Scrutinizing Privacy in Multi-Omics Research: How To Provide Ethical Grounding for the Identification of Privacy-Relevant Data Properties[1]

C.W. Safarlou[2]; A.L. Bredenoord[3]; R. Vermeulen[4]; K.R. Jongsma[5]

Note: This is an Accepted Manuscript of an article published by Taylor & Francis in The American Journal of Bioethics on November 22nd 2021, available at: https://doi.org/10.1080/15265161.2021.1991041. Please cite the published version.

The outline of a framework for assessing privacy risks in multi-omic research and databases provided by Dupras and Bunnik is a valuable contribution to the literature on the ethics of omics data. They provide an empirically informed list of privacy-relevant omic data properties that help us to better understand the variety of privacy risks involved and present three steps that one can follow in order to assess "data properties and interrelation effects between omic types for the detection of privacy risks in multi-omic research and databases" (Dupras and Bunnik 2021, 2). In our commentary, we take a step back to discuss their framework in light of the normativity of the concept of privacy. We start by scrutinizing the underlying premise of normative-neutrality in the authors' approach, subsequently show how a normative understanding of privacy would collapse the distinctions that they attempt to make between 'intrinsic' properties, 'extrinsic' factors & contextual factors and end by proposing a new and primary step to ethically ground their framework.

Dupras & Bunnik write that their "model does not make any normative claims about privacy itself, i.e. about the acceptability or unacceptability of privacy risks, or about the proper level of data protection required for the responsible conduct of biological or health research" (Dupras and Bunnik 2021, 12). However, as privacy itself is a normative concept, the (non)acceptability of privacy risks and an assessment of the proper level of data protection do not exhaust the relevant normative claims. This becomes clear in the authors' cited broad definition of privacy as "freedom from unauthorized intrusion," because, to have an understanding of what constitutes an intrusion, one is required to make a normative claim about what privacy is (Dupras and Bunnik 2021, 3). Because the authors do not explicitly problematize this argumentative step, they are implicitly working with a normative understanding of what constitutes private information and thus are attempting to use a normative concept in a neutral way. In practice, this attempt can be seen in their evaluation of various omic data properties in their Table 1 (Dupras and Bunnik 2021, 8).

To illustrate the importance of making the normativity of privacy explicit, we start by analyzing two examples of omic data properties from the authors' Table 1. When considering whether the data property in question conveys observable phenotypic information, the authors do not comment on the data property's impact on the level of data sensitivity. But given that epigenomic information has the potential to improve the accuracy of facial portraits made on the basis of genomic data, we could imagine cases where a person does not want his or her former phenotypic information to become known because they have had plastic surgery to repair damage caused by an accident (Dupras and Bunnik 2021, 11). Another example from this table concerns whether a data property is determined (in part) by acts or behaviors conceived as willful, but here

[1] Correspondence: c.w.safarlou@umcutrecht.nl
[2] Department of Medical Humanities, Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands
[3] Department of Medical Humanities, Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands
[4] Department of Population Health Sciences, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
[5] Department of Medical Humanities, Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands

the authors do not comment on the impact that this property has on identifying power. Yet, if the acts or behaviors in question can be linked to a particular type of behavior, then the data property in question has impact on identifying power. For example, one can use high-resolution mass spectrometry (HRMS) to determine whether a person uses illicit drugs. This affects the impact of the data property on identifying power because one could then be on the lookout for people who use illicit drugs. In both examples, one is making normative claims about what constitutes a possible violation of privacy. But to make these types of judgements, one requires a normative understanding of human life (1) and an understanding of the case-specific capacity of data to represent information about human life (2).[6] Explicitly introducing these requirements to the authors' framework would ethically ground their framework and subsequently improve its capacity to judge the privacy-relevance of omic data properties

At this point in our argument, the authors could object by saying that these two requirements would necessitate an introduction of 'extrinsic' or contextual factors into their analysis, whereas they wanted to restrict themselves to 'intrinsic' data properties. Quite forcefully, they argue that:

> "…the most important contribution of our case-example comparison of privacy risks in genomics versus epigenomics is a better understanding of the role of 'intrinsic' omic data properties (i.e., properties associated with the type of biological variants under scrutiny) can play in raising or increasing some privacy risks." (Dupras and Bunnik 2021, 2)

However, the authors themselves note that making this distinction is difficult:

> "We find important to acknowledge here that it is very difficult – and possibly impractical – to draw clear-cut boundaries between intrinsic data properties and extrinsic factors which may also increase (re)identification risks and the level of sensitivity of the information potentially conveyed. For instance, […] when trying to distinguish abnormal from normal or disruptive from adaptive (or neutral) biological variants, complex questions may arise relating to the definition and meaning of concepts such as "disorder" or "disease."" (Dupras and Bunnik 2021, 11)

Here, we would like to push the authors' train of thought one step further because we believe that adopting a normative notion of privacy necessitates the collapse of the boundaries between 'intrinsic' data properties, 'extrinsic' factors and contextual factors. To even start to identify what they consider to be 'intrinsic' omic data properties and put them on the list of privacy-relevant omic data properties, one must already involve the two requirements that we mentioned above, namely a normative understanding of human life (1) and an understanding of the case-specific capacity of data to represent information about human life (2). Let us look at another omic data property from their Table 1: whether the data property is determined (in part) by acts or behaviors conceived as willful. To identify this property as privacy-relevant, one already has to conceptualize people as beings with the power to determine their own course of action (1) and understand that epigenetic information has the capacity to model information that is determined (in part) by acts or behaviors conceived as willful (2). The same holds true for less obvious data properties mentioned in their Table 1, such as whether a data property is ubiquitous among cell types and tissues. As the authors write:

> "High ubiquity (e.g. genotype ubiquity) increases the likelihood that sensitive information (e.g. high risk to breast cancer) can be revealed by cell types or tissues which are neither functionally relevant to the particular information (e.g., saliva sample), nor the primary research object." (Dupras and Bunnik 2021, 8)

[6] For a defense of these two requirements in the context of business ethics & economic theory, see (Safarlou 2021).

This explanation implicitly utilizes a normative understanding of human life due to the common knowledge that information about a high risk for cancer is sensitive information for any human being's life (1) and includes a description of the capacity of data to represent information about human life by means of its explanation of how cell types or tissues can reveal sensitive information (2). Consequently, because privacy itself is a normative concept, and a normative understanding of human life (1) and an understanding of the case-specific capacity of data to represent information about human life (2) form prerequisites for identifying the privacy-relevant omic data properties that are listed by the authors, the distinctions between 'intrinsic' properties, 'extrinsic' factors and contextual factors cannot be drawn and collapse.

We do not believe that the collapse of these distinctions poses a fundamental problem for the framework that the authors propose. Identifying privacy-relevant data properties that are described on a level that is closely associated with the type of biological variants under scrutiny is still highly valuable for the reasons the authors mention (Dupras and Bunnik 2021, 2–3). However, one cannot do so while conceptualizing omic data properties by means of the aforementioned distinctions and while using a conception of privacy that is treated as neutral with respect to what intrusions are. Instead, we would encourage them to ethically ground their framework by fully integrating it with a normative conception of privacy.

Consequently, we propose that a new and primary step should be included in their framework in order to explicitly tackle the aforementioned argumentative step that has to be made from data to privacy-relevance (in essence, from fact to value). Formulated as a question: "What can this data potentially tell us about a person's life?"[7] This question would be answered by going over the first column in their Table 1, potentially adding new entries to the list, and then writing "Yes" or "No" in the second column. Adding this step to their framework has several positive consequences. First of all, introducing this step helps to solve the argumentative lacuna that is mentioned in the beginning of this commentary (the lack of an explicit recognition of the normative claims made with respect to the normativity of privacy itself). Secondly, explicitly reflecting on whether there are any new privacy-relevant omic data properties stimulates the model's desired flexibility in a still-growing research area (Dupras and Bunnik 2021, 13). Thirdly, it bolsters the authors' case against genetic exceptionalism by calling for an explicit reflection on the informative value of data (as such) about human life and thus treating all data equally from the outset, without making a special or metaphysical exception for genetic data (Dupras and Bunnik 2021, 2).

Bibliography

Dupras, Charles, and Eline M. Bunnik. 2021. "Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases." *American Journal of Bioethics*. https://doi.org/10.1080/15265161.2020.1863516.

Safarlou, Caspar Willem. 2021. "How (Not) to Connect Ethics and Economics: Epistemological and Metaethical Problems for the Perfectly Competitive Market." In *Words, Objects and Events in Economics: The Making of Economic Theory*, edited by Peter Róna, László Zsolnai, and Agnieszka Wincewicz-Price, 91–101. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-52673-3_6.

[7] The authors formulate their steps as questions in the last section of their paper (Dupras and Bunnik 2021, 13).