Personal Data Protection and Global Data Sharing

Mark Phillips
Centre for Genomics & Policy, McGill University
mark.phillips2@mcgill.ca

2019

INTRODUCTION

The core principles of personal data protection[1] have remained surprisingly stable throughout the first half-century of existence—especially given the dramatic technological change the world has experienced over the same period—but uncertainty and caprice are nonetheless currently the central features of the field's enforcement, interpretation, and development.

Data protection is driven by breach, vulnerability, and scandal. The Cambridge Analytica scandal has recently led to new regulation of personal data processing by political parties and social media services. Edward Snowden's revelations about mass surveillance led directly to the demise of the EU-U.S. Safe Harbor Framework, the key transatlantic data-sharing arrangement, when it was deemed inadequate by the European Court of Justice in 2015. In the science context, the demonstration by academics that a person's disease status could effectively be inferred based only on genomic summary statistics published in a case–control study led the National Institutes of Health to shutter open access to dbGaP, its Database of Genotypes and Phenotypes (Zerhouni and Nabel 2008). Similarly, reidentification of "anonymized" Australian healthcare data caused the government to push a law that would retroactively make it a criminal offence to attempt to re-identify government data (Phillips, Dove, and Knoppers 2017).

In our rapidly changing technological and regulatory context, risk cannot fully be eliminated. Barring a decision to forgo large-scale data sharing altogether, data-sharing initiatives must instead aim to leverage legal, technical, and organizational strategies to minimize the risk of scandal, breach, and adverse legal action, which ideally limit this risk to one that is very small.

This white paper focuses on issues related to data protection in general, and to the influential EU General Data Protection Regulation (GDPR) in particular. A large-scale data-sharing initiative should also consider potential tension with respect to other relevant areas of the law, notably intellectual property, though these fields are outside the scope of this paper.

[1] Although the notion of *data protection* is not in wide use in the United States, this paper is nonetheless centering on it because of its greater precision than "privacy", and because of the paper's focus on the EU *General Data Protection Regulation*.

The GDPR is of key importance to global data-sharing initiatives not only because of its unparalleled global influence, but also because of, on the one hand, its global territorial reach, including direct application to international entities who engage in certain forms of processing of the data of persons in the European Economic Area (EEA) including the United States, and on the other hand, its eye-catching fines of up to €20-million or 4% of annual worldwide turnover, whichever is greater.

From the general perspective just described, the following two sections of this paper address de-identification techniques, including differential privacy, and cross-border transfer of personal data.

DIFFERENTIAL PRIVACY AND DE-IDENTIFICATION

Identifiability is a fundamental concept in data protection. It is in fact possible to avoid regulation by data protection frameworks entirely by anonymizing personal data. This has made anonymization a particularly compelling compliance strategy when it is practical.

The de-identification framework set out in the U.S. *Health Insurance Portability and Accountability Act* (HIPAA) makes its standard clear. De-identification can be achieved either by removing all occurrences of eighteen specified fields (including names, telephone numbers, etc.), or by obtaining a detailed written opinion from a statistician declaring the risk of re-identification to be very small.

But most data protection frameworks, including the GDPR, abandon this type of clear standard in favor of a contextual, case-by-case approach to determine when anonymity has been achieved (Article 29 Data Protection Working Party 2007). This approach better reflects the reality that, on the one hand, even a name (e.g., "Fred") can be non-identifying absent any further context and, on the other hand, that any information, even data with no apparent meaning at all (e.g., a series of seemingly random bits) can reveal sensitive personal information when combined with other information that may be available (e.g., a decryption key). The drawback to this approach, of course, is that it is generally impossible in practice to be certain that data has been rendered anonymous prior to an official decision to that effect by a regulator or the courts or, in particularly unlucky cases, when evidence surfaces to demonstrate that it is possible to re-identify the data.

In addition to this drawback, since the mid-2000s, a series of clever re-identification attacks have been published whose overall effect has been to seriously erode confidence in the ability of anonymization to address contemporary data protection concerns outside of isolated circumstances (Ohm 2010).

New de-identification techniques have emerged in attempts to fill this gap, including differential privacy, *k*-anonymity, homomorphic encryption, secure multiparty computation, and secure enclaves. The legal status under European Union data protection law of several such techniques was explored by guidance published by the statutory body comprised of European data protection authorities, under the title *Opinion 05/2014 on Anonymization Techniques* (Article 29 Data Protection Working Party 2014).

According to this opinion, differential privacy will generally not render data anonymous to the degree that it can escape regulation by the GDPR. This conclusion is a consequence of the opinion's description of the implementation of a differential privacy strategy:

> [T]he data controller generates anonymised views of a dataset whilst retaining a copy of the original data. Such anonymised views would typically be generated through a subset of queries for a particular third party. The subset includes some random noise deliberately added ex-post. Differential privacy tells the data controller how much noise he needs to add, and in which form, to get the necessary privacy guarantees. (p. 15)

Because the fully identifiable dataset is retained by the data controller, it will generally remain technically possible to link the data back to individuals.[2] The traditional European data protection view has been that this inherently prevents the data from being considered anonymous.

But this strict approach has since begun to soften. In its 2016 *Breyer v. Bundesrepublik Deutschland* decision, the European Court of Justice held that it is possible for data to be considered anonymous from the perspective of entities who lack the legal means to access additional data necessary to re-identify data subjects, even if a third party is known to hold such data (European Court of Justice 2016). Mourby et al. (2018) recently articulated an interpretation of *Breyer* (at least as it applies in the UK) that is much more expansive, namely that anonymization under the GDPR is possible not only in situations where re-identification is technically feasible but legally impossible, but also whenever re-identification is not "reasonably likely", which they believe to be compatible with situations where re-identification is known to be both legally and technically feasible. For now, however, it remains dubious whether European law stretches this far.

But even if differential privacy or similar de-identification measures are incompatible with anonymization under the GDPR, this does not mean that it is a waste of resources to implement such measures. Not only do they contribute to minimizing the practical risk of breach and scandal, but thoughtful implementation will also generally further compliance with other GDPR obligations, such as its principle data minimization and obligations to implement appropriate technical and organizational safeguards. The opinion on anonymization referenced earlier does, however, suggest that certain limits should be respected when considering or implementing differential privacy techniques:

To limit inference and linkability attacks it is necessary to keep track of the queries issued by an entity and to observe the information gained about data subjects; accordingly, "differential-privacy" databases should not be deployed on open search engines that offer no traceability of the querying entities (Article 29 Data Protection Working Party 2014, 15).

In sum, even if large-scale data-sharing initiatives are generally unlikely to succeed in bringing their activities outside of the scope of the GDPR by anonymizing the personal data they control, they should nonetheless carefully consider whether and how to implement techniques to minimize identifiability at each stage of the data life cycle, and should regularly review their approach. This review should consider developments with respect to both the technical and legal state-of-the-art, paying particular attention to evolving GDPR case law and guidance released by the regulators themselves, such as the documents above, and anonymization handbooks (U.K. Information Commisioner's Office 2012).

CROSS-BORDER TRANSFER

Regulation of cross-border transfer of personal data has been an element of data protection since almost its earliest days (Phillips 2018). In this context, it is important to distinguish *transfer restrictions*, which require appropriate safeguards when personal data is transferred out of a legal jurisdiction to ensure that it will continue to receive appropriate protection elsewhere, and *data localization*, which is an absolute prohibition on transferring data out of a jurisdiction (or, occasionally, a requirement to keep a copy of the data within the jurisdiction irrespective of whether copies are also transferred elsewhere).

[2] In the words of the opinion: "avoid the mistake of thinking the data are anonymous for the third party, while the data controller can still identify the data subject in the original database taking into account all the means likely reasonably to be used" (p. 16).

This section focuses on transfer restrictions. Although data localization may conceptually appear to be more fundamentally in tension with international data sharing, in practice its effect is currently less pronounced. Although the number of localization laws has increased in the years following the Snowden revelations, their scope is generally limited to relatively narrow sectors (Wei 2018). One new and notable exception to this is China's new *Cybersecurity law*, which imposes a near-complete prohibition on the transfer of "critical information infrastructure" out of the country. The law defines this notion expansively, if not entirely clearly, but it seems to encompass a broad range of personal data (Chen and Song 2018). In recent years, international trade treaties have tended in the opposite direction, and increasingly aim to ban data localization restrictions within any country that is signatory (McLeod 2018).

Transfer restrictions, on the other hand, are intended to be compatible with the continued flow of personal data across borders, instead promoting (i.e., requiring) appropriate protection for the personal data even after it arrives at its destination. These restrictions tend to apply broadly to all personal data that is subject to a given data protection framework, such as in the GDPR, rather than to a specific sector.

Accordingly, before personal data can be transferred outside the EEA, the GDPR's transfer rules require that the transfer meet one of several conditions that it lists.

The GDPR's preferred transfer justification mechanism is an *adequacy decision*. This refers to a prior decision by the European Commission that has deemed the data protection framework to which the data will be subject at its destination to be "adequate" with respect to the level of protection provided by the GDPR. In the context of a global data-sharing project, this mechanism on its own will not fully satisfy the GDPR's transfer obligations, because it can only justify transfer to the relatively limited number of countries for which an adequacy decision has to date been granted (European Commission 2018).

If the global data-sharing infrastructure project in question is based outside of the EEA, however, it might nonetheless at least rely on an adequacy decision to satisfy the GDPR's transfer rules with respect to the transfer of personal data from the EEA to itself. If the project were based in the United States, the only framework that is currently considered adequate by the European Commission is the EU-U.S. Privacy Shield, a self-certification framework administered by the Department of Commerce. A U.S.-based data-sharing project could thus satisfy the GDPR transfer condition with respect to receiving the personal data of Europeans by going through the process of certifying itself as compliant with the Privacy Shield.

The remaining question, of course, would be the legality of such a project's onward transfers of the personal data it has received to the project's users around the globe.

The GDPR's transfer mechanisms, such as the Privacy Shield adequacy decision, interestingly appear to supersede the GDPR's global application. Recall that the introduction to this paper noted that an entity outside of the EEA that, for example, offers goods and services to people inside the EEA, will be directly subject to the GDPR with respect to the personal data it collects from those people. But if the entity outside the EEA instead receives that type of data on the basis of an adequacy decision from an EEA-based intermediary which instead offered such goods and services, then the GDPR appears not to directly apply to the non-EEA based entity: the framework that was approved as adequate (e.g., Privacy Shield) instead stands in for the GDPR.

Returning to the example of a large-scale data-sharing project based in the United States that was certified under the Privacy Shield, this would mean that when the project transfers its data onward to its users in other countries, the project will not subject to the transfer restrictions in the GDPR, but instead to those established by the Privacy Shield itself. The third principle in the Privacy Shield sets

out its rules regarding accountability for onward transfer, and centers on a requirement that the sender must first enter into a contract with the recipient aiming to maintain a similar level of data protection (U.S. Department of Commerce 2016).

In the absence of an adequacy decision, the GDPR's transfer condition may instead be satisfied by providing "appropriate safeguards", which in practice means one of the specific mechanisms listed in its Article 46. One such mechanism is to have the sender and recipient conclude a contract that incorporates "standard data protection clauses" that have been adopted by the European Commission, or adopted by a national supervisory authority and approved by the Commission. Another mechanism, in cases where a transfer out of the EEA nonetheless occurs within a single legal entity or certain related entities, is to put in place what the GDPR refers to as "binding corporate rules" to establish appropriate safeguards within the entity or group and to have them approved by the appropriate European data protection authority.

In the absence of the appropriate safeguards set out in Article 46 and if an adequacy decision is unavailable, the GDPR also sets out "derogations" in its Article 49 that can be used to satisfy its transfer rules as a last resort. These include explicit, informed consent of the people whose data is being transferred. However, guidance on the interpretation of this article—as well as arguably the text of the GDPR itself—suggest they should be resorted to only with significant caution as they are likely to "be interpreted restrictively so that the exception does not become the rule" (European Data Protection Board 2018).

FURTHER DATA PROTECTION ISSUES

The previous sections of this paper focused on two key areas of the GDPR on which I was explicitly invited to comment in the context of a global data-sharing initiative, namely identifiability and cross-border transfer of personal data.

A number of other data protection compliance issues are, however, likely to arise in this context. An entity to which the GDPR applies directly will have to consider other compliance issues, of which I will note two here. First, the GDPR allows personal data processing only when one of its lawful bases for the processing has been identified, which include the consent of the data subject (Article 6): an appropriate lawful basis should be chosen carefully. Second, projects that intend to process forms of personal data that Article 9(1) of the GDPR deems to be sensitive—which the GDPR refers to as "special categories of personal data"—the project will have to, in addition to identifying the lawful basis for processing personal data and meeting a transfer condition, satisfy one of the conditions described in Article 9.

However, if the data-sharing project determines that it is not directly subject to the GDPR because it has instead received data from Europe in the context of a transfer mechanism such as the Privacy Shield, as explained in the previous section of this paper, it is that mechanism whose provisions will instead need to be analyzed in detail. Although the Privacy Shield and other mechanisms that satisfy the GDPR transfer condition aren't required to provide a one-to-one correspondence with the content of the GDPR, some GDPR content is nonetheless likely to remain relevant. For example, personal data is defined by the Privacy Shield by reference to European data protection law, and so the section of this paper on identifiability will remain relevant under either the GDPR or the Privacy Shield.

CONCLUSION

Because of the GDPR and Privacy Shield are each recent arrivals in the world of data protection, little-to-no case law yet exists to assist in interpreting their provisions. This uncertainty further supports the suggestion in the introduction to this paper that global personal data-sharing initiatives should focus on minimizing their legal and practical risks, rather than insisting on eliminating them altogether. Although such initiatives should not hesitate to mobilize techniques to reduce the identifiability of the data they process, in most circumstances meeting the GDPR's threshold for complete anonymity will be impractical, at least without abandoning most of the initiative's intended benefits. In this context, consideration will have to be given to the optimal approach to satisfying the GDPR's transfer requirements, and in determining whether the GDPR will apply to the project directly, or whether it will instead be subject to a transfer mechanism such as the Privacy Shield.

REFERENCES

References

Article 29 Data Protection Working Party. 2007. *Opinion 4/2007 on the Concept of Personal Data*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf.

———. 2014. *Opinion 5/2014 on Anonymisation Techniques*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Chen, Yongxi, and Lingqiao Song. 2018. "China: Concurring Regulation of Cross-Border Genomic Data Sharing for Statist Control and Individual Protection." *Human Genetics* 137 (8): 605–615. doi:10.1007/s00439-018-1903-2.

European Commission. 2018. *Adequacy of the Protection of Personal Data in Non-EU Countries*. https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en.

European Court of Justice. 2016. *Patrick Breyer v. Bundesrepublik Deutschland*. http://curia.europa.eu/juris/document/document.jsf?docid=184668&doclang=EN.

European Data Protection Board. 2018. *Guidelines 2/2018 on Derogations of Article 49 Under Regulation 2016/679*. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_2_2018_derogations_en.pdf.

McLeod, James. 2018. "Data Localization Concerns in USMCA May Be Overblown." *Financial Post*. https://business.financialpost.com/technology/data-localization-concerns-in-usmca-may-be-overblown.

Mourby, Miranda, Elaine Mackey, Mark Elliot, Heather Gowans, Susan E. Wallace, Jessica Bell, Hannah Smith, Stergios Aidinlis, and Jane Kaye. 2018. "Are 'Pseudonymised' Data Always Personal Data?

Implications of the GDPR for Administrative Data Research in the UK." *Computer Law & Security Review* 34 (2): 222–233. doi:10.1016/j.clsr.2018.01.002.

Ohm, Paul. 2010. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* 57 (6): 1701–1778.

Phillips, Mark. 2018. "International Data-Sharing Norms: From the OCED to the General Data Protection Regulation (GDPR)." *Human Genetics* 137 (8): 575–582. doi:10.1007/s00439-018-1919-7.

Phillips, Mark, Edward S. Dove, and Bartha M. Knoppers. 2017. "Criminal Prohibition of Wrongful Re-Identification: Legal Solution or Minefield for Big Data?" *Journal of Bioethical Inquiry* 14 (4): 527–539. doi:10.1007/s11673-017-9806-9.

U.K. Information Commisioner's Office. 2012. *Anonymisation: Managing Data Protection Risk Code of Pracitce*.

U.S. Department of Commerce. 2016. *Privacy Shield Framework*. https://www.privacyshield.gov/EU-US-Framework.

Wei, Yuxi. 2018. "Chinese Data Localization Law: Comprehensive but Ambiguous." Accessed May 1, 2019. https://jsis.washington.edu/news/chinese-data-localization-law-comprehensive-ambiguous/.

Zerhouni, Elias A., and Elizabeth G. Nabel. 2008. "Protecting Aggregate Genomic Data." *Science* 322 (5898): 44a–44a. doi:10.1126/science.1165490.