FUTURE OF PRIVACY FORUM

VUB BRUSSELS PRIVACY HUB

Brussels Privacy Symposium 2020: Research and the Protection of Personal Data Under the GDPR

Symposium Report

Caroline Hopland, Hunter Dorwart, Gianclaudio Malgieri, Gabriela Zanfir-Fortuna and Rob van Eijk

2 March, 2021

The Future of Privacy Forum

In Europe, **the Future of Privacy Forum (FPF)** is an independent voice, maintaining its neutrality in any discourse. FPF is highly optimistic that social and economic good can be achieved through innovation in data and technology while also respecting privacy and data protection rights. We know we can make a difference in the dialogue needed to achieve just that.

FPF has built strong partnerships across Europe through its convenings and trainings for policymakers and regulators. FPF's transatlantic engagement helps regulators, policymakers, and staff at European Union data protection authorities better understand the technologies at the forefront of data protection law. FPF explains EU data protection and privacy law and the European Court of Human Rights legal framework to make them easily understandable for stakeholders in the U.S.. FPF hopes to bridge the gap between European and U.S. privacy cultures and build a common data protection language.

A space for debate and dialogue

FPF is a non-profit organization providing a space for debate and dialogue by:

- Sharing knowledge on European privacy and data protection law with its members
- Connecting a network of key players from corporations, NGOs, academics, civil society, and regulators
- Engaging with EU regulatory bodies and policymakers
- Being a respected voice in the media
- Advising corporations and policymakers regarding technological, privacy and data protection issues
- Offering regular peer-to-peer gatherings, workshops, Masterclasses, and training interventions in selected hotspots across Europe

Authors

Caroline Hopland

Hunter Dorwart

Gianclaudio Malgieri

Gabriela Zanfir-Fortuna

Rob van Eijk

Table of Contents

1. Introduction

On December 2, 2020, the Future of Privacy Forum (FPF) and the Brussels Privacy Hub of Vrije Universiteit Brussel (VUB) hosted the *Brussels Privacy Symposium 2020: Research and the Protection of Personal Data Under the GDPR,* convened by Jules Polonetsky, CEO of FPF, and Dr. Christopher Kuner, Co-Chair of the Brussels Privacy Hub. The Symposium brought together industry privacy leaders, academic researchers, and regulators to discuss data protection in the context of scientific research under the European Union's General Data Protection Regulation (GDPR) from various policy and technical perspectives.

Most notably, the panelists emphasized risks and vulnerabilities with respect to data protection in the scientific research context, highlighting issues with consent structures, artificial intelligence (AI) and machine learning systems during the Covid-19 pandemic, defining sensitive data, privacy enhancing technologies to be applied to research datasets, the role of international frameworks and of cross-border data flows, and certain risks of using data for research. EU policy makers, however, have provided useful guidance and introduced new frameworks to assist with data protection in the scientific research context. One of the relevant legislative proposals recently introduced by the European Commission is the Data Governance Act (DGA), which "aims to foster the availability of data for use by increasing trust in data intermediaries and by strengthening data-sharing mechanisms across the EU". It also proposes to promote "data altruism", allowing researchers access to larger datasets for their research. Overall, the Symposium focused on striking a balance between utility of research and privacy and data protection.

The keynote speakers included:

- Dr. Malte Beyer-Katzenberger, DG CONNECT, European Commission
- Cornelia Kutterer Senior Director, EU Government Affairs, AI, Privacy and Digital Policies at Microsoft Corporation
- Dr. Wojciech Wiewiórowski, European Data Protection Supervisor (EDPS)

The first panel explored *Complex Interactions: the GDPR, Data Protection and Research,* and it was moderated by Dr. Gianclaudio Malgieri, Associate Professor EDHEC Augmented Law Institute (Lille) and Affiliated Researcher LSTS VUB. Speakers included:

- Claire Gayrel, Deputy Head of Unit Supervision and Enforcement, EDPS
- Dr. Dara Hallinan, Legal Academic, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
- Dr. Ciara Staunton, Senior Lecturer in Law, School of Law, Middlesex University, London and Centre for Biomedicine, EURAC, Bolzano, Italy
- Dr. Henrik Junklewitz, Scientific Project Officer, Joint Research Center, European Commission

The second panel discussed *Using Sensitive Data in Research to Counter (Hidden) Bias and Discrimination,* and it was moderated by Dr. Gabriela Zanfir-Fortuna, Senior Counsel FPF and Affiliated Researcher LSTS VUB. Speakers included:

- Dr. Elettra Ronchi, Senior Policy Analyst, Organisation for Economic Co-operation and Development
- Dr. Paul Quinn, Professor, VUB
- Dr. Heng Xu, Professor of Information Technology and Analytics, American University
- Knut Mager, Head of Global Data Privacy, Novartis

2. Opening Keynote - European Union Data Governance Act Proposal

The Symposium began with a keynote from Malte Beyer-Katzenberger and Cornelia Kutterer, who discussed the recent Data Governance Act (DGA) proposed by the European Commission on November 25, 2020. The DGA targets four primary goals: i) setting forth conditions for the re-use of data held by the public sector; ii) creating a regulatory regime for data sharing services; iii) outlining the potential for data altruism through individuals donating their data to certain bodies for the public good; and iv) establishing a comprehensive governance framework including creating a European Data Innovation Board to provide clarification on the Act.

Beyer-Katzenberger explained the novel aspects of the DGA and helped clarify the Commission's overall thinking. As the first Act under the European Data Strategy, the DGA will attempt to tackle issues related to data sharing between the public and private sectors as well as establish a regulatory framework to guide the Commission's larger goals with respect to the digital economy. Indeed, Beyer-Katzenberger noted that the Act focuses on bringing out the tools, means, processes, and the possible intermediaries to facilitate the use of data in a way that allows data subjects to exercise more control and receive greater transparency.

While it will take time for EU policymakers to work out many of the specifics of the DGA, the proposed Act raises novel concepts and presents new terrain for compliance expectations across the board. The DGA does not aim to modify or come into conflict with the GDPR. Rather, the Act strives to make enforcement and compliance with the GDPR more effective while preserving the value of data and increasing its utility. Most importantly, the DGA suggests an alternative evolutionary path for the data economy that could help new products enter the market at a level playing field. Additionally, the DGA interacts with algorithmic machine learning and artificial intelligence systems with respect to scientific research. For example, by creating trust through the regulation of "data altruism", relevant and accurate personal data can be made available for the development of AI models.

2.1 ALGORITHMIC MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Beyer-Katzenberger stated that the DGA regulates corporate data holders who hold personal and non-personal data and looks towards their role in the larger ecosystem of connectivity, artificial intelligence (AI), and machine learning. One crucial element of this ecosystem is the potential use of large datasets held by the public sector for machine learning and next generation technological applications.

There are many scientific advantages in allowing corporate data holders to use various types of data held by the public sector for scientific purposes. For instance, public health data presents one of the largest opportunities for data sharing and scientific research. But such sharing also imposes many risks such as those related to privacy and discrimination. To this end, the DGA focuses on striking the right regulatory balance between these risks and benefits and aims to ensure such data sharing can happen in a controlled manner.

As Beyer-Katzenberger noted, the European Commission wanted to ensure that the public sector would engage in anonymization practices in appropriate ways and drafted the DGA to create a structured process for this purpose. Part of this process involves Member States setting up central, regional bodies, or sector-specific bodies that could provide regulatory oversight with the needs of users in mind.

In addition, the DGA also addresses issues surrounding cities and secure processing environments. It grants the public sector access to supervised access facilities or secure processing environments where researchers can use certain datasets for research. Such research includes calculating derivative factors about data, seeing algorithms and correlations patterns, or otherwise using the data for other scientific research purposes. Further access to data for research purposes could provide many benefits of these practices, as seen during the Covid-19 pandemic.

The DGA will also interact with upcoming regulatory proposals on AI by considering the potential risks posed by AI processing and services. Currently, many of the services that come to the market in Europe have never been trained on the meaning and definition of European data, leading to services that fail and contain numerous errors. Breyer-Katzenberg expressed that these systems should be able to work with the best data to avoid errors.

2.2 DATA ALTRUISM

Notably, the DGA also introduces the topic of “data altruism”, which allows individuals to make their personal data available to entities for a number of purposes, not the least including scientific research. Beyer-Katzenberger explained that this would help entities and research bodies gain access to stronger datasets and conduct more robust experiments and product testing. Indeed, the DGA attempts to make “data altruism” desirable to individuals by providing a specific way to facilitate it while still protecting data subjects’ rights.

This mechanism need not conflict with other legal tools to facilitate “data altruism” such as consent mechanisms under the GDPR. The DGA sets up a consent form that gives a standard type of language similar to recital 30 of the GDPR. But according to Beyer-Katzenberger the Act should also go further and promote dynamic consent by allowing data subjects to monitor what happens with the data. In addition, policymakers and industry should also create intuitive means for data

subjects to stay informed about what is happening with their data. Such means should also enable data subjects to easily revoke consent if they feel as though they signed up for something that is not actually in their best interests.

The DGA also attempts to recognize the professionalization of altruism, by supporting and encouraging NGOs and academics to allow non-profit organizations to organize data pools in a manner that could grow to useful sizes. Beyer-Katzenberger highlighted that citizens will want to give back to the community with respect to their data. Many communal spaces throughout Europe such as in Spain and France are already engaging in these activities. The DGA will not try to overregulate this space or prevent these activities from occurring.

Additionally, Beyer-Katzenberger noted that the DGA introduces additional concepts like "personal data stores," which refers to new services on the market that allow people to more effectively exercise their rights under the GDPR. Such services would involve "data operators" that provide technical tools to make it easier and more intuitive for individuals to submit access requests. These "data operators" would also allow individuals to give permission for other systems and entities to use their data.

2.3 CROSS-BORDER ACCESS TO DATA

Moreover, some parts of the DGA allude to cross-border access to data, including data sharing with entities outside of the European Union. Beyer-Katzenberger explained that these provisions are complementary to the GDPR provisions regarding international data flows, but aim to expand protection to commercially confidential information, trade secrets, and IP protected content. In doing so, the DGA aims to help companies who have put their data into a public sector database feel comfortable if there was ever a re-use of that data. The Act also strives to address problems related to commercially confidential information through regulatory mechanisms over data sharing service providers. Finally, guarantees to ensure that no governmental authority could access such data after that point for legitimate purposes will aim to ensure that certain protections will travel with the data, regardless of where it goes.

2.4 MORE PERSPECTIVE ON THE DATA GOVERNANCE ACT

The data sharing industry is nascent but has grown in recent years. Cornelia Kutterer highlighted how the DGA fits within this larger paradigm. The evolution of the data ownership debate has underscored the recent push towards "data altruism", as intermediaries and public organizations realize the benefit of pooling community data. For example, in light of the Covid-19 crisis, machine learning applications related to public health research have been instrumental in building epidemiological knowledge about the virus.

Kutterer stated that Microsoft shares the Commission's goal to enable a data sharing environment and advance Europe's digital goals and global competitiveness. However, she noted that it is important to see how this actually functions in practice in order to strike the right regulatory balance between competing interests. With respect to Microsoft's data sharing practices, Kutterer explained that the company has focused on allowing others to share data in a privacy respected manner, highlighted in the company's Open Data Campaign.

Microsoft is also working with the Open Data Institute and has invested in data sharing initiatives around differential privacy. These initiatives enable organizations to share information about datasets by describing patterns in it without releasing personal information about specific people. While it is important to regulate in this space, the Commission should also enable European companies to choose the best technology on the market in order to increase global competitiveness. These types of underlying technologies could help to address some of these privacy concerns.

3. Complex Interactions: the GDPR, Data Protection, and Research

The first panel of the Symposium focused on the complex interaction between the provisions of the GDPR, other data protection provisions and processing personal data for scientific research purposes. While the DGA aims to streamline data sharing in a controllable way, conversations around existing regulations like the GDPR underscore the challenges stakeholders currently face with respect to a multitude of issues. To address some of these issues, the first panel discussed the gap between the GDPR's safeguards and the derogations relating to the processing of personal data for scientific research purposes. In addition, the discussion also focused on the GDPR's framework limitations with respect to the collection of sensitive data and the sharing of such data across organizations and national borders. Finally, the panelists touched upon the numerous risks of sharing and processing sensitive health information for scientific research, as well as the current trends of consent structures for data subjects when conducting research.

3.1 CONSENT

Ciara Staunton first discussed the impacts of the GDPR on scientific research. She argued that given the benefits, there is currently a push towards open science and making data more accessible. However, when dealing with sensitive information (which is often health information), there are risks in the sharing and processing of this information. Such risks include individual and group privacy concerns, stigmatization and discrimination against individuals, groups and minority populations, and broad misuse of data. Other risks arise depending on the context in which researchers use the data and the vulnerabilities of the specific population in question.

Staunton noted that it is important to guard against these real and perceived risks to prevent corrosion of the public trust that scientific research partially relies on for its legitimacy. One major concern in this area is the adequacy of existing consent structures for the sharing and processing of sensitive health information for scientific research. Research ethics protect privacy by using informed consent and anonymization practices, but concerns exist regarding the degree to which consent is truly informed. What's more, Staunton argued that anonymization is not always desirable because the data can lose its value and data subjects may experience difficulty in withdrawing their data from the research itself.

The GDPR requires a focus on privacy at the very beginning of protocol development and throughout the lifecycle of the research project. Researchers must address compliance issues at the outset in a proactive manner while keeping in mind the utility constraints of the data for research purposes. The focus should be on how to best strike this balance between individual rights of the data subjects and the larger objectives of the research such as providing benefits to collective groups.

The purpose of the GDPR, Staunton added, is not to hinder research or restrict the sharing of data for research, but to ensure that personal data is processed for research in a manner that safeguards each research participant's privacy. For this, there are a number of derogations and exceptions provided for in the research context. Staunton noted that these safeguards are key when continuing this discussion, while other safeguards beyond the GDPR should be considered. As policymakers introduce new legislation, stakeholders should focus on the impact of the regulations on resources, finances, and costs. Such factors will be instrumental in complying with new data protection standards, which could unduly advantage well-resourced institutions at the expense of smaller projects.

To shade out these points, Claire Gayrel discussed the EDPS's January 2020, Preliminary Opinion on data protection and scientific research, which takes into account new and current trends in scientific research, consent in the context of research, and the increased role of private parties conducting research. EDPS believes there is a need for direction regarding how the GDPR should be understood in relation to scientific research. One trend, Gayrel explained, indicates that stakeholders view informed consent as a cornerstone principle of research ethics, providing for natural, lawful grounds for scientific research. In that sense, informed consent and data subjects' consent were inseparable and indivisible.

EDPS observed, however, that this approach began evolving and became less relevant over time. Researchers began focusing on the online environment that had limited direct contact with participants and using large scale genomic databases shared between multiple projects over long periods of time. These practices have questioned the regulatory idea behind informed consent and given rise to a new trend in medical research. This trend involves the practice of asking data subjects to consent to the further use of the data for research purposes without any additional restrictions on its usage. As Gayrel noted, the emergence of open access consent, where data subjects make their sensitive data available online underscores some of these developments and raises many ethical concerns for data protection going forward.

The EDPS has begun to address new ways to strengthen the conditions for data subjects to give and withdraw consent and question whether data subject's consent as lawful grounds for the posting of data is suitable. For instance, such mechanisms are less suitable when consent is required for an exchange of work or when obtained in vulnerable situations. The European Court of Justice (ECJ) has addressed this question and the requirements for specific consent in *Orange Romania*. In this case, the ECJ elaborated on the conditions for specific consent and also questioned the applicability of such requirements in the context of research where broad consent and increasingly broader consent is becoming the practice.

From these observations, the EDPS found that it was less relevant to consider informed consent and data subject's consent as a single and indivisible requirement. This, Gayrel explained, is why the new trends in research involving open data frameworks raise questions as to the desirability of specific types of consent and the tradeoff of treating consent as either a lawful ground for processing or a safeguard. The EDPS believes that dialogue with the ethics community and research organizations could help clarify these questions and define conditions where consent is applicable.

Claire Gayrel also described the EDPB's set of Guidelines in April 2020, which provide guidance on data processing during the covid-19 pandemic. Guidance was needed after numerous clinical trials, including cross-border clinical trials, commenced very quickly due the pandemic. The Guidelines address issues around consent and transparency, provide guidance on international data transfers, and point to derogations as a possible legal basis for transfers if no adequacy decision is in place. Gayrel argued that consent is not the most suitable ground for processing personal data for many research projects, but that it is valuable when all of its conditions are met. Gayrel noted that the EDPB will provide more guidance on these issues.

In relation to this, Dara Hallinan focused on how consent and data protection laws like the GDPR interact with research on both the policy and academic levels. Hallinan first argued that the basic function and importance of consent is often overlooked; a lacunae that often manifests in the description of consent as a safeguard. He explained that consent is the mechanism by which the data subject realizes the right to information self-determination. By contrast, a safeguard is an effort to channel behavior in order to minimize risk. As he put it, when there is an informational self-determination right in relation to some aspect of research processing, considering consent as a safeguard is a misleading use of the concept of a safeguard and ends up bringing confusion to the unique function of consent and undermining the significance of data protection in terms of the underlying rights.

Moreover, Hallinan raised concerns with treating consent in data protection law as a free standing entity completely detached from other areas of law and ethics and the practical circumstance in which research is conducted. This approach opens up the possibility that data protection could contradict other areas of law and ethics and with the reality of a processing situation. For example, if the criterion of 'freely given' in consent rules differs in clinical trials and data protection law, it could raise more ambiguity as to the meaning of the criterion in both situations. Hallinan argued that even though the Clinical Trials Regulation requires research subjects to give consent to participate in the research, that consent is different from consent to process their personal data under the GDPR (under which, however, other legal bases could be used to process personal data for clinical trials purposes). Hallinan highlighted the need to reconcile the various different approaches to consent in relation to various data processing situations and to align these with the reality of the processing operations in place today.

Finally, Hallinan also touched upon the issue of the degree to which the specifics of whether, how, and when consent as a personal data processing lawful ground for research should be determined by looking solely at data protection law. He stated that the GDPR, as omnibus legislation, provides little normative insight on the interaction between consent and research. He argued that stakeholders should develop a framework to fill the normative gaps in the GDPR by considering when and how consent should be legitimate and necessary in research, taking into account the vast discourse already present in research ethics.

3.2 AI ACCOUNTABILITY AND EXPLAINABILITY IN THE CONTEXT OF COVID-19

One growing area of importance for scientific research based on vast amounts of data is the use of artificial intelligence and machine learning. Henrik Junklewitz explained that due to increasing algorithmic complexity and the sustained need of large amounts of data, many issues can arise that touch upon the rights of data subjects. Junklewitz questioned whether the right of access outlined in Article 15 of the GDPR could guarantee that data subjects have adequate explanation of how researchers utilize algorithms in their research. He pointed out a number of fundamental questions to keep in mind including realizing algorithmic accountability, tackling biases, fairness, and transparency, and determining the circumstances where algorithms can actually provide what is legally required.

As Junklewitz noted, consent may be meaningless when the system being used for the data is not transparent or cannot provide a sufficient explanation to the data subjects. In machine learning development contexts, the lines between research stages, application stages, and production-ready stages can be quite blurry. When developers rush those systems from research stages into automated processing environments, there is a risk that they will not behave correctly, making consent about data use within those systems susceptible to ambiguity.

In the biomedical context, the Covid-19 pandemic has highlighted and exacerbated these issues, especially as recent technologies like contact tracing systems have been rushed to application stages. However, many panelists argued that there is a need for timely access to reliable data to track the spread of the virus and prepare for an effective response, and a need to develop diagnostic vaccines and therapies. The importance of access and the sharing of data to accomplish this end has raised many challenges for reconciling various data protection laws and standards.

Finally, as many panelists noted, Covid-19 is one of the first widespread pandemics in the digital era. While the proliferation of digital tools to respond to this crisis has greatly helped governments around the world, policymakers must also consider how such technologies will impact data protection and privacy rights. In the data protection context, safeguards will help promote trust and must be part of any public health response. Any limitations on these rights that exist have to be necessary, proportionate, timely and transparent.

Henrik Junklewitz also discussed some of the current capabilities and limitations for research during the covid-19 pandemic. Junklewitz pointed out the vulnerabilities within new AI systems for the use of data for covid-19 research and contact tracing. These systems have limitations, and as these increasingly complex systems use large datasets and health data, the limitations and challenges of these systems must be fixed. While these systems are extremely capable and have been trained in specific tasks so they can show superhuman performance, they will not replace complete decision making, especially not in the high risk scenario as a clinical situation, because they have only been trained on specific data and specific situations.

There is also an issue regarding correlation with casualties, as most of these systems make large correlations. In certain scenarios, causal relationships might want to be explored. There is an increasing complexity of models that leads to less transparency and accountability. Junklewitz argued that these applications using existing GDPR models (like DPIA or privacy-by-design applying the principles of Article 5 GDPR) to address algorithmic accountability without necessarily opening the most inscrutable black-boxes must be discussed in the GDPR context and trustworthy systems are needed by introducing proper regulation. He stated a risk based approach is needed, especially as the health domain is a high risk domain.

3.3 SCHREMS II AND INTERNATIONAL DATA TRANSFERS (CROSS-BORDER SHARING OF DATA FOR RESEARCH PURPOSES)

International data flows - or cross-border sharing of data for research purposes - also pose a need for detailed work and research. Dara Hallinan discussed the impacts that the Covid-19 pandemic and the recent Schrems II decision by the Court of Justice of the European Union (CJEU) have had on international data transfers, including those needed for research projects in international collaborations. He noted that one of the most significant problems that stands as a barrier for international data transfers is in fact related to the national security practices of different countries.

Accessing data originating in Europe for national security purposes by the US government in a way that the judges did not consider proportional, was in fact the key concern that the CJEU had in the Schrems II judgment. This judgment invalidated the EU-US Privacy Shield and brought all personal data transfers from Europe to the US under significant legal uncertainty, including those transfers necessary for research purposes in cross-border projects. But there is no consensus among democratic countries about how processing of personal data should relate to national security, there is no consensus on what is proportional or not. Hallinan questioned why differences in approaches to national security have to disrupt all other forms of personal data-based collaboration between states, particularly in relation to scientific research. He suggested that it may be worth considering the degree to which research processing could be treated as a separate matter, apart from national security processing, in considering international transfers. In this regard, perhaps a solution could be specific international agreements for cross-border sharing of data for research purposes, or other types of consensus among countries to allow free flow of data, including personal data, for academic research purposes.

4. Using Sensitive Data in Research to Counter (Hidden) Bias and Discrimination

The second panel discussed the complexities surrounding the definition of sensitive data as well as the rules that should apply to the processing of such data for research purposes.

4.1 DEFINING SENSITIVE DATA: A NEW APPROACH FOR THE BIG DATA ECOSYSTEM

Paul Quinn pointed out that the data processing environment has changed rapidly in the past few years, which has raised new questions regarding the proper definition of sensitive data. In part, the emergence of interconnected data ecosystems has changed the landscape of how practitioners and researchers work with data.

Quinn emphasized how comparisons to personal data provide a good illustration as to the implications of these changes. Many professionals have discussed the complexities of defining personal data in the era of big data, the appropriate scope of data protection laws, and core concepts such as anonymization, pseudonymization and encryption. While these topics have received much attention, there is a lot less literature on how policymakers should define sensitive

data as a general concept. Both the GDPR and previously the EU Directive for Data Protection define sensitive data through a context-based approach. This means that the level of sensitivity of personal data does not depend on the purposes (or the general activity) of the data controller, but on the abstract possibility to infer sensitive information from those personal data.

For instance, the Article 29 Working Party (which later evolved into the EDPB) referred to "health status" to determine whether information could qualify as health data. In the modern big data world, an extraordinary wide range of data can give an indication of health status. Under this approach, information recorded on mobile phones such as walking distance, caloric intake, and screen time could create probabilistic conclusions about current or future health status. This means that many big datasets could contain health data even when it is not intuitive due to increased computing power, artificial intelligence, and interconnectivity.

Quinn noted that these changes could undermine current legal provisions governing sensitive data if such complementary data becomes ubiquitous. The widening scope of sensitive data presents several challenges going forward. On the one hand, a broader concept of sensitive data may make regulation less effective by being overbroad. On the other, as controllers continue to process sensitive data, provisions governing such data are necessary to mitigate harms resulting from the use of the data such as discrimination and stigimization. In addition, controllers may attempt to circumvent using sensitive data altogether by using proxies.

In a paper with Gianclaudio Malgieri, Quinn proposes to shift away from a *context-based* approach to a *purpose-based* understanding of defining sensitive data to strike a balance between these conflictual considerations. Under this approach, controllers would first ask a threshold question of whether the data is reasonably likely to be sensitive. If no, the controller should further determine whether there is an intention to draw sensitive conclusions from the data. Such an approach could be used for other contexts besides health data that give rise to sensitivity such as sexual orientation, immigration status, or other characteristics of vulnerable populations, and would help address both the risk of "inflation" of the concept of sensitive data, and the risk of circumventing the rules of sensitive data through proxies.

4.2 POTENTIAL RISKS OF PRIVACY ENHANCING TECHNIQUES ON MASKING INEQUALITIES IN HEALTH RESEARCH

In addition to challenges around defining sensitive data, privacy enhancing techniques may also raise concerns around the use of data for research purposes. Heng Xu discussed at length the key findings of her recent research which examined the effects of anonymization and differential privacy on health disparity detection.

Generally speaking, the notable tradeoff between privacy and utility can also apply to themes around disparities emerging from the use of sensitive data such as identifiers based on race, gender, geolocation, income, and education. To address privacy problems emerging from the use of this data, researchers apply various privacy protection techniques such as anonymization and differential privacy with the hope that the datasets can be used in a less intrusive manner.

However, Xu's research indicates that the application of these techniques can mask important statistics for vulnerable groups, which makes detecting health disparities in these datasets more difficult. In other words, if health researchers work with datasets that have received privacy

enhancing techniques, the utility of detecting health disparities decreases. For instance, Xu pointed out that noise inserted into population counts in data sets could reduce the accuracy of mortality estimates between different racial groups. This, in turn, could affect the researcher's understanding of the health disparities across different racial groups by making it harder to find trends in the data.

In addition, Xu also emphasized that different privacy enhancing techniques produce different impacts on disparity detection. As a general taxonomy, Xu highlighted that there are two common types of privacy enhancing techniques. First, techniques centered around data removal, such as anonymization, aim to remove parts of the dataset that could potentially identify individuals. The second type, by contrast, refers to mechanisms that center around noise insertion. For instance, differential privacy techniques insert carefully designed noise that blocks the identification of the individual while allowing discovery of certain summary statistics.

In order to measure the different impact of these mechanisms, Xu also examined different disparity recognition techniques across sociology and epidemiology literature. The two most popular techniques involve discovering statistical separation of subpopulations through either the degree of separation as determined by both the mean difference and the standard deviation or the mean difference only. With regard to the standard deviation technique, if researchers detect two or three standard deviations across different sub-populations for a given health-related question, it could indicate a disparity. By contrast, with respect to the technique centered on mean comparison, researchers use real variation to detect disparities by calculating the difference between the mean outcomes for different subpopulations in a dataset.

While these separation techniques can be used to operationalize disparity detection in data sets, the interplay between these detection mechanisms and the different privacy enhancing techniques reveal notable patterns about the effectiveness of privacy enhancing techniques on health disparity detection. Data removal techniques such as anonymization, according to Xu, tends to produce more false positives for disparity recognition, while noise insertion techniques (e.g., differential privacy) rarely produce any false positives.

Given the impact of these techniques on disparity detection, policymakers should consider how legislative mandates to regulate sensitive data may produce unintended consequences. For instance, legislation that incentivizes health providers to broadly apply privacy enhancing techniques could harm the ability of researchers to detect disparity. Rather, regulators should properly assess the different impacts of privacy protection on different subpopulations and disparity detection before crafting mandates.

4.3 HEALTH RESEARCH AND THE COVID-19 PANDEMIC

While recent changes to the data ecosystem have both called into question the definition of sensitive data and raised statistical concerns around health disparity detection in research, the Covid-19 pandemic catalyzed more changes to the intersection between health research and data protection. As Knut Mager explained, Covid-19 was an accelerator for many recent developments concerning new ways of cross-border collaboration in the area of health research. In particular, collaboration between private and public entities on the global scale has been instrumental in creating epidemiological insight and developing vaccines and therapeutic remedies for the virus.

But as Mager pointed out, in order to preserve the flexibility and networking of academic and public/private research, policymakers must provide more legal certainty. In the context of Europe, the potential creation of data spaces underscores this need and should take into account a few considerations. For instance, policymakers, civil society, and other stakeholders should determine the boundaries of self-determination with regard to health data and societal benefits and find ways to ensure that data sharing for research purposes continues to follow democratic processes.

To this end, Mager recommended that Member States make use of the derogations in the GDPR in a coordinated manner to create further legal certainty. Industry and academics should complement this approach by creating best practices and codes of conduct around data sharing for health research. While this has been floated for quite some time, Mager pointed out that many stakeholders are currently working on codes of conduct to help regulate behavior that falls outside the scope of government regulation.

Going forward, codes of conduct must be meaningful in scope and enforceable but also receive strong regulatory oversight. Mager proposed that a European Health Data Institute could ideally create standards and interoperable specifications for data sharing to provide coordination. Indeed, evidence shows that robust cross-border collaboration requires common approaches to datasets that facilitate interoperability and foster educational efforts.

Stakeholders could work within existing initiatives and help scale these initiatives rather than impose a EU-wide standards-setting project. For instance, current collaboration between Estonia, Finland, and Portugal regarding health records could provide a scalable model and a template for best practices. Finally, Mager suggested that policymakers should consider sector specific adequacy.

4.4 INTERACTION BETWEEN THE OECD PRIVACY GUIDELINES AND HEALTH DATA

In addition to EU-specific initiatives, policymakers can also look to frameworks developed by intergovernmental organizations (IGOs) for guidance on how to treat sensitive data with respect to scientific research. For example, issues around cross-border flows of sensitive data formed a large part of the debate around the creation of the Organization for Economic Cooperation and Development's (OECD) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (Privacy Guidelines) in 1980 and its subsequent revision in 2013.

Elettra Ronchi discussed the history of these guidelines and detailed the OECD's ongoing review of whether the guidelines need further revision. She pointed out that the OECD created the Privacy Guidelines to address the twin concerns around the threat of privacy from more intensive use of personal data and the risk to the global economy from restriction to the flow of information. One central tension inherent in striking the balance between these two concerns arose from the scope of sensitive data and the appropriate regulatory framework of the guidelines. Indeed, this tension can be seen in debates around whether the guidelines should be of a general nature or whether they should be structured to deal with different types of data or activities.

As Ronchi highlighted, the experts agreed that it was impossible to identify a set of data that would universally qualify as sensitive across different nations and cultures. Building from this realization, the Privacy Guidelines suggest that different protective measures should apply to different categories of personal data, depending on the data's nature and the context in which controllers collect, store, process or share the information.

In one sense, the OECD Privacy Guidelines centers its treatment of sensitive data around the concept of context, a point Quinn alluded to earlier in the panel. But in another sense, the OECD has begun to acknowledge that the degree of sensitivity will also depend on purpose. For instance, the 2013 revision to the Privacy Guidelines explicitly recommends that controllers take into account the purpose and the context of processing data when deciding the sensitivity of data and any subsequent restrictions on transfer. In addition, the Privacy Guidelines also include an enhanced focus on the role of accountability to foster appropriate governance and institutional arrangements.

Ronchi also pointed out that while the analytical framework for sensitive data is beginning to change, there still remains a number of challenges regarding the Interplay between privacy and the transfer of health data. Despite the potential benefits of data sharing, research indicates that data sharing has not reached its full potential. While the implementation of the OECD Privacy Guidelines across the world has created space for harmonization and interoperability, the multiplicity of global privacy regimes has also produced a lot of uncertainty for governments, businesses and researchers.

In particular, localization measures and different approaches to health data continue to deter cross-border transfers and stifle research collaborations. Indeed, such restrictions have delayed the flow of critical data for a whole range of research including, as Mager highlighted, the area of biomedical research. There have also been reports indicating that uncertainty in the GDPR with respect to implementing derogations has also created difficulties for research between Member States. Therefore, the need to enable greater data access and sharing is paramount. Such a need has been on the agenda of multiple institutional forums including the G20 which has promoted the concept of “Data Free Flow with Trust.”

Finally, Ronchi indicated that policymakers must look beyond mere legal compliance and take a risk-balancing approach between the benefits associated with health sharing data and the risks to privacy. Such an approach requires transparency and understanding many factors including the expectations of individuals, the public perception of health data sharing, and the role of the larger data ecosystem. More governance frameworks for harmonization that take into account these factors could overcome some of the challenges for data sharing. Indeed, echoing Xu, making data more accurate and complete could also reduce existing harms resulting from inaccuracy or incompleteness.

5. Closing keynote - Dr. Wojciech Wiewiórowski, European Data Protection Supervisor

In its January 2020, Preliminary Opinion on data protection and scientific research, discussed above, the EDPS recommends intensifying the dialogue between Data Protection Authorities and ethical review boards for a common understanding of which activities qualify as scientific research, on codes of conduct for scientific research, and on closer alignment between EU research framework programs and data protection standards. Wojciech Wiewiórowski, the European Data Protection Supervisor, aimed to further explain the reasoning behind this opinion, and provided closing remarks on the topics discussed during the panels. Many of the questions discussed

during the Symposium, he noted, differ on which type of science is presented and researched. The Supervisor argued that he does not see this type of conversation happening with historians who now have to process data about human affairs. Overall, however, he expressed that we must return to the overarching questions of what "science" is, what "research" and "researcher" are; as nowadays the "researcher" may be playing another role in society, such as working for private entities to implement the scientific research.

Wiewiórowski also noted that the EDPS stresses that solidarity is part of its overall mission to discuss how data can be used for humankind; and is not just a question about data protection, but also one of information sharing and flows. He argued that the focus should be on trying to protect the human being, not the data, and this opens the doors to privacy tradeoffs. The EDPS attempts to operationalize data sharing and are answering questions from legislatures on the various initiatives.

Further, he explained that the EDPS is cautious about "data altruism" and believes terms like this one will present problems in the long-term, particularly when applied to sensitive data. The EDPS also believes that the concept of "sectoral adequacy" for international data transfers, like a sectoral adequacy only for research purposes, to be quite dangerous. He believes, however, that the data protection community should have a conversation about it. Conversation around research can also apply to other contexts such as workplace privacy and diversity questions. Furthermore, Wiewiórowski was confident on codes of conduct acting as useful internal sectoral rules, but believes that they might prove ineffective as tools for international data transfers (unless evidence shows the opposite).

6. Conclusion

The panelists dissected the various risks and vulnerabilities with respect to data protection in the scientific research context, highlighting the many issues which have been brought to the forefront during the Covid-19 pandemic. The pandemic shed light on issues around consent structures, AI and machine learning systems, sensitive data definitions, privacy enhancing techniques on datasets, the role of international frameworks, and discovered certain risks of using data for research. The panelists provided thought provoking questions, ideas and solutions to these issues. One particular means, discussed in great detail during the Symposium, is the DGA, as it regulates data sharing between private and public sectors, and introduces "data altruism" to grant researchers access to larger datasets. Each issue and solution presented by the panelists, however, highlighted the need to strike a balance between the utility of research and privacy and data protection, and the Symposium overall aimed to determine that middle ground and see how we can get there.

To learn more about FPF in Europe, please visit fpf.org/about/eu.

The Symposium was sponsored by:

TOMTOM BigID

1400 Eye Street, NW, Suite 450
Washington, DC 20005

fpf.org