working in education technology and digital infrastructures have tried to showcase these e-tools (Center for Open Science; UGCNETPAPER1; Zayapragassarazan). In many cases, researchers introduce the e-tools along the research process beginning with identifying the problem, proceeding to the research design, the collection and analysis of data, the interpretation, the presentation, and ending with the evaluation.

Here we examine the e-infrastructures used in each research process. Typical e-infrastructures used in each research process along with the functionalities of the e-infrastructures, especially enabled by digitization and the Internet, are listed below. Additionally, their impact is described in abbreviated form in parentheses. These kinds of impact are to be analyzed and interpreted in the next section.

The e-infrastructures and their impact listed in Table 1 are far from complete. However, almost all typical features are included; enough so that the possible DTs of research activities can be discussed and interpreted.

**Table 1: E-infrastructures used throughout the research lifecycle and their impact**

a) Identifying the problem, research design
- **Search tools**, especially for article discovery—Web of Science, Scopus, Google Scholar
  - ✓ Instant search among a massive amount of articles. *(efficiency, broad range, serendipity)*
  - ✓ Relevant articles recommended and citation information available. *(efficiency, impact, quantification)*
  - ✓ Knowledge maps available. *(efficiency, network)*
- **Reference management tools**—Endnote, Zotero, Mendeley
  - ✓ Effective organization of references, formatting of references. *(efficiency)*
  - ✓ Linking to data. *(efficiency, synergy)*
  - ✓ Sharing references with collaborators. *(efficiency, network, collaboration)*
- **Communication tools** for research discussions—e-mails, ResearchGate, Slack
  - ✓ Instant and simultaneous communication. *(efficiency)*
  - ✓ Communication between people who have never met before. *(efficiency, network, serendipity, outreach)*

b) Acquiring funding
- **Research grant application system**
  - ✓Digitalized. *(efficiency)*

c) Collecting and sharing data
- **Sensors**, sensor controllers, calibration tools
  - ✓ Automated, high-speed sampling. *(efficiency, Big Data)*
- **Life logs** on the Internet—Cookies
  - ✓ Acquiring people's behavioral data for the first time. *(efficiency, Big Data, personal data, personalization)*
- **Web questionnaires**—Google forms, SurveyMonkey
  - ✓ Questionnaire survey from a broad range of people. *(efficiency, broad range)*
- **Crowd-sourcing, citizen science**—Galaxy Zoo, Foldit, eBird
  - ✓ A broad range of research areas and leveraging human resources. *(efficiency, acceleration of research, broad range, public)*
- **Databases, data repositories**—Datacite, Dataverse, GenBank, ICPSR, UK Data Archive
  - ✓ Using data that you have not created. *(efficiency, acceleration of research, broad range, serendipity, multidisciplinarity)*
  - ✓ Comparative research across datasets. *(metascience, reproducibility)*
- **Data sharing**—Dropbox, Google Drive
  - ✓ Sharing data with collaborators. *(efficiency, openness, acceleration of research, network, collaboration)*

d) Analyzing data
- **Open-source codes**—GitHub, Jupiter Notebook
  - ✓ Sharing and collaborating on codes. *(efficiency, network, collaboration)*
- **Cloud computing platforms, SAAS**—AWS
  - ✓ Leveraging computing power on demand. *(efficiency, Big Data, computing power, license, on-demand charging)*

- **Disciplinary platforms**—NFFA, IMBE
  - ✓ Community portal, collaboration, sharing data and information. *(efficiency, network, collaboration, information base, community, platform companies)*
  - ✓ Analyzing tools provided. *(standardization, computing power, advanced tools)*
- **Artificial intelligence** (AI)—TensorFlow, Scikit, Microsoft Azure
  - ✓ Machine learning. *(AI, pattern recognition, research without theoretical framework)*
- **Linking data**
  - ✓ Analyzing combined datasets. *(broad range, multidisciplinarity, innovation, new research aspects)*
  - ✓ Adding persistent identifiers. *(identifier, linking information)*

e) <u>Organizing and interpreting outputs</u>

- **Figures, charts, tables, presentations**—Excel, Powerpoint, CAD
  - ✓ Drafting figures, charts, and tables. *(efficiency)*
- **Research data management**—OSF, DMPTool, DMP Online
  - ✓ Organizing and managing data. *(organizing, accountability, reproducibility)*
  - ✓ Planning data management. *(organizing, accountability, transparency)*
  - ✓ Time stamping and access control. *(integrity, security)*

f) <u>Dissemination of research output</u>

- **Preprint server, SNS tools**—ArXiv, ResearchGate, Twitter, Facebook
  - ✓ Sharing of research output at an early stage. *(swiftness, recognition, acceleration of research)*
  - ✓ Sharing of research output amongst informal circles including the public. *(swiftness, recognition, public, informal, new value system, linking with outside media)*
- **Journal platforms**—OJS, iThenticate
  - ✓ Open peer-review, post-publication peer-review, cascading peer-review. *(openness, swiftness, efficiency, new value system)*
  - ✓ Paper submission, peer-review, publication. *(efficiency, platform companies)*
  - ✓ Open access journals. *(openness, swiftness, new business model, author-pays model)*
  - ✓ Data journals. *(new value system, data science, data curators)*
  - ✓ Integrity check. *(integrity, accountability)*
- **Conference management platforms**—OpenConf, EasyChair, Eventbrite
  - ✓ Registration and online payment. *(efficiency, workflow)*
  - ✓ Paper submission, peer-review, conference proceedings. *(efficiency)*
  - ✓ Online programs, PRs, brochures. *(efficiency, linking with outside media)*
- **Data sharing**—Dropbox, Google Drive
  - ✓ Sharing data with the public. *(efficiency, openness, transparency, accountability, evidence-based, acceleration of research, network, collaboration, social problem solving, innovation, security, privacy)*

g) <u>Evaluation and impact</u>

- **Bibliometrics**—Citation, h-index, impact factor
  - ✓ Quantified impact. *(quantification, impact, evidence-based, publish or perish, chasing impact, world university rankings)*
- **Social impact**—Altmetrics
  - ✓ Automated collection of social impact. *(social impact, public, new value system)*

## 3. Interpreting the impact of e-infrastructures

The impact of the e-infrastructures identified in Table 1 (in parentheses) are included and categorized by different developments among e-infrastructures in Table 2. It should be noted that some kinds of impact appeared more often than others. The impact "efficiency" appeared in almost all features, which is understandable as e-infrastructures usually replace the physical workflow with electronic systems and make the workflow efficient, swift, and organized (A).

Once such e-infrastructures become effective in the workflows, and content has been gathered, the standardization and quantification stage begins. People try to set standards across e-infrastructures, starting to count and compare the practices. This, in turn, results in unpredicted consequences that we now see in academe. Research metrics—such as the publication number, citations, h-index, journal impact factor—have made it possible to compare academics in a quantitative and seemingly objective manner, which was not the case in the past. These metrics are easy to understand even for the lay public including governments. Thus, they have enabled the public to have a say in what is happening in academe. Governments have started to set goals for world university rankings and put pressure on universities and researchers.

The metrics are easy-to-understand tools for academics, too. Prior to the emergence of metrics, only experts in the specific discipline or subfield could evaluate which research was exceptional or not. Other researchers were not able to judge the quality of the research. As researchers are working in silos more than ever in the ever-evolving research sphere, this has caused serious problems. Even colleagues in the same department could not assess each other's ability, for example. Soon after researchers started to look at the metrics and judge their fellow researchers, however, these metrics became indispensable tools in faculty hiring and promotion processes. This practice has resulted in the so-called "publish or perish" phenomenon, causing research fraud and scientific misconduct along the way. Quantification through e-infrastructures created a completely new value system in academe (B).

The openness of the Internet and e-infrastructures, too, initiated developments in academe. The Internet allowed people to share content almost at no cost and instantly, resulting in massive content and information being available on the Internet. As anyone was free to search for any content, serendipitous discoveries were possible, which could never have happened in the physical world. It provided grounds for new research. This openness also enabled collaboration between researchers who work in different locations. Many international forms of collaboration evolved.

It also allowed communities on the Internet to form using e-platforms, serving as information bases and communication tools, providing an underpinning for research by making data and analytical tools available.

This openness also allowed the public to communicate and interact—sometimes also to collaborate—with academics. Either academics set the research agenda and include citizens as supporters (called "citizen science"), or citizens take the lead by setting the social issue to be addressed and draw on academics support. Of course, the interaction between academe and the public is not always cozy: As academe drastically expanded in scale after World War II and a never-ending competition in the realm of "science and technology" (S&T) evolved, S&T budgets have become such a burden on taxpayers' that academe was eventually forced to hold itself accountable to the public. The openness of the Internet served as a window that led to an increase in accountability, transparency and academic integrity.

At first, seemingly without any direct influence on research activities, service providers turned into platform companies. Today, computing power and research tools are provided over the Internet, journals are now being hosted on the publisher's platform. These companies offer their services through licenses and on-demand charges, which is a different model than charging for physical goods. As digital content can be easily made available on the Internet—which makes it difficult to charge access fees—new business models such as the freemium-model, author-pays models and others were invented. However, businesses are still struggling to find an appropriate income model in the open access world (C).

Although it may still be inconspicuous, e-infrastructures are also influencing the area of research: In the first place, e-tools were introduced to accelerate research. Automated sampling devices and analytical tools were developed. With the advent of the cloud, software came to be provided on the Internet (SAAS: Software as a Service) and researchers obtained access to high-computing power and advanced tools that they could not afford as individuals. Some numerical simulations were only made possible through such high-computing power. On the other end of the cutting-edge extreme, the Internet made resources available through which people from various backgrounds could work on multi-disciplinary topics, social problems, and new aspects of research. Policymakers started to hope for innovations through the openness of the Internet.

The massive amount of data accumulated through the Internet and e-infrastructures also opened the door for data-intensive science. People started to talk about the new possibilities of Big Data and actually started to work with it. Artificial intelligence (AI), using these massive data, opened up completely new ways of conducting research. The traditional way of conducting research was to set up a hypothesis based on a certain theory, and then test the hypothesis, while AI approaches the data through pattern recognition and draws conclusions from the patterns. No hypotheses or theories are needed. As academic disciplines, in general, are embodiments of certain theories, the research approach posed by AI is a possible disruptor in the way research is conducted. Still, as AI needs massive amounts of data for machine learning, which is still not available in many disciplines, it needs to wait for the impact it may have.

In addition to the evolvement of AI, metascience, the so-called research on research, is also gaining ground. In the past, researchers set up a hypothesis, collected data, analyzed it, and formed a research output; now, with many research outputs accumulated as articles and underlying data, researchers can also choose to analyze across similar research outputs and analyze the commonalities and differences. This is another new way of doing research in the digital era (D).

Even though not fully utilized in the research activities, e-infrastructures and the Internet collect personal data that could be analyzed. Some researchers have started to analyze Twitter and other transaction records. However, only a few researchers come close to tapping the full potential. Personal data can also be used to assist in research activities. Recommendations or personalized settings are examples. While recommendation services for research articles already exist, other services still need to be worked out (E).

**Table 2. Impact of e-infrastructures categorized by different developments in e-infrastructures**

| |
|---|
| **A) Developments of electronic systems**<br>– Efficiency, swiftness, organizing, workflow.<br>– Quantification, standardization.<br>– Identifier, linking information, linking with outside media.<br>– Reproducibility. |
| **B) Developments upon quantification by e-infrastructures**<br>– Impact, recognition, evidence-based.<br>– Publish or perish, chasing impact, world university rankings. |
| **C) Developments through the openness of the Internet**<br>– Openness, serendipity.<br>– Broad range, network, collaboration, outreach, informal.<br>– Information base, community.<br>– Public, accountability, transparency, integrity.<br>– Platform companies, license, author-pays model, new business model. |

**D) Developments in research area**
- Synergy, acceleration of research.
- Computing power, advanced tools.
- Multidisciplinarity, social problem solving, innovation, new research aspects.
- Big Data, data science, data curators.
- AI, pattern recognition, research without a theoretical framework.
- Metascience, new value system.

**E) Developments regarding the relation of persons and e-infrastructures**
- Personal data, personalization, privacy.
- Security.

## 4. Concluding remarks

In order to shed light on possible DTs in research activities, the e-infrastructures used throughout the research process were laid out and their functionalities and impact were examined. The impact for each functionality were grouped by different developments of e-infrastructures and interpreted. Although the impact for each functionality were extracted independently, the reorganized impact formed a coherent story of what is happening in the research sphere in the digital era.

Several DTs—seen as phenomena in which new value systems become established—can be observed. The quantified research metrics and the chasing these indices, as seen in the "publish or perish" model, is one notable disruption in academe. Another disruption observed would be in the business world in the struggle for new business models in the digital era that impact the broad resources that the Internet draws together. Researchers are working in larger groups with people from mixed backgrounds more than ever before. The availability of massive data and other resources, such as articles and e-tools, opens up new research areas such as data-intensive science, multi-disciplinary science, social problem solving, metascience, etc. Moreover, even though there is some way to go, personal and personalized data provide huge possibilities to find new research topics and new ways to navigate through the research process.

These symptoms are to be watched and strategically approached in the coming digital era.

## References

Center for Open Science. Official Website. https://cos.io/our-products/osf/ (accessed 31 October 2019).

UGCNETPAPER1. Official Website. https://ugcnetpaper1.com/application-of-ict-in-research/ (accessed 31 October 2019).

Zayapragassarazan, Z. 2014. ICT Tools for Research and Publications. https://www.slideshare.net/zprazan/ict-tools-for-research-and-publications (accessed 31 October 2019).

# Part III:
# New research objects and perspectives through interconnectivity

# Insights into the Digital Transformation From the Fields of Medicine and Healthcare in Japan

*Susanne* Brucksch

The digital transformation exerts wide influence on various parts of society as well as on the Social Sciences and Humanities (SSH), including the way research is conducted. I argue that the study of digital tools and their potential applications requires reflection within the social sciences to better understand the manifold implications of digitalisation while also paying sufficient attention to the mode of operation of digital technologies. Science and Technology Studies (STS) "focusing on the co-construction and intersection of technology and modern society" provide a promising research frame to address the process of digital transformation and its societal implications (Brucksch, Wagner 2016, 6). Basically, the digital transformation implies the "manipulation of digital information in 'computerised' forms". We need to understand digital technology as a human-made infrastructure, which is never just a neutral or value-free object "to which society simply responds" (Faulkner 2009, 14–15). Instead, the digital transformation should be "rather perceived as a process wherein social actors with varying visions, values, and concepts of usage inscribe their ideas into product designs and reconstitute a specific socioeconomic order within sociotechnical infrastructures" on many different layers (Brucksch, Wagner 2016, 6). I have suggested elsewhere to address (a) the material or physical dimension, (b) the social dimension, and (c) the regulatory, organisational and financial dimension alongside (d) the digital dimension in the study of domestic healthcare technologies (Brucksch, Schultz 2018). Building on these results, this commentary offers insights from the fields of medicine and healthcare in Japan to highlight some aspects of the relationship between digital technology and modern society.

The concept of digital transformation is frequently associated with processes such as digitalisation, virtualisation, and connectivity, all of which can be observed in the medical and healthcare fields. Connectivity implies that medical devices such as instruments, machines or apparatuses are no longer only located in their physical place of application. They are also situated in a digital sense with regard to the "data and information that is collected, stored, analysed and conveyed via those artefacts" (Petersen 2019, 41, 44). In other words, medical devices and other healthcare technologies are increasingly interconnected by information and communication technologies (ICTs). A good example is the Smart Healthcare House developed by the building company Sekisui House. It emerged from the concept of smart houses as energy-saving units. Another example would be monitoring systems (*mimamori sensā*) employed in elderly care by several healthcare providers. They are often embedded in assistive devices such as the Robot Assist Walker RT.1 (see Brucksch, Schultz 2018, 27–28, 37–40). These devices acquire, transmit and analyse user data such as bio-information or profiles of motion. While such monitoring systems primarily intend to serve users' needs, the surveillance function, which is not fully controlled by the user, raises concerns about privacy and data protection.

Medical imaging technologies and digitalised versions of patient records are prominent examples for the processes of virtualisation and connectivity in the medical context. Digital imaging contributes to the way human bodies are displayed, constructed and utilised during diagnosis and treatment. Japan shows a particularly high number of medical imaging technologies per capita, with ratios of 112 computed tomography (CT) scanners and 55 magnetic resonance imaging (MRI) units per 1,000 capita, far above the OECD averages of 27 and 17 respectively in 2017 (OECD 2019, 193). However, such "computerized tomography, ultrasound, PET scanners, and magnetic resonance imaging" entail "not photographic captures of reality but mathematically constructed representations of structures or metabolic functions" (Hogle 2008, 847). Related issues pertain to potential limitations in data acuity, availability of high-tech apparatuses, trained specialists or radiation protection in clinical workplaces. Recently, CT, MRI, and PET (positron emissions tomography) have become more equally distributed throughout Japan while many specialists such as pathologists or radiologists remain concentrated in metropolitan regions. Not surprisingly, this asymmetric distribution of devices and specialists has inspired many telemedicine projects such as telepathology and teleradiology over the past years (Matsumoto *et al*. 2015, 7–9; Park 2010, 34–35, 43). Another example from Japan is the Smart Cyber Operation Theatre, which is a "computer-assisted system based on dedicated information technology for guidance of the complex neurosurgical procedures and easy assessment of their results" at Waseda University (Iseki *et al*. 2012, 1). This example illustrates how advanced digitalisation and virtualisation accompanied with computer-assisted guidance may support surgery practices at different locations but also lead to automatic diagnosing and a growing distance between physician and patient.

Owing to the shortage of medical and healthcare staff, several regions and municipalities in Japan have introduced telehealth networks meanwhile to sustain medical services by employing ICT infrastructures. To understand the situation and the underlying challenges, we compared two cases of telehealth networks, namely the ism-Link network of Shimoina District in Iida City, Nagano Prefecture, and the Fukui Medical Net (see Brucksch, Schultz 2018, 43–53). These networks were introduced with the aim of realising a regional comprehensive healthcare system (*chiiki hōkatsu kea shisutemu*) on the ground. In the case of Iida City, the challenge was to bridge the distances between the urban centre and the mountainous outskirts. In the case of Fukui, the goal was to connect acute treatment at hospitals with aftercare and long-term care at home provided by care facilities in Fukui Prefecture. In both studies, local health institutions began to collaborate more intensively with each other only after telehealth networks had been introduced. However, concerns also came to light such as the cost for maintenance and renewal of infrastructure or lack of user acceptance among medical professionals and care staff, who preferred conventional means of communication such as facsimile and telephone. These cases demonstrate that the introduction of digital technologies such as telehealth networks are complex endeavours. They face not only technical, but also various social and organisational challenges and may cause deviations in the end from initially set cost reduction targets. It proves the case that multi-disciplinary perspectives provided by SSH are necessary if we seek to better understand and implement sociotechnical solutions.

The provision of online diagnosing of new patients became available only after legal reforms had been enacted in early 2020. During the pandemic caused by the novel coronavirus, the MHLW (Ministry of Health, Labour, and Welfare) listed over 10,000 medical and dental clinics accepting ordinary patients for online diagnosing provided remotely through digital devices like smartphones (Mainichi Japan, April 24 and 25, 2020). This change was deemed necessary because, although by early June only 12 % of all infections had occurred at clinics and hospitals, these accounted for 24 % of all deaths related to Covid-19 in Japan. As in-patients with cancer, blood disorders, advanced age and so forth are particularly vulnerable, many hospitals began to restrict or suspend medical examination for out-patients in an effort to prevent further infection (Asahi Shimbun, July 02, 2020; Mainichi Japan, June 08, 2020). At present, MHLW considers online diagnosing for new patients as a temporary measure until the pandemic ends (Mainichi Japan, April 24 and 25, 2020). Over time, however, online diagnosing and prescription might become a permanent option, necessitating further research from the SSH on the wider implications for Japan's health system.

Digital infrastructures may shift the initiative for medical treatments from practitioners and nurses to patients. Similarly, they may also provide alternative channels for remote psychological support for people suffering from serious health conditions (Brucksch 2020, 147–148). Hogle (2008, 844) argues that digital technologies and smart medical devices "can also lead to new professional groups when specific skills are required". This involves changes at the individual- and organisational levels affecting elderly patients, medical staff, nurses and healthcare workers as well as medical facilities and long-term care organisations. In a study of tele-monitoring devices in cardiology in the Netherlands, Oudshoorn (2008, 277) observed shifts in professional roles, changes in medical authority and health practices. She writes, "[w]ork previously performed by cardiologists, general practitioners, or nurses is delegated to individuals [patients] with no previous experience or education in these matters". In other words, the digital transformation influences work routines, professional roles and health systems on various levels, which entails improved options of aftercare, but has ramifications beyond the sheer promise of alleviating labour shortages and containing rising healthcare costs as well.

As a result of the digital transformation, "policies and programs in healthcare are oriented towards algorithmic medicine, drawing on big data analysis assisted through the use of AI, including machine learning, along with data generated by citizens themselves" (Petersen 2019, 41, 44). These developments also direct public subsidies, research funding as well as legislative processes. Hogle (2008, 849) points to the relationship between the state, public health, and ICT infrastructures by saying that "an elaborate information system that collects data on many aspects of human life on an ongoing basis and can be mined for a variety of purposes is essential to the state's interest in the health and well-being of citizens, which are also concerns for the good of the state". An ICT infrastructure can thus grow into a means of governance for a "capitalist regulatory state" (Faulkner 2009, 7), particularly in health systems confronting ageing, rising health costs, shortages in the medical and nursing workforce and an uneven provision of health services across regions. At the same time, the digital transformation in the field of medicine and healthcare continues to drive R&D and technological innovations and push for new regulations in the field of data protection and privacy.

In Japan, the government's basic approach to coping with ageing, depopulation in non-metropolitan regions, rising healthcare cost, labour shortages in healthcare and elderly care has been outlined by the political framework of a *Society 5.0*, a "human-centred society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space". To achieve this goal in the field of medicine and healthcare, the vision suggests "AI analysis of big data spanning diverse types of information, including personal real-time physiological data, healthcare site information, treatment/infection information, environmental information, and the following will be realized" (CaO n. d.). By putting forward such a vision, the government refers to goals such as autonomous living for the elderly, through solutions such as tele-communication via robotic devices, patient monitoring and automatic detection of states of illnesses. Increasing the linkage between health institutions to generate synergies and cost reduction in terms of financial and human resources, and reducing the need for human care and nursing interventions by providing robotic care devices (connected via digital networks) are considered important solutions to achieve this end. The overall idea is to extend the healthy life span and limit the need for care at the end of life, thereby also reducing the costs of ageing for society. In other words, behind the aforementioned contents and measures of this governmental vision, there are more diverse interests shining through the general wording while other options might go unnoticed. More specifically, we can observe a techno-positivistic perspective accompanied by economic priorities and avoidance of increasing healthcare staff instead of equally considering alternative options such as migration of healthcare specialists to Japan or improvement of wages and working conditions to address the labour shortage in elderly care.

To conclude, the research by disciplines such as STS will continue to focus on the intersection between technology and modern society and their co-productive nature. The digital transformation and its societal ramifications in fields such as medicine and healthcare, however, raise new questions, which need to be critically assessed. For instance, we might ask which uncertainties remain regarding algorithmic-generated displays of human bodies; are there any ethical boundaries for digital appliances in the medical field; which purpose do digitalised versions of patient records serve in the management of hospitals and in complex health systems; how do online tools influence practices of diagnosing, medical treatment and nursing care; who cares for the data security, privacy and dignity of individuals; and what are the vested interests of the various stakeholders behind and beyond the political visions of algorithmic medicine. The digital transformation leads to a research lacuna for the SSH. The analysis of intersections between digital technology and society extends well beyond the field of medicine and healthcare. Addressing these topics, the four aforementioned dimensions – material/physical, social, regulatory/organisational/financial and digital – may offer a promising framework. Finally, the notion that how we perceive physical or social phenomena depends on the tools by which we observe them, applies not only to the use of digital technologies in the medical field. It also holds true for digital tools employed in the SSH. In other words, our perception of society changes with the instruments used to observe and to describe social conditions. Hence, we also need to reflect critically on our research and the digital tools employed for their analysis.

## References

Asahi Shimbun. 2020. *Director explains hospital caught off-guard by cluster infection*. https://www.asahi.com/ajw/articles/13508743 (accessed 30 July 2020).

Brucksch, S. and C. Wagner. 2016. Introduction to the Technikstudien — Science & Technology Studies (STS) Research Initiative on Japan. *ASIEN*, 140: 5–21.

Brucksch, S. and F. Schultz. 2018. *Ageing in Japan. Domestic Healthcare Technologies. A Qualitative Interview Study on Care Robots, Monitoring Sensor Systems, and ICT-based Telehealth Systems*. https://leidenasiacentre.nl/wp-content/uploads/2018/09/Final_Report-30.08.2018.pdf (accessed 31.10.2020).

CaO (Cabinet Office). n. d. *Examples of Creating New Value in the Field of Health and Caregiving (Society 5.0)*. https://www8.cao.go.jp/cstp/english/society5_0/medical_e.html (accessed 13 November 2019).

Faulkner, A. 2009. *Medical Technology into Healthcare and Society*. London: Palgrave Macmillan.

Hogle, L. F. 2008. Emerging Medical Technologies. In: Hackett, E. J., Amsterdamska, O., Lynch, M. E. and J. Wajcman (eds.). *The Handbook of Science and Technology Studies*. Cambridge: MIT Press, 841–874.

Iseki, H., Muragaki, Y., Tamura, M., Suzuki, T., Yoshimitsu, K., Ikuta, S., Okamoto, J., Chernov, M. and K. Izumi. 2012. *SCOT (Smart Cyber Operating Theater) project: Advanced medical information analyzer for guidance of the surgical procedures*. https://waseda.pure.elsevier.com/en/publications/scot-smart-cyber-operating-theater-project-advanced-medical-infor (accessed 16 November 2019).

Mainichi Japan. 2020. *Online dental examination approved in Japan amid coronavirus fears*. https://mainichi.jp/english/articles/20200424/p2g/00m/0fe/085000c?mode=print (accessed 30 July 2020).

Mainichi Japan. 2020. *Japan list 10,000 clinics offering telemedicine for new patients*. https://mainichi.jp/english/articles/20200425/p2g/00m/0na/025000c?mode=print (accessed 30 July 2020).

Mainichi Japan. 2020. *Nearly 1/4 of all Japan coronavirus death connected to in-hospital cluster: Mainichi analysis*. https://mainichi.jp/english/articles/20200608/p2a/00m/0na/009000c?mode=print (accessed 30 July 2020).

Matsumoto, M., Koike, S., Kashima, S. and K. Awai. 2015. Geographic distribution of radiologists and utilization of teleradiology in Japan: A longitudinal analysis based on national census data. *PLoS One*, 10, 9: 1–14. doi:10.1371/journal.pone.0139723.

OECD. 2019. *Health at a Glance 2019: OECD Indicators*. Paris: OECD Publishing.

Oudshoorn, N. 2008. Diagnosis at a distance: The invisible work of patients and healthcare professionals in cardiac telemonitoring technology. *Sociology of Health & Illness*, 30, 2: 272–288.

Park, S. 2010. Geographical characteristics of telemedicine in Korea and Japan. *Geographical Review of Japan Series B*, 83, 1: 32–46. doi:10.4157/geogrevjapanb.83.32.

Petersen, A. R. 2019. *Digital health and technological promise. A sociological inquiry*. Milton Park et al.: Routledge.

# Politics in the Age of Connective Cynicism

*Fabian Schäfer*

The digital transformation (DT) of society has not only changed the way humans interact, but has also created new possibilities to study society and its political and cultural ramifications. The new discipline of computational social science has opened up new research fields, including approaches that try to model or simulate, sometimes even predict, human or social behavior, based on the analysis of large to even "big" datasets (e. g., data acquired from Social Media, search engines, or GIS). Needless to say, the all-encompassing DT necessitates adequate epistemological models and theories as well as methodological (computer-assisted) approaches, especially with regard to understanding the function and meaning of Social Media (SM). However, it is important to emphasize that social interactions on SM are not equal to or even remotely representative of "offline" social behavior. Different from other proponents working in the field of CCS, I argue that particularly with regard to SM, we are not merely dealing with *sociality online*, but with a whole new and unique form of *online sociality*, namely,

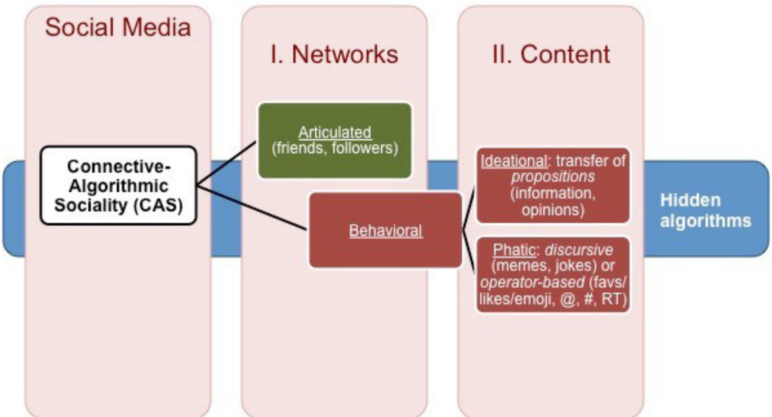
> a unique social sphere of **connective interactions mediated by algorithms**, based on the specific (determining, yet also enabling) affordances of social media platforms, which shapes our perceptions and has a growing impact on the outcome of sociocultural and political events.

It is only with the help of new computational approaches and methods, as they are currently developed in DH, CSS in particular, that it is possible to grasp the manifold mechanisms and impacts of what I therefore call **algorithmic-connective sociality (ACS)**.

Firstly, one needs to be aware of the fact that users and researchers are dealing with *hidden algorithms* (Diakopolous 2015; Latour 1999; Pasquale 2015). On SM, users are (voluntarily or involuntarily) interacting not only with real humans, but also with algorithms, such as filter algorithms suggesting information and potential friends, or machine learning algorithms, for instance, those used in software to control bots ("friendly" chatbots and "malicious" social bots alike). Therefore, we knowingly or unknowingly interact equally with human and non-human users and content. As researchers, we can study these hidden algorithms only as "black boxes" by analyzing the output they generate, thus reverse engineering their function. Based on large-data empirical analysis, we can thereby obtain at least circumstantial knowledge of what these algorithms are designed to do.

Secondly, I argue that the larger share of communication on SM is of a *connective* nature, and not "communicative" in the traditional sense. I use sociologist Kitada Akihiro's concept of "connective sociality" (*tsunagari no shakai-sei*) to describe this specific form of social interaction. Kitada (2005) originally coined this term in his landmark study *Warau Nihon no nashonarizumu* (Japan's Sneering Nationalism) to describe a self-referential communicational style having emerged online towards the end of the 1990s, particularly among *otaku* and Internet right-wingers (*netto uyoku*) in Japan. Other scholars from outside Japan have emphasized the importance of connectivity or connective sociality as well (cf. van Dijk 2013, Hepp et al. 2008).

**Figure 1**




On SM, ACS has two forms, one at the level of *networks* and the other at the level of *content*. At the level of networks, one can differentiate between *articulated* and *behavioral* networks (cf. boyd and Crawford 2012). Articulated networks resemble offline social ties, as they are explicitly articulated and publicly visible forms of SM such as friends lists on Facebook or followers on Twitter. These can be analyzed and visualized by applying basic statistical methods of network analysis or by qualitatively studying online profiles (i. e. virtual ethnography). The second form of ACS takes place rather at the level of content, namely language or audiovisual information. Posts on Facebook or tweets on Twitter can create networks amongst users sharing or liking a piece of information posted by others. It is important to add that behavioral networks can extend far beyond the articulated user networks, since posts or tweets can potentially be shared outside the articulated network. Put differently, to share or like content within the semi-private sphere of the articulated networks means to potentially and unintentionally share this information with other users via our more unstable behavioral networks.

With regard to what might therefore also be called behavioral connectivity, the "ideational" or "diaological" use of language or audiovisual information, namely, communication aiming at the transfer or exchange of propositions (information, opinions or attitudes), can be further differentiated from its "phatic" use. Phatic communication, which Kitada has therefore also described as "autotelic", is not primarily aiming at conveying a message or information. Anthropologist Bronislaw Malinkowski (1923: 315), who originally coined the term already in the 1920s, described phatic communication (or "communion", in his own words) as "a type of speech in which ties of union are created by a mere exchange of words". Put differently, the autotelic function of phatic communication lies in signaling (discursively or non-discursively) that one is still *being* connected in order to maintain social relationships. The habit of mutually greeting each other in the neighborhood or a village on a daily basis, which takes place by means of the exchange of signals (words, signs), is a perfect example of phatic "communion". Anybody who has grown up in a small village knows the importance of phatic "communion", since unwillingly forgetting or willingly neglecting to greet somebody in this setting can have severe consequences, even leading to the virtual expulsion from the village community.

Something that cannot be stressed enough is the fact that SM has integrated phatic communication into its very platform architecture by means of *functional operators*, such as Like-buttons, @-marks, hashtags or Retweets, that can be executed simply by one click or the use of a single symbol. Papacharissi (2015: 34, 35) therefore calls these operators "conversational" or "adressivity" markers. Since phatic communication on these platforms is as habitualized as in real life, it is also a highly emotional and affective behavior. Hashtags are a very good example of the functionality of phatic communication on social media, since they "combine conversationality and subjectivity in a manner that supports both individually felt affect and collectivity" (Papacharissi 2015: 27). It is in these operators or markers that the peculiar "connective affordance of social media" (Papacharissi 2015: 9) can be found, since to communicate on SM often does not necessarily mean to use language, but rather to phatically express affirmation or discontent with others.

From a political point of view, these phatic affordances of SM have facilitated the almost instantaneous generation of connective consent or discontent (or sometimes only a "fiction" of consent or discontent, created with the help of armies of trolls or social bots) to marginal or even radical political attitudes and opinions on social media. Bennett and Segerberg (2013: 32) have described the mobilizing capacity of SM as "connective action", namely as a "typically far more individualized and technologically organized set of processes that result in action without the requirement of collective identity framing or the level of organizational resources necessary to respond to opportunities."

## The digital transformation of politics: social bots & hate speech

I argue that certain contemporary ideological or political phenomena, like the global rise of right-wing nationalism and populism, cannot be explained without understanding the techno-social mechanisms of ACS described above. I will give two examples from my own research, namely social bots and hate speech, to explain this connection.

### *a) Social bots and computational propaganda*

Computational propaganda (i. e., the orchestrated attempt to manipulate public opinion or the outcome of elections via fake news or bots etc.) poses a fundamental threat to the political sphere, where SM has turned into a ubiquitous tool of political campaigning and mobilization. As already mentioned, SM can facilitate the mainstreaming of marginal and even extremist political positions based on the predominantly phatic mode of communicating. Normalizing extremist positions lies at the heart of what right-wing populist or neo-nationalist strategists call "metapolitical" strategy. Put differently, the fact that SM has become so important in politics plays into the hands of its enemies, since radical opinions or "alternative truths" or hate against political or ideological opponents can be phatically disseminated with the help of social bots or troll armies very easily.

In our research on the activity of social bots on Twitter during the 2014 General Election in Japan (based on a sample of 542,584 tweets, keyword-filtered with election-related keywords like parties, candidates and buzzwords, and collected three weeks before and shortly after the election), we were able to show that moderate Internet right-wingers (*netto uyo*) attempted to support PM Abe Shinzō's hidden nationalist agenda in the online-sphere with the help of bots (Schäfer et al. 2017). The aim was to manipulate public opinion by choking off or muddying certain political issues, or by creating trending topics and thereby pushing these on the political agenda. Although it is impossible to identify who is behind this activity (members of LDP's official online support group, the *Jimintō Net Supporters Club* (J-NSC), are very likely candidates since they seem to overlap in part with moderate Internet right-wingers). This kind of computational propaganda is representative of (right-wing) connective action, namely a connective network of human users and non-human botnets that was used to push or highjack certain hashtags in particular. As our research results can show, the most salient strategy was to verbally discredit politicians from oppositional parties or journalists from liberal media outlets during the electoral campaign period, often by simply adding the hashtag or word "anti-Japanese" (*han'nichi*) next to the name of a politician or journalist in a tweet, thereby publicly expelling a person from the "natural" community of "the Japanese", turning him or her into a traitor against his or her own people.

To summarize, one can argue that bots have become effective due to the linguistic and functional narrowness of phatic communication, because they can easily exploit the affordance structure of SM. However, it is important to emphasize that the effectiveness of bots is not only based on the technological improvement of the algorithms used in the software to control bots or whole botnets, but stems from the narrow communicational (phatic) behavior of the typical human user as well. In this regard, it has also become more and more difficult for researchers to detect bots in the behavioral network of often-indistinguishable human/nonhuman activity on SM. To rephrase our results more drastically, if the behavior has become so indistinguishable, eventually it does not matter if it is bots, very cleverly imitating human behavior, or human users, having become increasingly bot-like, that are behind these computational orchestrated attacks.

### *b) Hate speech against female politicians*

Hate speech is commonly defined as "language that is used to expresses [sic] hatred toward a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group" (Davidson et al. 2017). This includes misogynist or sexist language as well as racially discriminatory terms or slurs. Being an intermediate but asymmetric space of semi-privacy or semi-publicness for the persons being attacked by hate speech and the ones expressing their hate alike, with the former making themselves vulnerable by their online appearance and the latter publicly abusing somebody without revealing their true identity, the collapse of the clear distinction between private and public on SM has facilitated the spread of hate speech online.

In our study of hate speech against female politicians on Japanese Twitter, we intentionally collected our data (a dataset consisting of 9,449,645 words) in a time period when there was no major election looming or ongoing (namely, January through mid-April 2018) to gain insights into the magnitude of everyday verbal abuse against female politicians on SM. In the first step of our study, we conducted a manual sentiment analysis of a randomized sample of 50 tweets for each of the four quantitatively most salient female politicians in our dataset. The result of our sentiment analysis shows that only 4.5 % of the 200 tweets mentioning these four politicians conveyed a positive sentiment. In the next step, we analyzed the tweets including negative sentiments in greater detail. In all four cases, one-third to almost one half of the negative tweets contained abusive language. Since all four female politicians belonged to oppositional parties such as the CDP, one can argue that it is more likely to receive tweets conveying verbal abuse and hate speech than female members of the LDP.

Verbal abuse and hate speech against female politicians is mostly based on gender and sexuality, namely the outward appearance or an alleged misbehavior that does not match the conservative image of a woman. Moreover, misogynist or sexist verbal abuse and hate speech is often intersectionally co-occurring with nativist comments or even racist slurs. Similar to our analysis on the activity of bots, verbal attacks on female politicians followed the logic of nativist exclusion from the national community of “the Japanese” by addressing them as “anti-Japanese” or “traitors” (*baikoku*). Again, it is obvious that the harshest forms of hate speech and verbal abuse are coming from racist and misogynist Internet right-wingers, with some of them being supportive of certain female LDP members as long as they are neoconservative hardliners (such as former LDP politician Koike Yuriko).

The orchestrated attacks on female oppositional politicians are yet another example of right-wing connective action, having serious consequences on the political sphere and public discourse. Besides being experienced as psychologically abusive on a personal level, orchestrated and constant misogynist and racist hate speech against female politicians must be considered a hindering factor for women to stay or go into politics. Moreover, besides these direct political effects, online hate speech also contributes to the normalization of misogyny and racism, thereby shifting the normative borders of what one is allowed or not allowed to say publicly towards the radical extremes. This verbal barbarization of language in public discourse, which can be observed all over the world, contributes to the creation of a climate of polarization, which is also one of the aims of the metapolitical strategy of the New Right or right-wing populists.

## Conclusions: politics in the age of *connective cynicism*

To summarize the consequences of the DT for the political sphere with regard to the emergence of ACS, I argue that we are currently living in an age of "connective cynicism". It is only through the lens of this concept that one is able to grasp the connection between the techno-social emergence of ACS and the ideological rise of neo-Nationalism and right-wing populism not only in Japan, but globally. It is not by accident that Abe with his populist nationalism, the New Right and AfD/FPÖ in Austria / Germany, or the Alt-Right and Trump share not only a very similar nativist/racist and misogynist/anti-feminist exclusivist rhetoric, but are also very successful in employing SM in their political campaigns.

The concept of connective cynicism can explain the interdependence of a) the media-technological and b) the ideological dimensions of the ongoing digital (connective-algorithmic) transformation of politics, namely the interrelated co-occurrence of a certain ideological *zeitgeist* and the birth of technological affordances of SM. As already mentioned, it is important to bear in mind that the socio-technological affordances of social media platforms, particularly the mobilizing power of the highly affectual and emotional side of platform-integrated phatic communication, has contributed to, and maybe even facilitated the recent outburst of populist or nationalist sentiments as well as their instrumentalization by right-wingers and populists.

Secondly, I understand cynicism here as "enlightened false consciousness", namely a "modernized, unhappy consciousness, on which enlightenment has labored both successfully and in vain," a consciousness that "no longer feels affected by any critique of ideology", since "its falseness is already reflexively buffered" (Sloterdijk 1988: 5). If seen from this perspective, it becomes obvious that right-wing populists or neo-right-wingers apply cynicism as a political strategy. Cynics, one could argue, are not only well aware of the fact that what they are doing or saying is amoral or politically incorrect, but are doing so precisely because it is the case, strategically bargaining on the public outburst that there actions create, thereby creating the publicity they want to achieve. ACS, with its highly frequent and affective phatic communication is perfectly suited for this strategy. Moreover, we find this cynical strategy or attitude not only to be a central element in right-wing politics or populism, but also in the way Internet right-wingers "justify" the transgressive racist or misogynistic humor we find in verbal abuses and hate speech against female politicians or ethnic minorities. Right-wing populists and Internet right-wingers share a strategy of "tactical self-denial" (Tobias Weiss), namely a cynical strategy of self-renunciation by claiming that what they have said wasn't actually meant.

In this sense, both groups are launching constant attacks on the liberal or left-leaning political and intellectual establishment with their connective-cynical strategy, forming a mutual discursive opportunity structure in which the misogynist/racist cynical humor of *netto uyo* and Abe's (and others') right-wing populist politics have successfully interlinked. We are able to observe the long-term consequences of the dominance of connective cynicism in politics already today, namely the normalization of anti-feminist and racist or nativist discourses and terminologies or the growing distrust in the mass media and the political system in general. In the case of Japan, this has become obvious by the strategic verbal exclusion of common antagonistic enemies (the political opposition or liberal media outlets) from the "natural" community of "the Japanese" or Japanese society, by adding the hashtag "anti-Japanese", or by calling them "disgraceful adults" (*hazukashii otona*), as Abe has done in the past in one of his Facebook posts. Right-wing trolls or bots have become the ideal weapons in this ongoing culture war since the phatic communication (i. e., the use of a language that is reduced to the repetition of short slogans or the systematic strategy of performative verbal exclusion from a community) can be easily exploited by bots, thereby directly or indirectly supporting the connective-cynical strategy of political right-wingers.

## References


Bennett, W. L. and A. Segerberg. 2013. *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge: Cambridge University Press.

boyd, d. and K. Crawford. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15, 5: 662–79.

Davidson, T., Warmsley, D., Macy, M. and I. Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media.* Palo Alto, CA: AAAI Press.

Diakopoulos, N. 2015. Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. *Digital Journalism*, 3, 3: 398–415.

Fuchs, T. and F. Schäfer. 2020. Normalizing Misogyny: Hate Speech and Verbal Abuse of Female Politicians on Japanese Twitter. *Japan Forum*. doi.10.1080/09555803.2019.1687564.

Hepp, A., Krotz, F., Moores, S. and C. Winter (eds.). 2008. *Connectivity, Networks and Flows: Conceptualizing Contemporary Communications*. Cresskill: Hampton Press.

Kitada, A. 2005. *Warau Nihon no 'nashonarizumu'. [Japan's Sneering 'Nationalism']*. Tokyo: NHK Books.

Latour, B. 1999. *Pandora's Hope: Essays on the Reality of Science Studies.* Cambridge et al: Harvard University Press.

Malinowski, B. 1923. The Problem of Meaning in Primitive Languages. In: Odgen, C. K. and I. A. Richards (eds.). *The Meaning of Meaning. A Study of the Influence of Language Upon Thought and of the Science of Symbolism. Supplementary Essays by B. Malinowski and F. G. Crookshank*. Harcourt: Brace & World, 296–336.

Papacharissi, Z. 2017. *Affective publics: Sentiment, technology and politics*. New York: Oxford University Press.

Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Schäfer, F., Evert, S. and P. Heinrich. 2017. Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism and PM Abe Shinzō's Hidden Nationalist Agenda. *Big Data*, 5, 4: 294–309. doi10.1089/big.2017.0049.

Sloterdijk, P. 1988. *Critique of Cynical Reason*. Minneapolis: University of Minnesota Press.

Van Dijck, J. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.

# Part IV:
# Shifting boundaries of the social sciences and humanities (SSH)

# Collaborative and constructive research design in the digital age

*Cornelius* SCHUBERT

For some parts of the social sciences, the digital age dawned in the 1980s. It was the time when computers had just been introduced into workplaces and organizations on a large scale. Much to the surprise of the software engineers in charge, the users were reluctant to adopt the new technologies and many systems failed to be integrated into established workflows. Looking for answers, the software engineers turned to social scientists, mainly coming from ethnomethodological and interactionist traditions, in order to learn about the reasons why their systems failed in daily work practice. Likewise, the social scientists were keenly interested to learn about the transformations in work and cooperation occasioned by the new digital technologies. This mutual interest eventually led to the founding of a new interdisciplinary field between computer and social science: Computer Supported Cooperative Work (CSCW, Hughes et al. 1991; Schmidt/Bannon 1992).

I argue that the mutual engagement of computer and social scientists is a key relation to be addressed when thinking about the transformations of research designs in the digital age. I will show this not for cases of big data or machine learning, but for qualitative approaches such as grounded theory and qualitative data from participant observations and interviews. In case of CSCW, the engagement of computer and social science mainly lead to the transfer of qualitative methods and praxeological concepts into systems design. The digitization of workplaces thus led to an increase of qualitative research designs focusing on relations between social organization and technical infrastructures in praxis. In other words, the research designs at this dawn of the digital age largely drew on established approaches coming from anthropology and sociology. Up until today, CSCW draws heavily on methods and concepts from social science in order to design digital technologies (Rohde et al. 2017).

My main question is in which modes social sciences may engage with the digital transformation at the level of research designs. In contrast to the established modes of *synthesis* and *service*, I will argue for a *symmetrical* approach (cf. Barry et al. 2008). The case of CSCW can be seen to follow a logic of "integrative-synthesis" (Barry et al. 2008, 28) between computer and social science, culminating in the fusion of disciplines and forming novel hybrid fields such as "socio-informatics" (Wulf et al. 2018). However, critical voices have also pointed out that, in many cases, the collaboration between computer science and social science may take the shape of a "subordination-service mode" (Barry et al. 2008: 28), in which social science is reduced to providing "implications for design" (Dourish 2006) for software engineering. Indeed, in many cases, social scientists are called upon in a "response mode" (Ribes/Baker 2007 drawing on Strathern) to study the effects of digital information infrastructures without being involved in the design at all. The symmetrical approach conceives interdisciplinary collaboration in a different manner. It resonates with an "agonistic-antagonistic mode" (Barry et al. 2008, 29), in which both computer and social science maintain distinct disciplinary interests while at the same time engaging in close interdisciplinary exchanges. We refer to it as a *symmetrical approach* (Schubert/Kolb 2020), since the research designs of both social and computer science are closely mapped onto each other and brought into a productive tension.

The symmetrical approach draws on two developments. First, processes of digitization continue to spark interdisciplinary collaboration across many fields (Goulden et al. 2017). This may include disciplines, sub-disciplines or scholars who do not share a common history in collaborative work that fields like CSCW can draw upon. The collaborative challenge then resides in defining a productive "trading zone" (Galison 1996) that enables mutual benefits for the involved parties. This is a "going concern" (Hughes 1971, 52–64) of all interdisciplinary collaborations in the sense that such trading zones are "collective arrangements [that] are made and unmade" (ibid., 53). Second, the mutual history of computer and social sciences in fields such as CSCW has initiated fruitful areas of overlap concerning concepts and methods. Especially, it seems, qualitative research approaches such as grounded theory share some intriguing resemblances with design approaches such as object-oriented programming (see Bryant 2017). They both interlink empirical observations with conceptual abstraction, thereby working their way from the bottom up instead of prescribing top-down measures of theory or technology. Thus, we see that such approaches, while having marked differences, share basic similarities that can be used to bring the respective fields together.

Aligning research design and technical design in a symmetrical fashion furthermore entails various sets of shared assumptions that help to facilitate heterogeneous collaboration (see Schubert/Kolb 2020: 7–11 for an extended discussion). First, it rests on two general assumptions. Assumption 1a holds that a symmetrical approach should be impartial to either social or technical prioritization (cf. Callon 1986) and thus should avoid the biases of either technology-centered or human-centered design (Berg

1998). Assumption 1b states that research design in the social sciences and technological design in computer science may be understood as variations of a general type of problem solving, specifically as processes of "inquiry" (Dewey 1938). The second set of assumptions pertains to the "agonistic-antagonistic mode" (Barry et al. 2008: 29) of collaborative research. Assumption 2a acknowledges disciplinary discontinuities between social and computer science and seeks productive ways to exploit the "great divide" (Bannon 1997) by establishing a trading zone that enables the development of individual disciplinary insights. Assumption 2b holds that such trading zones will in turn allow collaborative research to push back on the respective disciplines by providing new insights that emerge from the ongoing interactions (Dourish 2006). The third set of assumptions pertains to the concrete configuration of inter- and transdisciplinary research. Assumption 3a states, as already mentioned, that some research designs in social science are compatible with specific modes of designing technology and thus provide the grounds for a continuous iterative engagement through mutual empirical and conceptual work (Bryant 2017). Assumption 3b builds on this by pointing out that such collaborative work requires high-frequency and long-term interaction by all parties, not subdividing the majority of tasks into temporally or functionally separate units, but running them in parallel throughout all phases of research.

Last, but not least, this symmetrical approach to collaborative research design does not only offer new modes of engagement between social and computer science in the digital age, it may also enable social science to become a constructive agent in technology design. Social science then goes beyond the established formats of observation, analysis, and critique and places itself into the thorny habitat of design. It moves into an uneasy position in which results from social research may become manifest in technological artefacts for better or worse. This 'design mode' extends the transformations of research design in the digital age. The main challenge of a symmetrical design mode is to enable research designs in social science to maintain their full epistemic potential and not reduce them to mere "implications for design" (Dourish 2006).

Instead of providing convenient shortcuts for integrating concepts and methods from the social sciences into computer science and vice versa, a symmetrical design mode requires mutual and ongoing reflections of research designs. In the long run, it calls for an active positioning of social sciences in the digital age and in a manner in which they are able to make specific contributions (cf. Collins 2018 for a recent discussion) to societal discourses as well as technical infrastructures. The symmetrical approach holds that collaborative research can facilitate contributions in three respects: first, a deeper understanding of digital dynamics in societal transformations; second, the potential to push back on social science methods and concepts from transdisciplinary collaboration; third, the involvement of social scientists themselves in the digital transformation through a constructive design mode of social research.

## References


Bannon, L. J. 1997. Dwelling in the "Great Divide". The Case of HCI and CSCW. In: Bowker, G. C., Star, S. L., Gasser, L. and W. Turner (eds.). *Social science, technical systems, and cooperative work: Beyond the great divide*. Mahwah: Lawrence Erlbaum, 355–377.

Barry, A., Born, G. and G. Weszkalnys. 2008. Logics of interdisciplinarity. *Economy and Society*, 37, 1: 20–49.

Berg, M. 1998. The politics of technology. On bringing social theory into technological design.

*Science, Technology, & Human Values*, 23, 4: 456–490.

Bryant, A. 2017. *Grounded theory and grounded theorizing: Pragmatism in research practice*. Oxford: Oxford University Press.

Callon, M. 1986. Some elements of a sociology of translation. Domestication of the scallops and the fishermen of Saint Brieuc bay. In: Law, J. (ed.). *Power, action and belief: a new sociology of knowledge?* London: Routledge, 196–233.

Collins, H. M. 2018. Studies of Expertise and Experience. *Topoi*, 37, 1: 67–77. doi:10.1007/s11245–016–9412–1.

Dourish, P. 2006. Implications for design. In: Grinter, R. (ed.). *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press, 541–550.

Galison, P. 1996. Computer simulations and the trading zone. In: Galison, P. and D. J. Stump (eds.). *The Disunity of Science. Boundaries, Contexts, and Power*. Stanford: Stanford University Press, 118–157.

Goulden, M., Greiffenhagen, C., Crowcroft, J., McAuley, D., Mortier, R., Radenkovic, M. and A. Sathiaseelan. 2017. Wild interdisciplinarity. Ethnography and computer science. *International Journal of Social Research Methodology*, 20, 2: 137–150.

Hughes, E. C. 1971. *The sociological eye. Selected papers*. Chicago: Aldine.

Hughes, J., Randall, D. and D. Shapiro. 1991. CSCW: Discipline or Paradigm? In: Bannon, L. J., Robinson, L. and K. Schmidt (eds.). *ECSCW 91 – Proceedings of the Second European Conference on Computer-Supported Cooperative Work*. Amsterdam: Kluwer, 309–323.

Ribes, D. and K. Baker. 2014. Modes of social science engagement in community infrastructure design. In: Steinfield, C., Pentland, B. T., Ackerman, M. and N. Contractor (eds.). *Communities and Technologies 2007: Proceedings of the Third Communities and Technologies Conference, Michigan State University 2007*. London: Springer, 107–130.

Rohde, M., Brödner, P., Stevens, G., Betz, M. and V. Wulf. 2017. Grounded Design – a praxeological IS research perspective. *Journal of Information Technology*, 32, 2: 163–179.

Schmidt, K. and L. J. Bannon. 1992. Taking CSCW seriously. *Computer Supported Cooperative Work*, 1, 1–2: 7–40.

Schubert, C. and A. Kolb. 2020. Designing technology, developing theory. Towards a symmetrical approach. *Science, Technology, & Human Values*, 1–27. doi:10.1177/0162243920941581.

Wulf, V., Pipek, V., Randall, D., Rohde, M., Schmidt, K. and G. Stevens (eds.). 2018. *Socio-Informatics. A practice-based perspective on the design and use of IT artifacts*. Oxford: Oxford University Press.

# Social Sciences and the Humanities in the age of Big Data. What is at stake?

*Martina* Franzen

Not digitization as such, but the emergence of digital data is widely seen as a catalyst for far-reaching socio-technological change (Schönberger & Cukier 2013; Süssenguth 2015; Geiselberger & Moorstedt 2013; Franzen 2019). Datafication has thus been characterised as "a new paradigm in science and society" (van Dijck 2014: 198). While science in particular has always produced data and analyzed them for the purpose of knowledge acquisition, the transformation of analogue events into digital data is now a ubiquitous and comprehensive process based on new communication and information technologies (Franzen 2018). Since the machine's performance depends on the volume and variation of training data, Big Data is key for machine learning. "The world of data has gone from being analogue and digital, qualitative and quantitative, transactional and a by-product, to, simply, BIG." (Uprichard 2013: 1)

The big question about the transformation of academic science by Big Data, however, can hardly be answered empirically. That would require a plethora of qualitative and quantitative studies that take into account the diversity of science in disciplinary cultures, including historical classifications. However, this paper argues that in order to be able to assess the potential for transforming science in general and the social sciences and humanities (SSH) in particular through Big Data, it is indispensable to not only consider the internal change in science (paradigms, methods and data) but also to look at the relationship between science and society. The decisive question then is to what extent datafication blurs established boundaries.

The demarcation of science from non-science is the dominant feature when it comes to characterising science and society relations. The 'boundary work' that professional scientists invest in to draw the line between professional science and parascience is only one intriguing example (Gieryn 1983). Science is a self-referential system that reproduces itself through publications that are received and cited by peers to create a network of scientific publications, which allows for a reflexive recourse to its own operations (Stichweh 1987). The social mechanism of quality control is peer review. Whereas peer review is used on different levels and for different purposes from individual promotion to research funding, the prototype of peer review is actually the pre-publication review of manuscripts in order to inform publication decisions (Hirschauer 2004). With the digitalisation of scholarly communication, the space limitations that were inherent in the printed medium have disappeared. Not only have mega-journals (such as PLOS ONE) emerged—which operate on the principle of "publish first, filter later"—, the medium of text has also been expanded by publishing research data in specific data repositories, in data journals or as add-ons for papers on journal websites (Franzen 2014). While the publication of the underlying data was triggered by replication problems in the experimental sciences (Franzen 2016), making data that has been collected available to the public in order to enable multiple usages has also become relevant in the SSH. Open Data is one facet of the credo of open science as currently incentivized by science policy, in particular by the EU research funding policy (European Commission 2016; Franzen 2019). The plea for open data goes hand in hand with the trend towards data-centred approaches in science. The promise of Big Data is that more data can lead to more knowledge, even if some premises are questionable (boyd and Crawford 2012). Nevertheless, some authors speak of a paradigm shift in science towards a data-driven approach (Kitchin 2014). Others question the idea of science by proclaiming (in a provocative way) the "end of theory" (Anderson 2008) in favour of statistical pattern recognition in data, as the example of decoding the human genome in the famous race between publicly funded research and the private sector suggests. Anderson even goes so far as to end his controversial article in Wired with the appeal: "It is time to ask: what can science learn from Google?" (ibid.)

Conversely, the SSH never tire of stressing that data does not speak for itself. In their view, critical data studies on the one hand, and qualitative and hermeneutic approaches to understanding social processes on the other, are still indispensable. The social sciences, however, in particular are in danger of seeing their traditional empirical methods overtaken. If one takes the survey instrument as an illustrative example, the methodological pitfalls are obvious: the survey is always only a selective representation of a population and at the same time, as a reactive instrument, not free of biases. If, firstly, it were possible to evaluate data from more people, and if, secondly, this data were not artificially collected, it might be possible to generate more knowledge about social processes or attitudes. Such user-generated data is currently being produced automatically and continuously through online participation in social networks such as Facebook, Twitter, or Instagram. That said – and here lies the real problem for the social sciences – unaffected by the trend towards Open Data as incentivized by science policy, companies such as Facebook are not subject to the requirement to publish or release their data (and rightly so for data protection reasons, as the recent scandal about Facebook's cooperation with Cambridge Analytica has impressively shown). The automatically generated data on usage behaviour could, however, provide relevant insights for the social sciences. But academic scientists normally do not have access to the data; data sovereignty is reserved for companies that use the data for their commercial purposes. (Social) data might be best characterized as "boundary objects" (Star & Griesemer 1989) to which science no longer has exclusive access.

In addition, the social sciences are increasingly relying on digital methods to try out new ways of gaining knowledge. Particularly when it comes to text mining in SSH as a supplement to classical content analysis or hermeneutic methods, the advice of computer scientists or the emerging group of data scientists is often sought to read out the data and make it fruitful for further analyses. This is where new forms of research cooperation are created, in which the SSH scholars are partly dependent on the expertise from outside to advance their research. To what extent these are reciprocal alliances between SSH and Date Science would be worth investigating.

Somewhat similarly, in the natural sciences' user-generated data play an increasing role when it comes to scientific problem-solving. This is referred to as Citizen Science: non-scientists participate in the scientific production of knowledge. The role of citizens in the corresponding projects is often limited to that of data providers, laypeople often acting as human sensors (Haklay 2013), although the normative concept of Citizen Science is much more comprehensive (Franzen 2019; Franzen, Kloetzer, Ponti et al., in press). One of the successful crowd science projects is *Foldit*: With a gamification approach, *Foldit* achieved much-noticed successes in decoding unknown structures of individual proteins without requiring the citizen scientists to have any biochemical expertise. At the so-called biennial "protein-folding Olympics", however, another winner emerged in 2018. It was neither professional scientists nor the Citizen Science Project Foldit that decoded the structures of the proteins most efficiently. Surprisingly, the winner was Google with its latest algorithm, *AlphaFold* (Sample 2018). *AlphaFold* was trained with a neural network on thousands of known proteins until it was able to predict 3D structures from amino acids alone. Should we say that the former Wired chief editor Chris Anderson (2008) turned out to be right and that it is time to ask what science could learn from Google? It is clear that the companies of the Internet age base their success on large amounts of data. If data is the `new oil` that everyone is talking about, the question of data sovereignty will become a question of power – beyond its original context of creation and beyond system-specific boundaries. Against this background, Lev Manovich's interpretation of the social implications of the increasing datafication of society in 2011 seems obvious. He differentiates between three different classes in the realm of Big Data: "those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have the expertise to analyze it". (Manovich 2011: 10)

This differentiation is suitable to apply it to science as such: access and power to handle large amounts of data will possibly prove more and more decisive not only for successfully studying research subjects but also for competition between researchers, between research institutions, and even countries.

If we bring the lines of development outlined above together and focus on the initial question of the relationship between science and society, several constellations regarding the datafication of knowledge production emerge. There are indications of new types of relationships – in part cooperative, but above all, competitive – between:

1) SSH and data sciences;
2) professional scientists and Citizen Scientists;
3) man and machine;
4) science and business.

For theoretical purposes, too, it must, therefore, be a matter of taking socio-technical change seriously and of going beyond the theory of functional differentiation that demarcates science from the economy or politics if these recent developments are to be adequately described. For a solid scientific-sociological analysis, it is also necessary to take into account the special cultural characteristics of the natural sciences on the one hand and the SSH on the other. There are also differences within the SSH that are worth examining in detail. As there is still much to be done to grasp the societal implications of digitalisation, one thing is clear: Dedicated research on the changes in knowledge production in the digital society opens the prospect of emerging changes at the level of society as a whole. In this respect, it seems worthwhile to continue to follow the processes of knowledge production and knowledge reception, not least to understand the transition to a digital society.

## References

Anderson, C. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired. https://www.wired.com/2008/06/pb-theory/ (accessed 27 October 2019).

boyd, d. and K. Crawford. 2012. Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15, 5: 662–679.

Dickel, S. and M. Franzen. 2015. Digitale Inklusion: Zur sozialen Öffnung des Wissenschaftssystems. *Zeitschrift für Soziologie*, 44, 5: 330–347.

European Commission. 2016. Open innovation, open science, open to the world – a vision for Europe. http://www.openaccess.gr/sites/openaccess.gr/files/Openinnovation.pdf (accessed 25 January 2018).

Franzen, M. 2014. Grenzen der wissenschaftlichen Autonomie. Zur Eigengesetzlichkeit von Publikationskulturen. In: Franzen, M., Jung, A., Kaldewey, D. and J. Korte (eds.). *Autonomie revisited – Beiträge zu einem umstrittenen Grundbegriff in Wissenschaft, Kunst und Politik. Zeitschrift für Theoretische Soziologie*, 2. Sonderausgabe: 374–399.

Franzen, M. 2016. Science between Trust and Control: Non-Reproducibility in Scholarly Publishing. In: Atmanspacher, H. and S. Maasen (eds.). *Reproducibility: Principles, Problems, Practices and Prospects*. Hoboken: John Wiley & Sons, 468–485.

Franzen, M. 2018. Die digitale Transformation der Wissenschaft. *Beiträge zur Hochschulforschung*, 40, 4: 8–28.

Franzen, M. 2019. Zum Wandel der wissenschaftlichen Wissensproduktion durch Big Data: Welche Rolle spielt Citizen Science? In: Musik, C. and A. Bogner (eds.). *Digitalization and Society. Österreichische Zeitschrift für Soziologie*, 1. Sonderheft: 15–35.

Franzen, M., Kloetzer, L., Ponti, M., Trojan, J. and J. Vicens. 2020, in press. Machine Learning in Citizen Science: Promises and Implications. In: Vohland, K., Land, A., Ceccaroni, L. et al. (eds.): *The Science of Citizen Science*. Springer.

Geiselberger, H. and T. Moorstedt (eds.). 2013. *Big Data – Das neue Versprechen der Allwissenheit*. Berlin: Suhrkamp.

Gieryn, T. F. 1983. Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review*, 48, 6: 781–795.

Haklay, M. 2013. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In: Sui, D., Elwood, S. and M. Goodchild (eds.). *Crowdsourcing Geographic Knowledge*. Dordrecht: Springer, 105–122.

Hirschauer, S. 2004. Peer Review auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation. *Zeitschrift für Soziologie*, 33, 1: 62–83.

Houben, D. and B. Prietl (eds.). 2018. *Datengesellschaft. Einsichten in die Datafizierung des Sozialen*. Bielefeld: transcript.

Kitchin, R. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1: 1–12. doi:10.1177/2053951714528481.

Manovich, L. 2011. Trending: the promises and the challenges of big social data. In: Gold, M. K. (ed.). *Debates in the Digital Humanities*. http://www.manovich. net/DOCS/Manovich_trending_paper.pdf (accessed 15 December 2019). doi:10.5749/minnesota/9780816677948.001.0001.

Sample, I. 2018. Google's DeepMind predicts 3D shapes of proteins. *The Guardian*. https://www.theguardian.com/science/2018/dec/02/google-deepminds-ai-program-alphafold-predicts-3d-shapes-of-proteins (accessed 2 December 2019).

Schönberger, V. and K. Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murry.

Stichweh, R. 1987. Die Autopoiesis der Wissenschaft. In: Baecker, D., Markowitz, J., Stichweh, R., Tyrell, H. and H. Willke (eds.). *Theorie als Passion. Niklas Luhmann zum 60. Geburtstag*. Frankfurt a. M.: Suhrkamp, 447–481.

Süssenguth, F. (ed.). 2015. *Die Gesellschaft der Daten. Über die digitale Transformation der sozialen Ordnung*. Bielefeld: transcript.

Star, S. L. and J. R. Griesemer (1989): Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19, 4: 387–420.

Uprichard, E. 2013. Big Data, Little Questions? *Discover Society, 1*. http://discoversociety.org/2013/10/01/focus-big-data-little-questions/ (accessed 27 October 2019).

van Dijck, J. 2014. Datafication, dataism and dataveillance. Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12, 2: 197–208.

# Knowledge production and the role of SSH in the digital age

*Franz Waldenberger*

How does the digital transformation affect the social sciences and humanities (SSH)? As some of the previous contributions convincingly show, the digital transformation raises new research questions, provides new research tools and upgrades research infrastructures. But there also more profound effects. To fully grasp the more fundamental implications forces us to reconsider our understanding about the nature of knowledge, why, where and how it is produced and how this relates to the way we perceive and conceptualize ourselves and society (Precht 2020: 18). It is in this wider context that I want to assess how the digital transformation affects the SSH. To do so within the limits of a short essay, I will selectively borrow concepts spanning from philosophy, evolutionary biology/psychology, sociology and economic growth theory to develop some ideas about the nature, logic and implications of knowledge production in the digital age and its main driver – big data and AI (BDAI).

## The knowledge producing society

Given the many different research interests associated with "knowledge", there is no commonly accepted definition of the term. In the following essay, I will very generally define knowledge as the ability to receive, process and send information. In this broad sense, knowledge can be understood to be embedded in quite different units of analysis: bacteria, plants, animals of all kind including human beings, but also computers and robots as well as social organizations like bee swarms, corporations or societies. For biological organisms and social organizations, the common characteristics also extend to the "purpose"[1] of knowledge, namely, to adjust to an environment in order to increase the probability of survival

[1] "Purpose" represents a teleological term typical of the sense-making human mind. Taken from a neutral perspective, what we observe are outcomes consisting of adjustments and survival rates. There does not need to be an intrinsic purpose. Still, to understand and make sense of what we see, we may in our interpretation attach a purpose to it.

(Dawkins 1989).[2] The content and "success"[3] of knowledge is the outcome of an ongoing interaction with an environment within which information is exchanged and in which action is taken. Embeddedness implies that the survival rate not only depends on the information handling capacity of knowledge, but also on the quality of the environment and, especially, its stability. One can only adjust to a sufficiently stable environment. In this respect, stability and sustainability appear to be closely linked.

Intelligence is the ability to not only apply knowledge purposefully, but to also intentionally change it (https://www.merriam-webster.com/dictionary/intelligence). If knowledge is intentionally changed, we speak of learning or knowledge production. The aim is to acquire more, better or more valuable knowledge. Evolution shows that the ability to learn or to change knowledge, does not require a conscious mind. Evolutionary processes can achieve the same with knowledge embedded in organisms although unintentionally and in many cases more slowly.

Human intelligence has been increasingly engaged in purposefully creating new knowledge, especially with the onset of secularization, when the belief that our conditions could be better improved by actions based on rational thought than by prayers to a divine and hopefully benevolent creator became the new social norm (Harari 2016: 259). Human knowledge creation has from early stages extended far beyond improving the adjustment to a given environment. It has significantly impacted and altered our natural environment as exemplified by land cultivation, animal farming, urbanization and transregional transport and communication infrastructures. The accelerating and lasting human transformation of the earth has prompted scholars to name the period starting around the mid-20th century the Anthropocene (Waters et al. 2016).[4]

The term knowledge society, introduced to characterize the reliance of modern societies on scientific knowledge (Böhme and Stehr 1986), is somehow misleading since all societies have always been based on knowledge. The reliance on science or the scientific method and the resulting implications are better captured by what I call "knowledge producing society". However, this again needs to be qualified in two respects. Firstly, production focuses on a specific kind of knowledge, namely, scientifically created, technological knowledge (TK), which is imper-

---

[2] Computers and robots do not per se use their knowledge in such a "selfish way," unless they are programmed to do so.

[3] "Success" like "purpose" is just another sense-making interpretation. See footnote 1.

[4] The newly proposed classification has yet to be approved by the respective geological authorities, i. e., the International Commission of Stratigraphy and the International Union of Geological Sciences.

sonal, highly standardized and can therefore be easily copied and scaled. Secondly, most of the production of TK occurs outside non-profit-oriented research institutions. It is driven by commercial interests.

The dynamics underlying knowledge production rely both on characteristics of TK and the logic of profit-oriented production regimes. Industrial production can be understood as the transformation of matter using energy and applying TK (Boulding 1978). In the process of production matter is transformed and energy is used up. It is thus only TK that can be increased. More or better TK will enable us to improve production processes, for example, by increasing energy efficiency, but it may also lead to the discovery of more and better materials or new sources of energy. TK can not only be increased, once created, it can be applied any number of times without being used up. These unique properties make it the perfect engine of economic growth (Romer 1990). To run the engine means to exploit the scalability of TK. It requires the geographic expansion of the economic system. The knowledge producing society is intimately linked to globalization.

A fundamental aspect of TK is its dispersion (Waldenberger 2019). As the social stock of TK grows, it surpasses the cognitive capacity of a single brain. This is made possible by division of labor, which creates another important link between the knowledge producing society and globalization. The more people participate in an economic system, the more brains can be involved in the production and application of increasingly specialized TK. In this process, each of us will know less and less of the totally available TK. However, even though we only possess and command an ever-tinier part, we can still benefit from the increasing stock of TK as it is embedded in the products and services we consume. The miraculous result has been made possible by institutions like markets and money. They have supported the transformation of human knowledge production into societal knowledge production, making society the knowledge producing "organism".

## Big data and artificial intelligence – a new stage of societal knowledge production

A core factor and main driver of knowledge production in the digital age are big data and artificial intelligence (BDAI). The term AI may well be questioned (Gabriel 2018: 31). Even deep or self-learning algorithms are still the product of human intelligence. Future generations of universal or general AI may write their own specialized AI programs. Some visionary proponents predict that in the foreseeable future AI will be able to fully emulate the human mind and finally surpass it, as it will not be limited by the size of a human skull (Kurzweil 2012: 123). Whether this proves to be true and what it implies are questions that have inspired and will continue to inspire many discussions, books and movies in the genres of science, fiction and their intersection, science fiction. For the time being, self-learning algorithms do not understand what they are doing and why they do what they do. The understanding, the motivation, purpose and sense-making has to be provided by the human mind.

BDAI basically stays within the framework and follows the logic of TK production. It is largely driven by commercial interests, even more so as one of the major ingredients, namely, big data, is often collected by and in the possession of private companies. Developing and applying AI requires highly specialized knowledge, leading to a further dispersion of TK. Still, more than with any other new technology, advances in AI are made readily available to masses of consumers often free of charge as they are installed in widely diffused ICT devices such as smartphones.

## Side effects – widening imbalances

The advances achieved by our knowledge producing societies have generally been associated with progress, i. e., a steady improvement of our living conditions. But their very success is increasingly casting doubt as to whether we are heading in the right direction (Precht 2020: 11). The commercially driven exploitation of human, and lately artificial, intelligence in the creation of TK has had undesirable side effects on individuals, society and the environment. They relate to three major imbalances.

*Imbalance between technological and non-technological knowledge*. Technological knowledge is only one kind of knowledge. Its accelerated growth has outpaced learning and adjustment processes in other fields, like emotional or social knowledge, needed to keep our mental health and to improve or sustain social relations. Advances in TK have allowed us to develop and produce weaponry sufficient to destroy our planet, which dangerously surpasses our ability to resolve conflict without war. Modern society has liberated the individual from the social bonds and career paths that used to be determined by dint of one's birth. We are still subject to resource constraints and various kinds of social pressure, but legally we are free to choose where and with whom we want to live, how to configure our social relations and how to plan our careers. However, we hardly explore and make full use of such choice sets and instead leave these fundamental life course decisions to chance or fate. In general, the overwhelming number of choices that modern society offers in many spheres of life often result in more anxiety than happiness (Schwartz 2009).

*Imbalance between technological knowledge and societal intelligence*. Societal knowledge production needs to be governed by societal intelligence. Given the dispersed nature of TK, societal intelligence cannot be centralized and unified. It is instead embedded in and supported by various institutions such as market competition, political competition, legal and administrative regulations, public media, civil society and academia. These institutions, most of which bear strong national characteristics, provide the knowledge infrastructure underlying a decentralized governance of checks and balances. While these institutions have also benefited from TK production, they have done so to a lesser extent due to their public good's character. Improvements in the provision of public goods are generally difficult to appropriate privately. As a result, there will be little private incentive to invest in research and development to advance societal intelligence. Our capacity to assess and control the risks inherent in new TK has therefore not been able to keep pace with the accelerated growth of commercially driven TK. The issue is aggravated by the fact that the high level of profitability generated by successful technological innovations provides businesses with resources to selectively influence societal governance mechanisms to their advantage.

*Ecological disequilibria*. TK production has enabled us to increasingly interfere with and transform our natural environment. The resulting unforeseen and unaccounted for ecological disequilibria are manifested in pollution, decreasing biodiversity and climate change. Especially the latter is now endangering our very survival. Again, as with societal intelligence, TK production could, in principle, have also been directed towards the preservation and improvement of our natural environment as some of it has. But the environment, too, represents a public good. Environmental investments have been dwarfed by commercially driven TK in negligence of environmental side-effects. Recent efforts and claims by companies to support the sustainable development goals can hardly make up for the damage already done.

## Governing societal knowledge production – the role of the SSH

As BDAI is mainly driven by commercial interests, its further acceleration of TK production is likely to aggravate the above-mentioned imbalances. We already see how our national legal frameworks are lagging behind in resolving essential questions raised by globally operating BDAI like the definition of property rights over data, the protection of privacy or the ethical questions related to what Harari refers to as "hacking humanity" (Harari 2016: 383).

As with other TK, BDAI could help us enhance our emotional and social intelligence. It could also be developed to support knowledge infrastructures or to resolve ecological disequilibria. Whether and to what extent research and applications of BDAI will be channeled in these directions crucially depends on whether there are sufficient incentives to do so. If the social benefits of correcting these imbalances can be privatized, commercial interests will lead the way, although they will need to be guided by appropriate regulations to ensure that private and social value creation coincide. Where profit motives cannot be expected, resources need to be mobilized in other ways, for example, by utilizing public budgets or by securing funding from private donors. The hope that the conditions for a responsible development and application of BDAI can be met depends crucially on the quality of the institutions that make up the knowledge infrastructures underlying societal intelligence and its decentralized governance of checks and balances.

It is here that SSH can be seen to play a vital role. That they could and should play an important role in societal intelligence is well expressed in recent national research programs:

> *Embedding SSH research across Horizon 2020 is essential to maximise the returns to society from investment in science and technology.* (EU, Horizon 2020, Social Sciences and Humanities, https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-and-humanities)

> *Understanding Society – Shaping the Future. BMBF programme for the humanities and social sciences (2019–2025)* (German Federal Ministry of Education and Research, https://www.bmbf.de/upload_filestore/pub/Understanding_Society_Shaping_the_Future.pdf)

> *Social sciences and humanities research addresses critical questions about who we are as human beings, what we need in order to thrive in complex and challenging times, and where we are headed in the years ahead. It enhances our ability to understand and creatively respond to complex individual, social, cultural and economic issues.* (Social Sciences and Humanities Research Council of Canada, https://www.sshrc-crsh.gc.ca/about-au_sujet/publications/strategic-plan-strategique-2016-eng.pdf)

However, it is too naïve to think that by just stating appropriate goals and allocating larger amounts of money, SSH will be able to live up to these expectations. In fact, when addressing questions of societal intelligence, we confront a fundamental paradox. How can it be possible to design or at least purposefully influence a system, which we know is necessarily characterized by dispersed knowledge? Would the design not require some centralized knowledge beyond the capacity of any part of the system? This fundamental complexity needs to be seriously addressed and sufficiently answered if we want to improve the governance of our knowledge producing societies.

To do so would first of all require new transdisciplinary and holistic approaches to our understanding of human knowledge and societal knowledge production. Traditional human-centered, rationality and science-focused theories of knowledge leave out essential representations and qualities of knowledge (Narby 1998). They also tend to only account for knowledge produced in Western cultural contexts. A holistic approach must include non-scientific sources of knowledge and societal intelligence in other cultures. The dispersed nature of TK, and its implications for societal complexity as well as the possibilities and limitations of a decentralized governance regime of checks and balances would be another important research field. Here Herbert Simon's conception of the architecture of complexity and Hayek's idea of spontaneous order can provide useful starting points (Waldenberger 2019). Knowledge infrastructures and their role of linking, integrating and aggregating dispersed knowledge should also prove a useful concept.

Transdisciplinary research has long been on the agenda of national research policies. However, deeply rooted disciplinary boundaries continue to exist in undergraduate and graduate curricula, universities, science associations and the scientific media and conference landscape. They continue to shape academic career paths and research choices.

Like in the case of TK, SSH knowledge production, too, has resulted in knowledge dispersion.[5] However, in contrast to TK, this has not produced the benefits of division of labor for two interrelated reasons. Firstly, SSH knowledge is less standardized. It is more situational, i. e., conditioned by the personal, social and cultural context of the researcher. This limits its transferability and scalability. Secondly, SSH theories and concepts are generally contested. They tend to compete with other research findings rather than complementing them, adding to them or building on them.

To overcome the "dilemma of specialization" in SSH (Hayek 1967), we have to not only recognize the importance of SSH for societal intelligence, but to also understand the conditions of knowledge production in SSH and how respective incentive systems would need to be adjusted in order to enhance transdisciplinary research.

To sum up, addressing the more fundamental question as to how digital transformation impacts the role of SSH in knowledge production, I

[5] There seems to be a natural tendency within steadily growing social systems as well as organisms in general towards further differentiation (Wortmann 2012).

suggest three nested answers. Firstly, knowledge production in SSH can and is expected to play an important role in strengthening the knowledge infrastructures needed for the governance of TK production in the digital age. Secondly, to fulfil this role and live up to the expectations, the SSH need to seriously study the conditions of societal knowledge production. To be prepared to do so, they must, thirdly, understand and adjust the institutional framework underlying knowledge production in the SSH.

## References

Böhme, G. and N Stehr (eds.). 1986. *The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations*. Berlin: Springer.

Boulding, K. E. 1978. *Ecodynamics. A New Theory of Societal Evolution*. Beverly Hills: Sage Publications.

Dawkins, R. 1989. *The Selfish Gene*. Oxford: Oxford University Press.

Gabriel, M. 2018. *Der Sinn des Denkens*. Berlin: Ullstein.

Harari, Y. N. 2016. *Homo Deus. A Brief History of Tomorrow.* London: Vintage.

Hayek, F. A. 1967. The Dilemma of Specialization. In: Hayek, F. A. (ed.). *Studies in Philosophy, Politics and Economics*. Chicago: The Chicago University Press, 122–132.

Kurzweil, R. 2012. *How to Create a Mind. The Secret of Human Thought Revealed*. New York: Pinguin Random House.

Narby, J. 1998. *The Cosmic Serpent. DNA and the Origins of Knowledge*. London: Orion Publishing Group.

Precht, R. D. 2020. *Künstliche Intelligenz und der Sinn des Lebens*. München: Goldmann.

Romer, P. M. 1990. Endogenous Technological Change. *Journal of Political Economy*, 98, 5: 71–102.

Schwartz, B. 2009. *The Paradox of Choice. Why More Is Less*. New York: HarperCollins e-books.

Waldenberger, F. 2019. Einige Überlegungen zu den Möglichkeiten und Grenzen staatlicher Regulierung in einer durch Arbeitsteilung geprägten Wissensgesellschaft. In: Baums, T., Remsperger, H., Sachs, M. and V. W. Wieland. (eds.). *Zentralbanken, Währungsunion und stabiles Finanzsystem. Festschrift für Helmut Siekmann*. Berlin: Duncker & Humblot, 621–635.

Waters, C. N. et al. 2016. The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science*, 351, 6269: 1–10.

Wortmann, H. 2012. A simple Evolutionary Model of Social Differentiation. *Zeitschrift für Soziologie*, 41, 5: 375–391.

# Postscript: Comparative and historical perspectives on the digital transformation

*Harald Kümmerle*

This postscript may help the reader to reappreciate the contributions in this volume by placing them in a broader spatial and historical perspective. Although there is no explicit reference to *digital spatiality*, the term provides a good heuristic when considering the usage and impact of digital technology in various regional contexts. The four headings, under which the articles are grouped, roughly correspond to the four perspectives on *data acceleration regimes* proposed by historical accounts of how societies reacted to sudden surges of information.

Big Data is considered to represent both a recent as well as a global phenomenon. However, historically, there is more continuity in Big Data than one would expect. Since the introduction of digital computers, the actual algorithms and operations performed with and on data have evolved in a continuous fashion; a clear line when Big Data became a driving factor cannot be drawn and doing so would not provide much insight. Also, developments related to Big Data have not been uniform across societies. Just as the title of a very influential edited volume "Raw Data is an Oxymoron" (Gitelman 2013) indicates, there is no data devoid of the context in which it is produced and used. It is through comparison and especially through case studies that the cultural embeddedness of Big Data and of the DT as a whole can be fathomed.

The digital realm is articulated with spaces, which are shaped by geopolitical, sociocultural, and other spatial dimensions (Gray and Driscoll 1992). It must be noted that the Internet, too, is not spatially homogenous. In different regions, the Internet is experienced differently: the development of the material infrastructure, political regimes and regulation regarding for example age and copyright exert strong influence on Internet usage. Moreover, differences in usage patterns (e. g., which social networks are dominant or which languages are used) lead to differences in how users interact with each other virtually. Thus, it makes sense to say that there are geographic variations in how the Internet is constituted (Roth 2016, 156, 160–61). For academia, too, the research literature accessible digitally varies by nation, university, or even university institute depending on which licenses are bought. Accounting for such differences enables an understanding of *digital* *spatiality* while maintaining a link to physical space (Gairola and Roth 2019).

Just as the digital transformation (DT) will bring about a stronger interdependence between material and digital spatialities in smart cities (Roche 2017), the university of the future will not simply be an extension of today's university in which access to research literature and data is, in some sense, "easier". When the DT is interpreted as a data revolution which, granted, usually has a normative connotation, it is considered imperative to aggregate data and to denigrate data sets that are not suited to this end as "stranded data" (Kitchin 2014, 156). A statement made by the Alliance of German Research Organizations in 2010 declaring that "research data with assured quality form a pillar of scientific insight" (*Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis*) (Allianz der deutschen Wissenschaftsorganisationen 2010) points in a similar direction. However, such datafication of the academic infrastructure may stifle methodic pluralism. How we position ourselves with regard to such normative aspects of the DT is closely connected to how we envision the digital space. There should be no single perspective. Having scholars from different national and disciplinary backgrounds reflect on the implications of the DT for the SSH is more relevant than ever.

Big Data can also be situated historically. In particular, the following four perspectives have been proposed from a historical perspective with regard to how data sets can be assigned bigness: the technological perspective (the data is difficult to process "manually"), the open-endedness perspective (the stream of data extends continuously), the relational perspective (the data is related to other data, its nature is "networked"), and the paradigmatic perspective (the data is of a size that allows for data-driven arguments). In certain constellations, these four dimensions influence and amplify each other, resulting in the emergence of Big Data. Technological advances enable the production of new open-ended data streams. The data streams, in turn, are subject to standardization, and relationships between them become visible. Finally, new knowledge can be gained from the possibilities that these networked data provide – in some cases this leads to a paradigm shift, which in turn drives further technological changes (Kaplan and di Lenardo 2017, 2).

There are in fact historical constellations, which have been referred to as *data acceleration regimes*, similar to the processes currently associated with the DT – if not in scale and scope, then at least qualitatively. In Mesopotamia, for example, keeping up with the rising complexity in the circulation of goods and population meant developing new administrative standards. The emerging information system played an important role in the longevity and influence of the empire. For similar reasons connected to circulation, a transformation in the data practices in the Roman Empire took place, leading to new writing techniques and innovations in the documentation of events (Kaplan and di Lenardo 2017, 2–3). For many such *data acceleration regimes*, detailed studies do in fact already exist; the point here is that knowing that, we can see that many if not most phenomena of the DT do have a precedent. Thus, studies that look at how governments are shaping the *data acceleration regime* created by the DT, for example, the Japanese government through its initiative "Society 5.0" launched in 2016 (Waldenberger 2018), might benefit from applying both a comparative as well as historical perspective.

The data acceleration caused by the DT has created the impression that there is "too much to know" (Waldenberger 2019). But this information overload, too, has historical precedents, perhaps most prominently in the Renaissance (Blair 2010, 57). As it has been argued convincingly, the emergence of book printing in Europe alone does not offer a satisfactory explanation. In China, printing existed for centuries without being considered a cause of information abundance. Rather, in Renaissance Europe a multitude of cultural processes and events, like the reevaluation of the classics of antiquity and the religious schism, and of economic and political factors, like the discovery of the New World, radically changing spatiality in various ways, played a role. Similarly, developments with no apparent connection to digital technology have to be taken into account when interpreting contemporary perceptions of information overload. During the Renaissance, scholars devised new methods of note-taking and managing excerpts, in the process reconsidering what kind of information was worth preserving and disseminating, especially through reference books (Blair 2010, 47, 60–61). While it may not be completely up to us to decide how we accommodate the DT, the outcomes will depend on how we shape material and discursive practices in the present.

## References

Allianz der deutschen Wissenschaftsorganisationen. 2010. *Grundsätze zum Umgang mit Forschungsdaten*. http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten. (accessed 01 September 2020).

Blair, A. 2010. *Too Much to Know: Managing Scholarly Information Before the Modern Age*. New Haven: Yale University Press.

Gairola, R. K. and M. Roth. 2019. Cyber Zones: Digital Spatialities and Material Realities across Asia. *Asiascape: Digital Asia,* 6, 1–2: 4–16. doi:10.1163/22142312–12340099.

Gitelman, L. (ed.). 2013. *"Raw Data" Is an Oxymoro*n. Infrastructures. Cambridge, et al.: MIT Press.

Kaplan, F. and I. di Lenardo. 2017. Big Data of the Past. *Frontiers in Digital Humanities,* 4: 12. doi:10.3389/fdigh.2017.00012.

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles: Sage.

Roche, S. 2017. Geographic Information Science III: Spatial Thinking, Interfaces and Algorithmic Urban Places–Toward Smart Cities. *Progress in Human Geography,* 41, 5: 657–66. doi:10.1177/0309132516650352.

Roth, M. 2016. Herausforderungen für die Japanforschung im Zeitalter der digitalen Medien. In: Richter, S., Clart, P. and M. Roth (eds.). *100 Jahre Ostasiatisches Institut der Universität Leipzig, 1914–2014*. *Leipziger Ostasien-Studien*, 19: 145–164.

Waldenberger, F. 2018. Society 5.0. *Konrad-Adenauer-Stiftung: International Reports*, 1: 48–55.

Waldenberger, F. 2019. Governance in the Age of Ignorance: The Role of Knowledge Infrastructures. *G20 Magazine*. http://digital.thecatcompanyinc.com/g20magazine/japan-2019/governance-in-the-age-of-ignorance-the-role-of-knowledge-infrastructures/ (accessed 01 September 2020).

# List of authors

Susanne Brucksch is principal researcher at the German Institute for Japanese Studies (DIJ) in Tokyo. Before, she was senior research fellow at Freie Universität Berlin and visiting scholar at Waseda University in 2016. She obtained her PhD in Japanese Studies at the Martin Luther University Halle-Wittenberg (MLU). Her research covers medical technologies, innovation, STS, policy studies, inter-organizational collaboration and environmentalism. At present, her research focusses on "Technical Innovation and Research Collaboration in Japan: The Biomedical Engineering Sector". She also serves as chair of the advisory board of the German Centre for Science and Innovation (DWIH, since 2019) Tokyo as well as member of the board and co-organizer of STS Section, German Association for Social Science Research on Japan (VSJF).

Martina Franzen is a postdoctoral fellow at the Institute for Adcanced Study in the Humanities (KWI) in Essen (Germany). She is a sociologist by training, received her PhD from Bielefeld University and is specialised in science studies. Before joining the KWI, she was a research associate at WZB Berlin Social Science Center in the research group Science Policy, a lecturer at the Humboldt-Universität zu Berlin and a research associate at the Institute for Science and Technology Studies (IWT). Her most recent publications include: Reconfigurations of science communication research in the digital age. In: Handbook of Science Communication edited by Annette Leßmöllmann, Marcelo Dascal, Thomas Gloning. Berlin/Boston: de Gruyter (2020). Die digitale Transformation der Wissenschaft (The digital transformation of science). In: Beiträge für Hochschulforschung 4/2018.

Miho Funamori is associate professor at Information and Society Research Division, National Institute of Informatics (NII). She is also strategy manager at Research Center for Open Science and Data Platform within NII. Having worked for ten years at the headquarter of The University of Tokyo giving advice to the administration, she specializes in international higher education policy and university management, and is knowledgeable in the higher education sphere domestically and internationally. Prior to joining the academia, she has worked at the Office for International Cooperation, Ministry of Education, Culture, Sports, Science and Technology (MEXT). She is expert member at the Japan Association of National Universities, and board director of FORCE11 since 2019.

Harald Kümmerle is senior research fellow at the German Institute for Japanese Studies (DIJ) in Tokyo. He studied mathematics at the Technical University of Munich (TUM), Japanese Studies at the Martin Luther University Halle-Wittenberg (MLU), and Japanese as a Foreign Language at Keio University in Tokyo. His doctoral thesis (Japanese Studies; defended in 2019) concerned the institutionalization of mathematics as a science in Meiji- and Taishō-era Japan. During the time as a doctoral researcher, he was junior visiting research fellow at Keio University, visiting research fellow at the Centre for Science Studies at the German National Academy of Sciences Leopoldina and research fellow at the MLU Halle-Wittenberg. His interests include the digital humanities and the public discourse on the digital transformation in Japan. Currently, he also is an associated researcher at the Graduate University for Advanced Studies, SOKENDAI.

Yukio Maeda is professor of Political Science at Inter-Faculty Initiative in Information Studies, The University of Tokyo. He is also senior researcher at Center for Constructing Data Infrastructure for the Humanities and Social Sciences, Japan Society for the Promotion of Science. He earned his Ph. D. in political science from Department of Political Science, University of Michigan in 2001. His research interests include (1) public opinion and voting behavior, (2) approval ratings for prime ministers, (3) methodology for public opinion polling and social survey, and (4) data preservation and dissemination in social sciences. He was a member of Social Science Japan Data Archive from 2006 to 20017. He also served as a Japanese local committee member for many international academic surveys, including Comparative Study of Electoral Systems (CSES), Asian Barometer, World Value Survey (WVS) and East Asian Social Survey (EASS).

Christian Oberländer is professor for Japanese Studies at the Martin Luther University Halle-Wittenberg (MLU). He studied medicine, Japanese studies and public administration at the University of Bonn, Keio University and The University of Tokyo as well as Harvard University. His research interests include pre- and postwar Japanese politics, society and social security as well as the history of science and medicine in Japan. Among other topics, he has conducted research on the history of Japanese traditional Kanpo medicine in the $19^{th}$ and $20^{th}$ centuries, on the introduction of nursing care insurance in Japan during the 1990s and on Japan's international relations in the $20^{th}$ century. Recently he served as PI in the collaborative project "Informatics Instruments for the Social Sciences" at the Institute for Political Science and Japanese Studies and at the Institute of Informatics of the MLU Halle-Wittenberg.

Dr. Atsuko Sano is an e-learning specialist at the National Women's Education Center (NWEC) of Japan. She obtained her DBA (Doctor of Business Administration) in social design studies at Rikkyo University in Tokyo. Her doctoral dissertation focuses on integration courses for immigrants in Germany based on her research there. She worked as an e-learning planner at a private company for six years, as well as an assistant technical staff at The University of Tokyo (mainly at the Information Technology Center) for two years before joining the NWEC. She received a four-year KAKEN (Grants-in-Aid for Scientific Research) funding for researching digitalization and gender equality in Germany in 2018. She was recently invited to speak at the panel "AI and Gender" at a conference of the International Society for Gender Studies (held on September 1, 2019 at Rikkyo University). More information about her research activities is available here: https://researchmap.jp/sanoatsuko/?lang=english

Prof. Dr. Fabian Schäfer (PhD Leipzig University, 2008) holds the chair in Japanese Studies (Modern and Contemporary Japan) at the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). His research interests include the digital transformation of the public sphere, computational social sciences, media history, and modern Japanese (media) philosophy.

Cornelius Schubert is a lecturer for innovation studies at the University of Siegen. He studied sociology, psychology and English linguistics at the University of Kassel and the Technische Universität Berlin. He specialises in science and technology studies, medical and organizational sociology as well as qualitative research methods. He has conducted research on human-technology interactions in medicine, focusing on coordination, distributed agency and improvisation in high-tech settings. He also studied global innovation networks in the semiconductor industry, looking at the creation of organisational fields and technological paths. In addition, he is interested in the impact of computer simulations on predictive knowledge practices as well the proliferation of social innovations as policy tools. He currently is a PI in the project "Visually Integrated Clinical Cooperation" within the Collaborative Research Centre "Media of Cooperation" as well as in the Research Training Group "Locating Media" at the University of Siegen.

Franz Waldenberger is director of the German Institute for Japanese Studies (DIJ) in Tokyo. He is on leave from Ludwig-Maximilians-Universität Munich (LMU) where he holds the professorship for Japanese Economy at the Munich School of Management and the Japan Center. He received his doctorate in economics from the University of Cologne. His research focuses on the Japanese Economy, Corporate Governance and International Management. At DIJ he initiated the research program "Risks and opportunities in Japan – challenges in face of an increasingly uncertain future". He has published numerous articles and books on the Japanese economy. He is editor in chief of "Contemporary Japan" and member of the editorial board of other Japan and Asia related social science and economics journals. He was visiting professor at Hitotsubashi University, Osaka City University, Tsukuba University, The University of Tokyo and Shimomura Fellow at the Research Institute of Capital Formation of the Development Bank of Japan. He is member of the German Japan Forum and member of the board of the Japanese German Business Association (DJW).