The Right to Read Without Being Read: Research Ethics in the Study of Digital Reading Behaviour

Karolina Andersdotter

Abstract:

The impact of digitalization on privacy is obvious in many parts of society, including the collection of behavioural data of readers who access literature through digital means (e-books, audiobooks).[1] This data can provide a lot of personal information about the reader (e.g. reading speed, time spent reading, opinions and interests linked to one's book selection) and in combination with other collected data creates a very detailed picture of a person's lifestyle and movement patterns.

In this position paper, I discuss ethical implications related to the use of large commercial data sets consisting of sensitive personal data in humanities and social sciences research. The ethical implications are explored through the lens of two case studies on digital reading behaviour. By raising ethical questions related to the study of reading, user consent, and legal certainty in the fast-developing information society, I present issues that there is an urgent need for the academic community to discuss to make sure good research practices are in place even when using new types of data and digital methods.

I highlight current privacy and power asymmetry issues between the stakeholders in the research process, especially the users-turned-research subjects, and argue that the research community must assume a larger ethical responsibility when applying novel data-driven methods.

Keywords
Research ethics, privacy, reading studies, behavioural surplus

---

1   Throughout this paper I use digitalization to denote the digital transformation of and its impact on societies, services, and objects.

Introduction

Historically, books and reading have been regarded as crucial for people's ability to participate as active citizens in a society.[2] This role of books and reading is reiterated in Swedish library legislation that connects the development of a democratic society to the dissemination of knowledge.[3] The importance of private reading is emphasized as a historical precondition for rational-critical debate as well as ingrained in human rights.[4]

The book market and the reading habits of people are impacted by the rapid development of technology. The digitalization of the book medium—as text, hypertext, sound, or immersive experience—has led to changes in the reception of and interaction with literature. Legislation does not always manage to keep up with the rapid technological development; meanwhile the digital book market and the reading human meet at an intersection that raises new questions about roles and rights.

Digitalization offers great opportunities to develop commercial services by utilizing the behavioural data that is a by-product of using digital services. This data can be used, among other things, to tailor the consumption experience according to the reader's supposed preferences, as well as to make comprehensive analyses of the reader's consumption behaviour that previously required separate data collection. But digitalization also impacts human rights compliance, for instance freedom of expression and protection of privacy.[5] This has been noted in relation to the EU legislative processes on, for example, data protection, copyright, and AI regulation.[6]

From a privacy point of view the collection of behavioural data about readers provide a lot of personal information about the reader. In combination with other data this information may give a very detailed picture of a person's lifestyle and movement patterns, and this is detrimental to privacy. The challenge of balancing preservation of privacy with keeping access and use of information open and efficient has not gone unnoticed in library and information science research.[7]

---

2 Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (Polity Press, 1989), 36–37.

3 SFS 2013:801, *Bibliotekslag* [Swedish Library Act], http://rkrattsbaser.gov.se/sfst?bet=2013:801.

4 Habermas, *The Structural Transformation of the Public Sphere*, 158; UN General Assembly, "Universal Declaration of Human Rights," 217 (III) A (United Nations, 1948), accessed September 2, 2024, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

5 SFS 1994:1219, *Lag om den europeiska konventionen angående skydd för de mänskliga rättigheterna och de grundläggande friheterna* [Swedish statute incorporating the European Convention on Human Rights]. http://rkrattsbaser.gov.se/sfst?bet=1994:1219.

6 European Commission, "Data Protection in the EU," accessed 4 July, 2023, https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en; European Commission, "Questions and Answers – New EU Copyright Rules," updated 4 June, 2021, accessed 5 September, 2024, https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_2821; European Commission, "A European Approach to Artificial Intelligence | Shaping Europe's Digital Future," accessed 3 October, 2023, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

7 David Bawden and Lyn Robinson, "Essentially Made of Information: Concepts and Implications of Informational Privacy," *Information Research* 24, no. 4 (15 December, 2019), https://informationr.net/

When scholars use commercial reading data to study the behaviour of readers, an additional layer of complexity is added that brings tension to the balance of the right to reading in private, commercial interests, and the researcher serving the public interest. In this paper, I present and discuss two Swedish studies of digital reading behaviour based on commercial user data. While the points of reference are Swedish guidelines and research practices, the questions on studies of digital reading and accompanying ethical concerns have bearing also in an international context, as the selling and reading of books are international phenomena.

Problem statement

In this position paper, I aim to raise awareness about the ethical problems that arise when using large commercial data sets in humanities and social sciences research. The issue of behavioural data and privacy has been discussed by many from a general privacy and ethics perspective. The method of using commercial user data in studies conducted by university affiliated researchers is still a quite new occurrence, at least in fields traditionally associated with the humanities (including my own field, library and information science, that exists in the borderland of humanities and social sciences). There are quite a few studies using these kinds of data, for instance on food consumption based on loyalty card data, where "[i]nformed consent was not required […] and not possible to obtain as all data were anonymized."[8] I will come back to this argument on informed consent later in the paper and explain why I think it is problematic. Overall, it seems to me the general discussion about the ethical implications of using commercial data for academic research – especially in relation to sensitive personal data and privacy – is lacking in the scholarly community, though not non-existent.[9] Assuming the use of commercial user data is becoming increasingly common in academic research it would be wise to begin this discussion sooner rather than later.

By conducting a close reading of the data collections practices and ethical considerations of two Swedish research projects that use data from commercial book services I highlight ethical challenges to consider when forming responsible research practices in future data driven research projects in the humanities and social sciences.

These issues are likely to become more common as commercial (and non-commercial) data collection and accessibility increase. We need a discussion about what ethical risks arise when using commercial personal data in research. Who should be responsible for mitigating these risks? The researcher? The research support services? The data provi-

---

        ir/24-4/colis/colis1913.html.

8   Cf. Stephen D. Clark et al., "Dietary Patterns Derived from UK Supermarket Transaction Data
     with Nutrient and Socioeconomic Profiles," *Nutrients* 13, no. 5 (May, 2021), https://doi.org/10.3390/
     nu13051481; Victoria Jenneson et al., "Exploring the Geographic Variation in Fruit and Vegetable Pur-
     chasing Behaviour Using Supermarket Transaction Data," *Nutrients* 14, no. 1 (January, 2022), https://
     doi.org/10.3390/nu14010177.

9   Remy Stewart, "Big Data and Belmont: On the Ethics and Research Implications of Consumer-Based
     Datasets," *Big Data & Society* 8, no. 2 (1 July, 2021), https://doi.org/10.1177/20539517211048183.

der? The research examples are currently relatively limited, which builds momentum to create a sound research practice before an unintentionally careless one is consolidated.

I will start this position paper with presenting an overview of how books as a medium and the book market are changed by the technological development of society, how this change affects readers, what implications follow when large data sets containing personal data are being used as research material, and lastly offer a brief overview of Swedish research ethics guidelines on these topics. Then follows a presentation of the ethical reasonings in two Swedish case studies that I will use as a backdrop for further discussion about current and future challenges of privacy and research ethics.

The two case studies, Elisa Tattersall Wallin's dissertation *Sound Reading* and the article *Time to Read* (co-written with Jan Nolin) and Karl Berglund's articles *Introducing the Beststreamer, Is Backlist the New Frontlist?* (co-written with Ann Steiner), and book *Reading Audio Readers: Book Consumption in the Streaming Age* (partially built on reworked versions of the aforementioned texts) use data from Bookbeat and Storytel respectively. Both of these audiobook subscription services are available in multiple markets in Europe as well as the rest of the world.[10]

Background

Law, technology, society, and culture are increasingly being studied in interdisciplinary constellations. They are so affected by one another that an interdisciplinary approach is sometimes needed to shed light on certain issues, for instance from the perspective of my field, library and information science. In this section I will highlight the following areas to provide context for my critical discussion: the digital transformation of books and the book market, the study of reading, and commercial data collection and privacy.

The reading of audiobooks and e-books has increased explosively in Sweden. Both sales and library loans are on the rise and although it is difficult to say whether it is due to trend or availability, audiobooks and e-books now account for a significant part of the book market.[11]

---

10 Elisa Tattersall Wallin, *Sound Reading: Exploring and Conceptualising Audiobook Practices among Young Adults* (Högskolan i Borås, 2022), http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-27165; Elisa Tattersall Wallin and Jan Nolin, "Time to Read : Exploring the Timespaces of Subscription-Based Audiobooks," *New Media and Society* 22, no. 3 (2019), http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-21524; Karl Berglund, "Introducing the Beststreamer: Mapping Nuances in Digital Book Consumption at Scale," *Publishing Research Quarterly* 37, no. 2 (June, 2021): 135–51, https://doi.org/10.1007/s12109-021-09801-0; Karl Berglund and Ann Steiner, "Is Backlist the New Frontlist?: Large-Scale Data Analysis of Bestseller Book Consumption in Streaming Services," *Logos* 32, no. 1 (25 May, 2021): 7–24, https://doi.org/10.1163/18784712-03104006; Karl Berglund, *Reading Audio Readers: Book Consumption in the Streaming Age* (Bloomsbury, 2024).

11 Svenska bokhandlareföreningen and Svenska förläggareföreningen, "Bokförsäljningsstatistiken 2022 ... och utvecklingen de senaste fem åren (2018-2022)," (Svenska bokhandlareföreningen and Svenska förläggareföreningen, 2023), 50–51, accessed September 2, 2024, https://forlaggare.se/wp-content/uploads/2023/02/Bokforsaljningsstatistiken-2022.pdf; Kungliga biblioteket, "Bibliotek 2022 : Sveriges offentligt finansierade bibliotek," (Kungliga biblioteket, 2023), accessed September 2, 2024, https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-707.

The reading of audiobooks and e-books enables new types of research as it is possible to collect more data about the reading process than by observing a reader with a physical book. Since the law cannot keep up with technological developments, previously clear legal boundaries between commercial actors, researchers, and research subjects are being challenged. In the humanities, data driven methods and automatic processing of large data sets with potentially sensitive personal data have traditionally been uncommon and research ethical practices are not as thoroughly discussed and established as compared to, for instance, medical research.[12]

Digitalization has a large impact on books as well as on other media. The nature of digital objects generally differs from physical objects by, among other things, being editable, interactive, possible to access and to modify by means of other digital objects, and distributed.[13] These inherent properties can be technically and legally limited in various ways (for example licensing agreements and encryption) in order for the digital objects to be sellable in the market, for example through copyright restrictions for sharing and copying. In general, however, digitalization has increased accessibility to digital media, for instance due to digital distribution channels that enable immediate access to a book in audio or text format. Digitalization also enables literature distribution platforms to customize the experience for the readers through data collection of search history, book collections, time spent reading or listening, social contacts, and so forth.[14] Similar to Web 2.0 (the participative or social web), there is now a reading 2.0 based on user interaction.

The digital distribution channels responsible for this change (like many other companies operating in the digital sphere) have interests that go beyond serving the needs of their reading customers. There is a large market for the kind of user data generated by digital reading.[15] Digitalization opens completely new opportunities to study the habits, patterns, and preferences of readers and this information can be used to develop the book market.

---

12 Vetenskapsrådet, *Good Research Practice* (Vetenskapsrådet, 2017), 39; E. M. Jones et al., "DataSHIELD – Shared Individual-Level Analysis without Sharing the Data: A Biostatistical Perspective," *Norsk Epidemiologi* 21, no. 2 (13 April, 2012), https://doi.org/10.5324/nje.v21i2.1499; Klaus Hoeyer, "The Ethics of Research Biobanking: A Critical Review of the Literature," *Biotechnology and Genetic Engineering Reviews* 25, no. 1 (1 January, 2008): 429–52, https://doi.org/10.5661/bger-25-429; "Etikprövning - så går det till," Etikprövningsmyndigheten, accessed 4 September, 2024, https://etikprovningsmyndigheten.se/for-forskare/sa-gar-det-till/.

13 Distributed in this context means that the object is seldom contained within a single source or institution. For instance, an electronic book in two different libraries is the same object, whereas a physical book would have two exist as two objects to exist in the two libraries. Similarly, a digital object may be copied without loss of quality, making the original file and the copied file indistinguishable (in essence, the digital object is not the file but the file's content), whereas a copy of a physical book would result in a new object. Jannis Kallinikos, Aleksi Aaltonen, and Attila Marton, "The Ambivalent Ontology of Digital Artifacts," *MIS Quarterly* 37, no. 2 (2013): 357–70.

14 Storytel, "Integritetspolicy för Storytel," last updated 12 September, 2022, accessed 5 September, 2024, https://web.archive.org/web/20240905122618/https://legal.storytel.com/se/sv/documents/privacy-policy?request_locale=se.

15 Jens-Erik Mai, "Big Data Privacy: The Datafication of Personal Information," *The Information Society* 32, no. 3 (26 May, 2016): 20, https://doi.org/10.1080/01972243.2016.1153010.

While the data collection model is common in commercial e-book and audiobook services, it is worth noting that it is technically feasible to develop e-book services that do not collect data about the reader; this applies to, for example, Sweden's national library's e-book application *Bläddra*.[16] The application is an example of privacy by design, an approach to system development that aims to consider personal privacy throughout the development process.[17] This approach is also partially reflected in the EU's General Data Protection Regulation's (GDPR) *Article 25: Data protection by design and by default*.[18]

Technological development creates and decides on new conditions for the book market. New needs and visions arise from both users and digital distribution channels, and these are met in whole or in part through innovation in the book market. Both copyright and other information-related legislation is affected by the digitalized book medium, and legislators are challenged to keep up.

The study of reading

It is relevant to study how information in general and literature in particular is made available in new ways, to see how older formats are renewed, and understand how this affects the social practices of readers.

Historian Robert Darnton wrote that "reading remains the most difficult stage to study in the circuit followed by books" and publishing researcher Claire Squires confirmed this statement and underlined that "patterns of consumption do not strictly mirror patterns of reading".[19] With this in mind, one can assume that there is a great interest in exploring reading behaviour, both from a research and a market point of view. Digital reading provides an opportunity to undertake such studies on an unprecedented level.

Squires brought up Amazon's purchase of the website Goodreads as an example of how valuable this type of user data can be and explained that the problem for both researchers and publishers is getting access to this data. Squires wrote that

> it is crucially important to continue the discussion both of the methods
> for collecting and analyzing data and considering in which hands – or
> with the digital environment, whose servers – such information is held.

---

16   Kungliga biblioteket, "Bläddra," last updated 3 May, 2024, accessed 5 October, 2023, https://web.archive.org/web/20231005153806/https://www.kb.se/samverkan-och-utveckling/biblioteksutveckling/nationella-minoriteters-bibliotek/bladdra.html.

17   Ronald Hes and John J. Borking, *Privacy-Enhancing Technologies: The Path to Anonymity*, Achtergrond-studies en Verkenningen (Registratiekamer, 1998), 7.

18   Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance), 119 OJ L § (2016), accessed 4 September, 2024, http://data.europa.eu/eli/reg/2016/679/oj/eng.

19   Robert Darnton, "What is the History of Books?," *Daedalus* 111, no. 3 (1982): 74; Claire Squires, "The Global Market 1970– 2015: Consumers," in *Companion to the History of the Book*, ed. Jonathan Rose and Simon Eliot (John Wiley & Sons, 2019), 610–11. In this context, unlike the rest of this paper, Squires' use of *consumption* refers to the act of buying books rather than taking part of the content of a book.

> This is a question of market information and the scholarly desire for knowledge, but also an ideological one about the ownership of information.[20]

The data collection carried out by audiobook and e-book companies has been a cause for concern among researchers. Lawyer and information scientist Elizabeth Henslee and lawyer Meredith Mays Espino both noted that service providers who closely monitor the reading habits of their users can find out a lot of information about them that is not strictly related to reading. Knowing when and how a person reads, what bookmarks and margin notes they put in the text, and what settings they make in an application can provide information about things such as work schedules, sleep patterns, and vision problems.[21] Henslee and Mays Espino are not alone in their concern; for example a specific law was established in California in 2012 to protect readers' personal privacy in the digital environment, signifying an existing concern outside academia.[22]

Sensitive personal data consists of, but is not limited to, information about political opinions, religious or philosophical beliefs, information about health, or information about a person's sex life or sexual orientation. A person's reading preferences could provide clues or lead to conclusions about this type of personal data, which is a reason why public libraries generally do not keep records of library users' previous loans – in Sweden, the library confidentiality is even included in the legislation.[23] It is also a reason for commercial services to keep this data; the behavioural data can be used by itself and in combination with data from other sources to make analyses and conclusions. This data can enhance the services, providing a better user experience, but it can also be used beyond this purpose.[24]

The academic discussion about reading in different formats and ways, and how to study these phenomena has engaged scholars from various fields.[25] Digital reading behaviour

---

20 Squires, "The Global Market," 611.

21 Elizabeth Henslee, "Down the Rabbit Hole: E-Books and User Privacy in the 21st Century," *Creighton Law Review* 49, no. 1 (December, 2015), http://hdl.handle.net/10504/84523; Meredith Mays Espino, "Sometimes I Feel Like Somebody's Watching Me . . . Read?: A Comment On The Need For Heightened Privacy Rights For Consumers Of Ebooks," *UIC John Marshall Journal of Information Technology & Privacy Law* 30, no. 2 (1 January, 2013), https://repository.law.uic.edu/jitpl/vol30/iss2/3.

22 Reader Privacy Act, Pub. L. No. 1.81.15., 1798.90 California Civil Code § 1798.90 (2012).

23 SFS 2009:400, *Offentlighets- och sekretesslag* [Swedish Publicity and Privacy Act], 40 kap. §3. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400_sfs-2009-400 (accessed 4 September, 2024).

24 Ella Horttanainen, "From Customer Data to Customer's Data : Reverse Use of Customer Data as a Tool for Value Co-Creation" (Jyväskylän yliopisto, 2021), https://jyx.jyu.fi/handle/123456789/78665; Ann Steiner and Torbjörn Forslid, "Bokhandeln mellan kultur och ekonomi," in *Litterära värdepraktiker*, ed. Torbjörn Forslid et al. (Makadam, 2017), 67–125.

25 Tully Barnett, "Distributed Reading: Literary Reading in Diverse Environments," *Digital Humanities Quarterly* 12, no. 2 (2018), http://www.digitalhumanities.org/dhq/vol/12/2/000389/000389.html (accessed 1 October, 2024); Iben Have and Birgitte Stougaard Pedersen, "Reading Audiobooks," in *Beyond Media Borders, Volume 1*, ed. Lars Elleström (Springer International Publishing, 2021), 197–216, https://doi.org/10.1007/978-3-030-49679-1_6; Anna Lundh, Åse Hedemark, and Linnéa Lindsköld, "Critical Studies of Reading: Consolidating an Emerging Field of Research," *Information Research* 27 (25 October, 2022), https://doi.org/10.47989/colis2232; Anne Mangen, "The Digitization of Literary Reading," *Orbis Litterarum* 71, no. 3 (2016): 240–62, https://doi.org/10.1111/oli.12095.

has often been studied but mainly with qualitative methods.[26] Such methods were also called for when Kuzmičová et al. highlighted the need for further studies of fiction reading from mobile phones.[27]

Wanting to read, consenting to be read

As mentioned, reading habits can reveal a lot about a person. The International Federation of Library Associations and Institutions expressed caution against this development already in 2013:

> Expanding data sets held by governments and companies will support the advanced profiling of individuals, while sophisticated methods of monitoring and filtering communications data will make tracking those individuals cheaper and easier. Serious consequences for individual privacy and trust in the online world could be experienced.[28]

The increased collection and use of personal data in the digital market is described by social psychologist and philosopher Shoshana Zuboff as surveillance capitalism; they explained that

> although some of these data are applied to product or service improvement, the rest are declared as a proprietary *behavioral surplus*, fed into advanced manufacturing processes known as "machine intelligence", and fabricated into *prediction products* that anticipate what you will do now, soon, and later.[29]

Zuboff elaborated further to say that this behavioural surplus becomes a product in itself, meaning that a user of a digital service may become a product as well. Similarly, library and information science scholars Diana Floegel and Philip Doty warned that this may also be used to shape people's behaviour to achieve commercial, political, or security goals, and that, as a systematic phenomenon, this could threaten democratic processes and individual autonomy.[30]

---

26 Esmeralda V. Bon and Michael Burke, "Devices, Settings and Distractions: A Study into How Students Read Literature," in *Pedagogical Stylistics in the 21st Century*, ed. Sonia Zyngier and Greg Watson (Springer International Publishing, 2022), 183–206, https://doi.org/10.1007/978-3-030-83609-2_8; Have and Pedersen, "Reading Audiobooks"; Lotta C. Larson, "E-Books and Audiobooks," *The Reading Teacher* 69, no. 2 (2015): 169–77, https://doi.org/10.1002/trtr.1371.

27 Anežka Kuzmičová, Theresa Schilhab, and Michael Burke, "M-Reading: Fiction Reading from Mobile Phones," *Convergence* 26, no. 2 (1 April, 2020): 333–49, https://doi.org/10.1177/1354856518770987.

28 IFLA, *Riding the Waves or Caught in the Tide? Navigating the Evolving Information Environment* (IFLA, 2013), 4, https://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report_v3.pdf.

29 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power* (Profile Books, 2019), 8. Zuboff's italics.

30 Diana Floegel and Philip Doty, "The Library/Surveillance Interface," *Proceedings of the Association for Information Science and Technology* 58, no. 1 (2021): 152–61, https://doi.org/10.1002/pra2.444.

Personal data collection in digital services is not unusual and the user has most likely consented to it by agreeing to the service's terms of services (ToS). Agreeing to the ToS is often a prerequisite for being able to use the service. These agreements can be long and complicated. For instance, a Swedish Storytel user needs to consult Storytel's ToS (~23,600 characters), which refers to Storytel's privacy policy (~25,500 characters), which refers to Storytel's cookie declaration (~13,000 characters) to find out what data is collected and what it is used for.[31] For a customer whose main objective is to read a book, that is a lot of information to comprehend and make a decision on.

Law scholar Nancy S. Kim argued that there is a problem with user agreements today because

> businesses, courts and technology create a coercive contracting env-
> ironment where one-sided legal terms are imposed upon non-drafting
> parties who literally have no choice but to accept them if they wish to
> participate in modern society.[32]

Kim noted that consumers often accept digital and printed contracts without reading them, and with the digital format of contracts, the physical limitation of too many sheets of paper is missing, which has led businesses to "use contracts to extract from consumers additional benefits that were unrelated to the transaction." This creates a kind of information asymmetry, a power imbalance between users and service providers, and while users are concerned about this, they "undertake little to protect their data."[33]

Two case studies of data-driven audiobook research

To contextualize the ethical dilemmas when using large commercial data sets with personal data I present and discuss two Swedish research projects as case studies. The studies by Tattersall Wallin and Berglund build on user data from the audiobook subscription services Bookbeat and Storytel.[34] They highlight how commercial data enable new types of research but at the same time challenge the ethical reasoning of the researchers.

---

31 Storytel, "Storytel Användarvillkor," last updated 13 September, 2023, accessed 5 September, 2024, https://web.archive.org/web/20240905122304/https://www.storytel.com/se/sv/documents/terms-and-conditions; Storytel, "Integritetspolicy för Storytel"; Storytel, "Om cookies – Storytel," accessed 5 September, 2024, https://web.archive.org/web/20240905123249/https://www.storytel.com/se/cookies.

32 Nancy S. Kim, *Wrap Contracts Foundations and Ramifications* (Oxford University Press, 2014), 4, https://doi.org/10.1093/acprof:oso/9780199336975.001.0001.

33 Kim, 51. See also Daniel Kerpen, Matthias Dorgeist, and Sascha Zantis, "Intersecting the Digital Maze. Considering Ethics in Cloud-Based Services' Research," in *Research Ethics in the Digital Age: Ethics for the Social Sciences and Humanities in Times of Mediatization and Digitization*, ed. Farina Madita Dobrick, Jana Fischer, and Lutz M. Hagen (Springer Fachmedien, 2018), 143–52, https://doi.org/10.1007/978-3-658-12909-5_15; Stewart, "Big Data and Belmont"; Susanne Barth and Menno D. T. de Jong, "The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review," *Telematics and Informatics* 34, no. 7 (2017): 1038–58, https://doi.org/10.1016/j.tele.2017.04.013;

34 Tattersall Wallin, *Sound Reading: Exploring and Conceptualising Audiobook Practices among Young Adults*; Tattersall Wallin and Nolin, "Time to Read"; Berglund, "Introducing the Beststreamer"; Karl Berglund and Ann Steiner, "Is Backlist the New Frontlist?"; Berglund, *Reading Audio Readers*.

It is noteworthy that both researchers have a background in library and information science, a field that concerns itself with information behaviour and information ethics and is situated within humanities and social sciences. I chose these studies as examples because they are unique in their scope, they are closely related to my own field, and they seem to me to have been well received by the library and information science community (without the ethical scrutiny that I would normally expect from this academic field). My argument does not regard only these case studies, but uses them as examples to raise more general questions about research ethics that are relevant for any researcher who considers working with large commercial data sets.

Research ethics and large data sets

Research ethics are ethical questions about the research topic, the methods, and the researcher's relationship to the research task, especially when the research involves people. Because of the scope of these case studies, issues relating to the collecting, handling, and sharing of information (including personal data) are included in the research ethics definition.

The availability of data sets that are large enough to be difficult to process manually has redefined humanities and social sciences research. Legal philosopher Helga María Lell argued that the challenge for these disciplines is to preserve the human being as an ontological value, and not simply see them as a statistical data representation or a quantified self.[35] Social scientists Jacob Metcalf and Kate Crawford discussed the ethical challenges in defining the boundaries of a human-subject in research; if a researcher acknowledges the person in a data set, they have to make a different ethical assessment than if they do not.[36] From the ethical point of view it is important to note that a quantified self, consisting of anonymized data points, can be re-identified as a person, meaning privacy remains a pivotal value even in large and seemingly anonymous data sets.[37]

According to the Swedish Research Council[38] "all research dealing with sensitive personal data shall be ethically reviewed, regardless of how the data has been collected and whether or not the researcher has obtained the participants' consent." The review board should "evaluate how the human rights and basic freedoms of those involved are treated in relation to the value of the research."[39]

---

35 Helga María Lell, "Human Rights and the Regulation of Anonymity. New Challenges to Law and Research," in *Research Ethics in the Digital Age: Ethics for the Social Sciences and Humanities in Times of Mediatization and Digitization*, ed. Farina Madita Dobrick, Jana Fischer, and Lutz M. Hagen (Springer Fachmedien, 2018), 119–27, https://doi.org/10.1007/978-3-658-12909-5_12.

36 Jacob Metcalf and Kate Crawford, "Where are Human Subjects in Big Data Research? The Emerging Ethics Divide," *Big Data & Society* 3, no. 1 (1 June, 2016), https://doi.org/10.1177/2053951716650211.

37 Luc Rocher, Julien M. Hendrickx, and Yves-Alexandre de Montjoye, "Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models," *Nature Communications* 10, no. 1 (23 July, 2019): 3069, https://doi.org/10.1038/s41467-019-10933-3.

38 The Swedish Research Council was also the funder of Berglund's four- year research project (2019-02829, 'Patterns of Popularity: Towards a Holistic Understanding of Contemporary Bestselling Fiction') of which Berglund's case study articles are an output.

39 Vetenskapsrådet, *Good Research Practice*, 30–31.

For research involving the handling of sensitive personal data, informed consent is normally required, "an exception is allowed, however: it is not necessary to inform research subjects if it is impossible, or if it would mean an unreasonably great work effort."[40] While pointing out general guidelines, funding agencies often leave it to researchers and research support services to figure out the details of these difficult decisions, causing a tension between a research subject's privacy and the advancement of research.[41]

The origin of the research data

To understand the ethical choices made by the researchers, I will explain the origins of the commercial user data turned research data of the two cases. The data comes from Storytel, a public company, and Bookbeat, part of the private company Bonnier Group.[42] Similar to other streaming services, their users agree to ToS, privacy policies, and related documents when signing up for the service. Storytel's privacy policy states that personal data may be used for research and that the legal basis for using this data is consent or contractual obligation.[43] An older version – the one in effect when the study using Storytel data was made – of the privacy policy does not mention "research", but it includes a paragraph about market research that is missing in the current version:

> Your personal data may also be processed in marketing and customer analyses, market research, statistics, business monitoring as well as business and method development by Storytel in order to develop and adapt the Service and its functions[44]

Similarly, Bookbeat's privacy policy does not mention research, but it states that personal data can be collected for market research purposes, for instance, by analysing e-mail interactions ("if the e-mail has been opened, if links in the e-mail have been clicked on, and time and place/city when the e-mail was opened by the recipient"). The policy clearly states that "if you do not wish to receive […] book recommendations [based on user data collection and analysis], BookBeat's service is less suitable for you."[45]

---

40 Vetenskapsrådet, *Good Research Practice*, 30–31.
41 Live Håndlykken Kvale and Peter Darch, "Privacy Protection throughout the Research Data Life Cycle," *Information Research* 27, no. 3 (15 September, 2022), https://doi.org/10.47989/irpaper938.
42 Storytel, "Om oss," accessed 11 September, 2024, https://www.storytelgroup.com/sv/om-oss; Book-Beat, "Ljudböcker och e-böcker. Prova gratis! – BookBeat," accessed 25 June, 2024, https://web.archive.org/web/20240625230537/https://www.bookbeat.com/se/about.
43 Storytel, "Integritetspolicy för Storytel."
44 "Dina personuppgifter kan även komma att behandlas i marknads- och kundanalyser, marknadsundersökningar, statistik, affärsövervakning samt affärs- och metodutveckling av Storytel i syfte att utveckla och anpassa Tjänsten och dess funktioner" (author's translation to English) Storytel, "INTEGRITETSPOLICY FÖR STORYTEL," updated 12 June, 2019, accessed 21 July, 2022. http://web.archive.org/web/20220721162819/https://www.storytel.com/se/sv/documents/privacy-policy.
45 "om e-postmeddelandet har öppnats, om länkar i e-posten klickats på, och tidpunkt och plats/stad då e-posten öppnandes av mottagaren"; "Önskar du inte få sådana rekommendationer på böcker […] är BookBeats tjänst mindre väl lämpad för dig." (author's translation to English) BookBeat, "Book-Beat Integritetspolicy," last updated 11 June, 2021, accessed 5 September, 2024, https://web.archive.org/web/20240905141858/https://assets.ctfassets.net/4s7izhcdroyy/1eS2r2ZPSnZea7JmxvMZye/06851fe11603669bb6fe09debad30741/privacyPolicySE_20210611.pdf.

Examples of data collected by Bookbeat are reading and listening history, title selection, saved books, rating of books, bookmarks, favourite authors/series, button presses, search history, and other choices in the app and on the website. The data can be shared with other companies within Bonnier Group and with companies that perform services for Bookbeat (for example Microsoft Azure, Google, Facebook, and Instagram).[46]

"Sound Reading" and "Time to Read"

The first case consists of Tattersall Wallin's article and dissertation (part I, the summary essay). The article explored the times (of day) and spaces (mobile, stationary) of audiobook readers in a specific target group and was included as an article in the dissertation. The ethical issues relating to the research data were raised in both the article and in the dissertation. The study investigated temporal patterns in transaction logs from Bookbeat.[47]

The researcher's collaboration with Bookbeat consisted of the company searching its logs in May 2018 based on four questions the researcher had provided and then sharing the search results with the researcher. The researcher described uncertainties with the data set, such as the total number of subscribers not being included in the data set (Bookbeat did not want to share their exact number of subscribers at the time of data collection due to competitive reasons). The researcher noted that "this is a possible issue when collaborating with industry, as they need to consider their business interests first."[48]

Tattersall Wallin wrote that while it is usually the researcher who is in control of the data collection process, "with digital methods this kind of control is not always possible" and noted that strategies are needed to navigate a collaboration. The researcher's strategy in this study consisted of creating trust between herself and the company, communicating the seriousness of the data collection so that "those who provided the data did a professional and precise job" and "ethical ground rules were established, including in relation to GDPR, so that both sides could feel comfortable, and that the privacy of users was respected."[49] An important circumstance of the data collection was that the university's lawyers had advised the researcher that Bookbeat should conduct the searches and only share anonymous, statistical data instead of raw data, since "an ethical concern common to transaction logs is that they may detail personal information, such as IP address, detailed search history or geographical information."[50]

It is important to remark that Tattersall Wallin made the ethical considerations that can be expected of a researcher. The researcher reasoned that "it would be highly difficult, if not impossible, to gather consent from all the tens of thousands of people who used the *BookBeat* service to listen to audiobooks during the time of the study"; that the study does

---

46 BookBeat, "BookBeat Integritetspolicy."
47 Tattersall Wallin, *Sound Reading: Exploring and Conceptualising Audiobook Practices among Young Adults*; Tattersall Wallin and Nolin, "Time to Read."
48 Tattersall Wallin, *Sound Reading: Exploring and Conceptualising Audiobook Practices among Young Adults*, 38–39.
49 Tattersall Wallin, *Sound Reading*, 39.
50 Tattersall Wallin, *Sound Reading*, 46.

not handle personal information but only user information on a group level; and that at the time of data collection, it was stated in the Bookbeat ToS that aside from being used for development, personal data could be shared with collaborators.[51] Tattersall Wallin's research is conducted within the limits of what the Swedish Research Council considers ethically correct; their guidelines state that consent is not required from people if it means that obtaining consent hinders the study.

In the article that the dissertation refers to Tattersall Wallin and Nolin wrote that they collected a "unique and very large dataset" from Bookbeat. They acknowledged weaknesses in the data collection process, such as the lack of control over the data set and limited transparency from the company, but stated that this was compensated by the scope of the data set, which would otherwise have been impossible to get hold of. They estimated that the number of users in the data set were 80,000, a large number compared to other studies of media habits where the number of respondents usually amount to a few thousand.[52]

In summary, Tattersall Wallin deemed it impossible to obtain consent and found that this was not a serious problem because firstly, the research data did not involve personal data, and secondly, the users had consented to personal data sharing through the ToS. This somewhat contradictory argumentation suggests that the data management is a quite complex issue to navigate and solve.

"Introducing the Beststreamer", "Is Backlist the new Frontlist?", and "Reading Audio Readers"

Berglund's research on patterns in the reading of bestsellers is the second case, here represented by an article, a book chapter, and a book which partially consists of reworked versions of the two aforementioned texts.[53] The research was based on a large data set from Storytel and the aim was to find out how book consumption differs in terms of streamed bestsellers and printed bestsellers as well as to "show the usefulness and considerable possibilities with computational approaches for digital publishing studies and contemporary book history" by using the platform's own user data.[54]

In the article, Berglund and Steiner stated that "a digital book trade needs digital methods if it is to be studied effectively" and argued that with the help of large data sets, scholars may answer questions about genre, interrupted reading, completion rate, and at what time books are consumed. Their results show that there are big differences between print and digital best sellers.[55]

---

51   Tattersall Wallin, *Sound Reading,* 46–47.
52   Tattersall Wallin and Nolin, "Time to Read," 473.
53   Berglund, "Introducing the Beststreamer"; Karl Berglund, "Strömmade bästsäljare : Litteraturkonsumtion i digitala prenumerationstjänster utifrån Storytels användardata," in *Från Strindberg till Storytel: Korskopplingar mellan ljud och litteratur,* ed. Julia Pennlert and Lars Ilshammar (Daidalos, 2021), 327–62, Berglund, *Reading Audio Readers.*
54   Berglund, "Introducing the Beststreamer."
55   Berglund and Steiner, "Is Backlist the New Frontlist?"

Berglund wrote that a majority of the research within publishing studies is qualitative and that a criticism of quantitative perspectives exists. The criticism questions what results large data sets can yield; it argues that large data sets can contribute to a measurement culture, and that it is unclear what the measurements in question actually say about reading behaviour. Berglund's critical discussion on methodological choices focused on these aspects, even though risks surrounding big data, privacy, and surveillance mechanisms were addressed in several of the cited references in the article.[56] In the more recent publication he expanded on the issue of unavailable raw data from an open science perspective, something that has consequences for reproducibility of research. He rightly stated that this is a problem also in many qualitative studies, where interview transcripts may be unavailable for scrutiny.[57] He also reflected on the data consisting of information patterns from individual readers and noted that he had mitigated the risk of privacy intrusion by not showcasing individual patterns as examples, instead opting for user groups and proxy readers to guarantee that "no individual reader can identify themselves."[58]

Regarding the research data (consisting of ~10 million data points) Berglund explained that there was "a cooperation agreement between Storytel AB and Uppsala University, that clarifies that access to consumption data is given on the condition that it is not disseminated further, and that Storytel has no views on what research is being conducted."[59] A critical reflection on the data can be found in Berglund and Steiner's article where they point out that the data set "only shows patterns from one actor in one national book trade, and that it is derived from a commercial company aggregating data from their users."[60] Berglund argued that "the feedback loop from readers back to the book streaming services is decisive, and something that is likely to have substantial consequences on publishing in years to come," highlighting the increased power of the book streaming services in the publishing ecology.[61]

In the article, Berglund emphasized the need to study digital content with digital methods and further stated that "the best way to approach streaming platforms is by investigating them on their own terms, in their own habitat." Even though publishing scholars "probably never will be able to get a hold of the algorithms that steer consumption behavior in digital platforms," he deemed his approach – studying the outcome of the algorithms in user interactions and consumption patterns – the second-best alternative.[62]

---

56 Cf. Simone Murray, *The Digital Literary Sphere: Reading, Writing, and Selling Books in the Internet Era* (JHU Press, 2018); Simon Rowberry, "The Limits of Big Data for Analyzing Reading," *Participations* 16, no. 1 (May, 2019), 237–57; Daniel Allington, Sarah Brouillette, and David Golumbia, "Neoliberal Tools (and Archives): A Political History of Digital Humanities," *Los Angeles Review of Books*, 1 May, 2016, https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/.

57 Berglund, *Reading Audio Readers*, 25-26.

58 Berglund, *Reading Audio Readers*, 26.

59 Berglund, "Introducing the Beststreamer"; "[E]tt samarbetsavtal mellan Storytel AB och Uppsala universitet, som klargör att tillgång till konsumtionsdata ges under förutsättning att den inte sprids vidare, och att Storytel inte har några synpunkter på vilken forskning som bedrivs" (author's translation to English) Berglund, "Strömmade bästsäljare."

60 Berglund and Steiner, "Is Backlist the New Frontlist?," 11.

61 Berglund, *Reading Audio Readers*, 37-38.

62 Berglund, "Introducing the Beststreamer."

Discussion

Considering the two cases, on the one hand, we might deem the researchers' due diligence sufficient. Consulting university lawyers has been part of their research process, and they follow existing agreements and guidelines to the best of their abilities. The researchers were aware of the novelty of their methods, and we might find that this dimension needs to be considered in their ethical reasonings. In both cases "digital methods" are presented as a panacea to the lack of control over research data. If using digital methods to study digital reading provides an opportunity to study reading on an unprecedented level, we might ask whether the lack of precedent means that extra care must be taken when developing methods to study the phenomenon (such as the private reading aspect), or whether the novelty of digital methods is important enough to deviate from a thorough ethical investigation that may not lead to a clear conclusion?

Looking at my first question on ethical risks, I find two main things to consider. First, there is the risk of sticking to existing criteria for good research practice rather than developing new criteria that are consistent with the problems raised by large data sets. This risk mainly concerns the collection of informed consent from the humans being studied.[63] Second, there is the risk related to the partially undefined rules and norms in the general society of which the researcher is part of. The researcher's ethical choices may influence how these rules and norms are defined, ultimately having an impact on how we as a society view the right to privacy. In this discussion, I focus on three main issues that encompass both risks: informed consent, research ethics as a matter beyond the law, and the privacy of the reader vs the study of reading.

Issues of consent

The first risk pertains to the researcher's evaluation of the research subject's privacy in relation to the research process. What are the researcher's responsibilities to the source of their research data and where do the responsibilities end?

The consent of the Bookbeat and Storytel research subjects in these studies was not ascertained by the researchers and not required by the funding body or university. The users are referred to as their quantified selves – data points – becoming what Lell warned about and highlighting the issue of defining boundaries of the human subject that Metcalf and Crawford wrote about.[64] Just as in the food consumption studies mentioned in the introduction, the researchers' reasoning largely seem to be that informed consent is not required from the users who serve as research subjects, based on their representation as anonymized data points.[65]

---

63 Vetenskapsrådet, *Good Research Practice*.

64 Lell, "Human Rights and the Regulation of Anonymity. New Challenges to Law and Research"; Metcalf and Crawford, "Where are Human Subjects in Big Data Research? The Emerging Ethics Divide."

65 Jenneson et al., "Exploring the Geographic Variation in Fruit and Vegetable Purchasing Behaviour Using Supermarket Transaction Data"; Clark et al., "Dietary Patterns Derived from UK Supermarket Transaction Data with Nutrient and Socioeconomic Profiles."

It should be noted that Tattersall Wallin considered the issue, but mentioned the impossibility of collecting consent from all users. In many cases, however, service providers regularly send out notices about ToS updates that users must accept to continue to use the service. If researcher and service provider are collaborators, perhaps it might be possible for the service provider to collaborate by sending out a request for consent to participate in an academic study through their application?

The consent that *does* exist through accepted ToS should be a compelling argument to ascertain extra consent from the users, considering the power asymmetry between service provider and user.[66] Such a starting point should prompt extra care from a researcher when they consider the origin of their research data. We also need to reconsider the concept of anonymized data as a guarantee for privacy when we know that de-anonymization is possible through technology and therefore a risk for the research subjects – whether they have consented to participate or not.[67]

Research ethics – a matter of conscience, not law?

In both cases, reassurances from university lawyers were part of the research process. Berglund seemed to consider his research project one step removed from the research subjects; the users have a relation to Storytel and Storytel has a relation to Uppsala University. Leaving the decision about whether the informed consent of users is needed to university lawyers could be seen as an example of what Kvale and Darch mentioned about unclear rules leaving decisions to research support functions.[68] Considering that this type of commercial research data may increase in future research projects there is a need to review this from a more nuanced perspective (perhaps especially from a user perspective) considering the stakes for privacy and the right to consent.

As previously stated, the lack of progress in legislative processes means some aspects of digitalization are ungoverned until they are not, and assumedly this is what university lawyers base their decisions on. It is therefore important to distinguish between legal and illegal practice, scholarly correct or incorrect practice, and ethically correct or incorrect practice. Sometimes the three are the same and sometimes not. When the legislative process is slow, something that might be a correct legal practice might be an ethically incorrect practice. The penalties for being incorrect is very different from one category to another which means that it is very important to be clear about the categories, and to reason about which one of them should be carrying weight in the research process.

In relation to the weighting and penalties for the different practices, we must discuss whether the researcher's ethical responsibilities are bigger if there is no specific law or pre-

---

66 Kerpen, Dorgeist, and Zantis, "Intersecting the Digital Maze. Considering Ethics in Cloud-Based Services' Research"; Stewart, "Big Data and Belmont"; Barth and de Jong, "The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review."

67 Rocher, Hendrickx, and de Montjoye, "Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models."

68 Kvale and Darch, "Privacy Protection throughout the Research Data Life Cycle."

cedent, or if they are free to experiment while waiting for a precedent. What is the course of action when research ethics conflict with the law, or if the law changes so that previous research ethical assessments are at odds with the new law? While I generally applaud the "don't ask for permission, beg for forgiveness later" paradigm, I remain unconvinced it should apply in research ethical assessments. I would argue that the researcher's responsibilities should include reflection on existing law and whether there might be reason to be more precise or careful than the minimum of what the law requires.

When making the ethical assessment for similar future research projects, one solution that could help give representation to the users is to consult the national data protection authority (GDPR obliges every EEA country to have one). Unlike university lawyers, whose main purpose should be to support the university and its staff, a data protection authority is specialized in data protection concerns. Maybe there is a need for a more fluid relation between scholarly and ethical practices to make this feasible; case by case examples may be difficult to immediately codify in research ethics guidelines, but they still require attention.

On a general note, I would also encourage research ethics to go beyond a checklist of fulfilled requirements. Guidelines are a common and helpful practice, but are they ultimately enough to respond to the complexities of research and society? Considering the question of how and how much a researcher should be responsible for ethical risks I would urge a discussion on the researcher's role in society. By having the privileged role of a scholar, one could argue that a researcher has more responsibility than most when dealing with the ethics of large data sets, the collection methods, and the data analysts' and algorithms' impact on people's continued behaviour.

If a researcher is responsible for making an ethical assessment of the digital services providing the data, this could lead to a rejection of using behavioural surplus as research data.[69] While this could be an important ethical stand to take, it could also put the researcher at risk because they pass up an opportunity that someone else may pick up. The action then becomes a principled stand at the cost of publication opportunities, and the personal data is still being used. This scenario would be an argument for collective action. Still, if everyone applies that train of thought, how is change ever going to happen?

Considering the researchers who argue the consent given from users may not be a result of true choice, but of a lack of choice, I would say making an ethical assessment is an important responsibility of a researcher.[70] Ultimately, there will be new legislation governing the collection and use of personal data, especially by third parties, and hopefully, this legislation will be in line with the human rights. But until then, researchers need to step forward and take responsibility for their potential research subjects.

---

69 Zuboff, *The Age of Surveillance Capitalism*.
70 Cf. Floegel and Doty, "The Library/Surveillance Interface"; Zuboff, *The Age of Surveillance Capitalism*; Kim, *Wrap Contracts Foundations and Ramifications*; Kerpen, Dorgeist, and Zantis, "Intersecting the Digital Maze. Considering Ethics in Cloud-Based Services' Research."

The privacy of the reader vs the study of reading

When trying to outline the emerging field of critical studies of reading, Lundh et al. described Berglund's and Tattersall Wallin's quantitative analyses as employing a "non-evaluative approach to reading activities and reading practices". While they acknowledge that "[b]eing able to access and read documents of various kinds is fundamental to numerous activities in schools, workplaces, and in people's private lives", they do not reflect further upon the relation between the private life of the reader in relation to research methods.[71]

Berglund's reflections on the reading data feedback loop touched upon many points that could be criticized with the book streaming services. Discussing the datafication of digital reading, he emphasized the fact that "book streaming services, through this data, know much more about readers than publishers and authors do, and more than the literary industry have ever known before."[72] This is a crucial point that can be debated from a privacy perspective and the challenge for researchers is to decide whether to participate in this debate or not. Whose responsibility is privacy?

Considering the strong links between privacy, a democratic society, library practice, and library and information science, it is strange that this debate is not more present in related research. I would argue that the privacy of the reader and the right to private reading should become more present in the study of reading as more data and methods become increasingly available. That something is *possible* to do does not necessarily mean it is *advisable.* This should be a guiding principle in research ethics assessments, especially when working with novel methods. As a human right, the sign value of privacy should perhaps not be underestimated, and I am tempted to argue that on a general level the safeguarding of principles related to human rights could help lead to a more resilient democratic society.

Conclusion

In this position paper, I have discussed the use of large commercial data sets in the study of reading. By raising ethical matters related to user consent and legal certainty in the fast-developing information society. In my argumentation I draw on examples from two case studies and present three specific topics that need to be discussed to ensure good future research practices when using new types of data and digital methods: the issue of consent, the flaws in general research ethics practices, and the balance between reader privacy and the need to study reading.

As the case studies show, there are no easy answers to these questions and therefore I do not present any recommendations. But I urge a discussion to take place with the extensive background of this paper in mind. The principles at stake go beyond the individual

---

71 Lundh, Hedemark, and Lindsköld, "Critical Studies of Reading: Consolidating an Emerging Field of Research."

72 Berglund, *Reading Audio Readers*, 38.

researcher's decisions and connect to the development of the information society as a whole. Nevertheless, the individual academic choices we make are intrinsically linked to how society develops and how we perceive information in relation to people.

I want to end by emphasizing the need of better balancing the interests of researchers, research subjects, and commercial research data providers. I want to highlight that currently one group of stakeholders, the research subjects, are not sufficiently included in the research process (they possibly even lack awareness they are part of one) and therefore their privacy is threatened. I argue that the research community (interdisciplinary *and* within disciplines) must assume more responsibility in this matter, and that research ethics must go beyond the borders of a research project to look at the full picture.

I want to reiterate what sociologist Remy Stewart wrote in his article on the ethics and research implications of consumer-based datasets: "These matters are not ones that can be properly addressed without collective consideration of its multiple complexities by scholars from a diversity of interests, backgrounds, and opinions."[73] We need a discussion where scholars of law, technology, humanities, and social sciences can weigh in on this complex matter, and learn from each other's experiences of working with large sets of sensitive personal data and ethical considerations in relation hereto. The questions are not simple, but the day academia recoils from difficult questions may be the day we are no longer relevant for society.

Disclosure statement

To the best of their knowledge, the author has no conflicts of interest pertinent to the contents of this article. The author comes from the same country as both case study authors. Additionally, she was awarded her MA in library and information science (LIS) in the same country and therefore has some connection to the Swedish LIS research community. Furthermore, the author and Berglund worked together as librarians at Uppsala University Library for two years (2018-2020).

---

73 Stewart, "Big Data and Belmont."

Bibliography

Allington, Daniel, Sarah Brouillette, and David Golumbia. "Neoliberal Tools (and Archives): A Political History of Digital Humanities". *Los Angeles Review of Books*, 1 May, 2016. https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/.

Barnett, Tully. "Distributed Reading: Literary Reading in Diverse Environments". *Digital Humanities Quarterly* 12, no. 2 (2018). http://www.digitalhumanities.org/dhq/vol/12/2/000389/000389.html.

Barth, Susanne, and Menno D. T. de Jong. "The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review". *Telematics and Informatics* 34, no. 7 (2017): 1038–58. https://doi.org/10.1016/j.tele.2017.04.013.

Bawden, David, and Lyn Robinson. "Essentially Made of Information: Concepts and Implications of Informational Privacy". *Information Research* 24, no. 4 (15 December, 2019). https://informationr.net/ir/24-4/colis/colis1913.html.

Berglund, Karl. "Introducing the Beststreamer: Mapping Nuances in Digital Book Consumption at Scale". *Publishing Research Quarterly* 37, no. 2 (June, 2021): 135–51. https://doi.org/10.1007/s12109-021-09801-0.

———. *Reading Audio Readers: Book Consumption in the Streaming Age*. Bloomsbury, 2024.

———. "Strömmade bästsäljare : Litteraturkonsumtion i digitala prenumerationstjänster utifrån Storytels användardata". In *Från Strindberg till Storytel: Korskopplingar mellan ljud och litteratur*, ed. Julia Pennlert and Lars Ilshammar. Daidalos, 2021. 327–62.

Berglund, Karl, and Ann Steiner. "Is Backlist the New Frontlist?: Large-Scale Data Analysis of Bestseller Book Consumption in Streaming Services". *Logos* 32, no. 1 (25 May, 2021): 7–24. https://doi.org/10.1163/18784712-03104006.

Bon, Esmeralda V., and Michael Burke. "Devices, Settings and Distractions: A Study into How Students Read Literature". In *Pedagogical Stylistics in the 21st Century*, ed. Sonia Zyngier and Greg Watson. Springer International Publishing, 2022. 183–206. https://doi.org/10.1007/978-3-030-83609-2_8.

BookBeat. "BookBeat Integritetspolicy". Last updated 11 June, 2021, accessed 5 September, 2024. https://web.archive.org/web/20240905141858/https://assets.ctfassets.net/4s7izhcdroyy/1eS2r2ZPSnZea7JmxvMZye/06851fe11603669bb6fe09debad30741/privacyPolicySE_20210611.pdf.

———. "Ljudböcker och e-böcker. Prova gratis! - BookBeat". Accessed 5 September 2024. https://web.archive.org/web/20240625230537/https://www.bookbeat.com/se/about.

Clark, Stephen D., Becky Shute, Victoria Jenneson, Tim Rains, Mark Birkin, and Michelle A. Morris. "Dietary Patterns Derived from UK Supermarket Transaction Data with Nutrient and Socioeconomic Profiles". *Nutrients* 13, no. 5 (May 2021): 1481. https://doi.org/10.3390/nu13051481.

Darnton, Robert. "What Is the History of Books?". *Daedalus* 111, no. 3 (1982): 74.

Espino, Meredith Mays. "Sometimes I Feel Like Somebody's Watching Me . . . Read?: A Comment On The Need For Heightened Privacy Rights For Consumers Of Ebooks". *UIC John Marshall Journal of Information Technology & Privacy Law* 30, no. 2 (1 January, 2013). https://repository.law.uic.edu/jitpl/vol30/iss2/3.

Etikprövningsmyndigheten. "Etikprövning - så går det till". Accessed 4 September, 2024. https://etikprovningsmyndigheten.se/for-forskare/sa-gar-det-till/.

European Commission. "A European Approach to Artificial Intelligence | Shaping Europe's Digital Future". Accessed 3 October, 2023. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

———. "Data Protection in the EU". Accessed 4 July, 2023. https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en.

———. "Questions and Answers – New EU Copyright Rules". Updated 4 June, 2021, accessed 5 September, 2024. https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_2821.

Floegel, Diana, and Philip Doty. "The Library/Surveillance Interface". *Proceedings of the Association for Information Science and Technology* 58, no. 1 (2021): 152–61. https://doi.org/10.1002/pra2.444.

Habermas, Jürgen. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Polity Press, 1989.

Have, Iben, and Birgitte Stougaard Pedersen. "Reading Audiobooks". In *Beyond Media Borders, Volume 1*, ed. Lars Elleström. Springer International Publishing. 2021. 197–216. https://doi.org/10.1007/978-3-030-49679-1_6.

Henslee, Elizabeth. "Down the Rabbit Hole: E-Books and User Privacy in the 21st Century". *Creighton Law Review* 49, no. 1 (December, 2015). http://hdl.handle.net/10504/84523.

Hes, Ronald, and John J. Borking. *Privacy-Enhancing Technologies: The Path to Anonymity*. Achtergrondstudies En Verkenningen. Registratiekamer, 1998.

Hoeyer, Klaus. "The Ethics of Research Biobanking: A Critical Review of the Literature". *Biotechnology and Genetic Engineering Reviews* 25, no. 1 (1 January, 2008): 429–52. https://doi.org/10.5661/bger-25-429.

Horttanainen, Ella. "From Customer Data to Customer's Data : Reverse Use of Customer Data as a Tool for Value Co-Creation". Jyväskylän yliopisto, 2021. https://jyx.jyu.fi/handle/123456789/78665.

IFLA. *Riding the Waves or Caught in the Tide? Navigating the Evolving Information Environment*. IFLA, 2013. https://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report_v3.pdf.

Jenneson, Victoria, Graham P. Clarke, Darren C. Greenwood, Becky Shute, Bethan Tempest, Tim Rains, and Michelle A. Morris. "Exploring the Geographic Variation in Fruit and Vegetable Purchasing Behaviour Using Supermarket Transaction Data". *Nutrients* 14, no. 1 (January 2022): 177. https://doi.org/10.3390/nu14010177.

Jones, E. M., N. A. Sheehan, N. Masca, S. E. Wallace, M. J. Murtagh, and P. R. Burton. "DataSHIELD – Shared Individual-Level Analysis without Sharing the Data: A Biostatistical Perspective". *Norsk Epidemiologi* 21, no. 2 (13 April, 2012). https://doi.org/10.5324/nje.v21i2.1499.

Kallinikos, Jannis, Aleksi Aaltonen, and Attila Marton. "The Ambivalent Ontology of Digital Artifacts". *MIS Quarterly* 37, no. 2 (2013): 357–70.

Kerpen, Daniel, Matthias Dorgeist, and Sascha Zantis. "Intersecting the Digital Maze. Considering Ethics in Cloud-Based Services' Research". In *Research Ethics in the Digital Age: Ethics for the Social Sciences and Humanities in Times of Mediatization and Digitization*, ed. Farina Madita Dobrick, Jana Fischer, and Lutz M. Hagen, 143–52. Springer Fachmedien, 2018. https://doi.org/10.1007/978-3-658-12909-5_15.

Kim, Nancy S. *Wrap Contracts Foundations and Ramifications*. Oxford University Press, 2014. https://doi.org/10.1093/acprof:oso/9780199336975.001.0001.

Kungliga biblioteket. "Bläddra". Last updated 3 May, 2024, accessed 5 October, 2023. https://web.archive.org/web/20231005153806/https://www.kb.se/samverkan-och-utveckling/biblioteksutveckling/nationella-minoriteters-bibliotek/bladdra.html.

———. "Bibliotek 2022 : Sveriges offentligt finansierade bibliotek". Kungliga biblioteket, 2023. Accessed September 2, 2024. https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-707.

Kuzmičová, Anežka, Theresa Schilhab, and Michael Burke, "M-Reading: Fiction Reading from Mobile Phones", *Convergence* 26, no. 2 (1 April, 2020): 333–49, https://doi.org/10.1177/1354856518770987.

Kvale, Live Håndlykken, and Peter Darch. "Privacy Protection throughout the Research Data Life Cycle". *Information Research* 27, no. 3 (15 September, 2022). https://doi.org/10.47989/irpaper938.

Larson, Lotta C. "E-Books and Audiobooks", *The Reading Teacher* 69, no. 2 (2015): 169–77, https://doi.org/10.1002/trtr.1371.

Lell, Helga María. "Human Rights and the Regulation of Anonymity. New Challenges to Law and Research". In *Research Ethics in the Digital Age: Ethics for the Social Sciences and Humanities in Times of Mediatization and Digitization*, ed. Farina Madita Dobrick, Jana Fischer, and Lutz M. Hagen, 119–27. Springer Fachmedien, 2018. https://doi.org/10.1007/978-3-658-12909-5_12.

Lundh, Anna, Åse Hedemark, and Linnéa Lindsköld. "Critical Studies of Reading: Consolidating an Emerging Field of Research". *Information Research* 27 (25 October, 2022). https://doi.org/10.47989/colis2232.

Mai, Jens-Erik. "Big Data Privacy: The Datafication of Personal Information". *The Information Society* 32, no. 3 (26 May, 2016): 192–99. https://doi.org/10.1080/01972243.2016.1153010.

Mangen, Anne. "The Digitization of Literary Reading". *Orbis Litterarum* 71, no. 3 (2016): 240–62. https://doi.org/10.1111/oli.12095.

Metcalf, Jacob, and Kate Crawford. "Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide". *Big Data & Society* 3, no. 1 (1 June 2016): 2053951716650211. https://doi.org/10.1177/2053951716650211.

Murray, Simone. *The Digital Literary Sphere: Reading, Writing, and Selling Books in the Internet Era*. Johns Hopkins University Press, 2018.

Reader Privacy Act, Pub. L. No. 1.81.15., 1798.90 California Civil Code § 1798.90 (2012).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), 119 OJ L § (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng.

Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre de Montjoye. "Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models". *Nature Communications* 10, no. 1 (23 July, 2019): 3069. https://doi.org/10.1038/s41467-019-10933-3.

Rowberry, Simon. "The Limits of Big Data for Analyzing Reading". *Participations* 16, no. 1 (May, 2019): 237–57.

SFS 1994:1219. *Lag om den europeiska konventionen angående skydd för de mänskliga rättigheterna och de grundläggande friheterna* [Swedish statute incorporating the European Convention on Human Rights]. http://rkrattsbaser.gov.se/sfst?bet=1994:1219.

SFS 2009:400, *Offentlighets- och sekretesslag* [Swedish Publicity and Privacy Act]. 40 kap. §3. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400_sfs-2009-400.

SFS 2013:801. *Bibliotekslag* [Swedish Library Act]. http://rkrattsbaser.gov.se/sfst?bet=2013:801

Squires, Claire. "The Global Market 1970– 2015: Consumers". In *Companion to the History of the Book*, ed. Jonathan Rose and Simon Eliot. John Wiley & Sons, 2019.

Steiner, Ann, and Torbjörn Forslid. "Bokhandeln mellan kultur och ekonomi". In *Litterära värdepraktiker*, ed. Torbjörn Forslid, Jon Helgason, Christian Lenemark, Anders Ohlsson, and Ann Steiner, 67–125. Makadam förlag, 2017.

Stewart, Remy. "Big Data and Belmont: On the Ethics and Research Implications of Consumer-Based Datasets". *Big Data & Society* 8, no. 2 (1 July 2021): 20539517211048183. https://doi.org/10.1177/20539517211048183.

Storytel. "Integritetspolicy för Storytel". Last updated 12 September, 2022, accessed 5 September, 2024. https://web.archive.org/web/20240905122618/https://legal.storytel.com/se/sv/documents/privacy-policy?request_locale=se.

———. "INTEGRITETSPOLICY FÖR STORYTEL" (previous version). Last updated 12 June, 2019, accessed 21 July, 2022. http://web.archive.org/web/20220721162819/https://www.storytel.com/se/sv/documents/privacy-policy.

———. "Om Cookies – Storytel". Accessed 5 September, 2024. https://web.archive.org/web/20240905123249/https://www.storytel.com/se/cookies.

———. "Om oss". Accessed 11 September 2024. https://www.storytelgroup.com/sv/om-oss.

———. "Storytel Användarvillkor". Last updated 13 September, 2023, accessed 5 September, 2024. https://web.archive.org/web/20240905122304/https://www.storytel.com/se/sv/documents/terms-and-conditions.

Svenska bokhandlareföreningen and Svenska förläggareföreningen."Bokförsäljningsstatistiken 2022 ... och utvecklingen de senaste fem åren (2018-2022)". 2023: 50–51. Accessed September 2, 2024. https://forlaggare.se/wp-content/uploads/2023/02/Bokforsaljningsstatistiken-2022.pdf.

Tattersall Wallin, Elisa. *Sound Reading: Exploring and Conceptualising Audiobook Practices among Young Adults*. Högskolan i Borås, 2022. http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-27165.

Tattersall Wallin, Elisa, and Jan Nolin. "Time to Read : Exploring the Timespaces of Subscription-Based Audiobooks". *New Media and Society* 22, no 3 (2019). http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-21524.

UN General Assembly. "Universal Declaration of Human Rights." 217 (III) A. 1948. Accessed 2 September, 2024. https://www.un.org/en/about-us/universal-declaration-of-human-rights.

Vetenskapsrådet. *Good Research Practice*. Vetenskapsrådet, 2017.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Profile Books, 2019.