JDSR 7(1) (2025) 1-19
JOURNAL OF DIGITAL SOCIAL RESEARCH
10.33621/jdsr.v7i148805

Digital transformation and research infrastructures

Promises and challenges of data-driven research in a Swedish context

Stefan Gelfgren and Coppélie Cocq

Umeå University, Sweden

🖂 stefan.gelfgren@umu.se

Abstract

Society is transforming due to changes in demographics, the environment, and technology, and thus faces multiple challenges. In this context, data coordination and access, collectively referred to as the digital transformation, are key to addressing anticipated societal tensions.

This interview-based qualitative study focuses on how researchers responsible for large-scale population-based research infrastructure view the opportunities and dilemmas in play in the intersection between data and personal privacy. The objective is to look beyond the glossy formulations of official strategy documents to see how the digital transformation (more specifically, data-driven research) is perceived from the active researcher's point of view, and what the intellectual negotiation process is like. What is of interest here is how the accessibility of register data is legitimized, and what developments and significant changes are simultaneously taking place. The research questions are:

1) How does the research community acknowledge the tensions and dilemmas between the possible risks and harms of large-scale, data-driven, population-based research, and its potential benefits?

2) How are the accessibility and coordination of research data justified and discussed by the research community, given the risks and potential, in relation to political and societal goals and policies?

With the contemporary Swedish research context as a point of departure, these research questions are addressed based on policy documents about digitalization, and on interviews with researchers.

Keywords: Digital transformation; digital humanities; surveillance culture; data-driven research; research ethics.

Published under a CC BY-SA license

1. Introduction

Today, the opportunities to use data to track and analyze processes and behaviors are enormous, growing, and promising in many ways, but they are not unproblematic. These opportunities are embraced at a large scale through different means, by different actors, and with different agendas.

This article discusses and analyzes how personal data in registers and databases are used and expected to be used in creating new knowledge. The focus is mainly on social sciences, humanities, and health research, and how it can help find solutions to pressing problems in politics, culture, and health. This article examines the tensions between the accessibility of data in relation to research, on one hand, and the legal and ethical limitations of this accessibility, as perceived and voiced by representatives of the research community, on the other.

The point of departure of this article is the contemporary Swedish research context. The article is based on the reading of policy documents addressing digitalization, using them to establish the Swedish framework, and on five semi-structured interviews with researchers involved, at a national leadership level, in large-scale population-based databases. It addresses the following research questions:

1) How does the research community acknowledge the tensions and dilemmas between the possible risks and harms of large-scale, data-driven, population-based research, and its potential benefits?

2) How are the accessibility and coordination of research data justified and discussed by the research community, given the risks and potential, in relation to political and societal goals and policies?

The objective of this study is to understand the context, prerequisites, and potential development of population-based, data-driven research. The potentials, risks, limitations, and legitimization of data-driven research are approached from the perspective of active research leaders in relation to policies and agendas advocating digitalization. In focus is how they, in their role as researchers, need to negotiate hopes, fears, legislation, curiosity, and research ethics.

With this article, we contribute a humanistic, qualitative, and phenomenological perspective to the discussion of how the actual users of the data, the researchers, relate to the digital transformation. This enables us to discuss the implications and impacts of this research development concerning issues such as privacy, personal integrity, and legal frameworks. This stands in contrast to research focusing on, for example, law, science and policy, or science and technology studies approaches, even though this article is informed by work done within these disciplines.

As implied by the formulation of the concept, data-driven research is usually driven by data rather than by research questions. Regarding the data and databases of interest here, the causality runs both ways: new and larger datasets enable researchers to ask new questions, while attempts to answer new questions (some of which are undefined beforehand) drive the compilation of new datasets. This is a development that will continue with ongoing developments within artificial intelligence (AI).

2. Point of departure: Digitalization and digital transformation

As noted by, for example, Sadowski (2019), data have become essential for contemporary society—for commercial enterprises as well as governments: "just as we expect corporations to be profit-driven, we should now expect organizations to be data-driven; that is, the drive to accumulate data now propels new ways of doing business and governance" (p. 1). The abundance of data in combination with the development of increasing computational power and AI enable new opportunities to conduct data-driven research. For example, Vey et al. (2017) claimed that "we are at the beginning of a revolution that is fundamentally changing the way we live and work, the so-called Fourth Industrial Revolution" (p. 23).

Similar arguments have also been made by, for example, the UN Development Programme (2019), which has claimed that "emerging [digital] technologies have the potential to advance sustainability and to lead to better development work" (p. 6). The OECD (2019) has similarly stated that it should "enhance access to data to drive digital innovation [and] promote interoperable privacy regimes to facilitate cross-border data flows" (p. 3). In its strategic research agenda, the European Commission (2019) has stated that issues related to future businesses, climate, education, health, security, etc., depend on the digital transformation and how it is handled.

The impact of these developments on research has been discussed in previous studies. In their *Humanities World Report*, for instance, Holm et al. (2015) identified massive and complex data as a major research area where digital research is addressed by humanities scholars (pp. 68–72). Pappalardo et al. (2021) observed that data-driven science is "changing the way research is performed" (p. 261), arguing that a new paradigm is emerging. The abundance of data is changing research in and across different disciplines, such as geography (Miller & Goodchild, 2015), medical chemistry (Lusher et al., 2014), and psychology (Jack et al., 2018). In recent years, we have seen major advances in AI, not least concerning the publication of the ChatGPT bot (and other generative AI services), spurring discussions of the opportunities and complications presented by contemporary technological developments (see, e.g., the petition from March 2023 to "Pause Giant AI Experiments: An Open Letter", initially signed by a thousand representatives of areas such as the tech industry and academia; Future of Life Institute, 2023).

Our interpretations are also informed by mediatization scholars such as Couldry and Hepp (2017), who observed that it is mainly private companies pushing the digitization, and ultimately the datafication (drawing on van Dijck, 2014), of society. In line with this, we note how the companies pushing digitalization will simultaneously own the data we all create, and that this raises questions of, for example, personal integrity and data ownership. Scholars such as Lyon (2017, 2018) and Zuboff (2019) view this development along similar lines, claiming that we live in a "culture of surveillance" (Lyon, 2018) or in a system of "surveillance capitalism" (Zuboff 2019). These scholars have seen how the digital footprints we all leave behind flow between, and are used by, businesses, banking systems, welfare authorities, the police and military, etc. (cf. Sadowski, 2019). Data are thus intertwined with our individual lives and behaviors, affecting our integrity and our perception of ourselves and our society. In this regard, Couldry and Hepp (2017) have claimed that we now live in a time of deep mediatization, meaning that the world and our relations are understood through media and data owned by companies and authorities, not by individuals.

A world of data indeed also offers new opportunities for research, but it also entails some rather problematic issues regarding, for example, personal privacy and legitimization.

This article focuses on the tensions between official digitalization agendas and concrete research practices, between the hopes associated with the digitization of society ("the digital transformation") and the related dilemmas of the datafication of society. More specifically, we examine how Swedish researchers responsible for large-scale population-based databases and related research infrastructure view the opportunities and dilemmas in play today, how research access to data is legitimized, and what developments and significant changes are occurring. The objective is to look beyond the glossy formulations of national and international strategy documents to see how the digital transformation (specifically regarding data-driven research) is perceived from the active researcher's point of view.

3. The Swedish case: The researchers' framework

Along with the other Nordic countries, Sweden has a long history of compiling data about its citizens (see, e.g., Andreassen et al., 2021; Ustek-Spilda & Alastalo, 2020, for an overview). For decades (even centuries), these data had mainly administrative and scientific purposes within the welfare state system, but today they have become an asset to exploit.

Similar discussions have been taking place in the Nordic countries. For instance, Tupasela et al. (2020) have observed, in the context of Finland and Denmark, the emergence of the notion and imaginary of a "Nordic data gold mine," and that the "logics of accumulation … reconfigure how the sources of this [sic] data are considered and imagined" (p. 2); such a discursive construct can be seen as "a necessary precondition" for articulating data policies ((Åm et al., 2021, p. 291 in a Norwegian context; see also Frank, 2020). Reutter and Åm (2024) have studied Norwegian policy documents to see how the discourse of hopes and technological determinism has been formulated. They have observed the Norwegian government pushing the digital transformation, noting that "datafication is accompanied by widespread beliefs that collecting and analyzing data can generate information and knowledge necessary for optimizing daily practices or for improving decision-making" (p. 1-2) and that "governments act as facilitators of digital markets" (p. 2).

In Sweden as early as the 17th century, clergy of the Church of Sweden (a state church until January 2000) registered Swedish citizens and kept track of birth and death dates, confirmation and marriage dates, names of families and relatives, and the ability to read and to understand the fundamentals of Christian doctrine (which were interwoven with the social order). These registers have been digitized and are now the basis of unique and world-renowned databases. In 1947, Sweden was the first country to introduce personal identification numbers to cover its whole population, initially for taxation purposes. Now the number is used whenever someone interacts with various state authorities, healthcare institutions, insurance companies, and banks and when ordering/buying various services. This gives Sweden (as well as other Nordic countries, as mentioned above) a unique position for register-based research at an individual level, especially if it is coordinated with health, welfare, and taxation registers (to mention a few).

Another contextual aspect of significance is the high level of trust in Sweden, as in the other Nordic countries, compared with large parts of the world. In the Nordic context, Sønderskov and Dinesen (2016) found "strong evidence of institutional trust influencing social trust" based on their analysis of datasets from Denmark. The Nordic countries are said to be "remarkable with respect to high levels of both social trust and, to a lesser extent, institutional trust" (Sønderskov & Dinesen, 2016, p. 187; see also Delhey & Newton, 2005; Zmerli et al., 2007), which are conclusions drawn from data from the World Values Survey (WVS) and the European Values Survey (EVS). Among groups earning the highest levels of trust, we find researchers and academia. And researchers are concerned with maintaining the level of trust they receive from the citizens.

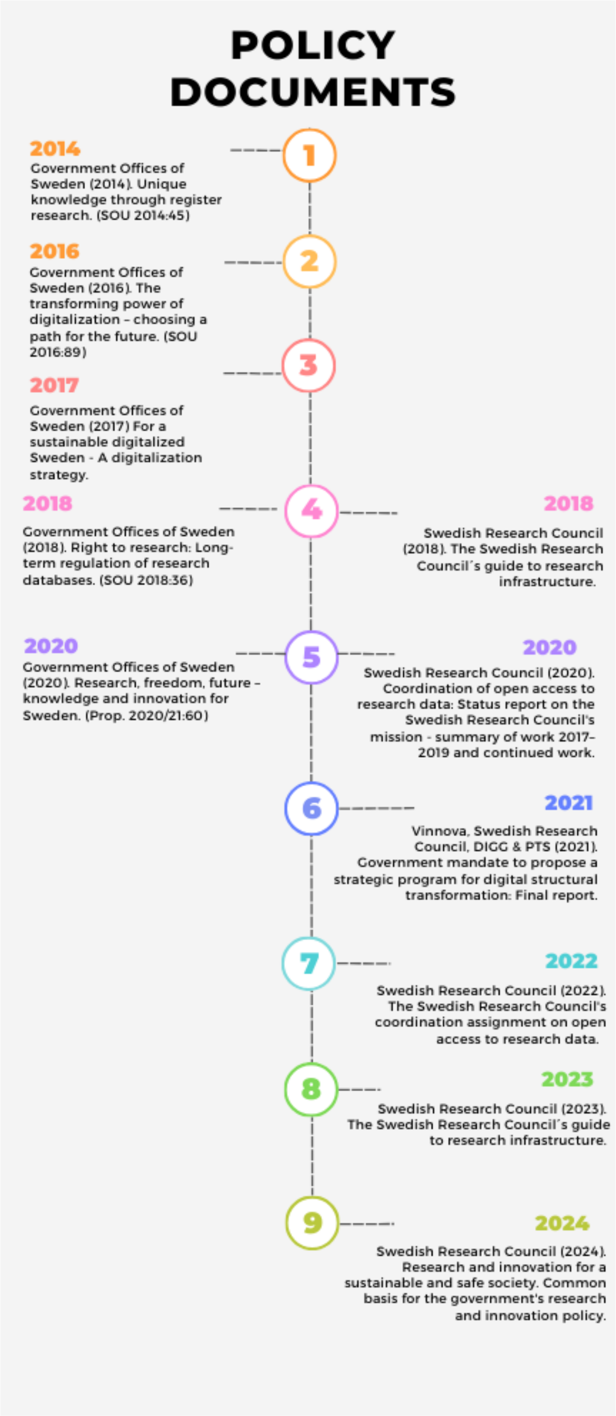
As early as the 1990s there was discussion of the high value of data based on the Swedish resident population, but ethical concerns were also raised early on concerning surveillance, and issues of privacy and integrity. "Integrity issues with the IT age were discussed by both politicians and journalists, and the [governmental] IT-Commission was no exception. But it never triumphed over the belief that problems would be overcome" (Bennesved, 2024). This line of reasoning still resonates in contemporary discussions, as we will see below.

4. Material and methods

This article is based on both Swedish policy documents, which introduce the overall framework and give context, and interviews with research leaders, which add depth and complexity to the discussion. These two levels are then contrasted in the discussion.

The first part of our material thus consists of strategy documents and agendas published in the last decade that address the process of digitalization at a national level. Here we have focused mainly on the Swedish government and the Swedish Research Council (Vetenskapsrådet), the main authority through which the government channels research funding.

Close reading was used as a first step in order to identify patterns and central themes in these policy documents, especially formulations related to the anticipated digital transformation. These policy documents provide the baseline to which the interviews then are contrasted.



2014
Government Offices of Sweden (2014). Unique knowledge through register research. (SOU 2014:45)

2016
Government Offices of Sweden (2016). The transforming power of digitalization – choosing a path for the future. (SOU 2016:89)

2017
Government Offices of Sweden (2017) For a sustainable digitalized Sweden - A digitalization strategy.

2018
Government Offices of Sweden (2018). Right to research: Long-term regulation of research databases. (SOU 2018:36)

2018
Swedish Research Council (2018). The Swedish Research Council´s guide to research infrastructure.

2020
Government Offices of Sweden (2020). Research, freedom, future – knowledge and innovation for Sweden. (Prop. 2020/21:60)

2020
Swedish Research Council (2020). Coordination of open access to research data: Status report on the Swedish Research Council's mission - summary of work 2017-2019 and continued work.

2021
Vinnova, Swedish Research Council, DIGG & PTS (2021). Government mandate to propose a strategic program for digital structural transformation: Final report.

2022
Swedish Research Council (2022). The Swedish Research Council's coordination assignment on open access to research data.

2023
Swedish Research Council (2023). The Swedish Research Council´s guide to research infrastructure.

2024
Swedish Research Council (2024). Research and innovation for a sustainable and safe society. Common basis for the government's research and innovation policy.

Figure 1. Overview of the documents analyzed (Authors' illustration).

To study tensions and dilemmas in relation to opportunities and challenges within the research community, we conducted in-depth interviews with established and active senior scholars involved in policy and decision-making processes at the national level. These interviews took the Swedish case as the point of departure. The scholars interviewed for this article were either heads of databases and related research infrastructure and/or were involved in policy and strategy work—all in disciplines that handle data from individuals, specifically the social sciences, humanities, and health research, and at the intersection between these traditional disciplines. They were also active researchers using the data obtained, for example, through and from the research infrastructure they led. They had extensive experience (30+ years) in their fields, and were chosen to cover a large interdisciplinary field. Gender and geographical distribution were also taken into consideration.[1]

Five interviews were conducted during the spring of 2020, after informed consent was received. Three interviews were conducted through video and two were conducted face-to-face; in all five cases, only audio-recordings were made (as preferred by the interviewees) and stored on the interviewer's computer. The interviews were conducted and transcribed in Swedish, and the translations into English are our own. As active researchers themselves, all participants were used to interviews, to the video format, and to being recorded, so no specific methodological problems were encountered. The interviews lasted 40–60 minutes and centered on questions related to the topic of this study, i.e., the opportunities and dilemmas associated with population-based databases and registry-based research. Questions were asked about issues such as the potential for doing research with large-scale datasets, and the risk of using data in a possibly individually intrusive way. These questions were based on the analysis of policy documents articulating the potential and opportunities offered by the digital transformation. Opportunities, in this context, included hopes and risks. Dilemmas emerged in the interviews when discussing the application and implementation (as a fact or effort) of the digital transformation in research.

Questions were asked about what opportunities there were and how they could be better utilized, about potential ethical dilemmas, the future, and perceived limitations. All interviewees expressed the desire that more researchers would use the register-based data, but all interviewees acknowledged that there were risks connected to the increased use of such data. Thus, the tensions and dilemmas this article seeks to explore were prevalent in the interview material.

The interviews were transcribed and coded based on a thematic analysis (Gray, 2002; Riessman, 2007) aligned with our research questions and our theoretical understanding of the subject. Recurring themes in all interviews were identified as potentials, risks, legitimization, and limitations. These themes are further analyzed and discussed in the empirical section below.

5. Theoretical framework

For this study, we drew on thematic discourse analysis. Our focus was on reflections and ideas rather than on technology and research activities per se. The aim was, as mentioned, to study how the use of various forms of register data is negotiated in the field of data-driven research (in the social sciences, health studies, and the humanities). The negotiation process is situated within the all-embracing political and societal discourse, which pushes research toward a digital transformation that is expressed and materialized in the political, policy, and strategy agendas. To understand this framework, within which the research discourse is situated, we were inspired by science and technology studies (STS). According to the STS scholars Biljer et al. (2012), the discourse has gained "momentum" and is supported by "organizations and people committed by various interests to the system" (p. 70). We accordingly lean more toward the sociology of studies than toward science policy studies (cf. Gläser & Laudel, 2016).

However, given the positions of the interviewees, they were also part of formulating research agendas. Here we were inspired by, for example, STS scholar Geels (2002), who saw how technological development was formulated at a macro level (here, political discourse), negotiated at a meso level (here, research discourse), and implemented/applied at a micro level (omitted here because this study was not interested in the actual implementation process), and noted that all three levels were interrelated. We saw how the researchers formulated a research discourse within and related to the political discourse.

Our approach also acknowledges the significance of the social and cultural contexts in which technologies take shape and are implemented. Such an approach is illustrated, for instance, by boyd & Crawford (2012), who defined "[b]ig data as a cultural, technological and scholarly phenomenon that

---

[1] Sweden is a small country, and Swedish academia is even smaller, so due to matters related to confidentiality, it is impossible to further specify the participants' background and still guarantee anonymity.

rests on the interplay of (1) Technology … , (2) Analysis … (3) [and] Mythology" (p. 663). In line with this perspective, our article takes the theoretical stand that it is important to consider the role of "the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy" (boyd & Crawford, 2012, p. 663). The concept of imaginaries (Lyon, 2018) inspires this approach by emphasizing the influence of what we think about what we do (see, e.g., Samuelsson et al., 2023)—in this case how imaginaries influence discourses and research practices.

6. Digitalization in Swedish policy documents

As early as 2011, the Swedish government formed a commission to formulate a digital agenda for Sweden, in order to foster the potentials brought by the digital transformation (Government Offices of Sweden, 2011). The commission's final report in 2016 was overtly positive toward digitalization, beginning with the commission chair's following claim:

> We live in exciting and interesting times! Digitalization is the most transformative societal process since industrialization. This development allows us to do entirely new things and to perform activities we previously engaged in, in entirely new ways. Our knowledge and understanding of humanity, society, and the environment will be transformed by the opportunities provided by the analysis of large amounts of data. Industrialization led to the development of the welfare society we have today, providing more people with the opportunity for a good life. Digitalization has the potential to develop a democratic and sustainable welfare society that we can hardly imagine today. (Government Offices of Sweden, 2016)

Still, the report acknowledges claims that societal benefits must be balanced against issues related to privacy risks. The digitalization strategy issued by the Swedish government the year after the commission's 2017 report states that Sweden should aim to be the best in the world when it comes to realizing the potential of digitalization. There are major potential benefits for the economy, employment market, and democracy; there are risks, though, connected to trust, democracy, and personal integrity (Government Offices of Sweden, 2017).

This strategy, focusing on the potential, is also evident in research strategies. In Government Offices of Sweden reports from 2014 and 2018, opportunities for register-based research are surveyed to find legal space in which to conduct such research—to balance potentials and questions related to privacy (Government Offices of Sweden, 2014, 2018). For example, the 2018 report mentions that "Sweden has a world-leading position when it comes to statistics regarding living conditions and health. Due to our individual-based registers, there are rich opportunities to perform successful research based on these. … To fully exploit these unique prerequisites, the more efficient use of existing registers and databases is needed" (Government Offices of Sweden, 2018).

In a Swedish government investigation from 2021, the authors, four Swedish research authorities (Vinnova, Swedish Research Council, DIGG – Myndigheten för digital förvaltning/Agency for Digital Government, & PTS – Post- och telestyrelsen/Swedish Post and Telecom Authority), acknowledged the urgent need for Sweden to strategically coordinate and use the potential of the vast amount of data it has accumulated (Vinnova et al., 2021). The driving forces are a competitive and sustainable economy, high-quality research supporting innovation, a desire to address societal challenges, and a drive to achieve accessible public administration supporting innovation and participation (p. 5). The process of digitalization is supposed to be "green, competitive, and centered on humanity" (p. 12). A few possible negative aspects were noted, with data management and storage being mentioned as a challenge concerning issues of personal privacy: "Digitalization brings new challenges … related to managing and storing data in relation to individual integrity. These challenges need to be addressed for the individual to trust in a digital society" (p. 14).

Related to the unique Swedish registers (often at the individual level), the digitization and coordination of individuals, and the possibility of using the Swedish registers is mutual *trust* among the state, the authorities, and fellow citizens. This is a challenge, and the strategic program for the digital transformation states that "these challenges have to be addressed for the individual to feel trust in relation to digital society" (Vinnova et al., 2021, p. 13).

The governmental research funding bodies the Swedish Energy Agency, FORMAS, Forte, the Swedish National Space Agency, the Swedish Research Council, and Vinnova were commissioned by the government (Government Offices of Sweden, 2023) to formulate a basis for the government's future research and innovation policy. In their report, data-driven research was deemed important, but these funders simultaneously acknowledged related privacy issues.

Today the Swedish research agenda acknowledges that the current digital transformation is changing the foundations of society. In the *Guide to research infrastructure 2023* (Swedish Research Council, 2023), digitalization is emphasized as a key component of future research addressing societal challenges, referring to the governmental research agenda. The *Guide* states that this must be done while taking account of ethics and privacy concerns. This tension comes through in formulations such as the following:

> The increased need for longitudinal studies of individuals also raises the potential conflict between the requirement to protect the integrity of sensitive personal data and the need for and opportunities of open science, where data may also be made accessible during peer review of scientific publication. (Swedish Research Council, 2023, p. 51)

The *Guide to research infrastructure 2018* (Swedish Research Council, 2018) mentions that it is essential to ensure personal privacy in order to maintain trust in and the legitimacy of research. As illustrated in the above quotation, the guiding research agendas articulate a push toward more openness and coordination of research data through open-access policies and data management according to the findable, accessible, interoperable, and reusable (FAIR) principles (see, e.g., Swedish Research Council, 2020, 2023), following a directive to the Research Council from the government in 2017. A Swedish Research Council (2022) report on the accessibility and coordination of open data also mentions that issues of personal integrity must be taken into account.

The policy documents illustrate a clear push toward digitalization within the political discourse in Sweden, with the benefits being highlighted while more problematic aspects play a complementary but subordinate role. The main discourse articulated in the political agendas is that the process of digitalization will give Sweden advantages in the areas of business, welfare, the environment, etc., all expressed and materialized in Swedish business and research agendas, and that, if not utilized, Sweden's potential as a digitally transformed society will be lost. This view (or imaginary) of digitalization is articulated in mainly positive terms and in a future-oriented discourse that depicts a better society using a somewhat utopian rhetoric (cf. boyd & Crawford, 2012). This is the political framework the research community works within.

7. Negotiations within the research discourse

As shown above, the policy documents largely favor an increasing degree of digitalization to address contemporary challenges through realizing the potential of data-driven science—manifested in research agendas and the implementation of the FAIR principles (meaning that data should be Findable, Accessible, Interoperable, and Reusable). Here we want to contrast and problematize these generalizations, setting them against the voices of researchers as they come through in the interviews, taking the discussion from a meta to a meso level.

Four main recurring themes were identified in our thematic discourse analysis of the interviews. These themes are illustrative of, and highlight, the tensions in play within the research discourse and structure the presentation of our results, namely: “Legislations”, “Limitations”, “Legitimizations”, and “Potentials and Risks”.

7.1 Legislations and guidelines – “If you look strictly at the law…”

The frameworks referred to by the interviewees include mainly ethical vetting and GDPR legislation. These frameworks are mentioned both as necessary in order to regulate data use and how research is conducted, and as limiting and maladapted in relation to recent changes in contemporary data and their uses.

The Swedish Ethical Review Authority is the national agency that reviews and grants ethics permits. Researchers are required to apply for ethical vetting in certain cases, for instance, if the research project involves sensitive information (i.e., in the context of the Review Act, information concerning ethnicity, political opinions, religious or philosophical beliefs, union membership, health, sexual life or orientation, and genetic and biometric data). In addition, some specific laws and ordinances regulate access to register data. For instance, confidentiality legislation regulates data that are worthy of special protection and must be kept confidential. The main rule is to guarantee public access to information, and the legislation delimits under what conditions it may be possible to access data despite their confidentiality.

The legal framework that regulates data use today is, according to the interviewees, necessary for the control of data and registers, as mentioned, for instance, by Interviewee 5:

> That is the reason why we have ethical vetting – to see if the benefits of the research conducted are great enough to disclose the data. It is such a balance we need to strike.

However, some interviewees had observed that current legislation was not adapted to contemporary conditions. Several interviewees mentioned the Swedish Ethical Review Authority as an important instance for the research process, but also identified areas that need to be developed and adapted. Knowledge of and conditions for managing social media data constitute a weakness of the authority mentioned by Interviewees 1 and 5. Another interviewee, number 4, had experienced that expertise was lacking in disciplines outside health and medicine. We were also told that opportunities for linking/merging data were limited due to legislation:

> Now we hope for new legislation. We do not really know what’s going on. There have been a number of investigations of the matter. … Right now we do not have any good legislation in this area in Sweden. For example, there is no good legislation for research databases that allows you to build a research database and make it available as infrastructure. You do that anyway, but there are, if you look strictly at the law, no such opportunities. (Interviewee 2)

The same interviewee commented that new legislation was needed “above all, to adapt to a reality that actually already exists.”

Worth noting is that several interviewees comment how the legislation and regulatory frameworks regarding data have changed over the years. The interviewed researchers described how research and research infrastructure have been – and still are – a driving force behind the development of laws and codes of conduct for data management. Therefore, the current situation is not fundamentally new, even though the scale of possibilities for conducting data-driven research has immensely increased, which continues to (again) push the boundaries of regulations.

Questions concerning register data and access to research data have, in the past as well, emphasized the tensions between legislation and research (Interviewees 1, 2, 3). For instance, Interviewee 3 stated: “From having been a bit controversial and suspect, collecting health data has become quite uncontroversial.” This quotation illustrates not only a change, but also how our attitudes toward the acceptance of data collection and data use have shifted.

The main dilemma identified by the interviewees was that an ethical and legal framework is important and necessary. The legal framework keeps research within ethical bounds and balances possibilities with what is reasonable and relevant to do. Simultaneously, and paradoxically, the framework is very difficult to implement, or the implementation limits the possibilities that current data-driven research can offer.

7.2 Limitations – "Great potential, if only the framework is adjusted"

The nature of data has indeed changed, and today data are digital and come with the possibility for different datasets to be coordinated and thus made interoperable. The amount of data that is produced and that can be collected is unprecedented, and the technical possibilities accessible to researchers for coordinating and analyzing data are rapidly developing. As Interviewee 2 straightforwardly said, "I think it is impossible to stop this". These new conditions not only put light on the tensions between legislation and data collection/use as illustrated above, but also actualize the ongoing challenges that contemporary research meets and addresses, given that the digital transformation within the political discourse has gained momentum, according to Bijker et al. (2012).

Our interviewees provided several examples of how researchers have striven to advance and adapt frames to contemporary conditions and needs. Interviewee 2 said that one vision or goal was to make as much data as possible accessible, but that current legislation was a hindrance. The role of researchers in identifying needs for changes and adapting the legal framework is nothing new. The relationship among researchers' expectations, the newly emerging opportunities, and the limits on what can be done was mentioned in several interviews. Interviewee 1, commenting on the establishment of data infrastructure (over 10 years ago) for which they were responsible, said:

> I think that the Ethics Review Board had a bit of a hard time seeing how that would fit in with legislation and practice. … it ended with us actually having to come to the board and answer questions.

Another interviewee shared a similar experience:

> When I started looking for infrastructure funds, those who sat in [the responsible board of a funding agency] did not understand. They did not seem to understand what it was, that it was even a question of infrastructure. (Interviewee 4)

Several interviewees returned to the fact that they, and the needs of the research community, are not understood properly by the law and by the vetting boards that restrict the use of registers and accessible data. The guidelines for the law and the vetting boards are based on the needs of medical research and were established during a time when data was relatively scarce, they say. The threat to privacy and integrity is overestimated, according to Interviewee 1, because it is never in the interest of researchers to single out individuals: the primary interest of research is large-scale patterns. The potential to use data in registers is huge, and such data could be used even better. Registers are built of anonymized data, but of course if different datasets are combined, a qualitatively new register is created, which needs to be assessed by vetting boards and must comply with laws and regulations.

The interviewees argued, for example, that as social media data exist and are used by companies, such data should also be used in research. Another reason to use new data, such as social media data, is the difficulties in collecting data using traditional methods, since people are now less likely to complete surveys and the like. The same difficulties in collecting data in traditional ways were mentioned by Interviewee 3, who said that people "do not want to give out their phone numbers, and they do not answer the phone, and they do not complete questionnaires, and so on." Here other forms of data can be helpful by pushing research toward new datasets offering new possibilities.

It is in the possibility of combining data from different registers and data sources that the real potentials can be found, and as Interviewee 4 said:

> One should be clear about that, that it is this connection, or interoperability, that makes data interesting, and also extremely sensitive in terms of personal privacy. There are so many aspects, different facets of a person's life, that can be put together into something complete. Yes, anyone can understand that this is potentially very sensitive.

Therefore, it is also important to have a clear legal framework and for researchers to understand that their activities are based on citizen trust, as claimed by Interviewee 4. Trust is thus highly valued by the researchers, and the risk of compromising it is carefully considered, especially in a country where trust in research and researchers is high, as mentioned in the introductory sections.

The interviewees shared the opinion that there were immense opportunities to perform new and innovative research in connection with digitalization. Merging datasets of various forms and making them searchable and interoperable would enable qualitatively and quantitatively new research, which would be interesting from the point of view of curious researchers. New questions could be asked and answered faster, on a larger scale and in real time. In particular, data that people share via, for example, social media and health apps, offer new potentials if they are combined with existing data on demographics, welfare, and societal infrastructure. Still, the risks of doing so were acknowledged.

7.3 Legitimization – "If private companies do it, why not us?"

The interviewees were well aware of the pitfalls and risks related to the use of population-based databases, but at the same time, all were aware that such data are already used in compiling and analyzing huge datasets within and related to businesses and companies. Global actors such as Google, Facebook, and Amazon were emphasized in this regard. It is well known that these actors gather and sell data to anyone interested based on the content that individuals provide them (cf. Lyon, 2017, 2018; Zuboff, 2019). This is both theoretically known and practically experienced, and the interviewees mentioned, for example, that advertisements based on their search history would appear shortly after searching for or buying particular items online. Rhetorically, they asked whether it would not be better and safer if researchers used similar tools and methods to conduct research for the collective good instead of for profit based on doubtful business models. Academic research is guided by laws, ethical guidelines, and best practices, which are considered to constitute a bulwark against the misuse of data. For example, "if we already have data regarding our movements [through mobile phones and health apps], why not use it to do some research too?" (Interviewee 1). "So, other actors, commercial actors [Google and Amazon were mentioned] surveil us in a very, yes, more intrusive way than researchers do" (Interviewee 3).

From the researchers' perspective, the other main reason for building and using large datasets is, as hinted above, the advantages individuals and society can gain from the related research results in areas such as health, the environment, or "epidemiology … and [overall] from a state financial perspective, on how to use our resources as efficiently as possible" (Interviewee 4). By connecting and coordinating different datasets, it is possible to gain insights into increasingly complex problems, which of course is both relevant and tempting. Contemporary society is facing large problems concerning health (accentuated by the Covid-19 pandemic) and the environment (accentuated and discussed in relation to what are considered contemporary extreme weather conditions), and here register-based research can help us understand causes and find solutions. Merging welfare registers capturing longitudinal living conditions with health registers and social media data could be a huge asset for researchers asking and answering new questions (Interviewees 1, 3, and 4).

However, as one interviewee mused, data-driven research is also about the sheer curiosity of researchers and the human impulse to explore whatever possibilities there are: "If we have Mount Everest, we need to climb it!" (Interviewee 2). If there are multiple datasets that could be enhanced and enriched by coordinating them to answer new questions, there are always people who, out of sheer curiosity, will want to try.

Finally, it was noted that quantitatively large datasets are always anonymized and that there is no interest in the individual persons from which the data are derived. Therefore, it is not even possible or in any way relevant to identify the individual persons behind the data. It is argued that the risk of compromising the privacy of individuals is low to non-existent.

The bottom line is that the benefits of using these kinds of data in research are much greater than the risks, and because businesses can and do use such data, why not use them in a responsible way guided by ethical guidelines and best practices?

As mentioned in the background section on the Swedish context, Sweden has a long tradition of procuring register data. This historical background was mentioned by several interviewees as a motivation for continuity or even as "a tradition" (Interviewees 1 and 3) in data harvesting, collecting, etc. Interviewee 3 mentioned that "in Sweden, we have a tradition of digitizing both the health and welfare systems, since we were early adopters in this regard."

The trust and responsibility conferred by citizens (Interviewees 1 and 2) also legitimize the use of register data. One interviewee emphasized that their work with databases and registers was based on "the trust we have received from society, i.e., to use these data in a good way" (Interviewee 1). Sweden is a high-trust country, so it is not surprising that this aspect is an important motivation (and prerequisite) for the development of registers and databases. Trust in researchers was also mentioned in contrast to other actors involved in collecting (big) data. For instance, we were told by Interviewee 2 that research institutions are better than private or commercial actors in this regard:

> [In another country] they have [name of a database], which is one of the really large population databases, which is then linked together with healthcare data. And there it is owned by a private company. And it does not feel so good. … So, I think, I believe very much in the public. I also believe in the structured accessibility of what research needs.

On a similar note, another interviewee shared a reflection about how "it is not research that is the problem here, but it is … it is, so to speak, states and commercial actors" (Interviewee 5).

Overall, the aspects of tradition, continuity, and trust, referring to the Swedish case, are considered important for motivating and legitimizing population-based research. The interviewees discussed legitimization in relation to potentials and to what other actors in the field do – especially in relation to the business sphere. Here the interviewees referred to the need for researchers to maintain people's trust in them, otherwise their research would lose legitimacy. These are key aspects of how pure research differs from, for example, the suspect interests of commercial actors and states in surveilling their citizens. Research ethics, the idea of doing research for the greater good of society, and the need to maintain a trustworthy position also legitimize the access to, and use of, these kinds of data.

7.4 Potent data and potential risks – "Potentially very invasive with regard to privacy concerns"

Although the potentials of compiling and using registers and databases for research outweigh the associated risks and threats, the interviewees discussed possible negative consequences. Some risks can be foreseen, but these risks are more related to actors other than researchers who have other agendas, such as commercial actors (mentioned above) and state-governed intelligence agencies.

The other side of the huge possibilities and potential of register-based research can be seen in the policy documents. The better the coordination and combination of datasets, the better the potential to do ground-breaking research; at the same time, the more careful the consideration of privacy issues should be. In Swedish register data, it is possible to reach "a very fine and detailed level," so research based on these data could "potentially [be] very invasive with regard to privacy concerns" (Interviewee 4). Here we are talking about the potential and potent combination of health data from hospitals and recurring health surveys, data on demographics and socioeconomic conditions, social media content, and data from wearables.

Published under a CC BY-SA license

Traditionally, data have been collected at recurring intervals, giving only snippets of information limited to capturing conditions as specific points in time. However, with a more constant flow of data from, for example, social media and health apps, in combination with, for example, AI, language translation, and facial recognition, data will reach a new level of continuity and granularity. This will entail both greater potential and greater risk – the latter perhaps less in a research context, but more in the hands of malevolent states, businesses, and political actors, as mentioned in the interviews. Concerns were also raised regarding the possible use of such datasets to shape public opinion and consumer patterns (which is already taking place and was a main concern, for example, in the Cambridge Analytica scandal in 2018 and in the US Special Counsel Investigation in 2017–2019).

Social media data and data shared through, for example, health apps and services, were singled out as both extremely potent and risky: "The data we share about how we move, what we say, what we think, and what we download, these data are authentic data on a whole different level" (Interviewee 1). On the other hand, it is difficult to understand why people generally freely distribute such data, as Interviewee 5 observed, fascinated by the volume of data people share apparently without concern about how they could be used:

> People share so much data about themselves, and so much sensitive data about themselves, at the same time as there are so many conflicts concerning data in relation to research … It does not make sense! (Interviewee 5)

One interviewee also mentioned that datasets not intended to be out in the open are made accessible under the provisions of what is referred to as open access. Publishers demand that researchers publish their data in order to make the research process transparent, but, for example, the so-called quality registers (or health data registers) based on medical records were never intended to be public. We have a situation in which the research discourse is promoting open access, whereas:

> The quality registers were built by healthcare actors, intended to ensure quality in the healthcare system, but now the demands to open these registers are increasing, which the medical professionals are a bit reluctant to do. (Interviewee 3)

The question is how to draw the line between research interests and personal privacy, as formulated by Interviewee 5, when the potential also entails risks. Again, larger datasets enable interesting research regarding complex and relevant questions, but at some point, the sheer amount of coordinated data might become problematic. Data collected or shared for other purposes than originally intended, for example, data from health service quality registers or people's social media data, are, if combined, a potential risk, although simultaneously offering huge potential.

8. Discussion

The themes of research "Legislations", "Limitations", "Legitimization", and "Potentials and Risks" are factors the interviewees mentioned that influence the research discourse and are thus negotiated. The research discourse takes these factors into consideration while negotiating with the political discourse regarding the terms and conditions of the research discourse (see Figure 2). This form of negotiation illustrates how digitalization as a socio-technical phenomenon is perceived, articulated, and implemented according to both utopian imaginaries – i.e., how research and society will benefit from it – and cautious imaginaries emphasizing privacy (see, e.g., Lehtiniemi & Ruckenstein, 2019; Lyon, 2018; Tupasela et al., 2020).

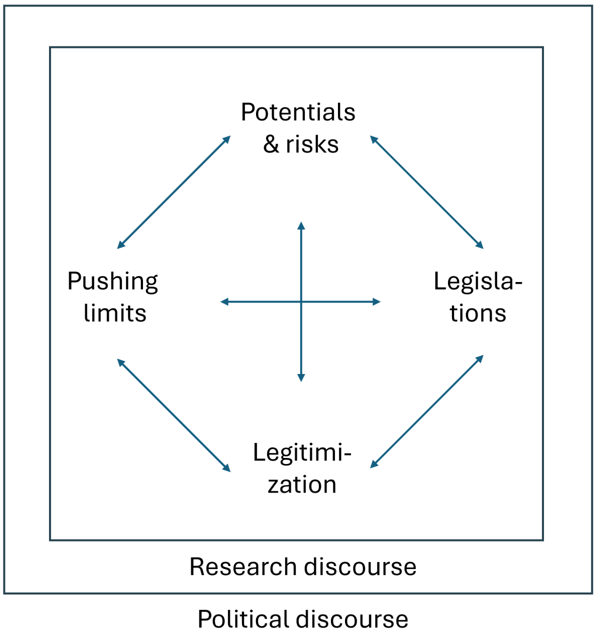


Figure 2. The negotiation process that occurs in the research discourse, related to the political discourse.

The political discourse, as mentioned in policy and strategy documents, encourages society and research to move toward increased digitalization. In Swedish policy documents, this is articulated as building on a unique tradition of collecting data about citizens—as an asset to explore and capitalize on – in line with the idea of a "Nordic data gold mine," as suggested by Tupasela et al. (2020).

However, taking account of discussions within the research community, as manifested in the interviews, it becomes clear that the digital transformation is not a predetermined one-way process that can be implemented without discussion of its consequences. Instead, it is a discourse that recalls the privacy paradox, which has been studied among ordinary citizens and users of digital media. Such a paradox, defined as the "discrepancy between individuals' intentions to protect their own privacy and how they behave in the marketplace" (Norberg et al., 2007, p. 101; see also, e.g., Kokolakis, 2017, for an overview), highlights the tensions arising when the user is "expected to trade the benefits that could be earned by data disclosure off against the costs that could arise from revealing his/her data" (Gerber et al., 2018, p. 229). More recent research indicates, however, that what is perceived as paradoxical at first sight can indeed be "partially interpreted and explained in terms of ethical and ideological considerations" (Cocq et al., 2020, p. 191).

Despite the overtly positive political discourse – both nationally and internationally – the interviewees expressed a more nuanced view of digitalization. The digital transformation of society in general, and of research in particular, was seen as bringing great potential and benefits, but also as needing to be balanced against anticipated risks (indeed, as noted in the policy documents but rarely dwelled on, or problematized, at length).

This discourse is present in all the above themes, and we find several contradictory aspects discussed in relation to the use of population-based research data of various forms. For example, the legal and ethical frameworks hinder researchers from doing whatever they want with the data. They are thus mentioned as important for ensuring that research questions are balanced against proper needs, relevance, and research ethics. The trust researchers feel they have from the citizens is also part of the discussion. Trust is valued and something they care about. Similarly, the interviewees saw huge potential in better computational power and new and interoperable data, but they also saw the legal framework as obsolete and inappropriate for the kind of research they wanted to do.

Ethical issues raised by data-driven technologies are being addressed and discussed in recent research, for instance, concerning the role and prerequisites of ethics committees reviewing research based on data-driven technologies in university contexts (Hine, 2021). Such research emphasizes the limitations and challenges that ethical committees encounter and underscores the need to develop effective systems of ethical governance. Also, Forgó et al. (2020) discussed the need for suitable infrastructural, organizational, and methodological principles when establishing ethical – legal frameworks in research. In a Swedish context, as illustrated and discussed under the themes of Legislations and Limitations, the work conducted by review boards and associated infrastructures in relation to ethics and legislation concerns understanding and adapting the current framework, which is being challenged by data-driven research.

The interviewees discussed the role of research as a curiosity-driven activity, and they were eager to find new solutions to existing and anticipated problems in research done for the greater good, and thus in line with both the research opportunities and potential breakthroughs that come with increased digitalization. Here the research discourse differs from, for example, the business discourse, which is primarily based on a profit-driven agenda. In the research discourse, the researchers discuss and negotiate risks associated with research conducted in relation to the very same process of digitalization.

The opportunities associated with data-driven research are balanced against its potential risks – i.e., addressing interesting and complex research questions versus threats to personal privacy. Ongoing and future research must be balanced and legitimized by invoking the great societal need for innovative research, and by comparing this need with what has already been done by other actors with commercial/financial interests. In contrast, the researcher's personal curiosity is a driving force that might push the boundaries of research.

9. Conclusions

Significant developments and changes have been occurring in relation to the use of research data, infrastructure, etc. This study casts light on the associated tensions and dilemmas, from the perspective of the research community, as these developments and changes have occurred in legislation and in practice. Recent generative AI developments highlight these tensions in relation to questions of the ownership of the data on which the AI models are trained (see, e.g., Lucchi, 2023; Samuelson, 2023).

Overall, we have a situation in which the political discourse has gained momentum and is pushing for digital transformation at the macro level. In the research community and research discourse, these tensions are articulated. Here we have aimed to nuance discussion of the digital transformation – a process that is not a predetermined one-way process, even though the political discourse has promoted the development of data-driven research, in line with the anticipated digital transformation of society. As the analysis of the policy documents clearly indicates, digital transformation and data-driven research are seen as keys to addressing contemporary challenges related to demographics, climate change, and democracy, also having the potential to give rise to new business models. The concept of open data in relation to technological advances is key to encountering the future, although issues such as personal privacy are acknowledged to be at stake.

Different interests and arguments were, however, discussed and balanced in relation to one another by the interviewed researchers. The potential to conduct new, important, and relevant research is acknowledged in the research discourse, but such research is not as straightforward or single-minded as is described in research policies and documents at the national and global levels. Analysis of the interviews shows that there are different reasons to legitimize the use of data and to coordinate and merge large datasets. The legitimization arguments expressed in our interviews centered on the facts that research using these forms of population-based data is beneficial for society and the greater good, that the risk of violating someone's privacy is low to none, and that it is better that such data be used by researchers working under ethical and legal guidelines than by businesses working for profit.

The forces in play concern the increased digitalization of society and what it means for research are, as we have shown, multiple. Policies, agendas, and political discourses are explicit, plain, and clear. The data considered here are also embraced by commercial actors, who are yet another driving force shaping how data are compiled and shared. However, we also see that the research community is a key actor in this process. In practice, the frameworks within which large databases and data infrastructure are developed and applied are constantly challenged by research and researchers, resulting in necessary adaptations.

This article has focused on the use of population-based data, discussing the contemporary and potential use of such data. However, to fully understand the opportunities and potentials of society's digital transformation – in research and in other areas of society – we must consider the owners and providers of such data, namely, the citizens. A related study conducted by our research group shows that many are concerned about their data being used without their consent, and thus adjust their online behavior to conceal their data or prevent it from being gathered (Cocq et al., 2020). The interviewees in this study also touched on the fact that it has become increasingly difficult to obtain data by having people complete questionnaires and voluntarily participate in research studies. Therefore, other data collection approaches are undertaken, and to obtain other forms of data, for example, through social media platforms. Our interviewees also mentioned that medical professionals who have compiled registers of health data (the quality registers) are hesitant to open these registers for research for other purposes than originally intended.

Ultimately, it is a matter of the legitimacy of the digital transformation, in research and elsewhere, and the question is whether potentially sensitive data should be allowed to "float around" and be used for various purposes other than originally intended as long as the intentions are good and relevant. This question was briefly touched on in the empirical material and selected policy documents, but to ensure continued trust and legitimacy in the future digital transformation of research, this article shows how these complex questions must be addressed more thoroughly.

Acknowledgements

This article is part of the project iAccept: Soft Surveillance – Between Acceptance and Resistance, funded by the Marcus and Amalia Wallenberg Foundation (MAW2016.0092).

References

Åm, H., Solbu, G., & Sørensen, K. H. (2021). The imagined scientist of science governance. *Social Studies of Science*, 51(2), 277-297. https://doi.org/10.1177/0306312720962573

Andreassen, R., Kaun, A., & Nikunen, K. (2021). Fostering the data welfare state: A Nordic perspective on datafication. *Nordicom Review*, 42(2), 207–223. https://doi.org/10.2478/nor-2021-0051

Bennesved, P. (2024). Surveillance, integrity and metadata in the information age The legacy of Sweden's ICT commission, 1994–2003. *Lychnos: Årsbok för idé- och lärdomshistoria,* 105–129. https://doi.org/10.48202/26116.

Bijker, W. E., Hughes, T. P., Pinch, T. J. (2012). *The social construction of technological systems: New directions in the sociology and history of technology* Cambridge, MA: MIT Press.

boyd, d., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Cocq, C.; Gelfgren, S.; Samuelsson, L.; Enbom, J. (2020) Online Surveillance in a Swedish Context: Between acceptance and resistance. *Nordicom Review*, 41(2), 179–193. https://doi.org/10.2478/nor-2020-0022

Couldry, N. & Hepp, A. (2017). *The mediated construction of reality*. Cambridge: Polity.

Delhey, J. & Newton, K. (2005). Predicting cross-national levels of social trust: Global pattern or Nordic exceptionalism? *European Sociological Review*. 21(4). 311–327. https://doi.org/10.1093/esr/jci022

van Dijck, J. (2014). Datafiction, dataism and dataveillance: Big Data between scientific paradigm and secular belief. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

European Commission. (2019) *Orientations towards the first strategic plan for Horizon Europe*. Accessed 10 Feb. 2021 from https://research-and-innovation.ec.europa.eu/system/files/2019-12/ec_rtd_orientations-he-strategic-plan_122019.pdf

Forgó, N., Hänold, S., van den Hoven, J., Krügel, T., Lishchuk, I., Mahieu, R., Monreale, A., Pedreschi, D., Pratesi, F., & van Putten, D. (2020). An ethico-legal framework for social data science. *International Journal of Data Science and Analytics*, 11(4), 377–390. https://doi.org/10.1007/s41060-020-00211-7

Frank, F. (2020). When an Entire Country is a Cohort. *Science* 287, 2398-2399. https://www.science.org/doi/10.1126/science.287.5462.2398

Future of Life Institute. (2023, 22 mars). Pause Giant AI Experiments: An Open Letter, Accessed 2 Feb. 2024 from https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy* 31(8). 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Gerber, N., Gerber, P. & Volkamer, M. (2018), Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, (77). 226-261. https://doi.org/10.1016/j.cose.2018.04.002

Gläser J, Laudel G. Governing Science: How Science Policy Shapes Research Content. *European Journal of Sociology*. 2016;57(1):117-168. https://doi.org/10.1017/S0003975616000047

Government Offices of Sweden (2011). *IT i människans tjänst—en digital agenda för Sverige (dnr N2011/342/ITP)*. See https://www.regeringen.se/rattsliga-dokument/kommittedirektiv/2012/06/dir.-201261.

Government Offices of Sweden (2014). *Unik kunskap genom registerforskning [Unique knowledge through register research]*. (*SOU 2014:45)*. Accessed 3 Dec. 2021 from https://www.regeringen.se/contentassets/2e42baeaa76e4e918dd852f2a3e43fca/unik-kunskap-genom-registerforskning-sou-201445

Government Offices of Sweden. (2016). *Digitaliseringens transformerande kraft—vägval för framtiden [The transforming power of digitalization—choosing a path for the future] (SOU 2016:89)*. Accessed 15 Feb. from https://www.regeringen.se/contentassets/f7d07b214e2c459eb5757cea206e6701/sou-2016_89_webb.pdf

Government Offices of Sweden (2017) *För ett hållbart digitaliserat Sverige—En digitaliseringsstrategi [For a sustainable digitalized Sweden: A digitalization strategy]*. Accessed 8 Aug. 2023 from https://www.regeringen.se/49adea/contentassets/5429e024be6847fc907b786ab954228f/digitaliseringsstrategin_slutlig_170518-2.pdf

Government Offices of Sweden (2018). *Rätt att forska: Långsiktig reglering av forskningsdatabaser [Right to research: Long-term regulation of research databases]. (SOU 2018:36)*. Accessed 3 Dec. 2021 from https://www.regeringen.se/49c0cd/contentassets/2570b2abef8c48f084bbddb48ed2300b/sou-2018_36_webb.pdf

Government Offices of Sweden (2020) *Forskning, frihet, framtid—kunskap och innovation för Sverige [Research, freedom, future—knowledge and innovation for Sweden]*. Prop. 2020/21:60. Accessed 23 May 2023. https://www.regeringen.se/rattsliga-dokument/proposition/2020/12/forskning-frihet-framtid--kunskap-och-innovation-for-sverige/

Government Offices of Sweden (2021). *Regeringsuppdrag att föreslå ett strategiskt program för digital strukturomvandling*. Accessed 3 Dec. 2021 from https://www.regeringen.se/48e3a9/contentassets/d9408f8eee6342b2b4aba3c52072dfcb/uppdrag-att-foresla-utformning-av-ett-strategiskt-program-for-att-mota-och-leda-i-den-digitala-strukturomvandlingen.pdf

Government Offices of Sweden (2023). Uppdrag till forskningsfinansiärerna att inkomma med analyser som underlag till regeringens forskningspolitik. (Dnr U2023/01317) Accessed 15 Feb. 2024 from https://www.regeringen.se/contentassets/5eee3b9b4a32457ea9a5fae3cf2e0bbc/uppdrag-till-forskningsfinansiarerna-att-inkomma-med-analyser-till-regeringens-forsknings--och-innovationspolitik.pdf

Gray, A. (2002). *Research practice for cultural studies: Ethnographic methods and lived cultures*. London: Sage.

Hine, C. (2021). Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation. *Research Ethics Review*, 17(4), 464–479. https://doi.org/10.1177/17470161211022790

Holm, P., Jarrick A., Scott D. (2015). *Humanities World Report*. Basingstoke: Palgrave Macmillan.

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-driven methods to diversify knowledge of human psychology. *Trends in cognitive sciences*, 22(1), 1–5. https://doi.org/10.1016/j.tics.2017.10.002

Kokolakis, S. (2017), Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, (64), 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. Big *Data & Society*, 6(1), Article 2053951718821146. https://doi.org/10.1177/2053951718821146

Lucchi, N. (2023). ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. *European Journal of Risk Regulation*, 1–23. https://doi.org/10.1017/err.2023.59

Lusher, S. J., McGuire, R., van Schaik, R. C., Nicholson, C. D., & de Vlieg, J. (2014). Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*, *19*(7), 859–868. https://doi.org/10.1016/j.drudis.2013.12.004

Lyon, D. (2017). Surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication*. 11, 824–842. https://ijoc.org/index.php/ijoc/article/view/5527

Lyon, D. (2018). *The culture of surveillance: Watching as a way of life*. Cambridge: Polity Press.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461. https://doi.org/10.1007/s10708-014-9602-6

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41, 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

OECD. (2019). *Going Digital: Shaping Policies, Improving Lives*. Paris: OECD Publishing.

Pause Giant AI Experiments: An Open Letter. Accessed 4 April 2023 from https://futureoflife.org/open-letter/pause-giant-ai-experiments

Pappalardo, Grossi, V., & Pedreschi, D. (2021). Introduction to the special issue on social mining and big data ecosystem for open, responsible data science. *International journal of data science and analytics*, 11(4), 261–262. https://doi.org/10.1007/s41060-021-00253-5

Reutter, L. & Åm, H. (2024) Constructing the data economy: tracing expectations of value creation in policy documents, *Critical Policy Studies*. https://doi.org/10.1080/19460171.2023.2300436

Riessman, C.K. (2007). *Narrative Methods for the Human Sciences*. London: Sage.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). https://doi.org/10.1177/2053951718820549

Samuelson, P. (2023) Generative AI meets copyright. *Science* 381,158-161. https://www.science.org/doi/10.1126/science.adi0656

Samuelsson, L., Cocq, C., Gelfgren, S., Enbom, J. (2023). *Everyday Life in the Culture of Surveillance*. Nordicom, University of Gothenburg. https://doi.org/10.48335/9789188855732

Swedish Research Council (2018). *The Swedish Research Council´s guide to research infrastructure 2018*. Accessed 3 Dec. 2023 from https://www.vr.se/download/18.7f26360d16642e3af996ff/1555326273381/Vetenskapsraadets-guide-infrastrukturen-2018_VR_2018.pdf

Swedish Research Council (2020). *Samordning av öppen tillgång till forskningsdata: Statusrapport i Vetenskapsrådets uppdrag—summering av arbetet 2017–2019 och fortsatt arbete [Coordination of open access to research data: Status report on the Swedish Research Council's mission—summary of work 2017–2019 and continued work.]*. Accessed 3 Dec. 2023 from https://www.vr.se/download/18.47291b121711f04ce6e3bb/1585815456838/Samordning%20av%20oppen%20tillgang%20till%20forskningsdata_VR_2020.pdf

Swedish Research Council. (2022). *Vetenskapsrådets samordningsuppdrag om öppen tillgång till forskningsdata 2022 [The Swedish Research Council's coordination assignment on open access to research data].* Accessed 15 Feb. 2024 from https://www.vr.se/download/18.72c4495e17f44b64443b03a/1647009787100/Samordningsuppdrag%20om%20%C3%B6ppen%20tillg%C3%A5ng%20till%20forskningsdata%20VR%202022.pdf.

Swedish Research Council (2023). *The Swedish Research Council´s guide to research infrastructure 2023*. Accessed 15 Feb. 2024 from https://www.vr.se/download/18.14f2b85918c8ae70d9922c20/1704736135760/Guide%20to%20research%20infrastructure%202023.pdf

Swedish Research Council (2024). *Forskning och innovation för ett hållbart och säkert samhälle Gemensamt underlag till regeringens forsknings- och innovationspolitik från Energimyndigheten, Formas, Forte, Rymdstyrelsen, Vetenskapsrådet och Vinnova*. [Research and innovation for a sustainable and safe society. Common basis for the government's research and innovation policy] Accessed 15 feb 2024 from https://www.vr.se/download/18.1e171f1718bcc1830fea98/1699964201384/Forskning%20och%20innovation%20f%C3%B6r%20ett%20h%C3%A5llbart%20och%20s%C3%A4kert%20samh%C3%A4lle-2023.pdf

Sønderskov, K. M., & Dinesen. P. T. (2016). Trusting the state, trusting each other? The effect of institutional trust on social trust." *Political behavior* 38(1). 179–202. https://doi.org/10.1007/s11109-015-9322-8

Tupasela, A., Snell, K., & Tarkkala, H. (2020). The Nordic data imaginary. *Big Data & Society*, 7(1). https://doi.org/10.1177/2053951720907107

UNDP. (2019). *Future forward: UNDP digital strategy*. Accessed 10 Feb. 2021 from https://digitalstrategy.undp.org/documents/UNDP-digital-strategy-2019.pdf

Vey, K., Fandel-Meyer, T., Zipp, J. S., & Schneider, C. (2017). Learning & Development in Times of Digital Transformation: Facilitating a Culture of Change and Innovation. *International Journal of Advanced Corporate Learning (iJAC)*, *10*(1). 22–32. https://doi.org/10.3991/ijac.v10i1.6334

Vinnova, Vetenskapsrådet, DIGG & PTS. (2021). *Regeringsuppdrag att föreslå ett strategiskt program för digital strukturomvandling: Slutredovisning* [Government mandate to propose a strategic program for digital structural transformation: Final report] https://www.vinnova.se/contentassets/b6f628d9450642068ce283db0f16381d/rapport-ru-kraftsamling-for-digital-strukturomvandling.pdf

Zmerli, S., Newton, K., & Montero, J. R. (2007). Trust in people, confidence in political institutions, and satisfaction with democracy. In *Citizenship and involvement in European democracies: A comparative analysis*. Eds. J. W. van Deth, J. R. Montero, & A. Westholm. New York: Routledge.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.