How the General Data Protection Regulation changes the rules for scientific research

STUDY

Panel for the Future of Science and Technology

EPRS | European Parliamentary Research Service

Scientific Foresight Unit (STOA)

PE 634.447 – July 2019

EN

How the General Data Protection Regulation changes the rules for scientific research

The implementation of the General Data Protection Regulation (GDPR) raises a series of challenges for scientific research, in particular for research that is dependent on data. This study comprehensively investigates the promises and challenges associated with the implementation of the GDPR in the scientific domain, with a special focus on the impact of the new rights and obligations enshrined in the GDPR on the design and conduct of scientific research. Furthermore, the study examines the adequacy of the GDPR's derogations for scientific research in terms of safeguarding scientific freedom and technological progress.

The study also provides policy options that delineate a pathway towards enhancing rather than stifling research, and facilitating privacy-preserving data-driven research under the provisions of the GDPR.

AUTHORS

This study has been conducted by the Health Ethics and Policy Lab, ETH Zurich, at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit within the Directorate-general for Parliamentary Research Services (DG EPRS) of the European Parliament.

The study was led by Prof. Effy Vayena. Under her supervision the following Lab members contributed to the report: M. Lenca, project coordination and scoping review; J. Scheibner, doctrinal analysis; A. Ferretti and F. Gille, scoping review and case studies; J. Amann and J. Sleigh, media analysis; A. Blasimme, review and policy options.

ADMINISTRATOR RESPONSIBLE

Mihalis Kritikos, Scientific Foresight Unit (STOA)
To contact the publisher, please e-mail STOA@ep.europa.eu

Acknowledgments

The authors would like to thank all who have helped with this research in any capacity, in particular the European Data Protection Supervisor's Office for the exchanges with Effy Vayena and the insightful comments they generously provided.

LINGUISTIC VERSION

Original: EN
Manuscript completed in July 2019.

PE634.447
ISBN: 978-92-846-5045-3
doi: 10.2861/17421
QA-04-19-501-EN-N

http://www.europarl.europa.eu/stoa (STOA website)
http://www.eprs.ep.parl.union.eu (intranet)
http://www.europarl.europa.eu/thinktank (internet)
http://epthinktank.eu (blog)

Executive summary

Scope of the study: the present study aims to provide a comprehensive assessment of the expected impact of the General Data Protection Regulation (GDPR) on scientific research in Europe and an early review of the measures taken by European research institutions to facilitate its implementation.

Methods: the study presents findings from a multi-method research approach involving four sequential components: a scoping review of the peer-reviewed scientific literature, a doctrinal analysis of the legal literature, a content media analysis and a case study analysis. Based on this mixed study design, a set of policy options is proposed to maximise the positive impact of the GDPR scientific research while minimising unintended adverse effects.

Results: the study results delineate a diverse, multifaceted and complex impact scenario. Study findings anticipate divergent or even contrasting impacts depending on the domain of science and the type of scientific activity under consideration. While several areas of normative ambiguity and potential concern from the perspective of researchers have emerged, our analysis anticipates that the GDPR will likely enhance a number of aspects of scientific research. These include data security, regulatory clarity regarding processor responsibilities and the transfer of data, research collaborations within the EU, and the autonomy and trust of data subjects. We found that the GDPR is not universally perceived as an enabler of research, due in part to regulatory ambiguities surrounding how it applies to research, and the perception of such applications as being more prohibitive than enabling. The study proposes a set of normative options to help resolve potential ambiguities and promptly address procedural concerns. These include suggestions as to how scientific research can find common ground with the new legal rules on data protection and how the scientific community can prepare for GDPR compliance, with a special focus on delineating regulatory, procedural and educational solutions.

Table of contents

List of tables

List of abbreviations

- AFIRRM: adaptivity, flexibility, inclusivity, reflexivity, responsiveness, and monitoring
- BaiLII: British and Irish Legal Informatics Institute
- BAM: binary alignment maps
- BBMRI: Biobanking and BioMolecular Resources Research Infrastructure
- CJEU: Court of Justice of the European Union
- CNIL: Commission nationale de l'informatique et des libertés
- CTR: Clinical Trials Regulation
- DNA: deoxyribonucleic acid
- DPA: data protection authorities
- DPD: Data Protection Directive
- DPO: data protection officer
- ECHR: European Court of Human Rights
- EDPB: European Data Protection Board
- EDPS: European Data Protection Supervisor
- EEA: European Economic Area
- ERA: European research area
- ERC: Ethics Review Committee
- ERIC: European Research Infrastructure Consortium
- EU: European Union
- FAIR: First Aid for Responsible Data Scientists
- FAQ: frequently asked questions
- GA4GH: Global Alliance for Genomics and Health
- GDPR: General Data Protection Regulation
- GER: Germany
- HIPAA: Health Insurance Portability and Accountability Act
- HRA: Health Research Authority
- HUDOC: human rights documentation
- ICO: Information Commissioner's Office
- INFO: information
- iPSC: induced pluripotent stem cells
- IRB: Institutional Review Board
- ISO: International Standards Organisation
- ISP: internet service provider
- MRC: Medical Research Council

- NAS: National Academy of Sciences
- NHS: National Health Service
- NSA: national security agency
- ORG: organisation
- Q&A: question and answer
- RRI: responsible research and innovation
- SNP: single nucleotide polymorphism
- STR: short tandem repeat
- UCL: University College London
- UK: United Kingdom
- VCF: variant call formats
- WGS: whole genome sequencing
- WordLII: World Legal Informatics Institute

1. Introduction

The recent implementation (25 May 2018) of European Union data privacy rules — the General Data Protection Regulation (GDPR) — has caused significant concern within the research community (Dove, 2018; Mourby et al., 2018; van Veen, 2018). These concerns, as raised by the STOA project call, stem from the possibility that the new data protection rules might prevent innovative research in the EU and, as a result, that the GDPR may act as a barrier to freedom of research. Balancing the conflicting rights to academic freedom and to informational self-determination, however, is one of the core tasks that scientific institutions need to manage. Genetic, biometric and health data are very sensitive types of information, the use and misuse of which has the potential to intimately affect individual subjects. For these reasons, there is a general prohibition on processing sensitive categories of personal data in the GDPR, unless certain conditions are met. Research-based organisations have also expressed concerns regarding the potential risk of fragmentation deriving from the possibility of Member States' derogations. These derogations may establish uneven conditions for researchers and pose challenges for research collaboration between Member States, and globally. There are also questions raised about the potential impact of the new data protection rules on international and global scientific research collaborations, with specific concerns and implications for data sharing.

To address these concerns, in response to the STOA project call 'How GDPR changes the rules for scientific research', this report aims to answer the following questions: What challenges does the GDPR bring to scientific research, particularly biomedical research? What measures, if any, have research institutions put in place to prepare for the GDPR? What can be done to prevent data protection regulation from stifling scientific research in Europe and by European institutions?

Ultimately, this project proposal will contribute to the understanding of how the European research community is coping with the GDPR. The analysis will delineate which implementation pathways are best suited to facilitating innovative scientific research in the EU. Subsequently, the analysis will lead to policy suggestions, findings that will be useful in informing the upcoming European Open Science Cloud and the ninth framework programme for research.

The GDPR and the data-protection landscape in Europe

The first proposal for the current General Data Protection Regulation was released by the European Commission in January 2012. This document and later proposals were voted on by the European Parliament on 12 March 2014. Following this vote, the Council agreed to a common approach on a revised text on 15 June 2015. On 15 December 2015, the European Parliament voted to pass the new data protection rules, which were then published on 4 May 2016. As mentioned in the previous section, although speculative, there has been significant academic commentary about the effects of the GDPR on scientific research. Despite the relatively recent implementation of the GDPR (25 May 2018), the European Union (EU) has a long history regarding data protection law. Before the founding of the EU, many European nations, such as the United Kingdom, France and Germany had detailed judicial

concepts concerning the right to 'informational self-determination'. This legislative history can be differentiated from that in common law countries such as the United States, Canada, Singapore, Australia and New Zealand. In these countries, the right to privacy has evolved piecemeal, either by reference to constitutional protections, legislative instruments or judicial pronouncements (Whitman, 2003).

The importance of data protection law in Europe was thrown into focus in the wake of the Second World War. More specifically, the use of surveillance technology by totalitarian regimes to commit crimes against humanity, along with advances in computer science, prompted consideration of the need for data protection legislation. The federal state of Hesse in Germany was responsible for passing the first data protection legislation in 1971. Soon after, in 1973, Sweden and Germany implemented national data protection legislation. The first supranational European regime, the Data Protection Directive of 1995, implemented many of the rights and legal obligations that now exist under the GDPR. The recognition of privacy is also built into EU supranational law, with Article 8 of the European Convention on Human Rights recognising a right to private and family life. Finally, there are many other European Union legislative instruments regulating clinical trials, human and tissue cell transfer and storage and database protection.

Legislative instruments, which are designed to protect specific rights and freedoms, have the potential to influence scientific research positively or negatively, and can clash with emerging policy perspectives. One key policy objective of scientific research in recent times has been the concept of open access to scientific data. Open access advocates argue that scientific data can be most effectively used and reused where it is not subject to any legal or technological limitations. This perspective is reflected in official policy documents from the American National Academy of Sciences (NAS). Initially, the NAS limited this data sharing ambit to meteorological, environmental and satellite data. However, the NAS's most recent Consensus Study Report, published in 2018 suggests that open access principles should be built into all scientific research projects by design. This perspective is also reflected in sections of the Treaty of the Functioning of the European Union, which establishes the European research area (Article 179 TFEU):

> 1. The Union shall have the objective of strengthening its scientific and technological bases by **achieving a European research area in which researchers, scientific knowledge and technology circulate freely**, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties.
>
> 2. For this purpose the Union shall, throughout the Union, encourage undertakings, including small and medium-sized undertakings, research centres and universities in their research and technological development activities of high quality; **it shall support their efforts to cooperate with one another, aiming, notably, at permitting researchers to cooperate freely across borders and at enabling undertakings to exploit the internal market potential to the full**, in particular through the opening-up of national public contracts, the **definition of common**

> **standards and the removal of legal and fiscal obstacles to that cooperation** *[emphasis added]*.

Therefore, by decreasing the legal and technical obligations associated with data transfer, regulators hope to decrease the overall economic cost of open data transfer. These open access principles have been somewhat controversially extended to biomedical and public health research. From one perspective, it can be argued that there is an 'ethical and scientific imperative' to share biomedical data for research purposes. From a scientific perspective, this can reduce the considerable cost of conducting repeated clinical trials, whilst the ethical imperative pertains to the fact that the majority of research subjects face harm and inconvenience without compensation. Accordingly, these participants should be rewarded by their data being made openly available to encourage broad contributions to the public good. Nevertheless, complete open access to scientific data can be fettered by other legal rights and freedoms. Under the previous data protection directive, the Court of Justice of the European Union (CJEU) decided key cases that have determined the scope of data protection law in Europe. The CJEU has also decided key cases with respect to the scope of other European instruments that can affect the sharing of data. Likewise, the European Court of Human Rights (ECHR) has created significant jurisprudence with respect to the related but separate right of privacy. In each of these areas of jurisprudence, there is very limited authority with respect to scientific research *per se*, however, the regulatory instruments and authorities that do exist can potentially indicate how European and national courts might reconcile these rights with the need to encourage open and unencumbered scientific research.

2. Methodology and study objectives

The present study has a fourfold objective (see Table 1):

1. To investigate the impact of the new rights and obligations of GDPR for scientific research and to provide an early impact assessment of the major challenges and opportunities that GDPR poses to scientific research, particularly biomedical research.
2. To identify measures that research institutions are putting in place to prepare for or cope with GDPR requirements and assess to what extent GDPR exceptions for science are sufficient.
3. To monitor media-driven public perceptions associated with GDPR in relation to scientific research.
4. To provide policy options about how scientific research can find common ground with the new legal rules on data protection and how the scientific community can prepare for GDPR compliance, with a special focus on regulatory, procedural and educational solutions.

The following study components and associated methodologies were selected to achieve the afore-listed study objectives:

- Scoping review of the relevant scientific literature and grey literature, including articles published in peer-reviewed scientific journals or conference proceedings, surveys, books, science magazines and also papers published as reports or pre-print papers in scientific repositories.
- Case studies: Purposive review of webpages of European universities or other research institutions.
- Doctrinal analysis: Comprehensive analysis of the legal doctrine including rulings by tribunals that meet their respective jurisdictions' rules to be cited as precedent.
- Media analysis: Content analysis of media documents including newspapers, magazines, yearbooks, etc.

Table 1 - Overview of objectives, methods and outcomes

Objective		Methods	Outcome
1.	To investigate the impact of the new rights and obligations of the GDPR for scientific research.	Doctrinal legal analysis on GDPR exemptions for scientific research Scoping review of primary and secondary literature	Ex ante impact assessment of the major challenges and opportunities that the GDPR poses to scientific research
2.	To identify measures that research institutions are putting in place to prepare for or cope with the GDPR requirements and assess to what extent the GDPR exceptions for science are sufficient.	Case studies Scoping review of primary and secondary literature Doctrinal legal analysis on GDPR	Overview of coping strategies in the research domain Assessment of the

		exemptions for scientific research	GDPR exemptions for science
3.	To monitor media-driven public perceptions associated with the GDPR in relation to scientific research.	Media content analysis	Synthesis of media-driven public perceptions
4.	To provide policy recommendations about how scientific research can find common ground with the new legal rules on data protection and how the scientific community can prepare for GDPR compliance, with a special focus on regulatory, procedural and educational solutions.	Empirically informed normative ethical analysis	Recommendations and option brief

These study components were selected based on the following methodological considerations and the tight schedule of the STOA call. First, the short time frame occurred between the date of implementation of GDPR and the completion of this report prevented any post hoc assessment. In particular, significant impacts can only be observed and measured within larger periods of time than this report allowed for. Therefore, we identified methodological strategies that could provide proxy information, which could be used in turn to generate plausible predictions and thereby enable a comprehensive ex ante assessment. The four methodologies were selected and combined under the assumption that each of them could provide relevant proxy information about the perceived impact of GDPR on scientific research. Each methodology could therefore provide a different, though complementary, perspective. Firstly, the scoping review method enabled us to monitor and review the expected impact of GDPR from the perspective of scientific researchers. Secondly, the doctrinal analysis examined the impact of the GDPR from the perspective of the legal doctrine and jurisprudence. Finally, the media analysis examined perceptions about the impact of the GDPR from the perspective of media outlets.

3. Scoping review

3.1 Design and methods

A scoping review of the relevant primary and secondary scientific literature was conducted. Six databases (Web of Science, Pubmed, IEEE Xplore, Scopus, Europe PMC and ProQuest) were searched to retrieve eligible publications. The scoping review is a review method aimed at synthesising research evidence and mapping the existing literature in a certain field of interest (Pham et al., 2014). Unlike a systematic review, scoping review methods are considered particularly useful when the topic has not yet been extensively reviewed or is of a complex or heterogeneous nature (Arksey & O'Malley, 2005; Pham et al., 2014).
Based on a successfully completed pilot-test, the following search strategy was deployed (for a visual organisation see Table 2):

> ((((((science[Title/Abstract]) OR research[Title/Abstract]) OR innovation[Title/Abstract]) OR researcher*[Title/Abstract]) OR scientist*[Title/Abstract]) OR science[MeSH Terms]) OR research[MeSH Terms]) OR innovation[MeSH Terms] AND ((GDPR[Title/Abstract]) OR General Data Protection Regulation[Title/Abstract]) OR Data protection[Title/Abstract]) OR Data protection directive AND ((((EU) OR Europe) OR European Union) OR Europe[MeSH Terms]) OR European[MeSH Terms] Filters: Publication date from 2015/05/25 AND (((((((((challenge*[Title/Abstract]) OR measure*[Title/Abstract]) OR impact*[Title/Abstract]) OR effect*[Title/Abstract]) OR preparation[Title/Abstract]) OR compliance[Title/Abstract]) OR impact*[MeSH Terms]) OR compliance[MeSH Terms]) OR governance[Title/Abstract]) OR public interest) OR implication*[Title/Abstract]) OR exemption*[Title/Abstract]) Filters: Publication date from 2015/05/25.

Query logic was modified to adapt to the language used by each engine or database. The timeframe of the search was defined as two years prior to and five months after GDPR's implementation date (May 25, 2018).

Database screening identified 1127 entries. Additional unstructured search and reference chaining retrieved a further 69 entries. A total of 1196 entries were imported into the Endnote literature manager software. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (http://prisma-statement.org/), three phases of filtering were performed independently by four researchers to minimise subjective bias. First, duplicates were removed both automatically, using the Endnote tool for duplicate detection, and manually, based on abstract screening conducted independently by the researchers. Second, eligibility assessment was performed independently by three groups of researchers on the remaining articles through title-abstract screening and, subsequently, full text screening. Diverging inclusion choices between the three groups of reviewers or

between the two reviewers within the same group were discussed in plenary with documented justification. This processed continued within the entire team until consensus was reached.

Based on the inclusion/exclusion criteria, studies included in the final synthesis had the following characteristics: (i) original peer-review articles, book chapters or conference proceedings; (ii) written in English, German, Italian, French, Spanish or Greek (languages spoken by the research team); (iii) published between May 25th, 2016 and October 26, 2018; and (iv) focused on an either prospective or retrospective impact assessment of the implications of GDPR for scientific research. Reviews, letters to the editors, business reports and dissertations were not included. Following the recommendations to enhance scoping study methodology delineated by Levac et al. (Levac, Colquhoun, & O'Brien, 2010), a data-charting form was collectively developed by our research team. This form was then used to determine which variables to extract from the review data.

Data analysis was performed both through numerical summaries and a thematic analysis. In the latter analysis, recurrent thematic patterns were identified through full-text screening and subsequent software-assisted coding. The coding phases were independently performed by four researchers. Once conceptually stable thematic patterns emerged from the codes, they were grouped together into a system of themes and subthemes. All entries were checked anew through an automated text search for the presence of the emerging themes. Following Braun and Clarke (Braun & Clarke, 2006), codes that did not seem to fit into any main theme were temporarily housed in a 'miscellaneous' group. These codes were subsequently either clustered into a new theme or reallocated to an existing thematic group after consultation. Internal consultation was iteratively performed among all members of our research team to integrate and validate our findings.

Table 2 - Logic grid for scoping review with keywords and terms or subject headings

Population	Intervention	Location	Outcome measure
AND	AND	AND	AND
Science	GDPR	EU	Challenge*
Research	General Data Protection Regulation	Europe	Measure*
Innovation	Data protection	European Union	Impact*
Researcher*	Data protection directive	European	Effect*
Scientist*		Europe (mh)	Preparation
Science (mh)		European (mh)	Compliance (mh)
Research (mh)			Impact (mh)
Innovation (mh)			Compliance (mh)
			Governance
			Public interest
			Implication*
			Exemption*

3.2 Presentation of the findings

Based on our inclusion/exclusion criteria, forty-five studies passed the three phases of filtering and were included into the final synthesis.

Data breakdown by year of publication shows a significant increase in the number of contributions on this topic over time. Four studies were published in 2016, seventeen in 2017 and the remaining twenty-four in 2018. This linear increase indicates a need to monitor this rapidly growing corpus of scientific literature and replicate the present synthesis at future intervals of time.

Data breakdown by study methodology indicates that no study to date has assessed the impact of the GDPR on scientific research using empirical methods. We hypothesise that this under-representation of empirical investigations is caused by the short time lapse between GDPR's implementation date and the date our literature review was conducted. Empirical studies require time to be designed and conducted, and furthermore, significant impacts of a certain intervention can only be observed and measured after a sufficient period time has passed since the introduction of that intervention. While empirical studies to assess the impact of GDPR on research are lacking, empirical approaches are being developed to facilitate GDPR compliance in the research domain. One notable example is Bialke et al. (2018), who have developed a digital consent management service tool which could be used in clinical and epidemiological studies. This tool is expected to be particularly useful in the context of multi-centre research initiatives (Bialke et al., 2018).

The body of literature included in the final synthesis primarily used the following theoretical methodologies: doctrinal analysis, comparative legal analysis, simulated scenario analysis, and conceptual analysis. A few studies were opinion/perspective articles or editorials. This finding is consistent with the fact that the large majority of the reviewed studies were ex ante projections and proactive evaluations of the impact of the GDPR.

The following recurrent themes emerged from our inductively conducted descriptive thematic analysis:

Domain of application. While a few studies discussed the implications of the GDPR on scientific research in general, most studies focused on its specific implications for biomedical research. Within this domain, the impact of the GDPR on genomics and public health research as well as research using biobanks was of particular concern. For example, Dias (2017) assessed the appropriateness of GDPR in the context of research involving genetic data, especially pharmacogenomics research. Ohmann et al. (2017) focused on the impact of the GDPR on clinical trials to reuse of data from clinical trials for future scientific research. Ohmann et al. underline the importance of consent to use data outside the protocol of the clinical trial and the right to withdraw that consent at any time. Furthermore, they consider mechanisms to review whether secondary analyses are appropriate and ethical (Ohmann et al., 2017). A focus on clinical trials, this time in conjunction with observational research, is also noted by van Veen (2018), who compares the consent requirements under the GDPR with those under the Clinical Trials Regulation. In particular, van Veen notes that consent requirements might change depending on

whether research relates to an observational study or a clinical trial. The article also compares the GDPR with an international regulatory instrument of the World Medical Association, the Declaration of Taipei. Although the Declaration required explicit consent, even for registries of medical data, there was a recent democratisation of the process which resulted in a change in the consent requirements (van Veen, 2018). de Leucona and Villalobos-Quesada (2018) focused on the impact of GDPR on research involving biobanks and human databases. In their view, 'the news brought about by this new act should further structure the European Research Area and balance the necessary investments to be planned for reaching legal compliance' (p. 294) (de Lecuona & Villalobos-Quesada, 2018). A similar focus on biobanks is evident in Morrison et al. (2017), who analysed, in particular, biobanks involving human induced pluripotent stem cells (iPSC) and their associated genetic and clinical information (Morrison et al., 2017). Finally, Kostkova (2018) has focused on disease surveillance research for public health (Kostkova, 2018).

Outside the health-related domain, a significantly smaller number of studies investigated the implications of GDPR for areas of scientific research such as disaster and crisis management research (Watson & Rodrigues, 2018), cyber threat intelligence (Sullivan & Burger, 2017), and digital networks (Weichert, 2018).

Risks/Benefits. The reviewed studies predict a number of risks and benefits associated with the implementation of the GDPR in science.

Expected risks include the possibility that the GDPR might place an excessive burden on researchers and research participants, increase the time needed to obtain IRB/ERC approval with consequent delays in project development, undermine attempts to develop a dynamic consent system, and potentially result in negative consequences for epidemiologic research caused by the GDPR's requirement to treat pseudonymised data as personal data.

The risk of excessive burden on researchers and research participants was emphasised by Mecenaite et al. (2017) in the context of research involving minors. According to their analysis, GDPR places an excessive burden on parents and children to make informed decisions about their personal data processing in the complex technology and data-driven environment. In their view, shifting the responsibility from parents to data controllers is a desirable strategy to better respond to the children's needs and expectations. Similarly, Morrison et al. (2017) have argued that GDPR is likely to increase the bureaucratic burden of sharing, at least on a large scale, human-derived iPSC cells and data because it will require 'additional administrative support such as the appointment of specialist Data Protection Officers' (Morrison et al., 2017).

Another area of concern identified in the literature surrounds the impact of the GDPR on research ethics. Authors expressed concerns about the impact of the GDPR on research ethics instruments such as informed consent and independent ethics review. For example, Simell et al. (2018) have identified the validity of old (e.g. pre-GDPR) consent as a possible threat to research. Specifically, if informed consent is used as a lawful basis for research, it must be assessed for validity. In addition, they postulate that the time taken to obtain ethics approval may also increase the delays for the completion of ethics review.

The possible negative implications of GDPR on the development of dynamic consent are discussed by van Veen (2018) in the context of observational research and clinical trials. Van Veen concludes that the purpose of the GDPR is to give data subjects more control over their data. However, patient rights might not be adequately balanced against the need for different forms of consent. In particular, van Veen argue that GDPR might undermine attempts to develop a dynamic consent system. The scenario may arise where a patient provides dynamic consent but then stops responding to notifications (van Veen, 2018). However, it remains unclear which aspects of GDPR might undermine the development of dynamic consent and how, leaving the claim largely unsubstantiated.

Timmers et al. (2018) identify an inherent tension between critical care research and data protection according to the GDPR. They observe that it is not possible to ask for the patients' informed consent to be enrolled in observational research at the point of admission to the hospital. Nonetheless, they emphasise that informed consent is the baseline to be enrolled in research with personal data, and therefore, they suggest that reliance on the GDPR research exceptions may be necessary for clinical care research (Timmers, Van Veen, Maas, & Kompanje, 2018). Shabani and Borry (2018) identify the shift towards treating pseudonymised data as personal data that requires protection as the greatest threat to research in the GDPR. This approach diverges from current approaches in epidemiology studies, for example, where coded data is heavily used on the assumption that it is non-identifiable. Changes to this definition, therefore, may significantly increase the involvement of bureaucracy that might be attached to the use of such data. Further, the authors argue that GDPR itself does not adequately define whether pseudonymised data will constitute personal data (Shabani and Borry, 2018), and in cases of national derogations, this lack of definition may affect cross border data sharing.

Expected benefits include increased cross-national harmonisation of data protection provisions with consequent facilitation of cross-national research collaborations, reduced occurrence of data breaches due to stricter and more proactive data privacy and security requirements as well as the possibility that increased control by data subjects might result in increased public trust.

Ho (2017) has observed that GDPR adopts a more research friendly approach compared to the previous Directive by including several derogations for consent and processing of data for secondary purposes (Ho, 2017). Dias (2017) evaluates GDPR as 'appropriate' in enhancing scientific research using genetic data, as the rules seem effective in building trust without undermining pharmacogenomics progress and its benefits to society. According to their analysis, the GDPR contributes to the progress of personalised medicine (enabled by pharmacogenomics research) and the GDPR provisions are adequate to boost the processing of genetic data for scientific research purposes under proper safeguards to support privacy and data protection, and thus is favourable both to the data subject and the public interest (Dias, 2017). The potential positive impact on scientific research of the GDPR has been further characterised by Chassang (2017) who concluded that the GDPR 'provides the equilibrium between the necessity of effectively protecting data subjects' rights in a digitalised and globalised world while allowing the processing of personal data, including sensitive data, for scientific research' (Chassang, 2017).

Measures taken and measures to take: Data analysis suggests that very few measures have been reportedly taken by the scientific community in the time period under investigation to ensure a positive impact of GDPR on scientific research. However, several studies indicate, in a prescriptive manner, which measures should be taken in order to achieve this goal. These prescriptive suggestions can be grouped in three main thematic families: technical, epistemic and governance-related solutions.

Technical solutions prescribe the enhancement of the fundamental digital infrastructure that enables data-sharing in the scientific domain. These include installing robust data management practices, developing researcher-friendly user software interfaces for GDPR compliance, and rendering algorithms used in research more amenable to ex post and ex ante inspection. For example, Morrison et al. (2017) have argued that a move 'toward digital tools for consent and engagement may offset some of the administrative burden of sustained interaction' (Morrison et al., 2017). Townend (2018) has proposed the efforts that have been reached with the GAG4H, the FAIR principles, the Personal Data Train and the blockchain as possible technical solutions. Goodman et al. have looked at the potential of machine learning algorithms and emphasised that research is underway in pursuit of rendering algorithms more amenable to ex post and ex ante inspection a number of recent studies have attempted to tackle the issue of discrimination within algorithms by introducing tools to both identify and rectify cases of unwanted bias (Goodman & Flaxman, 2017).

Epistemic solutions prescribe the promotion of data protection literacy and knowledge within the scientific community. These include the organisation of educational activities, specific training sessions and increasing the researchers' familiarity with best practices for data collection and handling (see details in section 5). For example, de Leucona & Villalobos-Quesada suggest that scientific education and public engagement are central to the creation and application of knowledge as well as to responsible research and innovation (de Lecuona & Villalobos-Quesada, 2018). Similarly, Marjanovic et al. identify 'good training and staff development programs' as complementary tools to both technical systems and governance in order to achieve GDPR compliance and maximize the impact of the GDPR for medical data-sharing (Marjanovic, Ghiga, Yang, & Knack, 2018).

Governance-related solutions encompass prescribed amendments to the governance architecture of scientific research. These include promoting the use of decentralised and/or citizen-own data sharing platforms, setting up more accountable scientific governance frameworks, enhancing ethics review mechanisms and establishing data handling best practices.

Governance-related solutions include both soft and hard-law measures at both the national and supranational level. These can be grouped in two main thematic families: national derogations and ethics review policies.

While the GDPR attempts to harmonise legal provisions to ensure adapted data protection in research, the field still remains widely regulated at the national level through national derogations. For example, between some EU member states there are divergences regarding the legal grounds for processing of

personal data without consent. According to Ho (2017), several challenges remain as the scope of scientific exemptions is still unclear, and the rules adopted by EU Member States have yet to be harmonised (Ho, 2017). Similarly, Shabani and Borry (201) have pointed out that how research exemption for research conducted in the public interest is derogated at the national level depends on the will of the signatories. They hypothesised that allowing member States to set further limitations on processing could undermine harmonisation of data protection within the EU (Shabani & Borry, 2018).

Nonetheless, Townend et al. (2018) have argued that despite the difficulties in national harmonisation of GDPR implementation, it is not impossible to achieve a harmonised framework for data sharing. As proposed by Glinos (2018), when concerting these harmonisation efforts, 'EU states thus must be vigilant that discretion given to them by the GDPR does not undermine interoperability'. This analysis emphasises that legislation is already adopted or in preparation in some EU states, such as the German Data Protection Amendment Act or the UK Data Protection Bill. They also underscore that 'interoperability is equally important at the global level; rules for personal data in research in one country may affect use of data from another GDPR is a landmark, but sharing of personal data for research across borders on a global scale will remain a technical, legal, and governance challenge—and opportunity—for the global science community' (Glinos, 2018).

From the perspective of research ethics policy, Ho has called in favour of setting up a more accountable governance framework that can work with existing ethics review mechanisms to provide oversight for all research, whether private or publicly funded (Ho, 2017). An important theme in this domain is the use of data without consent, in cases where such use is justified on the grounds of public interest. For example, Quinn (2017) recommends an exploration of whether the public interest requirement can be used to justify reuse of material without consent (Quinn, P., 2017). In particular, the author notes that the anonymisation of data required for compliance may render data useless. This loss of data is particularly concerning in light of 'waste data' discussions, where the inability to reuse data completely may necessitate the conducting of further clinical trials.

Some authors proposed to integrate multiple governance-related, technical and epistemic solutions as part of a comprehensive scheme. According to de Leucona and Villalobos-Quesada (2018), for example, GDPR will require researchers, developers and their institutions to integrate 'the concept of Responsible Research and Innovation (RRI), where governance, ethics, open access, public engagement, gender equality and scientific education must be central to the creation and application of knowledge' (p.296). Similarly, Sonja et al. (2018) observe that the sharing of health data could be improved if there was more focus on their collective collaboration and coordination. In particular, the report, which was based on a literature review and interviews, identifies the need for more collaboration between the health sector and other sectors, such as social care. In addition, technical systems to enable data sharing need to be accompanied by effective governance and management, systems interoperability, good training and staff development programmes, and clear guidelines and regulations (Marjanovic et al., 2018).

Open challenges. Data analysis revealed several areas of uncertainty currently affecting the implementation of the GDPR within the scientific domain. These include the global fragmentation of

data protection laws (including intra-EU national derogations to the GDPR) and the possibly negative impact on cross-border data linkage, the alleged ambiguity of the scope for research data without consent for public interest, the difficult applicability of strict anonymisation rules to biomedical research, as well as uncertainty around what constitutes 'scientific research' and 'best practice' in data handling.

Furthermore, authors identified additional areas of uncertainty related to the definition of data categories contained in the GDPR, especially genetic data. Sariyar et al. observed that although the GDPR frequently refers to genetic data, there is uncertainty as to what form of genetic data it applies to, as generally only a fraction of DNA is relevant for healthcare and research (Sariyar, Suhr, & Schlünder, 2017). The authors criticise the definition of genetic data in the GDPR as not adequately reflecting the 'exceptional nature' of genetic data, hence a differentiation between different categories of sensitive data might be required. This expansion is particularly pertinent for whole genome sequencing (WGS) data (and to a lesser extent methylation data) due to the potential for many more phenotypic inferences than non-genetic data. However, without further clarification it is uncertain how this definition might be used in practice and what impact it would have.

Hordern et al. (2016) make a similar observation regarding the alleged broad interpretation of the term 'health data' in the EU Data Protection Directive 95/46. In their view, the term, in most circumstances, was not connected to the provision of healthcare, requiring organisations to obtain explicit consent from individuals for its collection and use (Hordern, 2016). Therefore, they argue that one important test for the GDPR is to assess whether it clarified the interpretation of 'health data' and solved additional data protection compliance issues, for example, by identifying who is a controller, ensuring transparency, using health data for research purposes and keeping health data secure.

Townend et al. (2018) illustrate additional conceptual uncertainty inherent in the GDPR in the context of genomic research. In particular, Recital 33 suggests that broad consent is a necessity for medical research. However, this broad right is not guaranteed by Articles 4(11), 6(1)(a), (7) or (8), which indicate a broad reading of consent is required. With respect to further processing, Articles 5(1)(b) and 6(4) set out the principles with respect to further processing of data, but will not be enough, for example, where there are concurrent laws regarding medical confidentiality. Effectively, this requires an acceptable legislative provision to permit ongoing research. Thirdly, different jurisdictions have different limitations on what amounts to deidentification. The three areas that the author identifies as lacking consistency with respect to data protection are informed consent and anonymisation, compatible processing and reidentification of data (Townend et al., 2018). In their view, the GDPR has left it unclear as to what level of technical anonymisation is required, and what exceptions apply to research without the informed consent of the patient. Secondly, there is uncertainty as to the extent to which data which may or not have been consented for may be used for a related purpose. Thirdly, it is not clear from the GDPR whether data will be de-identifiable if it is accessed outside the data set where it is stored.

According to Reichel, while GDPR well-defines national and supranational principles on the transfer of data in the EU, it provides less certainty with respect to data transfer outside the EU (Reichel, 2017). The EU-US Safe Harbour regulation was overruled by the Court of Justice of the European Union (CJEU) in

the Schrems case. According to the CJEU, because some American companies such as Facebook were making the personal data of E.U. citizens available to U.S. government agencies, e.g. the National Security Agency (NSA), the fundamental privacy interests of E.U. citizens were not being protected. As Rothstein et al. have noted, as a result of this court decision 'there has been considerable speculation about what, if any, effect the case has on international collaboration in health research, in particular, the sharing of personal data by E.U. researchers with their U.S. colleagues' (Rothstein, 2016). This might create areas of uncertainty regarding the sharing of data within international collaboration consortia and supra-national research initiatives. Morrison et al. (2017) concur that 'the issue of transfer of data outside the EU is potentially the most challenging development' (Morrison et al., 2017).

Shabani & Borry (2018) note that there is also uncertainty as to what constitutes best practice for pseudonymisation, which, in turn, has an impact on how to share deidentified genetic data. They also argue that the GDPR might create a space for the processing of personal data that extends beyond the binary approach of consent or anonymization (Shabani & Borry, 2018).

4. Doctrinal analysis

This section of the report contains a legal analysis the GDPR to determine how it may impact scientific research in the European Union (EU). The GDPR is divided into a number of different chapters, which each assess different aspects of European data protection law. The first Chapter defines the territorial scope of the GDPR as applying to the processing of personal data of EU data subjects by entities established in the EU.[1] The second Chapter defines the principles for the processing of personal data. First, personal data must be processed lawfully, fairly and in a transparent manner (the 'lawfulness, fairness and transparency principle'). Second, personal data may only be collected for specified, explicit and legitimate purposes and not further processed for incompatible purposes (the 'purpose limitation principle'). Third, personal data must be adequate, relevant and limited to what is necessary to achieve processing (the 'data minimisation principle'). Fourth, data must be accurate and kept up to date (the 'accuracy principle'). Fifth, personal data that permits identification must be held for no longer than required (the 'storage limitation principle'). Finally, personal data must be securely stored to prevent breach (the 'integrity and confidentiality principle').[2] This Chapter also provides the grounds upon which personal data[3] and sensitive categories of personal data[4] may be lawfully processed. In particular, this Chapter emphasises the default ground for processing is the free, informed and express consent of a data subject.[5] The third Chapter then addresses the rights of the data subject. These include the right to information on processing,[6] to access data,[7] to rectify and erase data,[8] to restrict processing,[9] and to transfer their data.[10] The third Chapter also includes both the right to object to processing and the right to opt out of automated decision making.[11] The fourth Chapter addresses the authorities of controllers and processors. These include technical requirements, such as the need for data protection by design and default,[12] recording processing activities,[13] and security of processing.[14] These obligations also include organisational requirements, such as joint controller responsibilities,[15] notification of breaches,[16] conducting data protection impact assessments[17] and appointing data protection officers. The fifth Chapter establishes the grounds upon which personal data may be transferred to third party countries. The sixth and seventh Chapters then define the authority of supervisory authorities, including the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS). The eight Chapter sets out the remedies and penalties associated with the breach of the GDPR. Finally, the ninth

1 Regulation 2016/679, Article 3(1) (defines personal data); Article 3(2) (defines processing).
2 Ibid, Article 5(1)(a); Article 5(1)(b); Article 5(1)(c); Article 5(1)(d); Article 5(1)(e); Article 5(1)(f).
3 Ibid, Article 6.
4 Ibid, Article 9.
5 Ibid, Article 6(1)(a); Article 9(2)(a).
6 Ibid, Article 13 and 14.
7 Ibid, Article 15.
8 Ibid, Articles 16 and 17.
9 Ibid, Articles 18.
10 Ibid, Article 20.
11 Ibid, Articles 21 and 22.
12 Ibid, Article 25.
13 Ibid, Article 30.
14 Ibid, Article 32.
15 Ibid, Article 26.
16 Ibid, Article 33 and 34.
17 Ibid, Article 35.

Chapter sets out provisions relating to processing in specific circumstances, including for research purposes.

There are a number of competing policy considerations that must be balanced against one another. From one perspective, it is argued that there is an 'ethical and scientific imperative' to share personal data for research purposes. The ethical perspective pertains to the fact that the majority of research subjects face inconvenience and some risk without compensation for participation. Accordingly, these research subjects should be rewarded by their data being made openly available to encourage broad contributions to the public good. The scientific perspective springs from reducing the considerable cost of conducting repeated experiments. Further, the existence of data protection legislation may encourage the more transparent and open collection and processing of data (Sánchez & Sarría-Santamera, 2019). By decreasing the legal and technical obligations associated with data transfer, regulators hope to decrease the overall economic cost of, and consequently encourage, open data policies (Reichman & Okediji, 2011). More recently, biomedical, public health and sociology researchers have also been encouraged to engage in open access research. Nevertheless, complete open access to scientific data can be fettered by other legal rights and freedoms, including data protection (Dulong de Rosnay & Janssen, 2014). Under the previous Data Protection Directive, the Court of Justice of the European Union (CJEU) has decided key cases that have determined the scope of EU data protection law. The CJEU has also decided key cases with respect to the scope of other EU instruments that can affect the sharing of data. Likewise, the European Court of Human Rights (ECHR) has created significant jurisprudence with respect to the related but separate right of privacy. In each of these areas of jurisprudence, there is very limited authority with respect to scientific research *per se*. However, the regulatory instruments and authorities that do exist can potentially indicate how European and national courts might reconcile these rights with the need to encourage open and unencumbered scientific research.

This section of the report addresses the three research questions identified in the introduction through the lens of a legal analysis, and is split into two subsections. The first subsection describes a reproducible methodology for searching for case law and research articles that can help define the boundaries of rights under the GDPR as they apply to scientific research. It then purposively interprets each of the relevant articles of the GDPR sequentially as they apply to scientific research. It also uses the relevant case law search results to inform the interpretation of each section. The next subsection then analyses these results and considers how they apply to scientific research. First, it considers how the research exception within the GDPR operates and what types of research activities it covers. Secondly, it considers the impact of this exception on different parts of the GDPR, including direct and derogated exceptions. Thirdly, it then considers types of research to which the GDPR might not apply to, because they involve the use of either non-personal or anonymised data. Fourthly, it then addresses the sections of the GDPR that do not have scientific research exceptions, and which scientific researchers must therefore comply with. Finally, it addresses how other EU directives may influence scientific research or conflict with the GDPR.

4.1 Methodology

Case search strategy

As for the PRISMA guided systematic review defined in the first chapter, a number of search strings were used to identify the relevant cases included in this section. The three databases that provide open access to European case law are the BaiLII/WorldLII, EUR-LEX, and HUDOC databases. These databases index both European Court of Justice legislation and European Court of Human Rights legislation. In addition, BaiLII indexes national legislation (in the United Kingdom and Ireland). The next stage of developing a search strategy was to identify the exclusion criteria for the search strategy. The first exclusion criteria identified was the starting date for case law searching. As discussed in the introduction, the Data Protection Directive was passed by the European Parliament in 1995 and represents the first systematic attempt to codification of data protection law in Europe. Accordingly, 1995 was used as the earliest date for any case law search, providing approximately 24 years of precedent to examine.

The next stage of the exclusion process was to determine whether a case law search should be restricted to particular courts. The choice to include or exclude national courts from the search was complicated by (at the time of writing) the impending exit of the United Kingdom (UK) from the EU. The UK is a major source of both case law and case studies regarding compliance with European data protection law. This inference can be drawn from the systematic review in the first chapter, given the relatively large number of health data protection case studies in the UK. In particular, the *Data Protection Act 2018*[18] provides a useful comparison between European data protection law and how it has been interpreted by national courts. Both versions of the UK Act introduce the principles on data protection from the Directive and the GDPR verbatim. However, the English and Wales Court of Appeals has previously held that the data processing principles do not impose an absolute and unqualified obligation on data processors. Further, the UK law deviates from the European law with respect to consent requirements for medical research and public health matters significantly. Finally, once the UK leaves the EU, UK data protection law may continue to deviate from the standards set by the EU regulation (Taylor, Wallace, & Prictor, 2018).

Excluding courts from the scope of the search is also complicated by the different characteristics of legal reasoning in different European member states. In particular, a key element of legal reasoning involves the use of precedent, or a prior decision from another court, that might be used to solve the current case. In common law countries, courts are strongly bound by the doctrine of *stare decisis,* where previous decisions are binding on future courts. Decisions from the European Court of Human Rights and the European Court of Justice both feature a form of *stare decisis* or precedential authority. Further, the case of *Costa v ENEL*[19] established the principle of primacy of European law over national law. Accordingly, for the interests of completeness, a decision was made to include EU, Irish and UK case law as part of this report.

[18] Formerly the *Data Protection Act 1998* (UK) before the Royal Assent of the *Data Protection Act 2018* (UK).
[19] C-6/64 (1964); [1964] ECR 595

EU directives outside the GDPR

The next exclusion criterion was to select the different regimes that current apply to scientific research in the EU aside from the GDPR. First, there is Regulation No 536/2014 of the European Parliament on European Clinical Trial Regulations. This regulation is expected to repeal Directive 2001/20/EC on the Clinical Trials Directive in the second half of 2019. Further, the database that is required to support the new Regulation is expected to come into force in 2020. Both of these regulations are relevant to this analysis as they establish regulations with respect to informed consent and patient involvement in human clinical trials.

Secondly, there is Directive 2004/23/EC of the European Parliament on quality and safety standards for human tissues & cells for transplantation. The legislative framework encompassing this directive includes the associated Directives 2006/17/EC, 2006/86/EC, 2015/565 and 2016/566. Directive 2006/17/EC sets certain technical requirements for the donation, procurement and testing of human cells. Directive 2015/565 modifies these requirements. Directive 2006/86/EC sets traceability requirements, requirements for notification of serious adverse reactions and events, additional technical requirements for the coding, processing, preservation, storage and distribution of cells. Finally, Directive 2015/566 sets out rules for quality and safety standards of tissue.

Finally, this analysis was complemented by an analysis of relevant intellectual property directives. These included Directive 96/9/EC of the European Parliament on the Legal Protection of Databases. This specifically pertains to the *sui generis* database right, which extends copyright protection for compilations of data that would not otherwise qualify for copyright protection. However, the directive has been considered controversial for many years given the relatively narrow bounds of the use of databases for scientific and research purposes. These regulatory instruments were identified as important for this doctrinal analysis. However, to fully appreciate their operation, it is necessary to consider these instruments within the framework of European Human Rights law. The next section will address these influences of European Human Rights law on the instruments identified in this search.

European Court of Human Rights

The European Court of Human Rights (ECHR) is established by the European Convention on Human Rights. The ECHR's purpose is to hear any applications alleging that a state actor that is a member of the Council of Europe has breached the provisions of the Convention. In particular, the ECHR protects fundamental human rights, including the right to privacy, as recognised by Article 8 of the European Convention on Human Rights.[20] It should be noted that the jurisdiction of the ECHR is tied to membership of the Council of Europe. Therefore, states such as Switzerland, Russia, and Turkey that are not members of the EU are signatories to the European Convention. The ECHR is complemented by the Charter of Fundamental Rights of the European Union 2000/C364/01 (the Charter), which sets out

[20] **Article 8: Right to respect for private and family life -** Everyone has the right to respect for his private and family life, his home and his correspondence.

fundamental rights with respect to privacy and data protection.[21] It is possible to draw a link between Article 7 of the Charter and Article 8 of the Convention, as both pertain to privacy. However, there is no equivalent right to data protection under the Convention. Nevertheless, the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) does pertain to data protection. Precedent from the ECHR may therefore influence EU data protection jurisprudence (O'Leary, 2018). Accordingly, cases from the ECHR were included in this search. Finally, the Council of Europe also passed another treaty, the Convention for the Protection of Human Rights in Biomedicine, on 4 April 1997 (the 'Oviedo Convention'). This Convention specifically applies to the boundaries of human rights law in medical research. Specifically, Chapter II provides the relevant law with respect to consent. Although this Convention was targeted at the regulation of biomedicine, it also applies to the regulation of medical treatment and research. Accordingly, articles referring to all three instruments were included in this search.

Supervisory authorities

In addition to these case law materials, reports from the former Article 29 Working Party were retrieved. The Article 29 Working Party was established as part of the former Data Protection Directive, and served a key function in regulating and harmonising European data protection laws (Bignami, 2011; de Hert & Papakonstantinou, 2016). In addition to the Article 29 Working Party, the European Commission created the European Data Protection Supervisor (EDPS) as an independent advisory body. The EDPS monitors the processing of personal data by EU institutions and bodies, advising on policies and legislation surrounding privacy and ensuring supranational consistency in data protection (Gonçalves, 2017). The functions of the Article 29 Working Party have also been significantly expanded under the GDPR and the Working Party has been renamed the European Data Protection Board (EDPB). The GDPR also widens the power of different agencies to ensure consistency. In particular, Recital 135 and Article 63 provide that both the EDPB and the EDPS have significant interpretive power with respect to European data protection law (de Hert & Papakonstantinou, 2016). Therefore, where relevant, reports of the former Article 29 Working Party, the EDPB and the EDPS were referred to in this report. It should be noted that both the Article 29 Working Party and the EDPB have issued contradictory opinions (Moerel, 2011). Where appropriate, these contradictions will be noted and explained in further detail in this chapter.

Secondary resources

The final stage of the search strategy for this analysis was to identify secondary resources on the relevant regulatory instruments described above. As stated previously, there is considerable academic commentary on the scope and limitations of the GDPR. This commentary has been interpreted in the context of both case law and academic commentary on the former Data Protection Directive. Finally, a preliminary case law search with respect to the other regulatory instruments contained in this analysis revealed limited precedent on scientific research. Accordingly, a search for secondary materials, including journal articles and book chapters, was conducted to complete the doctrinal analysis. These articles included both legal articles, which considered the GDPR purely through a doctrinal lens, and

[21] In particular, Articles 1, 3, 7, and 8 are relevant in this context.

subject specific articles, which considered legal issues through the lens of a particular field. These articles were drawn from the databases Scopus, Web of Science, Directory of Open Access Books and Directory of Open Access Journals. These articles were not analysed and coded as part of a strict systematic analysis. Instead, they were used to justify the interpretation of the regulatory instruments under consideration.

4.2 Results

4.2.1 Scientific research exception

Under the GDPR, scientific research is considered a special category of data processing, subject to the safeguards that exist in Article 89(1). Edward Dove notes that the purpose of the scientific research exception is to 'instantiate into law what is already good scientific practice'. In other words, the existing institutional mechanisms (such as ethics review boards and peer review) act as an inherent safeguard against the misuse of scientific data (Dove, 2018). Recital 50 reinforces this exception, noting that the processing of data for research, archiving or statistical purposes should be considered as compatible and lawful processing under the GDPR. However, Article 89(1) imposes upon data processors and controllers a positive obligation to respect all the rights and freedoms of data subjects when relying on research exceptions. In addition, Article 89(1) requires data processors to ensure that technical and organisational measures are in place to ensure the principle of data minimisation. In particular, Recital 26 explains how Article 89(1) provides a graduated approach for data minimisation. The controller or processor should first determine whether they can use non-personal data or anonymised data for their processing tasks. If non-personal or anonymised data cannot be used, Article 89(1) provides pseudonymisation as an example of a safeguard to protect data subject rights and prevent the use of personal identifiers (Molnár-Gábor, 2018). The boundaries of non-personal data or anonymised data are described in further detail below.

Each of these categories of research are then defined within Recitals 157, 158, 159, 160 and 162. These recitals represent a significant step forward from those that existed under the previous Data Protection Directive, which did not give any guidance on what constituted scientific research (Dove, 2018). In particular, Recital 159 notes that scientific research should be interpreted broadly. This broad interpretation is reinforced by Recital 157, which recognises that useful scientific data can be collected from registries. These registries could include hospital registries for research into widespread medical conditions, as well social security registries for socio-economic research. Furthermore, Recitals 158 and 160 recognise that an exception may exist for the collection of personal data from living persons for archiving or historical purposes. This exception may also permit further processing for archiving with a view to providing historical information about totalitarianism or war crimes. Finally, and most contentiously, Recital 162 permits an exception for the processing of personal data with a statistical result, provided the result of processing is aggregate and not personal data. This recital effectively enables the use of big data analytics and machine learning tools for the processing of personal data (Paterson & McDonagh, 2018). However, Recital 162 also defines statistical processing as 'including processing for scientific reasons.' This phrasing extends the operation of any exceptions subject to Article 89(1) beyond scientific processing to statistical analysis conducted by private institutions. These

statistical purposes could include the processing of social media or purchase data, which in turn could be used for targeted advertisements or marketing research. In these circumstances, the same institutional safeguards may not exist to prevent abuse of the exception. The qualifier to only use on aggregate data may encourage data minimisation in statistical processing, as demonstrated by Apple's use of differential privacy to collect operating system error data (Pagallo, 2018). Nevertheless, due to concerns about the misuse of the research exception, scientific research organisations such as the BBMRI-ERIC[22] have suggested restricting exceptions subject to Article 89(1) to public interest research (Shabani & Borry, 2018). Therefore, legislative reform to clarify the scope of activities permissible to exceptions subject to the obligations imposed by Article 89(1) is a policy option that should be explored. The next subsection considers the articles of the GDPR which are subject to the scientific research exceptions pursuant to Article 89(1)'s positive obligations.

4.2.2 Impact of the exception on the GDPR

The GDPR offers three categories of research exception subject to the obligations imposed by Article 89(1). First, there are exceptions to the data processing principles and lawful grounds for processing. Secondly, there are exceptions to the data subject rights that are directly available under the GDPR. Thirdly, there is the potential for member states to implement scientific exceptions into national law.

Exceptions to the data processing principles and lawful grounds for processing:

In the GDPR, the data processing principles are defined in Article 5, and the lawful processing grounds are defined in Articles 6 and 9. Scientific researchers are required to comply with four of the data processing principles, irrespective of whether they are working in academic or scientific commercial research.[23] However, there are two exceptions to the data processing principles identified above for scientific or historical research purposes, archiving purposes or statistical purposes. First, there is an exception contained within the purpose limitation in Article 5(1)(b). The purpose limitation strictly limits the use of data to the purpose for which it was originally collected ('Article 29 Working Party Opinion 03/2013 on purpose limitation,' 2013). However, the exception within Article 5(1)(b) provides that processing for further research purposes pursuant to the obligations imposed by Article 89(1) is lawful. Accordingly, under this exception, medical data collected as part of hospitalisation could not be then used for research purposes without consent. By contrast, if data is collected as part of a research biobank, biobank administrators can use that data for other forms of scientific research (Simell et al., 2019). However, as for the previous Data Protection Directive, the onus is on the researcher or institution to demonstrate that the processing is for research purposes (Ruyter et al., 2010). Secondly, there is also an exception to the storage limitation in Article 5(1)(e). As discussed previously, the storage limitation principle prevents personally identifying data being retained for any period longer than necessary to complete processing. Article 5(1)(e) then provides that data may be retained for a longer period than would otherwise be necessary to complete processing, where future processing will be solely for

22 Biobanking and BioMolecular Resources Research Infrastructure-Europe Research Infrastructure Consortium.

23 Specifically, the lawfulness, fairness and transparency principle; the data minimisation principle; the accuracy principle and the integrity and confidentiality principle.

research purposes. Returning to the example of research biobanks described above, the storage limitation exception would allow the biobank to retain personally identifying data from tissue samples for future research projects (Chico, 2018). In concert with one another, these research exceptions would allow researchers to collect data and hold it for as long they wish, for any research purpose. As discussed above with respect to the scope of the exception, these organisations could include private research organisations.

Article 6 then sets the different lawful grounds that data processors and controllers can rely on to process personal data. Recital 40 suggests that the lawful ground for processing is consent pursuant to Article 6(1)(a). Article 6(1)(e) also permits processing to occur where processing is necessary for the performance of a task carried out in the public interest. Article 6(2) then permits member states to introduce national legislation that would apply to special categories of processing defined in Chapter IX. As discussed previously, these special categories of processing explicitly include scientific and other forms of research pursuant to Article 89(1). Because it relies on national law, Article 6(1)(e) may also create a small risk of regulatory fragmentation in the GDPR (Chen, 2016). In addition, Article 6(1)(f) permits processing where necessary to serve the legitimate interests pursued by the controller or by a third party. This exception would possibly encapsulate the processing of data for scientific or other research purposes. Nevertheless, Article 6(1) notes that 6(1)(f) shall not apply to processing carried out by public authorities in the performance of their tasks. This section would significantly limit the scope of Article 6(1)(f) in permitting scientific research, by forbidding public entities from relying on the research exception.

Outside Article 6, Article 9 imposes a prohibition on the processing of sensitive categories of personal data. Article 9(1) defines these sensitive categories to include personal data revealing racial or ethnic origin, political opinions, religious or philosophical data, and data concerning health or sex life. In addition, Article 9(1) expands these categories of sensitive related personal data to include genetic data, biometric data and data on sexual orientation. This definition was updated in the GDPR to recognise the advances in genomic technology that have occurred since the passage of the Data Protection Directive (Beyleveld & Taylor, 2007). Article 9(2) then describes a number of scenarios where the limits on the processing of this sensitive data may be lifted. In particular, Article 9(2)(j) notes that sensitive data may be processed for scientific research purposes subject to the obligations imposed by Article 89(1). However, these requirements do not necessarily mean that all data belonging to these sensitive categories of data must be processed in accordance with these exceptions. Articles 4(13), 4(14) and 4(15) define genetic data, biometric data and health related data respectively. These Articles explicitly note that each of these categories of data refer to personal data that can be used to identify a person or determine his or her health status. For example, whole genome sequence (WGS) data of one or more individuals is likely to constitute personal data, as it could be used to identify those individuals (Quinn, Paul & Quinn, 2018).

However, as for the Health Insurance Portability and Accountability Act (HIPAA), the GDPR adopts technology neutral standards with respect to other forms of genetic data. For example, genetic markers such as single nucleotide polymorphisms or short tandem repeats could be considered personal data if they are sufficient to identify a person. Further, the data stored in binary alignment maps (BAM), FASTQ

or variant call formats (VCFs) may or may not be considered personal data under the GDPR (Evans & Jarvik, 2018). As Kärt Pormeister notes, this ambiguity might have two, equally problematic effects. First, it may discourage researchers from relying on the exception to collect specific types of genetic data which may or may not constitute personal data. For example, researchers at smaller institutions who wish to identify different SNPs may not have sufficient administrative support. Secondly, in recognition of this ambiguity, researchers (particularly commercial researchers with sufficient administrative support) may simply rely on the exception to collect complete data from data subjects. Accordingly, this data can be continually processed without the need to seek consent from the patient again (Pormeister, 2017). Perversely, a potential effect of this exception would be to discourage smaller researchers from relying on the exception due to the administrative burdens of obtaining and protecting this data. By contrast, larger organisations (including private research companies) may have the financial resources to implement the necessary technical and organisational measures. This administrative burden may inadvertently lead to large research organisations dominating biomedical and biotechnology research. A potential solution to this conundrum may exist via Article 9(4) of the GDPR. This Article provides member states with the capacity to derogate stricter limits on the processing of genetic, biometric or health-related data. These stricter limits may be useful as a mechanism to further refine the research exception with respect to scientific research involving genetic, biometric or health related data. For example, both Finland and Italy have imposed heightened requirements in their data protection legislation for the processing of genetic data. However, any national derogations flowing from the GDPR also increase the potential of regulatory fragmentation in EU data protection law, which the GDPR was introduced to prevent (Chen, 2016). Accordingly, as more derogations are introduced in national law under Article 9(4), EU national legislatures should work together to ensure consistency.

Directly applicable research exceptions for data subject rights

The second category are Articles where directly applicable research exceptions are available. As stated in the methodology section above, in contrast to the Data Protection Directive, the GDPR is a Regulation. Accordingly, scientific researchers can directly rely on these exceptions without the need for these provisions to be implemented into national legislation. The first direct exception applies to the right of information where personal data has not been obtained from the data subject, but have instead been obtained from other sources. The data processor or controller, on request, must provide at least general information about where this data has been collected from. However, Article 14(5)(b) provides that the controller does not have to provide information to the data subject where appropriate safeguards are in place and it would otherwise significantly impede research. For example, some cyber security research involves collecting IP address information from machines that are suspected bad actors. Notifying these bad actors of the intention to collect these IP addresses would undermine the purpose of the study (Sullivan & Burger, 2017). The second direct exception applies to the particularly controversial right of erasure under Article 17 of the GDPR. The relevant CJEU authority is *Google Spain v González*. Google argued that their act of indexing data for search engine optimisation did not amount to processing of personal data. Google argued that even if they were data processors or controllers, the principle of proportionality would place responsibility for deletion on the website owners.[24] The CJEU considered

[24] C-131/12, paragraph 63.

the need to balance the fundamental right to data protection with the right to informational access, particularly where a search is carried out using an individual's name.[25] In addition, the CJEU noted the data processing principles of accuracy required controllers to rectify or erase to data where that data is no longer accurate.[26] Accordingly, the CJEU held that Google would be obliged to deindex data where that data was irrelevant, inaccurate or no longer necessary for processing. Further, the CJEU held that this obligation extended to situations where the publication of the data was lawful.[27]

In the aftermath of the *Google* decision, legal scholars and practitioners agreed the decision extended territorial liability for data processors with a minimal presence in the EU (Frantziou, 2015; Wolf, 2014). Accordingly, the *Google* decision caused widespread consternation amongst non-EU data processors as to their liability for erasure both within and outside the EU. In addition, there was concern that the decision might extend liability to the hosts of cloud computing software, which is discussed in further detail below (Hon, Millard, & Walden, 2011, 2012). However, a preliminary opinion issued by the CJEU in January 2019 significantly restricted the right of erasure under the GDPR. In *CNIL v Google*,[28] Advocate General Szpunar held that search engines are only required to enforce deindexing for searches conducted in the EU. This decision places a territorial limit on the application of the right of erasure. Although the case is in progress at the time of writing, research institutes should monitor the decision to consider its impact on scientific research, particularly cross border research (Mantelero, 2013). The GDPR helps answer these questions by creating a specific research exception for erasure pursuant to Article 17(3)(d). This exception allows researchers to ignore an erasure request where it would render impossible or seriously impair the processing of personal data for scientific purposes. Determining the legitimate scope of this exception requires balancing a number of separate considerations. On the one hand, Eugenia Politou and colleagues argue that imposing this requirement would place a significant administrative and technical burden on researchers (Politou, Michota, Alepis, Pocs, & Patsakis, 2018). On the other hand, Pormeister argues that this exception may be particularly problematic with respect to genetic data, where the informational potential of such data will grow over time (Pormeister, 2017). Therefore, this right must be read in consistency with the principles of proportionality under ECHR law. In other words, any legitimate grounds for refusing an erasure request must be proportionally balanced against legitimate reasons for retaining privately held data (Ambrose, 2014).

A similar interpretation should be given to the third direct exception which applies to Article 21 of the GDPR. Article 21 permits data subjects to object to the processing of their data, including profiling, subject to the lawful grounds contained in Article 6(1)(e) or 6(1)(f). This right to objection is an unconditional right, so a data processor or controller cannot ignore that request even where they have compelling reasons to continue processing (Dove, 2018). However, Article 21(6) permits a scientific researcher to ignore an objection request where the data processing is necessary for reasons of public interest. This requirement of public interest imposes a slightly different, and perhaps objectively higher, standard on data processors to the right of erasure. Nevertheless, both the right to erasure and object will be difficult to activate if a scientific researcher relies on the exception to the right of information

[25] C-131/12, paragraph 81-82.
[26] C-131/12, paragraph 72.
[27] C-131/12, paragraph 88.
[28] C-136/17.

(Wachter, 2018). There is therefore a possibility for these exceptions to be used in concert to continually reuse and store personal data from data subjects. In particular, where data has not been collected from the data subject by the data processor or controller, the data subject might be blocked from effectively exercising their rights. The only limit on continuous processing by the processor or controller would be the constraints imposed by institutional review boards. In addition, the proportionality principles contained within the European Charter and Council of Europe law might act as a backstop on processing. However, as a primary policy option, institutional review boards and data protection agencies should establish principles defining appropriate limits on processing.

Flexibility for nationally derogated research exceptions:

The final category of research exceptions are nationally derogated research exceptions. These research exceptions can be introduced by individual member states into national law pursuant to the obligations imposed by Articles 89(2) and (3) of the GDPR. Specifically, for personal data processed for *scientific or historical research purposes*, exceptions can be introduced for the data subject rights under Articles 15, 16, 18 and 21.[29] For example, Schedule 2 of the *Data Protection Act 2018* (UK) already includes these specific derogations (Chico, 2018). Where personal data is processed for *archiving purposes in the public interest*, Article 89(3) permits exceptions for the rights under Articles 15, 16, 18 and 21. In addition, Article 89(3) permits derogations for the notification obligation under Article 19 and the right to data portability under Article 20. It should be noted there is already a directly applicable research exception for the right to object to processing under Article 21(6) where an objection would render research impossible. Therefore, it may be possible for a particular jurisdiction to completely eliminate the right to object to processing for scientific, historical, or public interest research (Pormeister, 2017). However, Articles 89(2) and (3) require that national derogations are only permissible subject to the obligations imposed by Article 89(1). Further, Article 89(2) and (3) only permit derogations where the exercise of data subject rights would render impossible or significantly impede research. This threshold is a higher threshold than expected for the directly applicable research exceptions subject to Article 89(1). Specifically, Article 89(1) only requires compliance with specific safeguards for data processors and controllers to rely on research exceptions (Price & Cohen, 2019). By contrast, Articles 89(2) and (3) would require national derogations where the research in question would be impossible without limiting the rights of data subjects. For example, Fruzsina Molnár-Gábor gives the example of an exception to the right to access being derogated under Article 15 where it would completely prevent specific types of research (Molnár-Gábor, 2018). Perhaps in recognition of these possibilities, some member states have introduced stricter requirements than those imposed by Articles 89(2) and (3) for research exceptions. For example, the Croatian, Danish, Estonian, and Norwegian implementations of the GDPR limit statistical processing to public interest or official purposes. However, permitting derogations subject to Articles 89(2) and (3) could undermine one of the fundamental objectives of the GDPR; that is, to standardise data protection law throughout Europe. Permitting member states to introduce their own exceptions for scientific research could lead to the fragmentation of these rules across the EU (Chen, 2016). Therefore, as a policy option, the national legislators of member states should attempt to define consistent national exceptions as soon as possible (Vestoso, 2018).

[29] GDPR, Article 89(2).

4.2.3 Scientific processing outside Article 89(1) – non-personal data and anonymised/ pseudonymised data

Personal and non-personal data:

To comply with the principle of data minimisation, scientific researchers should first consider whether it is possible to conduct data processing without using personal data. However, personal data has been given a broad definition under the GDPR. In particular, personal data may be both volunteered or surrendered by individuals, or otherwise collected from these individuals through observation (Pearce, 2018). This broad definition may be of particular concern to social science researchers, who must be mindful of data protection law as their research becomes increasingly data driven (Vestoso, 2018). The scope of personal data becomes more complicated with respect to inferences that can be drawn from other data. In particular, the cases of *YS v Minister voor Immigratie, Integratie en Asiel*[30] and *Peter Nowak v Data Protection Commissioner*[31] are informative in this regard. The former case concerned the collection of personal data for processing of a residency permit application. After the data was collected, the responsible case officer at the Dutch Ministry for Immigration also wrote a set of minutes about the application. The minute contained information such as the details of the case officer, data relating to the applicant, and details of the documents submitted by the applicant. The minute also contained a legal summary of the case officer's assessment of the applicant's case. The applicants requested the Dutch Ministry of Immigration provide them with these details when their residency permit applications were refused. The CJEU contrasted the personal information supplied in the residency applications with the legal information contained in the application. Specifically, the CJEU argued that the purpose of the Data Protection Directive was to protect the rights of individuals with respect to their personal data, including the right to rectification. The inferences drawn from this personal data though did not fall within the definition of personal data under the Data Protection Directive, as they constituted the case officer's interpretation.

By contrast, *Peter Nowak v Data Protection Commissioner* concerned the applicant Nowak's marked examination paper. Nowak applied under Irish data protection law to request the paper from the Institute of Chartered Accounts of Ireland after he failed his examination for the fourth time. However, the Institute refused to provide him with the document on the grounds that it did not contain personal data. Nowak then appealed to the Data Protection Commissioner, and then to the High Court of Ireland, the latter of which referred the matter to the CJEU. Counsel for Nowak submitted a number of arguments as to why the personal data could be distinguished from the data under consideration in *YS v Minister voor Immigratie*. First, the answers reflected the candidate's understanding of a particular field (accounting), expressed in the candidate's handwriting. Secondly, these answers were collected to determine whether the applicant was competent to practice in a particular field. Thirdly, because the examination determined the applicant's competence, they had a material effect on the applicant's

[30] *YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S* Joined Cases C-141/12 and C-372/12
[31] C-434/16

rights.[32] The CJEU accepted these arguments, and added that not treating a marked examination paper as personal data would exculpate an examination authority from treating it as confidential. Accordingly, the CJEU permitted access to the examination papers, noting that such access did not extend to the rectification of incorrect answers (as discussed below).

Both *YS v Minister voor Immigratie* and *Nowak v Data Protection Commissioner* are relevant in determining whether inferences can constitute personal data. This determination is particularly pertinent for assessing whether the inferences drawn from large data sets might constitute the processing of personal data. As later sections will argue, the context in which personal data is generated is also relevant for determining whether it may have an impact on privacy or informational self-determination. A large data set of purchase information by itself may not constitute personal data. However, if an individual's purchase habits can be identified from the data set, it is then possible to draw inferences from this data that in turn could constitute personal data. In particular, it may be possible to identify whether that purchasers suffers from any diseases, or whether they are pregnant, based on the products they purchase (Price and Cohen, 2019). Potential guidance comes from the Recital 26 of the GDPR. This recital states that in determining whether a natural person is identifiable, account should be taken of the means to do so, including singling out. This recital suggests that even if big data algorithms do not reveal an individual's identity, they will still constitute personal data processing if they can single out that individual (Paterson and McDonagh, 2018). Further, given that data protection law is a safeguard for scientific research, scientific best practice should adopt a precautionary approach as a policy option. Accordingly, scientific research institutions should develop procedures to explain when non-personal data becomes personal data after processing. Further, education can help researchers develop normative rules for managing the generation of personal data through inferences.

Anonymised and pseudonymised data:

The next stage of the data minimisation approach requires processors and controllers to consider whether they can conduct processing with anonymised data as opposed to non-personal data. Recital 26 of the GDPR notes that the GDPR's requirements do not apply to anonymised data. However, it is important to note that anonymised data must be distinguished from pseudonymised data. As discussed by Paul Quinn, pseudonymisation generally involves removing identifying details from the data and placing these details in a separate file. This separate file is then held by the data processor. The identifying details are then replaced by a pseudonym. This process can be contrasted with anonymisation, where all identifying details are stripped from the data (Quinn, P., 2017). This definition is provided by Recital 26 of the GDPR. Critically, both Recital 26 and Article 89(1) establish that pseudonymised data **remains** personal data (and is subject to the GDPR), whilst anonymised data **is no longer** personal data. Unfortunately, there is still debate as to what standard of pseudonymisation is required to satisfy these requirements. With respect to the Data Protection Directive, Deryck Beyleveld and David Townend argue that data anonymisation constitutes a form of data processing (Beyleveld & Townend, 2004). Accordingly, the principles of personal information processing apply to anonymised data at the very least *before* it is anonymised.

[32] C-434/16, paragraphs 37 to 39.

Unlike other areas of scientific research under data protection law, there are some cases with respect to how anonymisation should been interpreted. In the UK, the *Common Services Agency (CSA) v Scottish Information Commissioner* dealt with sufficient pseudonymisation. The data in question had been subject to barnardisation, which involves randomly adding or subtracting values from some cells in the table where the data is stored. Barnardisation is frequently used by UK government departments (such as the department in the *CSA* case) as a form of information control. In the facts at hand, the CSA refused to release information regarding the incidence of leukaemia in the Dumfries and Galloway regions. The CSA argued that the data could be re-identified given the relatively small number of residents in the area. However, the House of Lords held that the barnardisation fulfilled the requirements for deidentification, as the recipients of the data would not be able to deidentify patients using that data. These findings were confirmed in *R (Department of Health) v Information Commissioner*.[33] Nevertheless, the CJEU has been silent as to the acceptable standard of identifiability and security required for pseudonymisation for research. As discussed by Miranda Mourby and colleagues, the closest analogous case exists with respect to cases surrounding the use of IP addresses (Mourby et al., 2018). In *EMI & Others v Eircom Ltd*[34] the Irish High Court considered, as a secondary matter, whether IP addresses were personal information. The High Court argued that on the facts, IP addresses would be used to identify particular piracy activity, as opposed to names or home addresses. Accordingly, the High Court concluded that these IP addresses did not constitute personal data.[35]

However, in *Patrick Breyer v Bundesrepublik Deustchland,* the CJEU reached a different conclusion. Breyer, a civil liberties activist, sought to access several websites operated by German Federal Institutions. These websites indicated that access operations were stored in log files on these websites. These included the web page or file accessed, the terms searched for, the time of access, and the IP address of the computer from which access was sought. The CJEU noted that IP addresses are essentially 'digital addresses' for a computer connected to a wide area network such as the Internet. These addresses are necessary to ensure that data is transferred to the correct recipients. The CJEU further noted that IP addresses could be divided into two categories; static and dynamic IP addresses. Static IP addresses are assigned to one computer, whereas dynamic IP addresses change each time a machine connects to the network. Germany argued that the dynamic IP addresses of Breyer's computer did not amount to personal data, as only Breyer's internet service provider (ISP) would have access to his connection details. Whilst the *Bundesgerichtshof* (Federal Court of Justice) agreed with Germany's argument, it was rejected on appeal to the CJEU. The CJEU held that Article 2(a) of the Directive explicitly referred to personal data that could be either directly or indirectly used to identify a person. Therefore, given a particular machine connected to a particular account belonging to a user could be identified by a particular IP address. In determining whether other classes of data counted as anonymous, the CJEU provided the following guidance:[36]
'[Personal data will be considered anonymous] if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.'

[33] [2011] EWHC 1430 (Admin).
[34] [2010] IEHC 108
[35] [2010] IEHC 108, paragraphs 25-6.
[36] C-582-14, paragraph 46.

Nevertheless, the CJEU adopted what they called a 'relative' approach in determining for whom an IP address would be personal data. In other words, for government web administrators, the IP address would not be personal data, whereas for Mr Breyer's internet service provider it would be.[37] It is likely that the *Breyer* decision will apply to the GDPR. However, the *Breyer* decision creates a different test for anonymisation than that provided by the Article 29 Working Party in its opinion paper (Ali, Khan, & Khan, 2002). The Working Party noted that amongst both national legislation and international standards, there is no consistency in what constitutes anonymisation. For example, Italian, German and Slovenian data protection legislation adopts the 'disproportionate effort' test established by the CJEU in *Breyer*. By contrast, the ISO 29100 standard defines anonymisation as a process by which:

'[P]ersonally identifiable information is irreversibly altered in such a way that a [data subject] can no longer be identified directly or indirectly, either by the [data controller] alone or in collaboration with any other party.'

Likewise, the French data protection law notes that just because disproportionate effort is required to reidentify the data does not mean that the data is anonymised. These requirements must also be read in context of *Rijkeboer*, which established that agencies that collect personal data must retain that data in an identifiable format. This retention is necessary to enable data subjects to exercise their rights of access and information (Beyleveld and Townend, 2004). Further, the Working Party conducted a technical analysis of the different anonymisation and pseudonymisation techniques available for personal data. In particular, the Working Party identified three potential risks that must be resolved for true anonymisation to occur. First, *singling out* involves identifying all records associated with an individual in a dataset. Secondly, *linkability* involves the ability to identify two records concerning the same data subject or a group of data subject. These records may be either in the same database or in two or more databases. Thirdly, *inference* involves deducing the value of an attribute with significant probability from the values of other attributes. The Working Party then compared the potential risks associated with different forms of anonymisation,[38] as well as pseudonymisation. However, the Working Party concluded that based on this assessment, all forms of anonymisation were vulnerable to at least one of the three risks described above. These risks are described in Table 1 below:

[37] C-582-14, paragraph 25.

[38] The forms of anonymisation considered by the Article 29 Working Party include noise addition, substitution, aggregation and K-anonymity, L-diversity, Differential privacy and Hashing/tokenisation. Noise addition works by adding or multiplying variables to provide confidentiality. Although noise addition reduces the risk of linkability and inference, it cannot prevent singling out of records. Substitution involves randomly substituting the input data with an alphanumeric value or a date. However, whilst probabilistic inference remains possible with substitution, singling out and linkability remain a potential risk (even if less reliable). Aggregation involves grouping a particular record with at least K other records so that they can no longer be singled out. L-diversity extends aggregation anonymisation by ensuring that all records that are aggregated have L aggregated values. Whilst this anonymisation prevents inference identification, linkability is still a possibility. Hashing involves replacing relevant personal data with a cryptographic hash, which can be easily reversed using a cryptographic key. Whilst this method prevents records from being singled out or linked, inference attacks can be conducted. Further, privacy may be compromised if an unauthorised user accesses both the key and the data. Finally, differential privacy involves changing certain parts of the relevant data so that individual users cannot be identified based on their records. Although differential privacy provides the greatest degree of protection, it may be still vulnerable to a lack of sufficient noise (Elliot et al., 2018; Ohm, 2009; Raghunathan, 2013).

Table 3 - Different forms of pseudonymisation and anonymisation and associated risks (reproduced from the Article 29 Working Party)

	Singling out	Linkability	Inference
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation and K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/tokenisation	Yes	Yes	May not

Accordingly, the Working Party concluded that no single form of pseudonymisation or anonymisation will perfectly prevent reidentification of data. The Working Party's opinions on the scope of anonymisation are consistent with Paul Ohm's observations in the US on the 'death of anonymisation' and the impossibility of perfect deidentification (Ohm, 2009). To this end, the Working Party suggested that the most appropriate anonymisation or pseudonymisation techniques should be determined on a case by case basis.

This report caused significant controversy in the legal literature as to the scope of anonymisation. Before the report was issued, the European Court of Human Rights considered the question of anonymisation in the case of *S and Marper v UK*. In that case, the ECHR recognised that genetic samples are 'intrinsically private' information and therefore tight controls over processing of that data are warranted.[39] The ECHR also recognised the development of future genetic technology is impossible to evaluate, suggesting that case law should not be used to stymie this development (Hallinan, Friedewald, & De Hert, 2013). In part, because the GDPR does not mandate the use of specific statistical anonymisation techniques, there is legislative flexibility for advances in anonymisation technology (Dove, 2018). However, the GDPR also explicitly classifies pseudonymised data as personal data. Combined with the Working Party's opinion, this classification may render data that has been subject to anonymisation as personal data (Mourby et al., 2018). As alluded to earlier in this report, the heightened standards for anonymisation also compound the difficulty in relying on scientific research as a lawful ground for processing. In particular, in some areas of social sciences research such as anthropology, relying on anonymisation may be impossible due to the very nature of the research activities being conducted (Yuill, 2018).

Yet another complicating factor remains the effect of CJEU's more pragmatic decision in *Breyer* and its implication for pseudonymised data transferred between parties. For example, Researcher A may have collected personal data from a set of data subjects as part of a research project. A set of key codes are then used to pseudonymise the data and remove all personal identifiers for the data subjects. With the consent of those data subjects, Researcher A then transfers the pseudonymised data to Researcher B,

[39] *S and Marper v United Kingdom* [2008] ECHR 1581, paragraph [104].

who does not have access to the keys, but only has access to the pseudonymised data. Mourby and colleagues argue that Researcher B does not have access to key codes, and therefore would need to expend disproportionate effort to reidentify the subjects. Mourby and colleagues (along with Mark Elliot and colleagues) argue that it is impossible to determine whether data has been anonymised or pseudonymised purely by examining it (Elliot et al., 2018). Instead, it is necessary to consider the data environment within which the data exists. In practical terms, this definition means that researchers should not only consider the nature of the anonymisation or pseudonymisation techniques, but the contracts associated with transfer. Therefore, the most effective policy option to resolve this conflict is the use of institutional policy to set an appropriate standard of anonymisation or pseudonymisation. In particular, technical security mechanisms (such as barnardisation) along with organisational measures (such as release of data without access keys) could be used to prevent access to the data.

4.2.4 Sections without scientific research exceptions

Data processing principles and the principle of accuracy:

Outside the data processing principles with exceptions subject to the safeguards imposed Article 89(1), there are still ambiguities as to how scientific researchers should comply with the data processing principles. In particular, Hoeren notes that there is a concerning ambiguity with respect to the accuracy principle under Article 5(1)(d) (Hoeren, 2017). As stated previously, the English translation of Article 5(1)(d) provides that data must be 'accurate'. In the relevant UK legislation, this implication has been taken further to establish the data controller will not be held liable where data is incorrectly entered by the subject.[40] In other words, the data controller will only be required to take reasonable steps to ensure the accuracy of the data. A similar interpretation is supported by the Czech, Danish, French, Hungarian, Italian, Portuguese, Romanian and Spanish translations of the GDPR. All of these translations use variants of the phrase accurate (presnost, exactitude, rigtighed, pontosság, exatidão, esattezza, exactitate and exactitude respectively). By contrast, in interpreting data protection legislation, German courts have consistently concerned themselves with 'factual correctness' (Hoeren, 2018). Such a conclusion can be drawn from the Dutch, German, Polish and Slovak translations use the words 'juistheid', 'richtigkeit', 'prawidłowe', and 'správnost' (correctness).

To this end, does the GDPR therefore set a standard of accuracy or a standard of correctness? Hoeren further argues that the latter definition might undermine big data research, as some data mining techniques are predicated on processing large quantities of raw data (Hoeren, 2017, 2018). However, as suggested by *YS v Minister voor Immigratie, Integratie en Asiel*, the obligation to ensure the data accuracy does not extend to inferences drawn from that data. In other words, whilst the principle of data accuracy applies to the data itself, any further decisions made using that data are not covered by the principle of data accuracy. This decision may provide something of a limitation on the obligations of data processors in a scientific context with respect to ensuring the correctness of the results drawn from personal data. For instance, a social scientist may conduct a survey of a particular voting population to gauge their political opinions. These voters may be uncomfortable to indicate their preference for a controversial

[40] *Smeaton v Equifax PLC* [2013] ECWA Civ 108, paragraphs 71-80.

candidate, and so may therefore give incorrect answers when responding to the survey. Applying a strict standard of correctness to personal data, the social scientist may have breached the processing principle of accuracy by failing to ensure this data was correct. Whilst the consequences of this application may not be severe for social sciences research, they could be significantly more so for biomedical and pharmaceutical research. Accordingly, as for the definition of inferences, the best policy option for resolving this ambiguity involves standard setting at the institutional level. In other words, each research department should set an appropriate standard for data correctness or accuracy depending on their responsibilities. For example, sociology researchers may receive more leniency for their data processing than their equivalents in biomedical or biotechnology research.

Relying on consent to conduct data processing:

As alluded to previously, it is still possible for researchers to conduct research by relying on the informed consent of the data subject. However, in these circumstances, the researchers will be bound by the same requirements as other data processors, as well as the positive obligations imposed by Article 89(1). For example, returning to the lawful grounds of processing specified by Articles 6 and 9, the default lawful ground for processing is the free, explicit consent of the data subject. Ultimately, this consent should be sufficient to allow a data subject to exercise their rights under the directive and withdraw from the project if necessary. However, this requirement for consent presents challenges of its own. In particular, for many scientific research projects it may be impossible or difficult to determine the purposes for which data may be used in advance. For example, when machine learning algorithms are used to collect data, the underlying trends that may be drawn from that data are not immediately apparent when the algorithm is designed. Therefore, explicitly stating the purposes for which data is being used may be impossible where the trends observable from that data are not apparent in advance (Butterworth, 2018).

Accordingly, some form of broad consent, or open consent where the data subject consents to both the present purpose of collection and future purposes, may be required (Dove, 2015; Hallinan & Friedewald, 2015). Nevertheless, there is a critical conflict in the GDPR over the availability of broad or general consent. On the one hand, Recital 33 of the GDPR appears to endorse broad consent as a legitimate form of consent by recognising that not all purposes can be identified during data collection. On the other hand, Recital 33 is contradicted by the Article 29 Data Protection Working Party's official report on informed consent. The Article 29 Working Party explicitly stated that Recital 33 does not dispense with the requirement for informed consent under the GDPR. Further, the Article 29 Working Party held that to satisfy the informed consent requirements, the data subject needs sufficient information to withdraw their consent (Li, Zhao, & Zhong, 2016). In concert, these requirements significantly narrow the scope for broad consent with respect to scientific research. As Paul Quinn and Liam Quinn notes, this interpretation conflicts with the interpretation of Recital 33. Further, Quinn and Quinn argue that a limitation to explicit informed consent would require researchers to seek explicit consent at each change in processing during the research project (Quinn, Paul & Quinn, 2018). In particular, there are research projects where the requirement to seek specific consent may undermine the success or failure of the research project. For example, the need to seek fresh consent may be undermined where consent has already been sought under the previous directive (Simell et al., 2019). As for the definitional uncertainty over sensitive data, this requirement may discourage reliance on consent, particularly for researchers

without the administrative support to seek consent again. Accordingly, the difficulty in seeking consent might drive researchers to rely on the scientific research exceptions subject to Article 89(1). However, as discussed previously, the obligations imposed by Article 89(1) have the potential to create two stages of regulation for small-scale and large-scale research institutes. On the one hand, uncertain requirements for consent will place a significant administrative burden on small-scale research institutes. These requirements may discourage small-scale researchers from attempting research where it may be impossible to seek broad consent for the specific project. On the other hand, large scale researchers will bypass consent as a lawful ground for processing, and instead seek to rely on an alternative ground to consent for processing. To this end, ethics committees should help guide the development of organisational strategies for when to rely on different grounds of lawful processing.

Conflict and ambiguities between data subject rights:

Outside the scientific research exceptions subject to Article 89(1), there are potential conflicts between relevant data subject rights and scientific research. In particular, Article 22, which guarantees the right to opt out of automated profiling, may be relevant for scientific research. Article 22 was transferred largely unchanged from Article 15 of the European Data Protection Directive, which in itself was borrowed from French data protection law. Initially, this processing exception was designed to protect data subjects from an abstract class of automated decision-making tools used for profiling. This profiling might then be used to affect 'access to important facilities, such as credit, housing or insurance (Veale & Edwards, 2018). Crucially, Article 22, as Article 15 before it, does not create an exception for automated processing for scientific research purposes. Therefore, all automated processing of personal data pursuant to Article 22 must occur either pursuant to a contract or based on explicit consent by the data subject (Dove, 2018). However, since the passage of the Data Protection Directive there have been enormous technical advances in artificial intelligence and machine learning algorithms. Accordingly, there is significant academic debate as to whether the current Article 22 provides sufficient protection to data subjects against the use of automated processing (Poullet, 2018). In particular, Izak Mendoz and Lee Bygrave note that, by the operation of Article 22(1), the right to object to automated processing consists of four parts (Mendoza & Bygrave, 2017):

(1) a decision must be or has been made;
(2) that decision is based solely on automated processing;
(3) the decision has either legal effects or similarly significant consequences; and,
(4) the basis of the decision was automated processing or profiling.

The first condition is relatively straightforward to satisfy, and requires a decision to have been acted upon. The third condition is context specific, and its operation with respect to research will depend on the nature of research being conducted. The legally contentious condition is the second, which requires that the decision be made solely on automated data processing. Although this provision has not been tested, the use of 'solely' suggests a relatively narrow scope of operation (Wachter, 2018; Wachter, Mittelstadt, & Floridi, 2017). Accordingly, if a decision maker has merely used automated processing as an aide, it may be possible to argue that decision will not fall within the boundaries of Article 22.

Scientific researches should note that this relatively narrow scope does not entirely exculpate scientific research from the scope of Article 22. However, research ethics committees should pay careful attention to the framing of research projects to determine exactly how decisions will be made within the scope of the project.

Data processor and controller obligations:

Chapter Four of the GDPR sets imposes specific obligations on both processors and controllers. As Dove observes it should be noted that the GDPR splits the obligations of processing into two groups. First, data controllers are persons or entities that determine the objective and means for processing personal data. EU case law has broadly interpreted what constitutes the objective and means for processing data. In *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH*,[41] visitor data to a Facebook fan page was collected by the page's administrator. In this case, the administrator has held to be a controller under the Data Protection Directive.[42] In the context of scientific research, these groups could include universities, research institutes, and private research companies. Articles 24 and 25 establish the responsibilities of the controller in guaranteeing data protection. In particular, Article 25 introduces a requirement that data controllers introduce technological means, or data protection by design and default, for guaranteeing security. The Article 25 requirements were introduced pursuant to an understanding that legal principles alone are not sufficient to guarantee sufficient data protection (Hornung, 2013). The broad definition of 'controller' in Article 4 also indicates the EU intends to define data controllers and their responsibilities broadly so as to provide maximum protection for subject rights (Brkan, 2016). Accordingly, researchers and universities should assume that when processing personal data, their activities render them data controllers. Further, researchers should plan for, and ethics review committees should mandate, data protection by design principles as part of any research proposals or approvals. This design should also be accompanied by controllers maintaining detailed records of processing,[43] conducting data protection impact assessments before each processing operation, and appointing data protection officers.

An additional issue arises with respect to the issue of collaborative research projects between institutions, which may fall under the auspices of Article 26**.** The Data Protection Directive previously implicitly acknowledged that the goals of processing could be determined by more than one legal entity (Van Alsenoy, 2012). However, Article 26 extends this joint liability by acknowledging that controllers will be jointly liable for processing decisions made by the other controllers. Further, there are no exceptions for scientific research purposes with respect to joint controllership. Accordingly, the principles of joint liability under Article 26 apply equally to collaborations between public institutes as they do between private research companies or public-private collaborations. Any research collaborations should therefore be subject to a joint contractual agreement between each of the two parties. This research agreement should identify and state define the appropriate obligations in

[41] C-210/16
[42] C-210/16, paragraph [44]
[43] GDPR, Article 25.

accordance with the obligations on data controllers described previously (Stalla-Bourdillon, Pearce, & Tsakalakis, 2018). The GDPR then considers the obligations of data processors, or entities that process data on behalf of the data controller. In the context of scientific research, these groups may include research collaborators, cloud processors or scientific computing centres (Dove, 2018). A crucial change under the GDPR has been the significant expansion of liability for data processors. Although Article 5(2) states the data processing principles are the responsibility of data controllers, Recital 76 establishes that liability vests not only with the controller, but all parties who are involved in the processing of personal data. Accordingly, data processors should ensure that they also comply with principles of data protection by design, as well as the need to monitor data processing. An example of the potential risks associated with the failure to implement these policies was the Royal Free Trust and Google DeepMind collaboration in the UK. In this case, the Royal Free Trust shared NHS patient records with Google DeepMind for the purposes of building an application for diagnosis. The Royal Free Trust argued it could rely on the principle of 'implied consent' under the *Health and Social Care Act 2012* (UK) to transfer this data to DeepMind. However, the UK Information Commissioner's Office (ICO) investigated and held that the Royal Free Trust did not have authority, pursuant to its contract with DeepMind, to transfer the data. These cases demonstrate the importance of researchers maintaining clear contractual relationships with potential processors for the management of personal data (Rumbold and Pierscionek, 2017).

Transfers of data under the GDPR

A key change introduced by the GDPR has been an update of the rules surrounding the cross-border transfer of data to non-EU countries (henceforth referred to as third party countries). Pursuant to Chapter Five of the GDPR, there are four grounds for the lawful transfer of data. First, transfer may occur on the basis of an adequacy decision. For researchers, this exception is a comparatively narrow mechanism to rely upon given the relatively small number of countries that have been recognised as providing adequate data protection.[44] In part, this relatively small number is due to the strict requirements for a country to meet the adequacy standards. These requirements include respect for fundamental human rights, the existence of effective supervisory authorities or international commitments the third country has entered into (Wagner, 2018). For example, within the EU it may be fairly straightforward to transfer data to researchers working in Switzerland by reason of the European Commission's adequacy determination. However, a lawful transfer to other countries will not be possible on the basis of an adequacy decision. Further, the decision of the UK to leave the EU (as of the time of writing) may require the UK to also seek an adequacy status. The period of regulatory approval required for this adequacy status may have a significant impact on shared research between the UK and the EU (Taylor et al., 2018). Accordingly, researchers and research institutes may have better fortune relying upon the second category of lawful grounds for transfer, which are provided by Article 46. However, these transfers must occur pursuant to any one of a number of safeguards. Further, existing contracts that are in place for the transfer of data may not be sufficient to meet these requirements. Although existing agreements are

[44] At the time of writing, the Commission has only recognised as adequate Andorra, Argentina, the Faroe Islands, Guernsey, Israel, the Isle of Man, Japan, Jersey, New Zealand, Switzerland and Uruguay. As for the US, only Canadian private organisations can seek an adequacy determination pursuant to the *Personal Information Protection and Electronic Documents Act* 2000.

valid pursuant to Article 46(5) of the GDPR, organisations that currently have in place approved transfer agreements should review these for compliance with the GDPR.

The third and fourth avenues for transfer are established by Article 49 of the GDPR. The first avenue for transfer pertains to specific scenarios for transfer pursuant to Article 49(1). For example, data may be transferred where explicitly consented to by the data subject, or where it is necessary for the performance of a contract. These mechanisms may be useful for small scale studies where it is possible to obtain the consent of the data subjects and where appropriate safeguards exist (Bu-Pasha, 2017) . For example, the transfer of a small subset of patients with a rare disease, with their consent, outside the EEA may be permitted under Article 49(1) (Dove, 2018). Further, as Quinn argues, the right to data portability may dovetail into an informed consent model for transfer, as it would require the patient to initiate the data transfer (Quinn, Paul, 2018). Nevertheless, this model of transfer is unlikely to be preferable for large scale research projects, or research projects where consent is difficult to obtain. It should also be noted that Article 49(1) only permits specific transfers where a prevailing public interest exists. Although such a public interest may include public health related concerns (such as the transfer of samples for the prevention of a pandemic), it does not necessarily include scientific research. Finally, Article 49(2) permits the limited transfer of data where the same guarantees are afforded in the recipient country. Accordingly, this analysis demonstrates that the most straightforward means for researchers to transfer personal data outside of the EEA is either via written agreement or through explicit consent of participants. In particular, any written agreements for transfer between institutions should be sufficiently flexible so as to protect both the legitimate interest in processing scientific data and the rights of research participants.

4.3 Other directives that may apply to scientific research

The Clinical Trials Directive

The relevant provisions of the Clinical Trials Directive pertain to the need to seek informed consent to protect clinical trial subjects. Article 2 first defines informed consent and ethics committees, which are two important concepts in the operation of the Directive. The Directive then turns to address the question of how clinical trial subjects need to be protected. These principles for protection include minors and incapacitated adults unable to give informed consent. In particular, Article 3(2)(c) of the Clinical Trials Directive uses the definition of 'consent' from the former Data Protection Directive for the processing of personal data. These provisions are broadly similar to the consent rules contained in Chapter V of the current Clinical Trial Regulations ('Protection of Subjects and Informed Consent'). The two exceptions to this rule are the introduction of new informed consent rules for cluster trials and situations where consent cannot be obtained beforehand. The former situation arises where the groups of individuals involved in a trial are randomised, as opposed to the individuals themselves. A cluster randomisation trial may be suitable in circumstances where individual randomisation is impossible or difficult to achieve. For example, if a drug was being trialled within a single clinic, the shared knowledge between the physicians at that practice might undermine the controlled nature of the trial. Therefore, two groups at two relatively similar clinical practices might be used to trial the drug so that there is

adequate separation (Cave, 2018; Tenti, Simonetti, Bochicchio, & Martinelli, 2018). These regulations are expected to come into force in 2020. It should be noted by researchers that Directive 2001/20/EC will remain in force for up to three years after the Regulation comes into force. Accordingly, clinical trial approvals will be valid for this period, regardless of whether they are authorised under Directive 2001/20/EC or the Clinical Trial Regulations (Tenti et al., 2018). Further, it is also necessary for researchers to consider whether the GDPR or the Clinical Trial Regulations apply. Recital 161 of the GDPR states as follows:

For the purpose of consenting to the participation in scientific research activities in clinical trials, the relevant provisions of Regulation (EU) No 536/2014 of the European Parliament and of the Council should apply.

However, Article 93 of the Clinical Trials Regulation provides that 'Member States shall apply Directive 95/46/EC to the processing of personal data carried out… pursuant to this Regulation'. The potential overlap is further exacerbated by the different grounds for informed consent established under the GDPR and the Clinical Trial Regulation. As stated previously, consent under the GDPR is predicated on the grounds of 'freely given, specific, informed, unambiguous, and explicit' consent by the data subject. By contrast, informed consent under the Clinical Trials Regulation referred to written consent sufficient for the participant to understand the risks and benefits of participation. In addition, Recital 27 of the Clinical Trials Regulation refers to Articles 1 and 3 of the Charter of Fundamental Rights of the European Union. Therefore, the Clinical Trials Regulation defines consent as a mechanism to protect the integrity and human dignity of the trial participant. By contrast, the GDPR focuses on free and informed consent of the subject as a legal ground for processing. To resolve this conflict, in January 2019 the EDPB issued an opinion addressing the overlap between the GDPR and the Clinical Trials Regulation. The EDPB determined that whether the GDPR or the Clinical Trials Regulation would apply depended on whether the use of the data constituted primary or secondary use. In this regard, the EDPB acknowledged their definition of primary use was different from the definition supplied in the Article 29 Working Party's opinion issued in 2013. In that opinion (which pertained to the purpose limitation data processing principle), the Article 29 Working Party held that data collection constituted primary processing. Any processing subsequent to data collection constituted further processing. This narrow definition was adopted to prevent data processors from using data for inconsistent uses after collection.

By contrast, the EDPB divided processing of clinical trial data into primary processing within the clinical trial protocol and secondary processing outside the protocol. Within processing data as part of a clinical trial protocol, the EDPB held that processing activities may fall under two further categories. First, data could be processed for safety reporting, inspection or retention purposes as part of a clinical trial protocol. The processing of personal data according to the protocol may be justified under Article 6(1)(c), as processing necessary to comply with legal obligations (that is, the Clinical Trials Regulation). In the alternative, processing for sensitive categories of personal data may be justified under Article 9(2)(i), as processing necessary for reasons of public interest with respect to public health. Secondly, data could be processed for purely scientific reasons subject to the clinical trial protocol. In this case, the EDPB held that any processing of clinical trial data must occur under one of the grounds for scientific

processing specified under the GDPR. These may include Articles 6(1)(a) or (9)(2)(a) for explicit consent, Article 6(1)(e) for reasons of public interest or Articles 6(1)(f) and (9)(2)(i) or (j) for scientific research ('EDPB Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR),' 2019).

However, if personal data from the clinical trial is processed for scientific purposes, the relevant standard of consent would be that imposed by the GDPR, not the Clinical Trials Regulation. Accordingly, scientific researchers would therefore need to ensure that any consent is freely given in accordance with the GDPR, and that the grounds for lawful processing are clearly articulated. This consent requirement also requires data processors and controllers ensure that the consent of participants is not overruled due to the power imbalance between the parties. Such a power imbalance may emerge where the data subject feels they must consent to the trial to receive ongoing access to medicine. Further, the EDPB held that any withdrawal of consent would apply to processing relying on consent as a lawful ground, and not processing on other grounds (namely Article 6(1)(c)). Regarding secondary use outside the clinical trial protocol, the EDPB held the need to seek reconsent from patients would depend on the compatibility of lawful grounds of processing. That is, if the processor or controller can demonstrate that their purpose is consistent with the original purpose of processing, they can reuse the data without the need to seek consent ('EDPB Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR),' 2019). The EDPB's opinion clarifies much of the ambiguity with respect to the overlap between the GDPR and the Clinical Trials Regulation. However, there remain outstanding questions. In particular, it appears illogical that a participant who consents to participate in a clinical trial might not be able to consent for data processing due to a potential power imbalance. Conversely, data processed via Article 9(2)(i) and (j) would be reuseable for scientific research. These exceptions could permit the ongoing storage and reuse of the data by privately funded pharmaceutical companies, without permitting data subjects the right to erase their data or object to processing. Therefore, both the EDPS and national regulatory bodies should establish how data might be reused for secondary scientific reasons pursuant to Articles 6 and 9 as described above.

Human tissue and cells directives:

The relevant sections on consent are contained in Chapter III of Directive 2004/23/EC, require compliance with all 'mandatory consent requirements'.[45] In addition, all transfer of data must comply with relevant data protection and confidentiality laws.[46] In addition, the technical directives (2006/17/EC[47] and 2016/566[48]) place technical requirements on the recording of consent and the handling of risks associated with adverse incidents. The Tissue and Cells Directives were introduced to reshape the regulatory landscape in Europe for tissue storage and exchange. The Tissue and Cells Directives were also an attempt to import fundamental European human rights laws into these regulations. As alluded to previously, these instruments include the European Convention on Human

[45] Directive 2004/23/EC, Article 13.
[46] Directive 2004/23/EC, Article 14.
[47] Article 2(5)(b), Annex III, Annex IV.
[48] Annex III

Rights and the Oviedo Convention. A key emphasis was placed on the importance of 'informed consent' for tissue donation. The rationalisation of this decision was to ensure that only actively participating patients were involved in the donation of tissue.

However, the Human Tissue and Cells Directives do not specify the *substantive* consent requirements for the collection of human tissue. As stated above, Article 13 of Directive 2004/23/EC only specifies that consent for the collection of tissue must comply with mandatory consent or authorisation requirements in the member state. Likewise, Article 14 of Directive 2004/23/EC require member states to take necessary measures to protection data, but does not state the necessary measures that are required. However, Annex A requires that informed consent must be given by a trained person in an appropriate and clear manner, covering the consequences and risks of transmission. In particular, neither of these provisions answer the question on data subjects can consent to future uses of tissue via broad consent or whether anonymisation is necessary (Favale & Plomer, 2009). Until this conflict with the GDPR is resolved, scientific researchers therefore should be mindful of any divergences in national legislation with respect to data protection and consent laws (Kaye et al., 2016). As discussed previously, the express purpose of the GDPR is to prevent the degree of fragmentation present in previous data protection instruments. However, it remains to be seen to what extent member states will introduce derogations for consent and scientific research. Therefore, both researchers and research ethics committees should consider potential consent requirements before planning any research involving the use of human tissues and cells.

Database Directive:

The *sui generis* database right is a contentious form of intellectual property protection, not in the least because of the investment it is designed to protect. In particular, the Database Directive provides a separate right to copyright protection that does not extend to copyrighted materials. As the next section will discuss, copyright protection in both common and civil law systems requires some investment or creativity to vest in a particular expressed work. Copyright cannot simply vest in facts per se because they lack the requisite creativity or originality. In the US, the Supreme Court in *Feist Publications v Rural Telephone Service Co*[49] established the 'sweat of the brow' doctrine that compilations of data lacked sufficient originality to qualify for copyright protection. Therefore, Feist Publications were entitled to copy telephone numbers from a telephone directory published by Following the *Feist* was heard by the Supreme Court, there was significant concern about ensuring that database creators were adequately compensated for their work.

Accordingly, Directive 96/9/EC was introduced to help incentivise the creation of compilations of data. Protection under the database directive extends to where there has been qualitatively or quantitatively a substantial investment in either obtaining, verifying or presenting the contents of a database. This substantial investment is then used to prevent extraction or reutilisation of the whole or a substantial part. In addition, the existence of a sui generis database right does not extinguish the copyright protection that may vest in the data. In other words, the copyright that vests in the data collected for

[49] 499 US 340 (1991).

the database is separate from the rights that vest in the structure of the database. The data contained within the database might not be original, but the structure of the database must be to satisfy the requirement. For example, data such as taxonomic data or biodiversity information in basic research will almost certainly not qualify for copyright protection because it lacks originality. However, if this data is arranged in an original fashion as part of a database, that database will qualify for protection under the *sui generis* right (Agosti & Egloff, 2009; Egloff, Patterson, Agosti, & Hagedorn, 2014). The *sui generis* right may also extend to compilations of human genomic or biomedical data that would not otherwise qualify for copyright protection.

Article 8 then establishes the terms of use for databases under the *sui generis* regime. First, Article 8(1) notes that the maker of a database cannot prevent that user from extracting insubstantial parts of the database. The insubstantiality of these components can be either qualitative or quantitative. Combined with Article 7, Article 8(2) then states that lawful users are prevented from extracting parts of the database that conflict with the normal exploitation of that database. What constitutes a conflict with normal exploitation very much depends on a case by case basis (Nettleton & Llewellyn, 2009). For example, in *Directmedia Publishing GmbH v Albert-Ludwigs-Universität*, the database in question was a compilation of 1,100 poems published between 1730 and 1933. In this case, the broad structure of the 'database' was protected to balance the relatively narrow protection afforded under copyright. This broad protection was also reflected in the broad protection of the extraction and the reutilisation requirement. Directmedia Publishing argued that because they did not technically copy the database, they had not breached the database right. However, the CJEU rejected this argument on the grounds that replicating functionality was sufficient.

Note that unlike copyright, where the originality of vests in the written work, the originality for the sui generis right vests in the structure of the database. For this structure to be sufficient original, it cannot be adopted purely for technical reasons. This principle was confirmed in the case of *Ryanair Ltd v PR Aviation BV*.[50] PR Aviation operated a website which allowed consumers to identify flights offered by low cost air companies. At the same time, Ryanair also made flight data available on their website. However, Ryanair's website required visitors to agree to terms of service that only allowed the website to be used for private purposes. Further, these terms of service prohibited the use of web scraping software to extract information from the database. Accordingly, the question before the CJEU was whether the database structure attracted sufficient originality for the sui generis right to vest in the database. The CJEU also interpreted Ryanair's contractual provisions in congruence with Article 15 of the Directive, which invalidates any contractual provision that contradicts Article 6 or 8. Therefore, Article 15 would prohibit any contractual terms that prevent the use of extracting minimal or insubstantial parts of the database. Accordingly, counsel for PR Aviation attempted to argue that the database attracted the protection of the *sui generis* right. If they established that the *sui generis* right did vest in the database, they would be able to argue that Ryanair's terms of service breached Article 15. However, the CJEU held that no database right vested in the database, and therefore PR Aviation had breached the terms imposed by Ryanair. Article 9(a) and (b) of the Directive then establish other exceptions to the *sui generis* right**,** namely for private research and scientific research, provided it is for a non-commercial purpose. It

[50] C-30/14 (2015)

is important to note that there has been no case law to determine the scope of this research exception and what qualifies as non-commercial. Therefore, basic research projects (such as text mining) may qualify under this research exception. By contrast, applied or use inspired research projects are less likely to qualify for the research exception (Truyens & Van Eecke, 2014). More importantly, the uncertainty over the extent of database rights with respect to scientific databases may threaten the exchange of open scientific knowledge (Egloff et al., 2014).

Another point of crucial overlap may exist between the rights guaranteed by the Database Directive and the rights of data subjects under the GDPR. A somewhat infamous example of a genomic database is the BRCA database compiled by Myriad Genetics, a spin off company of the University of Utah. Myriad Genetics has been criticised for its aggressive enforcement of its patents on BRCA-2 breast cancer genomic tests. When these patents were invalidated by courts in the US and Australia, Myriad Genetics shifted its business model to analysing the data it had collected from patients. Crucially, all of this data was retained as confidential information, and Myriad denied patients access to vital underlying information. Consequently, four former patients (led by the American Civil Liberties Union (ACLU)) argued Myriad's policies breached the Health Insurance Portability Act (HIPAA) Privacy Rule. These patients brought their complaint to the US Department of Health and Human Services, arguing that they were entitled to access to their data. Although the matter remains unresolved, patients may bring a similar case in the EU based on the rights to access, deletion or data portability (Sobolciakova, 2018). The matter is complicated by the fact that in the EU, an organisation such as Myriad could argue that their genomic database is protected by the database directive. Whether patients could access, transfer or delete their data would depend on whether that extraction or removal would be inconsistent with normal exploitation of the database (Bovenberg & Almeida, 2019). Accordingly, researchers should be aware of the potential database rights that attach to original compilations of personal data when creating or accessing such databases. Further, the decision in *Ryanair* demonstrates that researchers should be aware of the potential contractual conditions imposed on compilations of data that do not qualify for the database right. A possible solution to this conflict may exist via Article 23 of the GDPR. Article 23 allows for a member state to qualify any of the rights contained in Chapter Three where restrictions would be a necessary and proportionate measure in a democratic society. In particular, Article 23(1)(i) permits qualifications to data subject rights to safeguard 'the rights and freedoms of others.' In the context of the right to access or erasure, this exception could be used to provide priority to database rights relative to the GDPR. However, this resolution may not provide an adequate solution if the US Department of Health and Human Services ruled that privacy protecting rights should take precedence over data protection rights.

5. Case studies

The GDPR affects research conducted within research institutions across European Union member states and potentially beyond. To communicate the content of the GDPR to the research community and to make research GDPR compliant, it is anticipated that research societies and institutions in Europe (and beyond) develop information material targeting researchers and academic staff (Koščík & Myška, 2018). Within the research environment, research societies are understood to disseminate research funding and research processes on the national and international level.

To understand precisely how this information is presented to the research community as well as what the content of such information is, this chapter reviews a sample of case studies of universities and research societies from across the European Union. These case studies will be compared with the findings of previous sections in this report in the discussion section to understand how the guidelines of these research societies and universities respond to the scientific, legal and public debate.

As research is, to a large extent, also conducted outside of the academic environment, this chapter presents additional examples of GDPR compliance guidelines as developed by or for the industry sector.

5.1 Methods

An exploratory review was conducted, as suggested by communications with the European Research Council INFO Team, following an initial unsatisfactory attempt to select case studies in a systematic way by reviewing the Times Higher Education top ranked university for each EU Member State as well as reviewing the Science Europe Member Organisations. This methodology was considered appropriate as it appears that universities and research societies rarely publish GDPR related information targeting researchers in the public domain. To accommodate for this obstacle, case studies were identified by searching google.com for 'University AND GDPR' in English, German, French, Italian (languages spoken by the research team). We also purposively searched for relevant GDPR information that is publicly available and addressed to researchers in private research institutions. Since research is not only conducted at universities and research societies, industries were purposefully searched to identify GDPR related information targeting researchers in the public domain. In particular, we focused on industries that are dependent on personal data, such as the healthcare industry.

Case studies were independently searched by two authors. A first selection of case studies was discussed among all authors and narrowed down through an iterative process to the final selection of case studies, as presented below. The broad inclusion criteria are: institution presents GDPR related information addressed to researchers; institution is based in an EU Member state.

The case studies were compared against the following criteria: domain of application; measures taken and measures to take; risks/benefits of the measures taken; open challenges.

5.2 Research societies

As examples of research societies, The Medical Research Council, UK, and Health Research Authority, UK, both provide GDPR guidance for researchers and study coordinators (HRA, 2017; MRC, 2019). These guides are engaging to researchers and are well developed in terms of the information they present. Both refer to information published by the Information Commissioner's Office (ICO). Similar authorities to the ICO exist across the European Union (EDPB, 2019). These authorities offer guidance on GDPR compliance, but are not necessarily aimed at the research community. An example is the Irish webpage http://gdprandyou.ie/ developed by the Irish Data Protection Commission (DPC, 2019), which provides general information about the GDPR and compliance for individuals and organisations. The information is clear and presented with infographics and check lists so that non-legal experts can inform themselves about GDPR in general.

5.2.1 Medical Research Council, UK

Domain of application

The Medical Research Council (MRC) has issued a large variety of guidance documents concerning the implications of GDPR for research (MRC, 2019). For example, in the document 'Key facts for research' MRC, in collaboration with the UK Information Commissioner's Office, provides clarity about actions to take in order to comply with GDPR provision. This document covers a variety of topics:

- Data collection (needs to be lawful, transparent and fair)
- Data usage and storage (any personal data can be used for research and there is no requirement to delete research data)
- Consent to research (researchers do not need to re-consent participants)
- Use of Big Data (the use of big data for research is allowed if adequately motivated)
- Distinction between Genetic Data and Personal Data
- Data sharing (researchers can still share data with other researchers in line with confidentiality requirements)
- Transparency requirements (GDPR includes precise transparency requirements to better inform participants, so protocols need to be respected)
- Informing participants adequately

Benefits

The MRC offers researchers a variety of tools and learning resources, such as animations, quizzes and an interactive Q&A page, where researchers can submit their questions and receive a comprehensive answer. The different formats used to provide the complex information to researchers make the information more accessible. These user-friendly tools may be beneficial to deliver the basic knowledge about GDPR to researchers as well as those that are not familiar with GDPR at all.

Measures taken

In order to allow researchers to better familiarise themselves with GDPR, the MRC provides an animated video that introduces GDPR and explains the lawful basis for research. The MRC developed a quiz for 'any researcher or research support staff who collect, manage, handle or access information about people including those who support research activities, those who supply data to researchers, chief investigators and archivists. It may also be relevant / of interest to research and other governance managers. But it will not directly address all of their specific learning needs'. The learning objective of the tool is not only the practical considerations that researchers need to make to ensure they are working in line with GDPR, but also the general understanding of how GDPR impacts common research practices (e.g. approvals, peer review and other safeguards; the role of pseudonymisation; 'big data' studies). In addition to formal documents, the MRC provides articles and blog posts for researchers. In these documents, researchers can find practical recommendations on how to make sure that the data processed for research is lawful, and how to be fair and transparent to research participants. Furthermore, in the blog it is made clear that GDPR will not drastically impact research: 'The requirements [for GDPR] largely mirror current good practice in research, so shouldn't have a big impact on what you, as a researcher, already do. The new law demands that data processing is lawful, fair and transparent. Organisations that process personal data, or control its processing, are accountable for this, yet we all have a role to play'.

5.2.2 Health Research Authority, UK

Domain of application

The Health Research Authority (HRA), UK, developed a comprehensive guide on the implementation of the GDPR for health and social care research (HRA, 2017). The guide covers key aspects of the GDPR for health and social care research, such as:

- Consent
- Data controllers and personal data
- Transparency
- Safeguards
- Data subject rights
- Data Privacy Impact Assessments

Benefits

The benefits of the information provided is that professionals working in the field of research can find information written in plain and easy language. These guidelines can inform their research and help professionals to conduct research that complies with GDPR.

Measures taken

The HRA states that although the GDPR impact on individual research projects will be rather limited, HRA aims to provide guidance specifically for researchers and study coordinators managing research projects. The guidance includes sections about relevant definitions and key aspects of the GDPR for health and social care research, as well as best practice and questions. These sections highlight the fact that since GDPR is a rich and complicated legal tool, simplifications and explanations are required to make GDPR accessible to researchers without a legal background. The guide also contains a section explaining the practical changes that health and social care researchers need to make in order to comply with the new law. This section covers four main topics in detail (which also emerge as relevant in the scoping review analysis): consent, transparency, safeguards, and data subject rights. In general, this section suggests that the expectations for consent and safeguards (such as confidentiality and security) should already be met by researchers when respecting existing arrangements. Transparency seems to be the hardest principle to comply to under the new GDPR provision. Therefore, in order to help researchers and study coordinators to meet new GDPR requirement, HRA drafted and included in the guidelines a series of transparency wording templates. The templates focus on increasing transparency for public sector sponsors, commercial organisations and charity sponsors, but also include a template email for sponsors sharing a GDPR amendment as well as transparency information for NHS sites. The HRA offers an additional list of scenarios and transparency measures that apply where personal data is provided directly by data subjects. However, in cases where personal data is obtained directly from another controller, it might be not practical for researchers to provide transparency information directly to participants. In such cases, researchers can seek advice from their organisations and potentially rely on provision of transparency research exemptions. Finally, the guideline states that exemptions to data subjects' rights should be considered on a study by study basis. Therefore, the researcher should not offer or limit the participants' rights without taking account of the relevance of the rights to a particular project.

5.3 Universities

Several European universities provide information about how their university, as an institution, complies with the GDPR in relation to institutional data such as staff employment data or student data. However, this information does not include detailed research specific guidance (KU Leuven, 2019).

Some universities, for example, the University of Vienna, Austria and the Ludwig-Maximilians University Munich, Germany, refer to a Data Protection Officer (LMU München, 2019). The Data Protection Officer needs to be contacted in case of doubts concerning how to conduct GDPR aligned research or any other work involving personal data that needs to comply with the GDPR. Further, universities in Europe provide information about the GDPR in form of internal circular letter, yet the content is not as specific and guiding. Also, it seems that some of the guidelines are for internal use only. The University of Freiburg, Germany, provides a GDPR course that can be found via google.com, but the link guides to an internal page requiring log in details.

5.4 GDPR compliance guides for researchers

It appears that UK universities are at the forefront of providing their researchers information material online about what the GDPR entitles and what the GDPR's implications for research are. In this context, several UK universities' websites refer, in addition to their own information, to the Information Commissioner's Office documents about GDPR compliance (ICO, 2019).

The three cases below show three different approaches in terms of design and format of providing GDPR compliance guides for their research community, yet their content is similar.

University of Oxford, UK

The University of Oxford published online information material to help their researchers to understand the GDPR itself and what its implications for research are (University of Oxford, 2019a). This information is not legal advice, and therefore researchers are encouraged to seek specific advice in relation to their research projects.

In addition to the information targeted specifically at researchers, the university also provides general information about privacy and how the university as an institution complies with the GDPR. This information is provided by the University Administration and Services (University of Oxford, 2018) and is helpful to understand how the University complies with GDPR, but is not necessarily as informative for researchers as the information specifically written for researchers.

Domain of application
The online information online is divided into six areas, and provides a link to a Data Protection Checklist (University of Oxford, 2019b) and a Data Protection & Research document (University of Oxford, 2019a), as well as providing contact information for the Information Compliance Team. The six areas are focussed on:

- Data protection; exploring what data protection is and why is it important.
- Scope of the GDPR; when does the GDPR apply to a research project? Identifying whether the research project requires data processing, whether the research project involves personal data, and what is meant by special category personal data.
- Responsibilities under GDPR; explaining who is responsible for GDPR compliance, what the duties and obligations under GDPR are, and any further requirements.
- Transfer of data; if and how data can be transferred to parties outside of the European Economic Area.
- GDPR exemptions; describes the relevant research exemptions from GDPR.
- Practical considerations; highlights practicalities in relation to GDPR and research.

Furthermore, the University of Oxford issued a data protection checklist which applies to four overarching areas: transparency, data minimisation, security and other safeguards. The checklist is

intended to assist researchers while drafting information for participants and consent forms, as well as in completing the sections in the ethics application form referring to the managing and handling of personal and other research data (University of Oxford, 2019b).

Last, the detailed Data Protection and Research document covers the same content as provided online in a single document.

Leicester University, UK

Leicester University provides detailed internal guidance for researchers (University of Leicester, 2019b) as well as reference to research specific GDPR information as provided online by the Health Research Authority (HRA, 2017), the Medical Research Council (MRC, 2019), and NHS Digital (NHS Digital, 2018). Additionally, contact information for the *Leicester Clinical Trials Unit* and working group *Leicester GDPR for Research* is provided online (University of Leicester, 2019a).

Domain of application

The internal documents are titled as follows:

• What you can do to prepare for GDPR • Email management Quick Guide • Individuals Rights Quick Guide • Privacy Notice Quick Guide • Lawful Basis Quick Guide • Privacy Notice Template • Reviewing your files Quick Guide • Student Information Retention Quick Guide • GDPR and Procurement Quick Guide	• GDPR and Consent Quick Guide • Is your team ready for GDPR Quick Guide? • GDPR and Research • Data Quick Guide • GDPR Myth Busting • Data Mapping Quick Guide • OneTrust Data Mapping User Guidance • Understanding Data Breaches Quick Guide • GDPR FAQ

Considering the findings of the scoping review with respect to uncertainties raised by researchers about the possible burden of GDPR for research, Leicester University issued documents covering frequently asked questions and myth-busting. The main concerns covered relate to work flow disruption, data management and administrative burden.

Lancaster University, UK

Lancaster University has developed information material (Lancaster University, 2019), as well as referring to external links such as the ICO (ICO, 2019) as well as EU GDPR.ORG (EU GDPR.ORG, 2019). Compared to the case studies above, Lancaster University provides a more concise guide.

Domain of application

Lancaster University provides GDPR compliance information for researchers in regards to:

- Introducing the GDPR
- Explaining what counts as personal data
- The impact of the GDPR on research
- Legal compliance of data processing
- Explaining action needed to be fair and transparent
- Whether participants in active projects need updates
- Information regarding secondary used
- Responsibility matters

In addition, contact details to the Integrity, Ethics, and Governance team in Research Services and the university's Data Protection Officer are provided.

Comparing the case studies against each other, the benefit of providing this information to professionals working in the field of research is that the information is written in an accessible and engaging language, and so, these guidelines can inform their research and help professionals to conduct research that complies with GDPR.

The risk associated with the dissemination of these guidelines is that they may be interpreted differently. To overcome this risk, the universities encourage researchers to seek specific advice in addition to using the information provided online. Additionally, when comparing the guidelines against each other, there appears to be large overlap when it comes to the areas of applications, but the information differs in detail as well as scope. For example, Leicester University developed a wide range of guides covering research, as well as other areas relevant for the research process, whereas Lancaster University took a more concise approach. However, the question of which approach is more user friendly and expedient remains unanswered in this report. It is also likely that the contact points at the university have particular tasks and roles in the information process about GDPR compliance.

Measures taken and measures to take by Oxford University, Leicester University, Lancaster University

In all the three cases, the universities published online in the public domain a range of GDPR compliance relevant information targeting researchers. As described above, each university took a different approach to visualise and manage this information. In all cases, the information online includes references to external links as well as internal contact persons.

Considering the challenges to find similar information across different universities within the European Union, it is anticipated that it will be necessary to develop such guidelines in a more consistent and harmonised way for all universities and public research institutes located in European Union's Member States. Such guides will also be valuable for universities outside the European Union.

5.5 Short courses for GDPR compliance

The review revealed that university institutions across Europe organised courses, both formal and informal, to prepare staff for the new incoming GDPR legislation. Academic institutions feel the obligation to treat personal data with care and respect, following the provisions of GDPR.

For example, the University of Limerick, Ireland, provides an 'Introduction to the General Data Protection Regulation (GDPR)'. This introductory course, two hours length and free-of-charge, is recommended to all university staff. The course aims to increase awareness among staff members concerning the role of GDPR in strengthening the rights of individuals and regulating the responsibilities of organisations. The course also provides a practical understanding of data protection obligations and requirements as well explaining the necessary steps to comply with GDPR.

The University College London (UCL), UK, also offers an online GDPR training course, which provides staff and students with details about GDPR content and implications for research. As well as providing the theoretical background and descriptions of relevant GDPR aspects to research, this course provides a series of good practices and examples that could be useful to course participants (UCL, 2019). For example, a list of valid and invalid potential requests under Data Protection Legislation is provided during the course[51]. The course is accessible in three versions: (i) readable PDF course book, (ii) audio version, (iii) one-to-one training session with a member of the GDPR programme. Once either option (i) or (ii) is completed, the course participant contacts the GDPR team, and a member of the team will come in person to run and mark the assessment. If option (iii) is chosen, the assessment and grading is included in the one-and-a-half-hour one-to-one session. This UCL course also offers additional material concerning guidance for Data Protection Impact Assessment and for privacy notices, as well as advice on reporting an incident, and on using the 'legitimate interests' basis for processing personal data.

With the introduction of the GDPR, the European research infrastructure SoBigData launched a free online course, the FAIR course (First Aid for Responsible Data Scientists), aimed towards data scientists and all those who work with data administration. The course provides participants with the basic elements of the new regulation and suggests ethical considerations that need to be asked by scientists. The course covers an overview of the main ethical issues of data science, presents the main content of the GDPR, including the obligations of a data controller, 'privacy-by-design', and the law on intellectual property. At the time of this review, three modules of the course are accessible: 'Ethical framework for data scientists', 'A journey through the new European Data Protection Law' and 'Intellectual property rights and social media content'. Each module is accompanied by a final test that provides immediate feedback on the level of competence achieved. In the coming months, SoBigData plans to expand the

[51] These are: 'I want to see a copy of my HR file'. Correct, this is the right of access.
'My details are wrong. Please correct them'. Correct, this is the right to rectification.
'I would like copies of previous versions of the HR policy'. Incorrect, this does not involve personal data so would not be handled under Data Protection Legislation.
'Please remove my personal information from SITS'. Correct, this is the right to erasure.
'Do not disclose my personal data to the Mr Jones'. Correct, this is the right to restrict processing.
'Give me a copy of UCL's annual accounts'. Incorrect, this does not involve personal data so would not be handled under Data Protection Legislation.

course on offer with other specific modules, as a summary of the main techniques to manage data anonymously.

5.6 Research and innovation in the industry sector

The healthcare industry is particularly dependent on personal data (BCG, 2012). Therefore, this review targeted preliminarily examples from the healthcare industry. However, it appears to be difficult to find GDPR compliance guidelines online in the public domain that are written specifically with industry research in mind. There are broad guidelines aimed at explaining the GDPR to industry professionals, such as the guidelines issued by the German Bundesverband der Deutschen Industrie (BDI, 2017), and proposals by consulting companies such as McKinsey & Company which also serve to advertise their services in this area (McKinsey & Company, 2017). Other activities supporting researchers in navigating the GDPR include, for example, the European Federation of Pharmaceutical Industries and Associations which held a workshop on this issue and reported that National Data Protection Authorities, Health Authorities and Ethical Committees have different views about the interpretability of GDPR with respect to research. Similarly, MedTech Europe, a European alliance of medical technology industry associations, hosted a workshop concerning data protection in health research with a specific focus on clinical trials (MedTech Europe, 2018). The workshop acknowledged that post-GDPR, considerable uncertainty is introduced concerning the role of real-world data for health research, clinical trial sponsors and sites, and consent of the data subject to data processing. Therefore, a coherent guide is needed to tackle and solve emerging uncertainty.

It can this be concluded that while various industry associations discuss GDPR compliance and have taken steps to develop guidelines, to date, systematic guidance on how the GDPR should be interpreted is still missing.

6. Media analysis

Within the scope of this report, the authors sought to collect evidence on the public's perception of the General Data Protection Regulation (GDPR) and in particular its impact on scientific research. Understanding public opinion is of great relevance due to the influential effect it has on public policy-making, especially in cases where an issue is relevant to the public (Burstein, 2003) – as is the case with the GDPR. Public opinion is thus a crucial factor to take into consideration when discussing the potential challenges associated with the implementation of the GDPR and how they are likely to affect the design and conduct of scientific research in Europe.

A common method to assess public opinion is the conduct of a representative survey. However, given the limited timeframe and availability of resources, a representative survey was not practical nor feasible. Instead, a news media analysis was conducted to investigate whether and how the impact of the GDPR on scientific research is portrayed by European news media. Given the constraints, this was deemed the most appropriate approach to generate a better understanding of the public's perception of the GDPR and its impact scientific research.

The underlying rationale for this approach is set out by agenda-setting theory (McCombs & Shaw, 1972), whose theoretical foundation draws on psychological concepts of priming as discussed in work on cognitive processing of semantic information (Scheufele, 2000). In its original formulation, agenda-setting theory postulates that the amount of news media coverage on a particular issue has an impact on how salient said issue is perceived by the public. In other words, the more a topic is present in the news media, the more this topic will be present in the minds of the public. According to agenda-setting theory, media are thus considered to be actively shaping public priorities, rather than merely mirroring them (Brown & Deegan, 1998), implying a causal relationship.

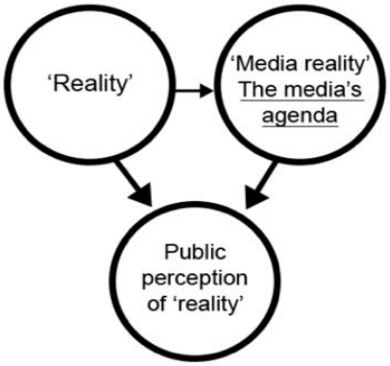

Figure 1. Agenda-Setting Theory (McCombs & Shaw, 1972)

While earlier work has argued that the media may not necessarily be able to influence *what* people think but merely determine *what they think about* (Bernard, 1963), this may in fact not hold true for the case of the GDPR. Indeed, it has been suggested that the less direct experience people have with a respective issue, the more likely they are to rely on news media for information and interpretation of that issue (Gene Zucker, 1978). Here, the concept of framing, which has been described as natural extensions of agenda-setting theory (Scheufele, 2000), needs to be briefly introduced. Framing refers to the way in which a particular issue is portrayed or *framed* by the media. In contrast to agenda-setting, which is concerned with the salience of issues, the concept of framing focuses on the salience of issue attributes (Scheufele, 2000).

In the case of the GDPR, it is reasonable to assume that only a small part of the public has direct experience with the GDPR in the context of scientific research. This, in turn, suggests that the media are

likely to play a crucial role in shaping public opinion in relation to the GDPR and its impact on scientific research by framing the issue in a particular way. For the purpose of this research, news media coverage is thus considered a proxy for public opinion.

6.1 Methods

A qualitative content analysis of news coverage around the GDPR was conducted in January 2019. News media articles were identified through NexisLexis, an online research database that encompasses international daily and weekly press. The search was performed by two researchers using the following search string: (('gdpr') or ('general data protection regulations') and ('research')) and date geq (2016-jan-01) and date leq (2019-jan-22).

A total of 476 news media articles were exported to an excel sheet for screening and analysis. Upon removal of duplicates, 455 articles remained to be screened for inclusion. To be included in the analysis, articles had to meet the following inclusion criteria: (i) news media article published in English; (ii) between 01.01.2016 and 22.01.2019; (iii) in a major European publication (see Appendix); (iv) covering aspects related to the GDPR and its impact on scientific research. Two researchers were involved in the screening and coding process to ensure intercoder reliability. For this purpose, a preliminary codebook was developed and continuously refined. The codebook is presented in Table 4.

Table 4 - Media analysis codebook

Themes	Description
Need for the GDPR	Articles covering issues related to - questionable practices of data collection/storage pre GDPR
Implementation	Articles covering issues related to the implementation of the GDPR - Challenges to implementation - Facilitators to implementation
Impact	Articles covering issues related to the impact of the GDPR - Positive impact - Negative impact

6.2 Results

A total of 17 news media articles from eight publications were included in the final analysis (see Table 5 and Figure 2). It is important to note that the majority of news media articles that were screened centred around the impact of the GDPR on marketing research practices in the private sector, addressing issues, such as, the (potential) negative consequences of the GDPR on marketing research.

In relation to the GDPR and its impact on scientific research, most of the articles under investigation addressed more than one theme, many of them referring to health-related data. News media coverage centred relatively uniformly around three themes: the need for GDPR, the implementation of GDPR and the impact of GDPR. Within which we identified five subthemes: (a) questionable practices of data processing pre GDPR, (b) facilitators of GDPR implementation, (c) the challenges involved in GDPR implementation, (d) negative impacts of GDPR, and (e) positive impacts of GDPR.

Table 5 - News media articles included in the final analysis, organised by date

No.	Article Title	Article Source	Date
1	National genome service	New Scientist	30.08.17
2	How AI could kill off democracy	The New Statesman	15.08.18
3	Is it time we kicked the social media habit?; As the Cambridge Analytica scandal plunges Facebook into crisis and more revelations emerge about data harvesting online, Richard Godwin says that there's growing evidence that online platforms actually make us unhappy	Belfast Telegraph	31.03.18
4	Ireland needs a national strategy for artificial intelligence	The Irish Times	28.06.18
5	Searching for answers in our genes; Prof Gianpiero Cavalleri, associate professor of human genetics and deputy director of the SFI Future Neuro Research Centre of Excellence at RCSI	The Irish Times	27.12.18
6	Facebook fined pre-GDPR maximum of ¬£500,000 by ICO over Cambridge Analytica	Computing	25.10.18
7	Cognitive industrial revolution is set to change everything	The Irish Times	25.05.18
8	Should your medical data be off the record?; Information can help save lives, but there is a debate as to whether it should be used elsewhere without consent	The Irish Times	24.08.17
9	University CIO: 'If I had a pound for every time I heard a piece of software can make you GDPR compliant...'	Computing	23.05.18

10	Outside the EU, Britain would be free to spend its money how it pleased; Letters to the Editor	The Daily Telegraph (London)	20.04.16
11	GDPR: Should you keep potentially useful data or delete it?	Computing	19.09.17
12	Why Brexit is dire news for research into cancer	Belfast Telegraph	15.06.16
13	Europe's Data Protection Law Is a Big, Confusing Mess	International New York Times	15.05.18
14	Facebook's fine 'could have been hundreds of millions'; Business Technology Intelligence ICO says scandal was serious breach that would have been dealt with severely under the new data regulations, reports Margi Murphy	The Daily Telegraph (London)	12.07.18
15	Can Facebook clean up its act?; The social media giant has assembled a team of experts to spot abuses and protect the company's reputation. But what do security experts	The Observer (London)	07.07.18
16	The healthcare CIO: Leading the transformation	Computing	02.11.18
17	UK data science can lead a global privacy-first mantra	The Daily Telegraph (London)	01.10.18

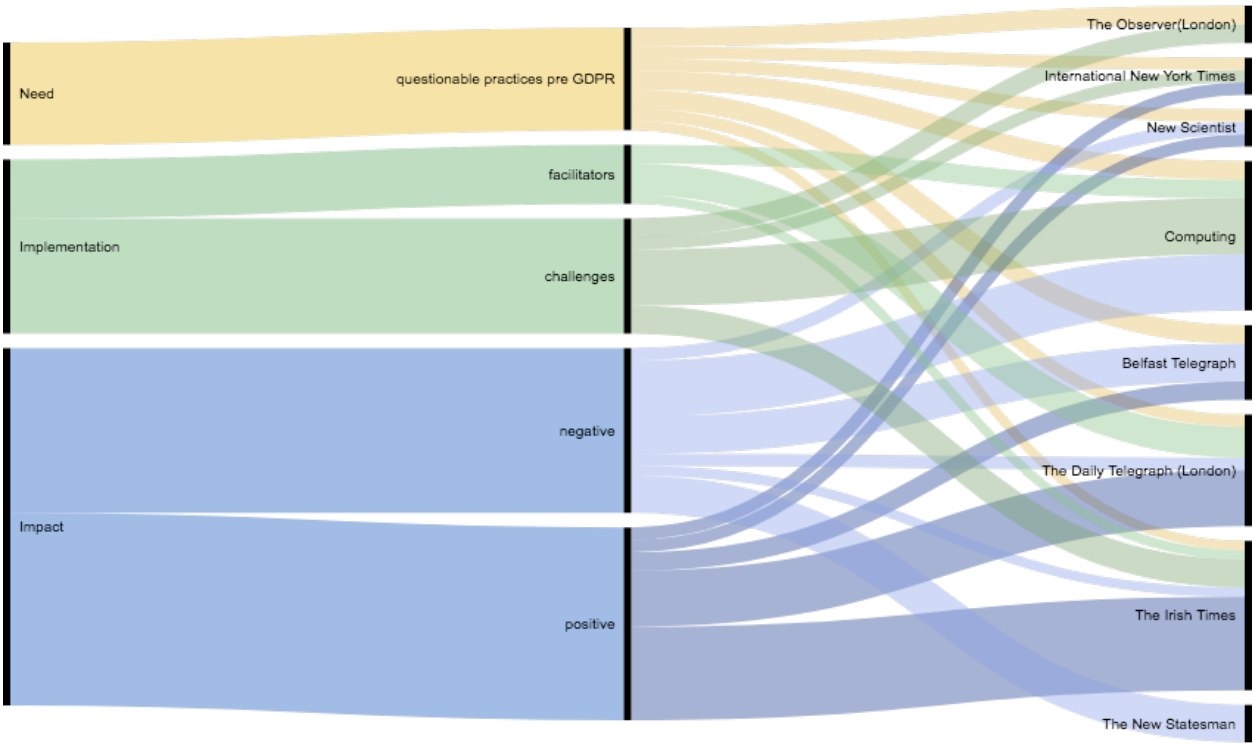

Figure 2- Alluvial diagram representing the correlations between themes, subthemes and sources of articles identified as relating to scientific research

Need: questionable practices pre GDPR

Articles coded as 'need' and subtheme 'questionable practices pre GDPR' cited cases where personal data had been misused by the research community or private sector. Such articles often listed instances of data misuse - 'the Trust had 'failed to comply with data protection law' and hadn't adequately informed patients as to how their data was being used. More widely, in July 2016, NHS England finally pulled the plug on its disastrous care.data programme.' (New Scientist, 2017). Articles also referred to events surrounding the Cambridge Analytica scandal 'Officials said they intend to audit of Cambridge University's psychometrics unit over concerns academics were using manipulative psychology techniques, applying them to Facebook data and possibly selling this to political consultancy groups, ahead of the European referendum.' (The Daily Telegraph London, 2018). It was emphasised how academic research may be used as a disguise to enable political parties to harness internet users' personal information: 'The defunct political consultancy was able to harvest information about millions of Facebook users without their consent when a third-party used a Facebook API to run an app called 'This is Your Digital Life', supposedly for academic research purposes.' (Computing, 2018).

With the impact of data misuse being the degradation of trust - 'The biggest risk for the sector, however, is probably the very problem that has sparked Facebook's current bout of introspection: trust.' (The Observer London, 2018). Additionally, the issue was portrayed as being collective– 'There is a growing realisation that our data is under attack.' (International New York Times, 2018).

Implementation: the challenges

Several articles presented information in relation to the challenges associated with the implementation of the GDPR, discussing factors that complexify being GDPR compliant. As one article put it - 'It's a challenge right around the world for research.' (Computing 2017). At the core of implementation challenges was the issue of the legislation being complex and not comprehensibility. 'No one understands the G.D.P.R. The law is staggeringly complex.' (International New York Times, 2018). The ambiguity of the legislation was also emphasised – 'What are often framed as legal and technical questions are also questions of values.' (International New York Times, 2018). More specifically, the 'day-to-day' implementation of the GDPR was identified as problematic – 'legal frameworks, particularly when they are long, complex and ambiguous, can't be the only or even the primary resource guiding the day-to-day work of data protection.' (International New York Times, 2018). In specific to the research community, the challenge of implementing GDPR was related to domain specific consent procedures - 'Getting informed consent from someone who donates their DNA is much trickier, as it isn't possible to know how an individual's data will be tested or used in the future, says Montgomery.' (New Scientist).

Implementation: the facilitators

Some articles presented information in relation to factors facilitating the implementation of the GDPR. In this context, one article referred to the leading role of UK in the field of privacy techniques 'Leading academics and researchers have been working on privacy techniques for over a decade. As home to some of the world's best and brightest privacy expertise, the UK has a great opportunity to take a leading position.' (The Daily Telegraph London, 2018). Reporting on the University of York's journey towards GDPR compliance one article referred to the installation of 'a project group for GDPR compliance' and 'excellent engagement across the organisation' (Computing, 2018). Similarly, another article presented a researcher's view on the need for 'a national response to GDPR', including guidance on the question of consent for processing of personal data, as well as, the need for individual institutions 'to develop their own code of conduct as to how they approach data protection from the beginning of projects rather than having it as a kind of tick-box exercise at the end of a project' (The Irish Times, 2017).

The role of authorities in providing guidance to universities was also addressed by another article: 'The higher education group said it would convene a working group to consider the privacy and ethical implications of using social media data in research, both within universities and in a private capacity.' (The Daily Telegraph London, 2018), as was the importance of collaboration with authorities 'A spokesman for the university told the NS Tech website yesterday: 'We will continue to cooperate fully with the Commissioner and will work with Universities UK as it explores the issues within the Higher Education sector around the emerging field of research using social media data.'' (The Daily Telegraph London, 2018).

Impact: the negative impact of GDPR

Many of the articles presented information related to the adverse effects the GDPR has or may have on the research and science community. As one article states, the GDPR has 'provoked significant concerns

on its potential (albeit inadvertently) to undermine clinical and translational research.' (Belfast Telegraph, 2016). Particular concern was given to academic research conducted using historical and longitudinal datasets and 'whether they will have to delete them on the grounds that they will not have the appropriate standards of explicit consent post-May 2018 to retain them' (The Irish Times, 2018). Here, one article emphasised the need for exceptions 'There will have to be exceptions. Let's say you're looking at immunisation rates, you have to have kept data over a long period of time. There'll be personal data you need to keep to so you can have comparisons over time.' (Computing, 2018.)

In other cases, articles detailed the consequences of non-compliance, specifically in regard to the amount institutions and organisations could be fined. 'Failure to protect data, or general non-compliance, can now lead to a penalty equivalent to 4 per cent of a company's global revenue.' (Computing, 2018.) One article also touched upon the hinderance of GDPR to progress and innovation highlighting the competitive advantage of non-European institutions that face comparatively few restraints when it comes to data usage. – 'There aren't pesky privacy or data protection laws to slow them down, such as the new GDPR rules in Europe' (The New Statesman, 2018).

Impact: the positive impact of GDPR

Articles identified as presenting the theme 'Impact' but rather in a 'positive' light, discussed how, due to the GDPR, public concerns and individuals' rights to privacy were being both acknowledged and protected by governments. 'The European Union has recently led the citizen's right to data privacy and right to be forgotten, through our GDPR legislation.' (The Daily Telegraph, London). Detail was also given as to how exactly governments responded with local legislation. 'GDPR has made everyone rightly aware of the importance of data privacy, and Ireland has drawn up a set of regulations for health research specifically.' (The Irish Times, 2018).

Articles particularly emphasised how the GDPR supports and promotes European rights to personal data and informational determinism, specifically in the context of health research - 'intensive lobbying by ministers and the research community has given us a final version that protects patients' sensitive health data while ensuring that vital research can continue.' (The Daily Telegraph London, 2016). As well, in these articles GDPR was portrayed as the key driver to a 'needed' change in data culture. 'The new law could replace a corporate-controlled internet with a digital democracy.' (International New York Times, 2018). GDPR was framed as a means to achieve interoperability, transparency and accountability. 'The new EU regulation would bring harmonisation, transparency and accountability to a very dense and complex area.' (The Irish Times, 2017)

7. Limitations

Each study component of this report presents limitations:

Scoping review: As any analogous review, this study component is subject to a potential risk of selection bias in the literature retrieval process. While we screened multiple scientific databases, additional entries might have been found in non-scientific repositories and/or additional scientific databases not included in our review protocol. This limitation was minimised in a twofold manner. First, unstructured searches of online search engines were performed to retrieve possible entries that were not indexed in the selected scientific databases. Second, citation-chaining was performed to retrieve relevant literature that was not identified through systematic methods. Another limitation is a potential selection bias pertaining to the language of selected articles. While we included articles written in English, German, Italian and French (languages known by the authors) into the screening phase, articles in other languages of the union were not included. Furthermore, since the keywords of our search were in English, articles that did not have a title, abstract or keywords in English might have not been retrieved.

Doctrinal analysis: As for any legal analysis, this study component is subject to a number of limitations. First, the scope of the case law included in the analysis was limited to EU, UK, and Irish precedent. The decision to focus on EU, UK and Ireland was made because divergences in national data protection law prior to the GDPR could undermine the interpretative aspects of the analysis. Further, it was reasoned that any significant decisions involving the impact of data protection law on scientific research would be appealed to the CJEU. However, this exclusion criterion could have also potentially removed cases from other jurisdictions that might have considered data protection law in the context of scientific research. Second, related to the first limitation is the relative absence of case law with respect to scientific research. In the case of the GDPR, this absence was expected given the relative recency of the new regime. However, with respect to the former Data Protection Directive, the majority of case law analysis utilised precedent that did not involve scientific research. These include cases used in the interpretation of consent-based processing, categories of personal data and data subject rights. By contrast, the scientific research exceptions were examined using textual analysis or with reference to academic literature. Therefore, it is possible that a future court may adopt a more pragmatic or proportionate interpretation of provisions of the GDPR than in this report.

Case studies: Following the constrains of the low number of applicable case studies, the findings of the case studies are based on a small sample. Hence, the generalisation or transfer of findings is not possible. The low number of case studies might be the result of a search language bias, as the research team did not cover all European languages. The main limitation is that institutions which might have developed guidance and training of their staff on GDPR matters may have done so internally and without making information publicly accessible, in which case we would not have been able to review the relevant material. A solution to this would have been utilising a different methodology whereby a questionnaire would have been sent to various institutions asking them to provide such information. We were unable to proceed with such an approach due to the tight schedule between the STOA call and the deadline for

delivery of the report. However, we still believe even this small number of cases provides substantial information about possible actions taken by institutions and their relevance and quality.

Media analysis: A number of limitations must be taken into consideration when interpreting the findings of this study component. First, only English language sources were included in the search for media coverage. Given that there are twenty-four official languages within Europe, results cannot be generalisable. Language bias also can account for why the majority of news media sources came from English speaking countries. Another bias was that only online news media was included in the search. Other potentially important sources of information for the public, such as social media channels and television, were not taken into consideration. Finally, although this study used news media coverage as a proxy for public opinion, the hypothesis generated within the scope of this research would need to be tested in a representative population survey to increase its robustness.

8. Discussion of the findings

At the time of writing, the GDPR has not been implemented for a long enough period to conduct a post hoc impact assessment. Significant effects of policy interventions only become observable and rigorously measurable after longer periods of time. The vast majority of EU member states (as well as EFTA countries such as Norway and Switzerland) have recently derogated or are still in the process of derogating the GDPR into national legislation. By contrast, our case studies identified a few embryonic measures that research institutions have taken to date in response to the GDPR (see section B). However, these are isolated cases and are hardly generalisable to the entire scientific community.

While post hoc impact assessments would be premature at this point in time, our findings provide a solid informational basis for a comprehensive ex ante impact assessment. Each of our four study components provides relevant proxy information about the expected impact of the GDPR on scientific research from the perspective of a relevant stakeholder category. These stakeholder categories include research scientists (scoping review), legal experts and public officers (doctrinal analysis), and the media (media analysis). Each of these stakeholder categories plays a critical role in shaping the public debate on GDPR and provides a critical perspective for assessing the impacts of this new regulation.

Our analysis delineates a diverse, multifaceted and complex impact scenario. As our findings illustrate, GDPR is expected to have divergent or even contrasting impacts depending on the domain of science and the type of scientific activity under consideration. In the next subsection, we will summarise which aspects of scientific research are expected to be enhanced by the GDPR based on our findings. Subsequently, we will discuss which aspects of scientific research are expected to be negatively affected or remain unaffected.

8.1 Potential impacts

The doctrinal analysis and the scoping review identified in the GDPR a number of areas of normative ambiguity and potential concern from the perspective of researchers. If not adequately addressed, these ambiguities and concerns could have a serious negative impact on scientific research. These negative impacts include: regulatory ambiguities raised by the new rights and obligations introduced by the GDPR, their compatibility with research requirements for obtaining consent and additional research burden for researchers and research institutions.

As discussed within the doctrinal analysis, the processing of personal data under the GDPR must be justified under a particular lawful ground. Whilst the default lawful ground for processing is consent, processing for public interest reasons is also provided as a ground for processing under Article 6. Furthermore, research is provided as an explicit ground for the lawful processing of sensitive categories of personal data pursuant to Article 9. Article 89(1) then provides the positive obligations that are incumbent on processors and controllers that seek to rely on exceptions in the GDPR for research processing. However, Article 89(1), in concert with Recitals 156, 157, 158, 159, 160, 161 and 162, provides a broad definition of different categories of research. These can include scientific research, historical

research, archiving in the public interest and statistical processing. On the one hand, this broad definition is designed to ensure sufficient flexibility for researchers to conduct a wide range of scientific research. On the other hand, this definition permits commercial scientific researchers and private research companies to rely on the exceptions provided by Article 89(1). These researchers may not comply with or have in place the same ethical and institutional safeguards as publicly funded academic researchers. More worryingly, the definition of 'statistical purposes' in the GDPR appears to include but not be limited to statistical processing for scientific research. Effectively, this broad definition would permit private companies conducting non-scientific research, such as marketing surveys, to rely on this exception.

Furthermore, there are concerns about the scope of the exceptions subject to the positive obligations imposed by Article 89(1). Firstly, there are two exceptions to the purpose and storage limitation data processing principles in Articles 5(1)(b) and 5(1)(e) of the GDPR. In concert, these exceptions allow data processors and controllers to retain and reuse data for research purposes beyond those specified at collection for longer than required to conduct processing. Secondly, Article 14(5)(b) creates an exception to the right of information for data that has not been collected by the data processor or controller. This exception allows a data processor or controller to refuse requests for information where data has been collected from a third party. Thirdly, Article 17(3)(d) creates an exception to the right of erasure for research where erasure would render impossible or seriously impair the purpose of research. Finally, Article 21(6) creates an exception to the unconditional right of objecting to processing where processing is necessary for the performance of the research in question. Further, Article 89(2) permits member states to derogate exceptions to the rights of access, rectification, restrictions on processing and objection into national law. In isolation, each of these exceptions is subject to the qualifications that they must be necessary or render impossible research. However, in concert these exceptions could be used to effectively deny data subjects their rights with respect to the GDPR. For example, logically it would be impossible for a data subject to request erasure of their data or object to processing without being able to access their data. These exceptions could encourage researchers to collect and process large quantities of data from data subjects without informing them of the purpose for processing.

The reverse problem may emerge when scientific researchers attempt to rely on consent as a lawful ground for processing. This problem is due to an interpretative conflict within the GDPR as to how consent must be given by the data subject for processing to occur. The GDPR operates on the presumption, via Article 4(11), that explicit, freely given and informed consent is the default requirement for the processing of personal data. However, this requirement primarily applies to the processing of personal data in a commercial or employment context. By contrast, the purpose of consent in a scientific research context is to act as a safeguard against abuse of the scientific method. In addition, at the beginning of a scientific research project, it may be impossible for the data processor to fully identify the reasons for which the data is being processed. To this end, Recital 33 of the GDPR permits researchers to seek consent for certain aspects of scientific research that is consistent with the ethical standards in the relevant field. However, this interpretation is in conflict with the opinion on consent under the GDPR issued by former Article 29 Working Party. In its 2017 guidelines, the Article 29 Working Party held that consent under the GDPR had to be sufficiently specific to allow a data subject to exercise their right of

revocation. The Article 29 Working Party went further and held that for scientific research, Recital 33 did not dispel the requirement for informed consent, and instead allowed it to be described in a more general fashion.

One solution to this conflict identified in the scoping review involves the development of dynamic consent as a means to gather consent from data subjects to conduct specific research. This dynamic consent could be achieved through technical means, such as email reminders sent to participants to allow them to opt in and out of consent. In the alternative, dynamic consent could allow participants a degree of decision-making freedom about when to consent to research. Claims that the GDPR might undermine the development of dynamic consent models seem insufficiently corroborated, hence will unlikely result in negative consequences for research projects that are planning to utilise dynamic consent. Interpretative ambiguities over consent also extend to the GDPR's interaction with other European regulatory instruments. In particular, the Clinical Trials Regulation, which will come into force in 2020, provides a separate standard of consent to the GDPR. Although the European Data Protection Board provides some clarity as to the supremacy of the GDPR's consent standard, this interpretation is complicated by national derogations. This requirement could potentially undermine compatibility with the GDPR and require review. There is also a need to update the Tissue and Cells Directive to ensure that it also supplies a compatible standard of consent with the GDPR. Researchers must also ensure that they have obtained the correct form of consent for a clinical trial or use of data for secondary research. These considerations highlight the need to recognise the GDPR as part of a broader regulatory framework for scientific research.

There are also concerns with respect to the definition of personal data and sensitive categories of personal data, as well as anonymised and pseudonymised data under the GDPR. With respect to the first category of data, the main area of concern relates to personal data drawn from inferences. The Court of Justice of the European Union (CJEU) has defined inferences connected to the data subject's personal data as also personal data. By contrast, inferences which are drawn using other sources, such as legal sources, do not count as personal information. The CJEU has argued that these inferences are available under the right to access to documentation, as opposed to legislation protecting personal information, and should not be treated as such. The impact that this distinction may have on research is yet uncertain and should be considered by the STOA Panel. In particular, some (but not all) big data algorithms rely on drawing previously unsuspected inferences using large quantities of data. Although these data may not necessarily be personal data, it may be possible to draw tangible conclusions about individuals. These conclusions would possibly qualify as personal data under the GDPR. This conclusion is supported by Recital 26 of the GDPR, which suggests that identification can include indirect identification through singling out. The question of singling out is pertinent to the definition of sensitive personal data, particularly genetic, health related and biometric data. On the one hand, genetic data that can be used to directly identify a person, such as whole genome sequence (WGS) data, will unquestionably be personal data. On the other hand, the question of identifiability becomes more complicated when considering genomic data that may or may not be used to identify a person. This data may include single nucleotide polymorphism (SNP) or short tandem repeat (STR) data. The question of identifiable data may

therefore need to be considered on a case by case basis. Care should also be taken to ensure that any genomic data is not only compatible with the GDPR, but other data protection regimes, such as HIPAA.

The question of singling out is also related to the definition of anonymised and pseudonymised data. Recital 26 notes that the GDPR does not apply to anonymised data. By contrast, pseudonymised data still counts as personalised data. Whilst anonymisation might be attractive as a means to bypass the GDPR's operation, there is still significant uncertainty as to what technologies will qualify as anonymisation. In particular, the former Article 29 Working Party's report in 2014 into anonymisation techniques suggests that no form of anonymisation can completely prevent reidentification in all circumstances. Attempts to anonymise data may also deprive researchers of valuable data, such as socio-economic data for use in longitudinal health trials. Accordingly, a determination of whether anonymisation or pseudonymisation should be made with respect to the context in which anonymisation or pseudonymisation is used. The STOA panel (along with the EDPS or the EDPB) should provide guidance as to which forms of anonymisation are appropriate for different fields of scientific research.

Outside of these negative impacts, the remaining ambiguities identified within the GDPR are likely to have a less severe impact or neutral impact on research. First, there is an ambiguity within the GDPR with respect to the standard required for the data accuracy processing principle pursuant to Article 5(1)(d). This ambiguity springs from the fact that some translations of the GDPR introduce a requirement for accuracy. Further, the UK courts have defined the accuracy processing principle as imposing a requirement to take reasonable steps to ensure the accuracy of data. By contrast, other interpretations of the GDPR impose a requirement of correctness. The problem with the latter interpretation is that the definition of correctness may vary between different disciplines. For example, the requirements for data to be correct may be significantly higher for biomedical research than they may be for anthropological research. However, this ambiguity does not necessarily need to be resolved through legislative reform. For example, it may be possible for research institutes and scientific organisations to work together to develop an appropriate standard for data correctness for their field.

Secondly, there are conflicts between, and ambiguities in, some of the rights guaranteed by the GDPR for data subjects. In particular, there is ambiguity with respect to the operation of Article 22, which grants data subjects the right to opt out of automated profiling. Because many scientific research projects rely on big data analytics to process personal data, this processing may be subject to Article 22. Further, because there is no exception with respect to scientific research and Article 22, there is no way for scientific researchers to avoid the obligations of this section. However, Article 22 only applies in situations where automated profiling is the sole measure used for a particular decision. Therefore, the scope of the automated profiling right is likely to be limited. For this reason, scientific research societies and institutes should work together to develop appropriate standards of automated profiling in particular fields of scientific research.

Nevertheless, based on our study findings, GDPR is expected to enhance a number of aspects of scientific research. These include data security, regulatory clarity regarding processor responsibilities

and transfer of data, collaborations within the EU, autonomy and trust of data subjects. As stated in Article 25 of the GDPR, one of the key changes elicited is that of privacy by design and by default. Researchers will now be obliged to consider data privacy considerations throughout project design and all stages of the research project along with the lifecycle of the relevant data process. This requirement imposes an obligation on controllers and processors:

> *'the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects' (GDPR, Article 25, §1).*

These requirements of stricter, proactive and more comprehensive data privacy and security requirements will likely enhance security of research data, the privacy of data subjects and reduce the chance of unreported data breaches. Protecting the privacy and security of personal data, especially health information, is a major obligation of research institutions. A US study revealed the occurrence between 2010 and 2013, of 949 breaches affecting 29 million records of protected health information. The same study predicted that:

> *'[with the widespread availability of] cloud-based services provided by vendors supporting predictive analytics, personal health records, health-related sensors, and gene sequencing technology, the frequency and scope of electronic healthcare data breaches are likely to increase' (Liu, Musen, & Chou, 2015).*

As cloud-based data storage, predictive analytics, electronic healthcare records and gene sequencing are all increasingly widespread technologies in Europe, it is reasonable to expect an increased incidence of data breaches involving health and other research-related information within the EU.

The findings of both the scoping review and the doctrinal analysis indicate the privacy and security provisions of the GDPR will reduce the frequency and severity of unreported data breaches. In particular, there may be a significant reduction in breaches involving research data, especially health information, in Europe. This prediction is currently open for early empirical verification. A statement released by the European Commission on January 25, 2019, claimed that there have been 41,502 data breaches reported since May 25, 2019. DLA Piper, a multinational law firm based in London, contended that the total number of breaches could be bigger as results above only account for voluntary contributions of 21, not all 28, EU member states. The firm reported over 59,000 data breaches since the GDPR's inception[52]. It is important to consider that a reduction in unreported data breaches does not necessarily imply a

[52] See: https://www.dlapiper.com/en/uk/insights/publications/2019/01/gdpr-data-breach-survey/

reduction in data breaches. In fact, the GDPR compels data controllers and processors to disclose any breaches of security without undue delay and within 72 hours to local data protection authorities (GDPR, Article 33). This requirement will likely to lead to a rise in reported data breaches, simultaneously reducing the number of data breaches that remain unreported. Some data protection authorities have highlighted this possible positive correlation between the rise in reported breaches and the introduction of the new data breach notification regime. This has been suggested, for example, by the UK Information Commissioner's Office (ICO) statistics. In December 2018, the ICO announced that it had received over 8,000 reports of data breaches since the GDPR took effect.

As a consequence of its capacity to harmonise national data protection legislations, the GDPR has the potential to drive the institutional convergence of data protection strategies. The doctrinal analysis identifies the transfer of data requirements in Chapter 5 as one area where the GDPR has significantly increased the prescriptiveness of European data protection law. Under the Data Protection Directive, the only mechanisms for transferring data were an adequacy decision for the recipient jurisdiction's data protection legislation, or explicit consent from the data subject. In the alternative, Article 49 provides a limited set of scenarios where data may be transferred without an adequacy determination. However, Article 49 does not provide explicit grounds for transferring data for scientific research purposes. Accordingly, the use of this exception for transfers for scientific research will be limited to circumstances where explicit consent has been sought from data subjects for the transfer. Yet another alternative exists pursuant to Article 46, where transfers of data may be permitted where the processor or controller offers equivalent protection via contractual mechanisms. This exception, pursuant to Article 46, provides a significantly more certain foundation for the transfer of data. Further, the introduction of specific organisational safeguards as an avenue for the transfer of data between organisations might encourage institutions to elaborate upon their own transfer agreements. The capacity to harmonise national data protection legislations is also expected to facilitate research collaboration among research groups within the EU. These include multi-national research collaborations and research consortia within the European Research Area (ERA), as provided for by the Framework Programmes for Research and Technological Development created by the EU. These collaborations will likely benefit from the GDPR in terms of intra-consortium data transfer and harmonised data protection rules. This potential benefit is particularly important in a research ecosystem that increasingly relies on cross-national research collaborations and data-driven science such as the ERA. Large-scale research initiatives such as those in the Horizon 2020 framework are on the frontline of research projects that will likely benefit from the GDPR.

Several provisions of the GDPR could generate positive societal effects by improving public perceptions of research. These include the privacy and security requirements introduced by the GDPR. They also include the requirements demanding affirmative consent from data subjects and the disclosure of information about where personal data are collected from the data subject and for which purpose. In particular, the GDPR could help promote trust in data-donation and sharing for scientific purposes among European citizens by empowering them with increased control over their data. In turn, data subjects would have a more affirmative role in data-related decision-making, enhanced data security safeguards and more transparent information about data uses. In addition, public trust is recognised as

'crucial for ensuring the legitimacy of current practices and systems of governance' (Aitken, de St. Jorre, Pagliari, Jepson, & Cunningham-Burley, 2016) and might be sustained through multiple facilitators (Adjekum, Ienca, & Vayena, 2017). Research has demonstrated that increased awareness about data uses and increased control of data subjects over their data are all predictors of increased trust in data-sharing and the participation of citizens in research. Likewise, minimisation of data abuses and associated harms, as well as measures to ensure confidentiality are considered predictors of increased trust (Aitken et al., 2016; Oliver et al., 2012).

While the other study components yielded a number of potential scenarios in relation to the impact of the GDPR on scientific research, news media coverage on this topic was scarce. Indeed, our analysis of major European news media indicates that the majority of news coverage in the selected time period focused on the impact of the GDPR on businesses. Accordingly, many of the news articles initially screened for eligibility centred around issues related to the collection and use of customer data for marketing research purposes, whilst few articles referred to the impact of the GDPR in the context of scientific research. This may, in turn, indicate that there is a lack of public awareness with regard to the impact of the GDPR on scientific research. One might further speculate that the general population may, in fact, not consciously distinguish between the different aims and purposes that drive data collection and analysis in the private sector versus the scientific community, thus unintentionally equating scientific and marketing research.

Moreover, when issues in relation to the GDPR and its impact on scientific research were presented by the media, these issues were mostly referred to in the context of health and biomedical research. This could imply that the public may not be provided by the media with sufficient information to consider the impact of the GDPR on research in non-health related research areas. A similar 'biomedical skew' was observable in the scoping review, where the majority of domain-specific peer-reviewed articles discussed the implications of GDPR for biomedical and public health. The reason for that might stem from a twofold fact. First, biomedical and other health-related data are generally perceived as among the most sensitive data types, hence entitled to the highest data protection standards. Second, especially in the media outlets, biomedical research might be unreflectively perceived as a synecdoche for the entire research domain.

Finally, it is plausible to expect that the GDPR will cause, at least in its initial phases of implementation, additional administrative burden on researchers. GDPR compliance will likely require research institutions to expand administrative support. GDPR compliance could require organisational approaches such as appointing specialist DPOs, specifying information about data protection and consent practices when seeking IRB/ERC approval and consultation with local DPOs. GDPR compliance could also require technical activities such as devoting more time to activities such as data anonymisation and pseudonymisation, and taking immediate action in case of data misuse. Under many circumstances, the cost of compliance might result in delays in project development and ethical approval and, less frequently, additional financial burden on research institutions. However, this possible increase in time and financial burden will likely be compensated by the benefits that higher data protection standards will imply for research. Furthermore, it is expected that, after familiarising themselves with the GDPR's provisions, institutions will overcome these predicted negative impacts and

develop adequate compliance mechanisms. Typically, compliance is more difficult in the earlier phases of introduction of a new regulation but tends to become easier over time as its provisions become entrenched and adequate coping mechanisms are adopted.

8.2 Open challenges

Uncertain impacts encompass both negligible changes from the previous Data Protection Directive and potential impacts that can be considered neither positive nor negative based on currently available information. These include impacts on data transfer outside the EU and the level of technical anonymisation.

While the GDPR is largely expected to facilitate cross-national data transfer within the EU, it is unlikely to have a direct significant impact on data transfer outside the EU or between inside-ERA and outside-ERA research institutions. As many authors have noted, the issue of data transfer outside the EU, as it usually occurs in international research consortia and supra-national initiatives involving both EU and non-EU partners, remains open. For data transfer between EU and US research institutions, the Court of Justice of the European Union (CJEU) overruling the EU-US Safe Harbour regulation generates additional uncertainty. In particular, it remains largely unclear on which conditions EU researchers may share personal data with US research institutes and hospitals pursuant to the Privacy Shield Framework.

In a similar manner, the GDPR permits the processing of pseudonymised data for uses beyond the purpose for which the data was originally collected. However, it remains uncertain what technical degree of pseudonymisation is required under the GDPR's provisions. In particular, both the CJEU and UK courts appear to have adopted the more pragmatic approach of classifying pseudonymised data as anonymised data when transferred without a key. Reducing uncertainty in this domain is highly important to clarify the conditions for sharing de-identified data, especially genetic data.

8.3 Measures taken by European research institutions

The various study components conducted for this report reveal that research institutions across Europe have not yet released into the public domain GDPR compliance guidelines for research in noteworthy numbers. Similar counts are observable for the industry sector. This paucity of publicly released information about measures taken by research institutions — including universities, research societies and research-pursuing companies — is equally noticeable across the scoping review, the doctrinal analysis as well as in the media.

Nevertheless, the report provides an overview of a small sample of purposely selected case studies that showcase how universities and research institutions in rare cases developed GDPR compliance guidelines for the research community. These preliminary findings suggest that the research community is preparing for GDPR compliance in a twofold manner. The first measure is through training and educational measures, such as organising short courses on GDPR compliance. Within industry and privately funded research, it appears that GDPR in general is widely discussed. However, GDPR

compliance guidelines targeting research conducted within the industry sector are not prominent in the public domain online.

In contrast to what can be found online in the public domain, the findings of the case study review suggest that institutions also provide guidance in internal online environments. However, these sources could not be accessed for review. The information presented online by the institutions consists of an educational as well as guiding components, as discussed below.

When comparing the case studies, it is evident that there is variation in the approaches taken. The universities and research institutions reviewed for this report developed guidelines in different formats, designs and degree of detail. However, all guidelines covered which are: introduction to the GDPR, Implications of the GDPR on the consent process, GDPR exemptions, responsible behaviour under GDPR and the role of the DPOs and the Data Protection Authorities (DPAs). Further, all guidelines suggest contacting DPOs or other personnel that can help to make research GDPR compliant. Also, it appears to be common practice to refer to national data protection officers for further detail.

Similar to the findings of the media analysis of this report, it seems that predominantly research societies working in the field of healthcare and biomedical research issue guidelines so far. When comparing the guidelines developed by universities against the guidelines developed by research societies, it seems that there is not much difference in content.

Universities across Europe offer GDPR compliance short courses, video training as well as DPO trainings. Further, education material about GDPR exists in different formats including self-check systems in quiz format that researchers can use to verify the compliance of their existing research projects with the GDPR. Altogether, these resources are targeted at the research community, and they aim at teaching researchers how to better familiarise and comply with the GDPR. In response to concerns and false assumptions about the GDPR and its consequences for research, institutions have also issued 'myth buster' documents as well as frequently asked questions (FAQ) documents.

9. Policy options

In light of the findings of our various study components, we contend that the research community would benefit from the development of harmonised guidelines across research institutions and research societies in the European Union. It is anticipated that national data protection offices together with the equivalent EU institution as well as representatives of the research community can take a leading role in developing standards that cover the relevant aspects of GDPR for the research community. Furthermore, the standards should address the concerns and uncertainties raised by the research community as presented above. These standards should inform the guideline-development process at research institution level in a way that the resulting guidelines, despite their potential context-specificity, cover the same areas and provide the same detail of information. For example, the BBMRI-ERIC has begun work on a European Code of Conduct for big data health research.[53] Additionally, such standards can inform researchers in the industry sector as well. This approach could contribute to a standardisation and harmonisation of GDPR compliance guidelines for the research community.

We suggest that guidelines should be developed by research institutions in collaboration with the national data protection offices and representatives of the research community including scientists and ethics review boards. In the development process and when drafting the guidelines, it will be pivotal to write these guidelines with the research community in mind. The guidelines need to be comprehensible for a broad audience with a diverse professional background, and be developed alongside further assessment as to which different formats (text, video, audio, gamification) are most useful to convey the necessary information to the research community.

Policy options should be considered at three main levels: at the level of regulations, processes and practices, as well as engagement and capacity building. A detailed overview of emerging issues and associated policy options (also including, for each of them, the type of reform and relevant responsible body) is presented in Table 6.

[53] See: http://code-of-conduct-for-health-research.eu/

Table 6 - Summary of the three tiers of recommendations; knowledge-based, technical, and regulatory.

Outlined in conjunction are the issues these recommendations resolve, the levels of reform suggested, and the responsible bodies identified.

* **EDPB** – European Data Protection Board. * **National DPAs** – National Data Protection Authorities.

Legislative/Regulation: legislative amendment of the GDPR, national instruments, or opinions issued by EDPB to clarify ambiguities.

Institutional/Normative: codes of practice or policies adopted by universities, research institutes, scientific societies or laboratories.

		Issues	Responsible body	Policy option
Regulatory		Potential conflict between the requirement for specific, informed and free consent, and the need for broad consent in scientific research.	+ EDPB + Research Inst. + Research Ethics Committees	Reconcile the requirement for specific, informed and free consent in the GDPR with the need for broad consent in scientific research and reconcile its definition with the requirements of consent in associated instruments.
		Broad interpretation of processing for statistical purposes that permits processing under the research exception subject to Article 89 (1) for non-scientific purposes.	+ EDPB + National DPAs	Clarify the exceptions subject to Article 89 (1) with respect to permitted processing for statistical and scientific purposes.
		Lack of clarity about what constitutes best practices for anonymisation and pseudonymisation.	+ EDPB	Establish data handling guidelines for the use and monitoring of anonymisation and pseudonymisation techniques within different contexts.
		Lack of clarity regarding the conditions for transnational transfers of personal data outside the EU in the context of transnational scientific projects.	+ EDPB + Research Inst.	Develop institutional guidelines that assist researchers involved in transnational transfers of personal data outside the EU in the context transnational collaborative scientific projects.
		Limited or inconsistent institutional guidelines for GDPR compliance produced to date by research institutions.	+ Universities + Research Inst. + Research Ethics Committees	Develop consistent institutional guidelines for GDPR compliance among researchers, with special focus on the grey zones between personal and non-personal data.
		Potential conflict between data subject rights under the GDPR and the protection of database rights under the sui generis database regime.	+ National DPAs + National Governments + EU Agencies	Resolve the conflict between data subject rights under the GDPR (the right to data portability and the right to access data) and the protection of database rights under the sui generis database regime.

		Lack of harmonisation of the national derogations of the GDPR	+ National DPAs + Research Inst. + Scientists	Coordinate the monitoring of derogations that apply to research and develop codes of conduct that address deficits of harmonisation
		Ambiguous interpretation of the accuracy data processing principle under Article 5(1)(d).	+ Universities + Research Inst. + Scientists + Academic Societies	Develop consistent standards of correctness and accuracy for all different domains of scientific research with respect to the data processing principles.
Procedural		Need for data management good practices.	+ Scientists + Academic Societies	Install robust data management practices.
		Lack of suitable data governance frameworks		Develop adaptive data governance frameworks
		Lack of standardisation about anonymisation and pseudonymisation.		Develop technical standards for anonymisation and pseudonymisation based on best practices.
		Scarcity of software tools for GDPR compliance that assist researchers.		Develop researcher-friendly software tools for GDPR compliance with special focus on open access tools for data portability.
Transitional & Capacity Building		Limited educational activities on data protection for researchers, students and training scientists.	+ Funding Agencies + Universities + Research Inst.	Organise educational activities & specific training sessions for data protection literacy among researchers, students and scientific trainees.
		Uncertainty about the administrative resources required for GDPR compliance.		Support more research on post hoc impact assessment.
		Perceptions in the scientific community about potential obstacles / burdens of GDPR compliance.		Monitor attitudes and develop tailored data protection literacy interventions.
		Limited media coverage of the rights and obligations of the GDPR in relation to research.		Raise awareness about the rights and obligations of the GDPR through activities for the general public.

9.1 Regulatory options

At the regulatory level, we suggest the following policy options:

- **Reconcile the requirement for specific, informed and free consent in the GDPR with the need for broad consent in scientific research and reconcile its definition with the requirement of consent in associated instruments.**

Article 6(1)(a) of the GDPR set free, informed and specific consent as the default ground for the lawful processing of personal data. Further, Articles 6(1)(a) and 9(2)(a) define consent as the default lawful grounds for the processing of personal data, and sensitive categories of personal data, respectively. In this context, Recital 32 establishes consent must be active, and not obtained via silence or a lack of objection. This standard of informed consent is appropriate for consumer transactions, where the purpose for processing is known in advance. By contrast, for some forms of scientific research project, it may be impossible to determine how collected data may be used. In addition, the increasing use of big data analytics and machine learning tools may mean that underlying trends may emerge from data during the course of processing. For this reason, Recital 33 of the GDPR notes that data subjects should be able to consent to specific areas of scientific research, rather than specific areas of processing. This consent would be valid under the GDPR, whilst still complying with the relevant ethical norms of the field of scientific research in question. Nevertheless, this interpretation must be read in the context of the former Article 29 Working Party's opinion of the nature of consent under the GDPR. On the subject of scientific research, the Article 29 Working Party held that Recital 33 did not dispense with the requirement of specific consent. Instead, the Article 29 Working Party held that Recital 33 allowed for specific consent where the research project could only be described in general terms. Prima facie, this interpretation creates a conflict with Recital 33's definition of consent for scientific research. Accordingly, a potential legislative policy option would involve amending Recital 33 so that it explicitly acknowledges alternatives to specific informed consent, including general consent to fields of scientific research. Alternatively, another policy option could be to collect dynamic consent from data subjects. Dynamic consent is a combined technical and organisational strategy to allow data subjects to granular consent to different forms of research.

As discussed above, the GDPR defines consent as free, informed and specific consent from data subjects. However, there may be other legal grounds for processing (including for research purposes, as discussed below). Nevertheless, the requirement of consent under data protection law has historically had a significant impact on the determination of consent in other directives. In particular, the former Data Protection Directive's definition of consent for the processing of personal data was used for both the Clinical Trials Directive and the Tissues and Cells Directive. The Clinical Trials Directive is set to be replaced by the Clinical Trials Regulation in 2020. However, the current European Data Protection Board has identified a conflict between the definition of consent under the GDPR and the future Clinical Trials Regulation. The former Article 29 Working Party gave patients being coerced into a clinical trial for fear of losing access to medication as an example of a lack of free consent. In 2019, the EDPB ruled that a distinction should be drawn for consent for primary purposes (that is, purposes associated with the

clinical trial) and secondary purposes (that is, research related purposes). With respect to the former category, a further distinction could be drawn between 'safety and reliability' purposes and consent for scientific purposes ('pure research activity purposes'). For safety and reliability purposes, the EDPB held that processing of personal data could be justified under the Clinical Trials Regulation pursuant to Article 6(1)(c).[54] Likewise, the processing of sensitive categories of personal data for safety and reliability can be justified under Article 9(2)(i).[55]

However, if research is processed for research activities, there are two possible grounds of consent. First, Article 6(1)(a) could be used in conjunction with Article 9(2)(a) for consent to be used as a lawful ground for processing. Secondly, Articles 6(1)(e) and 6(1)(f) could be used in concert with Articles 9(2)(i) or (j). These sections could be used to justify processing on the grounds of legitimate public interest, for public health or for research purposes. Nevertheless, for processing both within and outside a clinical trial protocol, the EDPB noted that the standard imposed for consent by the GDPR is different to the standard imposed by the CTR. If a researcher conducts a clinical trial, and wishes to reuse participant data for further research, they would need to obtain consent from the research participants. This consent would need to be compliant with the GDPR, and, where appropriate, would need to specify an appropriate legal ground for processing. Although the EDPB's report has provided some much-needed clarification with respect to this conflict, the Clinical Trials Regulation is yet to come into force. Member states will therefore need to modify their clinical trials legislation to ensure that they provide consistency with the standard of consent imposed by the GDPR. Further, national derogations of the Tissue and Cells Directive should also be updated to ensure compliance with the consent standard imposed by the GDPR.

- **Clarify the exceptions subject to Article 89(1) with respect to permitted processing for statistical and scientific purposes**

The GDPR explicitly permits research as a ground for the lawful processing of sensitive categories of data. Further, the GDPR creates a set of exceptions to the normal obligations that are imposed upon data processors and controllers. These exceptions are subject to the positive obligations to implement appropriate safeguards contained in Article 89(1). In Article 89(1) (along with the associated Recitals 157 to 162), research related processing, public interest processing and processing for statistical purposes are defined broadly. On the one hand, a broad interpretation of research is necessary to ensure that different fields of research are not unduly constrained by data protection laws. On the other hand, the exception at present permits both publicly funded academic research and privately funded commercial research. In addition, Article 89(1) also permits processing for statistical purposes, which is broadly defined to include but not be limited to scientific process. Accordingly, statistical processing for non-scientific, commercial purposes may be permitted under these exceptions. In these circumstances, adequate safeguards may not exist to ensure the protection of participant data. The broad scope of Article 89(1) is compounded by the effect of the exceptions which are available pursuant to the positive

[54] Article 6(1)(c) permits processing where such processing is necessary to comply with a legal obligation under the GDPR.
[55] Article 9(2)(i) permits processing of sensitive data where such processing would be necessary for reasons of public interest in the area of public health.

obligations it imposes. The limit on the right to information where data is collected by a third party may prevent the data subject from effectively exercising their right of erasure or objection. In turn, the exceptions to the right of erasure and objection could allow the data processor or controller can continually reuse data without the knowledge or consent of the subject. The exceptions to the purpose and storage limitation processing principles would also allow the controller or processor to store this data for a potentially indefinite period. Therefore, it is necessary to consider a number of policy options to prevent this scenario from eventuating. One policy option is to legislatively restrict the scope of the GDPR so as to narrow the forms of statistical processing that are permitted. This limitation could be achieved by requiring that statistical processing be conducted via a public interest requirement. In the alternative, research institutes and scientific societies could work collectively to develop an appropriate ethical standard to prevent the processing of data in an unethical fashion.

- **Establish guidelines for the use and monitoring of anonymisation and pseudonymisation technology within different contexts.**

Institutions should develop guidelines for the use of anonymisation and pseudonymisation technology within different contexts. In particular, there is a potential conflict within the relevant authority as to whether transferred data becomes pseudonymised data. The former Article 29 Working Party conducted a technical review and suggested setting a high threshold for anonymisation. The former Article 29 Working Party concluded that anonymisation could only be achieved where it was impossible to re-identify individuals from their data. The former Article 29 Working Party concluded that many 'anonymisation' techniques constituted pseudonymisation of data. However, both CJEU and UK case law have considered the use of anonymisation or pseudonymisation techniques where data is transferred from the processor to a third party. In particular, if the recipient of pseudonymised data does not have access to a mechanism to re-identify the data, that data should be considered anonymised. In light of this contextual requirement, institutions should develop guidelines for determining the status of transferred pseudonymised data. In addition, appropriate technical measures could be used to prevent access to re-identification mechanisms in the case of transfer.

- **Develop institutional guidelines that assist researchers involved in transnational transfers of personal data outside the EU in the context transnational collaborative scientific projects**

While the GDPR facilitates the transnational transfer of data within the EU, the normative conditions for transnational transfers of personal data outside the EU might be subject to regulatory ambiguity. This ambiguity might negatively affect the establishment and conducting of transnational collaborative scientific projects involving both EU and non-EU research institutions. As transnational scientific collaborations are essential to the scientific endeavour and key drivers of innovation, the development of institutional guidelines on transnational data transfer is highly recommended. These guidelines should assist and provide clear guidance to researchers involved in collaborative projects, including principal investigators, adjunct researchers, and staff.

- **Develop consistent institutional guidelines for GDPR compliance among researchers with special focus on the grey zones between personal and non-personal data.**

Research institutions should cooperate on a pan-European basis in order to develop consistent standards for the management and transfer of personal data. As the findings of both the doctrinal analysis and the scoping review illustrate, there are three areas in which guidelines could assist with GDPR compliance by researchers. First, institutional guidelines may assist with the management of personal data generated via inferences from non-personal data. In particular, machine learning algorithms and big data analytics techniques can generate unsuspected inferences from what would otherwise be considered non-personal data (such as health related data from purchase statistics). Because the GDPR adopts a broad interpretation of personal data, institutions should have clear guidelines on how researchers should respond when personal data becomes non-personal data. These guidelines could include mandatory reporting to institutional review boards upon identifying personal data, revising ethics approval, and introducing data protection by design where personal data may be inferred.

- **Resolve the conflict between data subject rights under the GDPR (the right to data portability and the right to access data), and the protection of database rights under the sui generis regime.**

This recommendation operates with respect to two issues. The first issue pertains to both inherent ambiguities within the data subject rights under the GDPR, particularly the right to opt out of automated profiling. Many scientific projects may use big data analytics and machine learning to categorise and profile different data subjects. Further, Article 22 does not provide a specific exception for research subject to the obligations contained in Article 89(1). However, it should be noted that there are a number of limitations on the right to opt out of automated profiling. In particular, the right does not apply if the data was processed with the data subject's explicit consent. Further, Article 22 only applies in circumstances where a decision is made solely using the automated profiling. In the context of scientific research, these requirements significantly limit the scope of Article 22's operation. Nevertheless, this limited scope does not exculpate researchers from liability. Potential policy options could include scientific research institutions and ethics committees developing appropriate standards for the use of automated profiling.

The second issue pertains to the conflict between the data subject rights and the sui generis database protection regime that exists in the EU. This regime provides protection for databases where those databases have an original structure. This regime also prevents third parties from taking a substantial portion of the data stored in that database. However, this regime may conflict with the rights guaranteed for data subjects under the GDPR. In particular, the rights to access, information and data portability could be particularly negatively affected by this right. This regime could also create inconsistencies with the equivalent right of access guaranteed by the Health Insurance Portability and Accountability Act (HIPAA) in the US. Therefore, scientific societies and research institutes should work collaboratively to determine appropriate limits on the database protection right.

- **Coordinate the monitoring of derogations that apply to research and develop codes of conduct that address deficits of harmonisation.**

The findings of both the doctrinal analysis and the scoping review highlight a widely held perception that there is a lack of harmonisation of the national derogations of the GDPR. This lack of harmonisation might result in regulatory ambiguities and ultimately undermine the transfer of data for scientific purposes. As noted earlier in this report, any national derogations flowing from the GDPR also increase the potential of regulatory fragmentation in EU data protection law. This regulatory fragmentation was a problem that emerged under the former Directive and which the GDPR was introduced to prevent. To prevent these risks, monitoring the spectrum of national derogations in each member state is critical. Furthermore, efforts need to be made to promptly address deficits of harmonisation at the international level. If derogations are introduced under Article 9(4), EU national legislatures should work together to ensure consistency. The national DPAs can have a leading role in facilitating this consistency-seeking harmonisation. In parallel, research institutions could contribute to the development of codes of conduct that address possible deficit of harmonisation and proactively provide consistent guidance for researchers. These codes of conduct could also form the foundations for corporate guidelines for the transfer of personal data across jurisdictions.

- **Develop consistent standards of correctness and accuracy for all different domains of scientific research with respect to the data processing principles.**

As we have seen in the previous section, another potential point of ambiguity within the GDPR is with respect to the accuracy principle of data processing in Article 5(1)(d). The English translation of this principle requires data to be kept accurate and rectified without delay where inaccuracy is discovered. Further, English authority suggests that data processors and controllers cannot be held liable where inferences are drawn from data, provided reasonable steps are required to ensure data accuracy. However, the translation of this word varies between different interpretation of the GDPR. For example, within the German interpretation of Article 5(1)(d), the relevant standard identified is one of 'correctness'. This interpretation is supported by previous interpretations of German data protection law, which have focussed on ensuring personal data is correct. In the face of this ambiguity, institutions should develop relevant standards of correctness or accuracy for different fields of research. For example, in social sciences research, it is recognised that there may be a respondent bias with respect to sensitive topics, such as political affiliation or health. However, as our scoping review has highlighted, the consequences of inaccuracy may be significantly more severe for biomedical research projects. Accordingly, researchers in these fields may be held to a significantly higher standard of reasonable accuracy.

9.2 Procedural options

Procedural options encompass policy options aimed at enhancing processes and established practices relating to data protection in the context of scientific research. These processes and practices include data management practices, technical standards for privacy-preserving data sanitisation, and technology-assisted regulatory compliance.

- **Install robust data management practices**

Managing data as a valuable resource is central to scientific research. The GDPR creates the adequate regulatory ecosystem for ensuring high data quality throughout the complete lifecycle of a particular research project. Installing robust data management practices for researchers can facilitate the successful implementation of the GDPR within the scientific community. Research institutions, academic societies and professional associations should be responsible for the development, promotion and enforcement of such practices. Based on the rights and obligations of the GDPR, key focus areas should be data integrity and data security. Furthermore, accountability for the adverse effects of poor data quality should be ensured through transparent data management practices. To ensure that adequate data management practices are installed and respected, research institutions might consider the appointment of data stewards as part of their organisational chart.

- **Develop adaptive data governance frameworks**

Compliance with data management requirements set forth by the GDPR emerged as one of the issues that will most likely affect scientific research practices under this new legal instrument. Over the last decade, scientific research has become increasingly more reliant on the collection, storage, analysis and re-use of personal data, especially but not limited to biomedical research. Such data now includes not only conventional health-related data but, to an increasing extent, new sources of data, including data generated by research participants themselves through smartphone apps, wearable devices or social media activities. In this respect, scientific research is now often conceptualised as a major component of the big data phenomenon. Recent technical developments in the field of artificial intelligence (AI) and automated data processing, such as AI-driven data mining, add a further layer of complexity to this phenomenon.
The general topic of data governance in the context of big data research is the subject of intense scholarly discussion. As a consequence, there is increasing awareness of the need to articulate a coherent set of governance principles that could channel the development of adequate oversight bodies and processes to handle the specific challenges raised by the data-centric nature of present-day scientific research. Numerous stakeholders in the research field are starting to devote efforts aimed at developing data governance guidelines that would align existing regulatory instruments – including the GDPR – with a broader set of ethical safeguards covering new data-related practices in the scientific space.

Vayena and Blasimme proposed the systemic oversight approach, a governance model that is better suited in meet the new demands of the data ecosystem (Vayena & Blasimme, 2018). Systemic oversight is an adaptive governance model that aims to provide an ethically robust common ground for data governance practices pertaining to research activities that employ big data, in the biomedical domain and beyond. The framework has six components covering a variety of newly emerging issues pertaining the impact of the collection and use of human big data in the context of research. The six components of systemic oversight are: adaptivity, flexibility, inclusivity, reflexivity, responsiveness, and monitoring (AFIRRM). In terms of implementation, the AFFIRM framework offers guidance into how oversight mechanisms, structures and processes shall be designed so as to meet the new demands of data governance in the field of data-driven scientific research.
The Wellcome Trust has also recently moved into a similar direction with a report highlighting four governance principles of dynamic oversight to grapple with the newly emerging setting of data-centric research[56]. According to this frame, oversight should be inclusive, anticipatory, innovative and proportionate.

- **Develop technical standards for anonymisation and pseudonymisation based on best practices.**

The GDPR recognises and promotes the privacy-enhancing effect of data sanitisation techniques such as data pseudonymisation and anonymisation. However, as the findings of our scoping review have highlighted, there is uncertainty among researchers about what constitutes optimal standard and best practice for, respectively, pseudonymisation and anonymisation. The development of technical standards is highly recommended to reduce this uncertainty and provide researchers with unambiguous and easy-to-apply technical requirements. Developing a check-list of specific data elements whose removal ensures pseudonymisation or anonymisation under the provisions of the GDPR might be a viable strategy to achieve this aim. Professional associations and academic societies could take a leading role in developing such harmonised technical standards.

- **Develop researcher-friendly software tools for GDPR compliance with special focus on open access tools for data portability.**

GDPR compliance might not necessarily be straightforward for research institutions and individual scientists. Therefore, the development of assistive software-based tools is a proactive measure to facilitate GDPR compliance and support researchers who prospectively seek to be GDPR-ready. Researcher-friendly software tools could also reduce the administrative burden on researchers. Finally, software tools ensure that the obligations of the GDPR are applied consistently with the requirements of consent in associated instruments (such as the Clinical Trials Regulation). This finding may hold particularly true if administrative management tools are used in coordination with digital consent management service tools.

[56] See: https://wellcome.ac.uk/sites/default/files/blueprint-for-dynamic-oversight.pdf

9.3 Transitional and capacity-building options

Transitional and capacity building options encompass the promotion of activities aimed at disseminating knowledge and building the capacity needed by research institutions to meet the challenges of GDPR compliance. These activities include expanding existing scientific curricula and training programmes with data protection literacy, conducting more extensive research on regulatory impact analysis and promoting awareness-raising activities among the general public. These activities are defined as *transitional* because their implementation is instrumental to facilitating the transition to the new GDPR regime. In particular, we suggest the following policy options:

- **Organise educational activities and specific training sessions for data protection literacy among researchers, students and scientific trainees.**

As the GDPR is a complex legal instrument which can affect the work of the research community, it is essential that researchers and other professional staff working with personal data understand the basic principles of GDPR, including its implications for data protection in the research context. As the case studies revealed, some universities across Europe developed short courses and other educational exercises to train research to become familiar with GDPR. These courses can range depending on the requirements of the given context, from a formal course to a less formal online format.

When developing educational and training activities for the scholarly community, it is essential to take the needs and preferences of the primary target audience into consideration. Given the time and resource constraints prevalent in academic environments, training activities should thus be targeted and time efficient. Moreover, it will be essential to make relevant information easily accessible and comprehensible for researchers from different fields and disciplines. In this context, the use of knowledge visualisation techniques (Tergan, Keller, & Burkhard, 2006) constitutes a promising approach. Knowledge visualisation is a methodology that can be used to translate complex, linear policy documents into more accessible, interactive formats that can reduce cognitive load and foster learning.

We further recommend for GDPR training to be mandatory for all scientific staff working with personal data to ensure participation in these activities. Such training could, for example, form part of designated training for new scientific staff upon arrival, also including early-career researchers.

- **Support more research on post hoc impact assessment.**

As our findings attest, there is uncertainty about the administrative resources required for GDPR compliance. While there is a plausible expectation that smaller research institutions might be disproportionally affected by the requirements for GDPR compliance compared to larger institutions, this hypothesis requires empirical verification. The same applies to the possibility that the time required to verify compliance and to obtain ethics review approval may increase as a consequence of the GDPR. Although ex ante assessments and theoretical analyses can provide insightful information, post hoc empirical studies are critical for comprehensively assessing the impact of the GDPR on scientific

research. For this reason, we recommend that further research in this domain is conducted in the near future. To this purpose, European funding agencies might consider opening targeted calls for proposals, possibly within the framework of existing EU Research and Innovation programmes and associated financial instruments.

- **Monitor attitudes and tailored data protection literacy interventions.**

Our findings reveal ambivalent and equivocal perceptions in the scientific community about the potential obstacles and burden imposed by the GDPR on research. The fact that some research institutions have adopted a 'myth-buster' approach is indicative of widely held misconceptions about the GDPR requirements. This broad range of attitudes suggests that there is a need to continuously monitor attitudes in the scientific community towards the GDPR in a systematic and longitudinal manner. This continuous monitoring could be implemented using social science methods such as surveys, and qualitative and ethnographic research methodologies. In particular, this research might be useful to identify, and, possibly, address persisting misconceptions about the GDPR. The outputs of this research could be targeted at information campaigns and educational programmes such as those proposed above. Finally, novel approaches to data governance, such as the systemic oversight approach Vayena and Blasimme (Vayena & Blasimme, 2018), might be well-suited to meet this demand for continuous monitoring.

- **Raise awareness about the rights and obligations of the GDPR through activities for the general public.**

As outlined earlier, findings of the media analysis indicate that there may be a lack of public awareness with regard to the impact of the GDPR on scientific research. Given that much of the media coverage focused on the consequences that the GDPR entails for business practices relating to marketing research and advertisement, we see a clear need to engage in communication with the public with respect to the collection and use of personal data in the contexts of scientific research. One way of doing this, is the development of public awareness campaigns.

Targeted campaigns should not only strengthen public understanding of the individual rights and obligations that are reinforced by the GDPR but should also aim to raise awareness with respect to the distinctions between scientific and marketing research. In this context, it will be central to create a better understanding of how personal data is used in the context of scientific research (as opposed to marketing research) and how data used in scientific research can contribute to the public good. This will, in turn, foster public trust and acceptance for the collection and use of personal data for scientific research.

An example of a public GDPR-awareness campaign is the 'Your Data Matters' campaign, launched by the UK Information Commissioner's office. The campaign provides a structured overview of the implications of the GDPR, including GIFs, short explanatory videos and other useful resources. It features practical templates and tips on exercising individual rights under the GDPR aimed at the general public. However, it does not specifically address the case of scientific research.

10. References

Adjekum, A., Ienca, M., & Vayena, E. (2017). What Is Trust? Ethics and Risk Governance in Precision Medicine and Predictive Analytics. *Omics, 21*(12), 704-710. doi:10.1089/omi.2017.0156

Agosti, D., & Egloff, W. (2009). Taxonomic information exchange and copyright: the Plazi approach. *BMC Research Notes, 2*(1), 53. doi:10.1186/1756-0500-2-53

Aitken, M., de St. Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics, 17*(1), 73. doi:10.1186/s12910-016-0153-x

Ali, M. Z., Khan, N. Z., & Khan, F. M. (2002). Advantages of digital photography record keeping in plastic surgery. *Journal of the College of Physicians and Surgeons Pakistan, 12*(10), 613-617.

Ambrose, M. L. (2014). Speaking of forgetting: Analysis of possible non-EU responses to the right to be forgotten and speech exception. *Telecommunications Policy, 38*(8), 800-811. doi:10.1016/j.telpol.2014.05.002

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology, 8*(1), 19-32.

Article 29 Working Party Opinion 03/2013 on purpose limitation. (2013, 2 April 2013). Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

BCG. (2012). The value of our digital identity. Retrieved from https://2zn23x1nwzzj494slw48aylw-wpengine.netdna-ssl.com/wp-content/uploads/2017/06/The-Value-of-Our-Digital-Identity.pdf

BDI. (2017). Die Datenschutz-Grundverordnung. . Retrieved from https://bdi.eu/media/themenfelder/recht/downloads/2017_Broschuere_DS-GVO.pdf

Bernard, C. (1963). The press and foreign policy. In: Princeton: Princeton University Press.

Beyleveld, D., & Taylor, M. J. (2007). Data Protection, Genetics and Patients for Biotechnology News & Views. *Eur J Health Law, 14*(2), 177-188.

Beyleveld, D., & Townend, D. M. R. (2004). When is Personal Data Rendered Anonymous? Interpreting Recital 26 of Directive 95/46/EC. *Medical Law International, 6*(2), 73-86. doi:10.1177/096853320400600201

Bialke, M., Bahls, T., Geidel, L., Rau, H., Blumentritt, A., Pasewald, S., . . . Hoffmann, W. (2018). MAGIC: once upon a time in consent management—a FHIR® tale. *J Transl Med, 16*. doi:http://dx.doi.org/10.1186/s12967-018-1631-3

Bignami, F. (2011). Cooperative Legalism and the Non-Americanization of European Regulatory Styles: The Case of Data Privacy. *The American Journal of Comparative Law, 59*(2), 411-461. doi:10.5131/AJCL.2010.0017

Bovenberg, J. A., & Almeida, M. (2019). Patients v. Myriad or the GDPR Access Right v. the EU Database Right. *European Journal of Human Genetics, 27*(2), 211. doi:10.1038/s41431-018-0258-4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Brkan, M. (2016). The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors? *Maastricht Journal of European and Comparative Law, 23*(5), 812-841. doi:10.1177/1023263X1602300505

Brown, N., & Deegan, C. (1998). The public disclosure of environmental performance information—a dual test of media agenda setting theory and legitimacy theory. *Accounting and business research, 29*(1), 21-41.

Bu-Pasha, S. (2017). Cross-border issues under EU data protection law with regards to personal data protection. *Information & Communications Technology Law, 26*(3), 213-228. doi:10.1080/13600834.2017.1330740

Burstein, P. (2003). The impact of public opinion on public policy: A review and an agenda. *Political research quarterly, 56*(1), 29-40.

Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review, 34*(2), 257-268. doi:10.1016/j.clsr.2018.01.004

Cave, E. (2018). EU Clinical Trials Regulation 2014: Fetter or facilitator? *Medical Law International, 18*(2-3), 179-194. doi:10.1177/0968533218799535

Chassang, G. (2017). The impact of the EU general data protection regulation on scientific research. *Ecancermedicalscience, 11*. doi:10.3332/ecancer.2017.709

Chen, J. (2016). How the best-laid plans go awry: the (unsolved) issues of applicable law in the General Data Protection Regulation. *International Data Privacy Law, 6*(4), 310-323. doi:10.1093/idpl/ipw020

Chico, V. (2018). The impact of the General Data Protection Regulation on health research. *British Medical Bulletin, 128*(1), 109-118. doi:10.1093/bmb/ldy038

de Hert, P., & Papakonstantinou, V. (2016). The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review, 32*(2), 179-194. doi:10.1016/j.clsr.2016.02.006

de Lecuona, I., & Villalobos-Quesada, M. (2018). European perspectives on big data applied to health: The case of biobanks and human databases. *Developing World Bioethics, 18*(3), 291-298. doi:10.1111/dewb.12208

Dias, R. D. M. A. (2017). The potential impact of the EU general data protection regulation on pharmacogenomics research. *Medicine and Law, 36*(2), 43-58.

Dove, E. S. (2015). Biobanks, Data Sharing, and the Drive for a Global Privacy Governance Framework. *The Journal of Law, Medicine & Ethics, 43*(4), 675-689. doi:10.1111/jlme.12311

Dove, E. S. (2018). The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. *The Journal of Law, Medicine & Ethics, 46*(4), 1013-1030. doi:10.1177/1073110518822003

DPC. (2019). The Data Protection Commission. Retrieved from https://dataprotection.ie/

Dulong de Rosnay, M., & Janssen, K. (2014). Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. *Journal of Theoretical and Applied Electronic Commerce Research, 9*(3), 1-14. doi:10.4067/S0718-18762014000300002

EDPB. (2019). List of Members. Retrieved from https://edpb.europa.eu/about-edpb/board/members_en

EDPB Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR). (2019, 23 January 2019). Retrieved from https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers-interplay_en

Egloff, W., Patterson, D. J., Agosti, D., & Hagedorn, G. (2014). Open exchange of scientific knowledge and European copyright: The case of biodiversity information. *ZooKeys*(414), 109-135. doi:10.3897/zookeys.414.7717

Elliot, M., O'Hara, K., Raab, C., O'Keefe, C. M., Mackey, E., Dibben, C., . . . McCullagh, K. (2018). Functional anonymisation: Personal data and the data environment. *Computer Law & Security Review, 34*(2), 204-221. doi:10.1016/j.clsr.2018.02.001

EU GDPR.ORG. (2019). GDPR. Retrieved from https://eugdpr.org/

Evans, B. J., & Jarvik, G. P. (2018). Impact of HIPAA's minimum necessary standard on genomic data sharing. *Genetics in Medicine, 20*(5), 531-535. doi:10.1038/gim.2017.141

Favale, M., & Plomer, A. (2009). Fundamental Disjunctions in the EU Legal Order on Human Tissue, Cells & Advanced Regenerative Therapies. *Maastricht Journal of European and Comparative Law, 16*(1), 89-111. doi:10.1177/1023263X0901600105

Frantziou, E. (2015). The Horizontal Effect of the Charter of Fundamental Rights of the EU: Rediscovering the Reasons for Horizontality. *European Law Journal, 21*(5), 657-679. doi:10.1111/eulj.12137

Gene Zucker, H. (1978). The variable nature of news media influence. *Annals of the International Communication Association, 2*(1), 225-240.

Glinos, K. (2018). Global data meet EU rules. *Science, 360*(6388), 467. doi:http://dx.doi.org/10.1126/science.aat9878

Gonçalves, M. E. (2017). The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward. *Information & Communications Technology Law, 26*(2), 90-115. doi:10.1080/13600834.2017.1295838

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision Making and a 'Right to Explanation'. *AI Magazine, 38*(3), 50-57. doi:10.1609/aimag.v38i3.2741

Hallinan, D., & Friedewald, M. (2015). Open consent, biobanking and data protection law: can open consent be 'informed' under the forthcoming data protection regulation? *Life Sciences, Society and Policy, 11*(1), 1. doi:10.1186/s40504-014-0020-9

Hallinan, D., Friedewald, M., & De Hert, P. (2013). Genetic Data and the Data Protection Regulation: Anonymity, multiple subjects, sensitivity and a prohibitionary logic regarding genetic data? *Computer Law & Security Review, 29*(4), 317-329. doi:10.1016/j.clsr.2013.05.013

Ho, C. H. (2017). Challenges of the EU general data protection regulation for biobanking and scientific research. *Journal of Law, Information and Science, 25*(1), 84-103.

Hoeren, T. (2017). Big data and the legal framework for data quality. *International Journal of Law and Information Technology, 25*(1), 26-37. doi:10.1093/ijlit/eaw014

Hoeren, T. (2018). Big Data and Data Quality. In T. Hoeren & B. Kolany-Raiser (Eds.), *Big Data in Context: Legal, Social and Technological Insights* (pp. 1-12). Cham: Springer International Publishing.

Hon, W. K., Millard, C., & Walden, I. (2011). The problem of 'personal data' in cloud computing: what information is regulated?—the cloud of unknowing. *International Data Privacy Law, 1*(4), 211-228. doi:10.1093/idpl/ipr018

Hon, W. K., Millard, C., & Walden, I. (2012). Who is responsible for 'personal data' in cloud computing?—The cloud of unknowing, Part 2. *International Data Privacy Law, 2*(1), 3-18. doi:10.1093/idpl/ipr025

Hordern, V. (2016). Data Protection Compliance in the Age of Digital Health. *Eur J Health Law, 23*(3), 248-264. doi:10.1163/15718093-12341393

Hornung, G. (2013). Regulating privacy enhancing technologies: seizing the opportunity of the future European Data Protection Framework. *Innovation: The European Journal of Social Science Research, 26*(1-2), 181-196. doi:10.1080/13511610.2013.723381

HRA. (2017). GDPR guidance for researchers and study coordinators. Retrieved from https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/

ICO. (2019). Guide to the General Data Protection Regulation (GDPR). Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

Kaye, J., Briceño Moraia, L., Curren, L., Bell, J., Mitchell, C., Soini, S., . . . Rial-Sebbag, E. (2016). Consent for Biobanking: The Legal Frameworks of Countries in the BioSHaRE-EU Project. *Biopreservation and Biobanking, 14*(3), 195-200. doi:10.1089/bio.2015.0123

Koščík, M., & Myška, M. (2018). Data protection and codes of conduct in collaborative research. *International Review of Law, Computers & Technology, 32*(1), 141-154. doi:10.1080/13600869.2018.1423888

Kostkova, P. (2018). Disease surveillance data sharing for public health: the next ethical frontiers. *Life Sciences, Society and Policy, 14*(1), 16. doi:http://dx.doi.org/10.1186/s40504-018-0078-x

KU Leuven. (2019). Privacy @ KU Leuven. Retrieved from https://admin.kuleuven.be/privacy/en

Lancaster University. (2019). GDPR: What Researchers Need to Know. Retrieved from https://www.lancaster.ac.uk/research/research-services/research-integrity-ethics--governance/data-protection/gdpr-what-researchers-need-to-know/

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science, 5*(1), 69. doi:10.1186/1748-5908-5-69

Li, X., Zhao, X., & Zhong, M. (2016). *Advancing public health genomics.* Paper presented at the Big Data and Information Security (IWBIS), International Workshop on.

Liu, V., Musen, M. A., & Chou, T. (2015). Data Breaches of Protected Health Information in the United StatesData Breaches of Protected Health InformationLetters. *JAMA, 313*(14), 1471-1473. doi:10.1001/jama.2015.2252

LMU München. (2019). Behördlicher Datenschutzbeauftragter der Universität München. Retrieved from https://www.uni-muenchen.de/einrichtungen/orga_lmu/beauftragte/dschutz/index.html

Mantelero, A. (2013). The EU Proposal for a General Data Protection Regulation and the roots of the 'right to be forgotten'. *Computer Law & Security Review, 29*(3), 229-235. doi:10.1016/j.clsr.2013.03.010

Marjanovic, S., Ghiga, I., Yang, M., & Knack, A. (2018). Understanding value in health data ecosystems: A review of current evidence and ways forward. *Rand Health Q, 7*(2), 3.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly, 36*(2), 176-187.

McKinsey & Company. (2017). Tackling GDPR compliance before time runs out. Retrieved from https://www.mckinsey.com/business-functions/risk/our-insights/tackling-gdpr-compliance-before-time-runs-out

MedTech Europe. (2018). MedTech Europe & IPMPC – Data Protection in Health Research Workshop. Retrieved from https://www.medtecheurope.org/news-and-events/news/medtech-europe-ipmpc-data-protection-in-health-research-workshop/

Mendoza, I., & Bygrave, L. A. (2017). The Right Not to be Subject to Automated Decisions Based on Profiling. In T.-E. Synodinou, P. Jougleux, C. Markou, & T. Prastitou (Eds.), *EU Internet Law: Regulation and Enforcement* (pp. 77-98). Cham: Springer International Publishing.

Moerel, L. (2011). The long arm of EU data protection law: Does the Data Protection Directive apply to processing of personal data of EU citizens by websites worldwide? *International Data Privacy Law, 1*(1), 28-46. doi:10.1093/idpl/ipq004

Molnár-Gábor, F. (2018). Germany: a fair balance between scientific freedom and data subjects' rights? *Hum Genet, 137*(8), 619-626. doi:10.1007/s00439-018-1912-1

Morrison, M., Bell, J., George, C., Harmon, S., Munsie, M., & Kaye, J. (2017). The European General Data Protection Regulation: Challenges and considerations for iPSC researchers and biobanks. *Regen Med, 12*(6), 693-703. doi:10.2217/rme-2017-0068

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., . . . Kaye, J. (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

MRC. (2019). GDPR resources. Retrieved from https://mrc.ukri.org/research/facilities-and-resources-for-researchers/regulatory-support-centre/gdpr-resources/

Nettleton, E., & Llewellyn, S. (2009). ECJ provides further guidance on the ambit of database right. *Computer Law & Security Review, 25*(5), 477-481. doi:10.1016/j.clsr.2009.07.002

NHS Digital. (2018). General Data Protection Regulation (GDPR) guidance. Retrieved from https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/information-governance-alliance-iga/general-data-protection-regulation-gdpr-guidance

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review, 57*(6), 1701-1778.

Ohmann, C., Banzi, R., Canham, S., Battaglia, S., Matei, M., Ariyo, C., . . . Demotes-Mainard, J. (2017). Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open, 7*(12). doi:10.1136/bmjopen-2017-018647

Oliver, J. M., Slashinski, M., Wang, T., Kelly, P., Hilsenbeck, S., & McGuire, A. (2012). Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. *Public Health Genomics, 15*(2), 106-114.

Pagallo, U. (2018). Algo-Rhythms and the Beat of the Legal Drum. *Philosophy & Technology, 31*(4), 507-524. doi:10.1007/s13347-017-0277-z

Paterson, M., & McDonagh, M. (2018). Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data. *Monash University Law Review, 44*(1), 1-31.

Pearce, H. (2018). Could the doctrine of moral rights be used as a basis for understanding the notion of control within data protection law? *Information & Communications Technology Law, 27*(2), 133-165. doi:10.1080/13600834.2018.1458449

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods, 5*(4), 371-385. doi:10.1002/jrsm.1123

Politou, E., Michota, A., Alepis, E., Pocs, M., & Patsakis, C. (2018). Backups and the right to be forgotten in the GDPR: An uneasy relationship. *Computer Law & Security Review, 34*(6), 1247-1257. doi:10.1016/j.clsr.2018.08.006

Pormeister, K. (2017). Genetic data and the research exemption: is the GDPR going too far? *International Data Privacy Law, 7*(2), 137-146. doi:10.1093/idpl/ipx006

Poullet, Y. (2018). Is the general data protection regulation the solution? *Computer Law & Security Review, 34*(4), 773-778. doi:10.1016/j.clsr.2018.05.021

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*(1), 37. doi:10.1038/s41591-018-0272-7

Quinn, P. (2017). The Anonymisation of Research Data - A Pyric Victory for Privacy that Should Not Be Pushed Too Hard by the eu Data Protection Framework? *Eur J Health Law, 24*(4), 347-367. doi:10.1163/15718093-12341416

Quinn, P. (2018). Is the GDPR and Its Right to Data Portability a Major Enabler of Citizen Science? *Global Jurist, 18*(2). doi:10.1515/gj-2018-0021

Quinn, P., & Quinn, L. (2018). Big genetic data and its big data protection challenges. *Computer Law & Security Review, 34*(5), 1000-1018. doi:10.1016/j.clsr.2018.05.028

Raghunathan, B. (2013). 16. Data Anonymization Techniques. In *The Complete Book of Data Anonymization: From Planning to Implementation* (pp. 171): CRC Press.

Reichel, J. (2017). Oversight of EU medical data transfers - an administrative law perspective on cross-border biomedical research administration. *Health and Technology, 7*(4), 389-400. doi:10.1007/s12553-017-0182-6

Reichman, J. H., & Okediji, R. L. (2011). When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale. *Minnesota Law Review, 96*(4), 1362-1481.

Rothstein, M. A. (2016). International Health Research after Schrems v. Data Protection Commissioner. *Hastings Center Report, 46*(2), 5-6. doi:10.1002/hast.539

Ruyter, K. W., LÕuk, K., Jorqui, M., Kvalheim, V., Cekanauskaite, A., & Townend, D. (2010). From Research Exemption to Research Norm: Recognising an Alternative to Consent for Large Scale Biobank Research. *Medical Law International, 10*(4), 287-313. doi:10.1177/096853321001000403

Sánchez, M. C., & Sarría-Santamera, A. (2019). Unlocking data: Where is the key? *Bioethics*. doi:10.1111/bioe.12565

Sariyar, M., Suhr, S., & Schlünder, I. (2017). How Sensitive Is Genetic Data? *Biopreserv Biobank, 15*(6), 494-501. doi:http://dx.doi.org/10.1089/bio.2017.0033

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society, 3*(2-3), 297-316.

Shabani, M., & Borry, P. (2018). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics, 26*(2), 149. doi:10.1038/s41431-017-0045-7

Simell, B. A., Törnwall, O. M., Hämäläinen, I., Wichmann, H. E., Anton, G., Brennan, P., . . . Perola, M. (2019). Transnational access to large prospective cohorts in Europe: Current trends and unmet needs. *N Biotechnol, 49*, 98-103. doi:10.1016/j.nbt.2018.10.001

Sobolciakova, A. (2018). Right of Access under GDPR and Copyright. *Masaryk University Journal of Law and Technology, 12*(2), [i]-247.

Stalla-Bourdillon, S., Pearce, H., & Tsakalakis, N. (2018). The GDPR: A game changer for electronic identification schemes? The case study of Gov.UK Verify. *Computer Law & Security Review, 34*(4), 784-805. doi:10.1016/j.clsr.2018.05.012

Sullivan, C., & Burger, E. (2017). 'In the public interest': The privacy implications of international business-to-business sharing of cyber-threat intelligence. *Computer Law & Security Review, 33*(1), 14-29. doi:10.1016/j.clsr.2016.11.015

Taylor, M. J., Wallace, S. E., & Prictor, M. (2018). United Kingdom: transfers of genomic data to third countries. *Hum Genet, 137*(8), 637-645. doi:10.1007/s00439-018-1921-0

Tenti, E., Simonetti, G., Bochicchio, M. T., & Martinelli, G. (2018). Main changes in European Clinical Trials Regulation (No 536/2014). *Contemporary Clinical Trials Communications, 11*, 99-101. doi:10.1016/j.conctc.2018.05.014

Tergan, S.-O., Keller, T., & Burkhard, R. A. (2006). Integrating knowledge and information: digital concept maps as a bridging technology. *Information Visualization, 5*(3), 167-174.

Timmers, M., Van Veen, E.-B., Maas, A. I. R., & Kompanje, E. J. O. (2018). Will the Eu Data Protection Regulation 2016/679 Inhibit Critical Care Research? *Medical law review*. doi:http://dx.doi.org/10.1093/medlaw/fwy023

Truyens, M., & Van Eecke, P. (2014). Legal aspects of text mining. *Computer Law & Security Review, 30*(2), 153-170. doi:10.1016/j.clsr.2014.01.009

UCL. (2019). GDPR. Retrieved from https://www.ucl.ac.uk/legal-services/sites/legal-services/files/part_a_ucl_gdpr_training_course.pdf

University of Leicester. (2019a). GDPR FAQ. Retrieved from https://www2.le.ac.uk/offices/ias/resources/policies/GDPR FAQ-1-1.pdf/view

University of Leicester. (2019b). GDPR. Retrieved from https://www2.le.ac.uk/offices/ias/resources/policies/gdpr

University of Oxford. (2018). Data Privacy. Retrieved from http://www.admin.ox.ac.uk/councilsec/compliance/dataprotection/

University of Oxford. (2019a). Data protection & research. . Retrieved from https://researchsupport.admin.ox.ac.uk/policy/data

University of Oxford. (2019b). Data protection checklist. Retrieved from https://researchsupport.admin.ox.ac.uk/policy/data/checklist

Van Alsenoy, B. (2012). Allocating responsibility among controllers, processors, and 'everything in between': the definition of actors and roles in Directive 95/46/EC. *Computer Law & Security Review, 28*(1), 25-43. doi:10.1016/j.clsr.2011.11.006

van Veen, E.-B. (2018). Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. *Eur J Cancer, 104*, 70-80. doi:http://dx.doi.org/10.1016/j.ejca.2018.09.032

Vayena, E., & Blasimme, A. (2018). Health Research with Big Data: Time for Systemic Oversight. *The Journal of Law, Medicine & Ethics, 46*(1), 119-129.

Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Review, 34*(2), 398-404. doi:10.1016/j.clsr.2017.12.002

Vestoso, M. (2018). The GDPR beyond Privacy: Data-Driven Challenges for Social Scientists, Legislators and Policy-Makers. *Future Internet, 10*(7), 62. doi:10.3390/fi10070062

Wachter, S. (2018). The GDPR and the Internet of Things: a three-step transparency model. *Law, Innovation and Technology, 10*(2), 266-294. doi:10.1080/17579961.2018.1527479

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76-99. doi:10.1093/idpl/ipx005

Wagner, J. (2018). The transfer of personal data to third countries under the GDPR: when does a recipient country provide an adequate level of protection? *International Data Privacy Law, 8*(4), 318-337. doi:10.1093/idpl/ipy008

Watson, H., & Rodrigues, R. (2018). Bringing privacy into the fold: Considerations for the use of social media in crisis management. *Journal of Contingencies and Crisis Management, 26*(1), 89-98.

Weichert, T. (2018). Health privacy in the age of digital networks. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 61*(3), 285-290. doi:10.1007/s00103-017-2686-7

Whitman, J. Q. (2003). The Two Western Cultures of Privacy: Dignity versus Liberty. *Yale Law Journal, 113*(6), 1151-1222.

Wolf, C. (2014). Impact of the CJEU's Right to Be Forgotten: Decision on Search Engines and other Service Providers in Europe: Case C-131/12 Google v. Agencia Española de Protectión de Datos (AEPO) and Mario Costeja Gonzalez, Judgment of 13 May 2014. *Maastricht Journal of European and Comparative Law, 21*(3), 547-554. doi:10.1177/1023263X1402100308

Yuill, C. (2018). 'Is Anthropology Legal?': Anthropology and the EU General Data Protection Regulation. *Anthropology in Action, 25*(2), 36-41. doi:10.3167/aia.2018.250205

11. Appendices

Appendix 1

The Major World Publications group file, MEPE, contains English full-text news sources from across Europe, which are held in high esteem for their content reliability. This includes the major newspapers, magazines and trade publications in English that are relied upon for the accuracy and integrity of their reporting.

Accountancy Age (UK)*
Airline Business
Baltic News Service
Belfast News Letter*
Belfast Telegraph
Belfast Telegraph Online
Brand Strategy*
Campaign
CMP Information
Computing
Contract Journal*
Control and Instrumentation*
Creative Review*
Daily Record and Sunday Mail
Design Engineering*
Design Week*
Electronics Weekly
Estates Gazette
Euromoney
Financial Director*
Flight International
Gazeta Mercantil Online*
Gazeta Wyborcza in English*
Global Capital Euroweek
Het Financieele Dagblad (English)*
Insurance Age*
International Money Marketing*
Investorschronicle.co.uk
ITAR-TASS
Management Today*
Marketing - UK*
Marketing Week
mirror.co.uk
Money Marketing
Moscow News*
MTI Econews
New Media Age*
New Scientist
Off Licence News*
Polish News Bulletin
Precision Marketing*
Process Engineering*
Professional Broking*
PR Week
Retail Week*
standard.co.uk
telegraph.co.uk
The Banker
The Business*
The Daily Mail and Mail on Sunday (London)
The Daily Telegraph (London)
The Engineer*
The Evening Standard (London)
The Grocer
The Guardian(London)
The Herald (Glasgow)
The Independent (United Kingdom)
The Investors Chronicle
The Irish Times
The Lawyer
The Mirror (The Daily Mirror and The Sunday Mirror)
The Moscow News (RIA Novosti)*
The Moscow Times*
The New York Times - International Edition
The Observer(London)
The People
The Prague Post
The Sunday Herald (Glasgow)
The Sunday Telegraph (London)
Utility Week*
What's new in Industry*

Appendix 2

Search strings for doctrinal analysis

The following section provides an overview of the different phrases used:

1. Search results citing case law:

The following search string was used to identify articles referring to case law:
ALL('<case name>')

For example:
ALL('*Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*')

This search string returned all findings on Scopus that referred to the case *Google Spain SL and Google Inc v. AEPD and Mario Costeja González*. In the alternative, a shortened case name was used:
ALL('*Google Spain SL. v AEPD and González*')

In the alternative, a reference to the case number was used. For example:
ALL('<case number>')

These search results were then limited to articles, books and book chapters.

2. Articles referring to the GDPR:

These articles were captured in addition to the articles that were gathered as part of the systematic review. Because this stage of the analysis involves inductive reasoning from case law, the search terms were broadened. Accordingly, the following search string was used to identify articles beyond the systematic review:
Scopus - ALL(gdpr OR 'General Data Protection Regulation')
Web of Science - TOPIC: (gdpr) OR TOPIC: ('General Data Protection Regulation')

These search results were then limited to articles, books and book chapters. In addition, these search results were limited to 2012 (when the first draft proposals for updating the Data Protection Directive were proposed by the European Commission).[57]

[57] Data Protection, European Commission (3 December 2012) < http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf>.

3. Articles referring to the Data Protection Directive:

These articles extended beyond those articles identified as part of the systematic review. In particular, the following search string was used to identify articles that had previously analysed the scope of the data protection directive. In addition, this search strategy also identified articles that had previous discussed national case law implementing the data protection directive.
ALL(('data protection directive') OR ('95/46/EC'))

These search results were then limited to articles, books, book chapters and reviews. The search string was extended beyond 2012 (when the Data Protection Directive was operational) to 1995.

4. Articles referring to the Clinical Trials Directive and the Clinical Trials Regulation:
ALL(('clinical trials directive') OR ('2001/20/EC'))

These search results were then limited to articles, books, book chapters and reviews. A similar search was conducted with respect to the Clinical Trials Regulation.

ALL(('European' AND 'clinical trials regulation') OR ('No 536/2014'))

These search results were then limited to articles, books, book chapters and reviews.

5. Articles referring to the Cells and Tissues Directives:

ALL((('2004/23/EC') OR ('2006/17/EC') OR ('2006/86/EC') AND ('Cells and Tissues') OR ('Tissues and Cells')))

These search results were then limited to articles, books, book chapters and reviews.

6. Articles referring to the Database Directive:

ALL('Database Directive' OR '96/9/EC')

These search results were then limited to articles, books and book chapters.

7. Articles referring to the Copyright Directive

ALL('Copyright Directive' OR '2001/29/EC') AND ALL('scientific')

This search string was chosen to focus on articles that have defined the boundaries of copyright law with respect to scientific inventions. These search results were then limited to articles, books and book chapters.

The implementation of the General Data Protection Regulation (GDPR) raises a series of challenges for scientific research, in particular for research that is dependent on data. This study comprehensively investigates the promises and challenges associated with the implementation of the GDPR in the scientific domain, with a special focus on the impact of the new rights and obligations enshrined in the GDPR on the design and conduct of scientific research. Furthermore, the study examines the adequacy of the GDPR exceptions for scientific research in terms of safeguarding scientific freedom and technological progress.

The study also provides policy options that delineate a pathway towards enhancing rather than stifling research, and facilitating privacy-preserving data-driven research under the provisions of the GDPR.

QA-04-19-501-EN-N

This is a publication of the Scientific Foresight Unit (STOA)
EPRS | European Parliamentary Research Service

ISBN 978-92-846-5045-3
doi: 10.2861/17421
QA-04-19-501-EN-N