Food | Consumer | Health
Designing a world-class infrastructure to facilitate research

Horizon 2020
INFRADEV-1-2014 - Design studies

RICHFIELDS Working Package 13
Deliverable 13.2

Ethical Design

At the Interface of Ethics for Big Data and
the European Union's General Data Protection Regulation

Date delivered:
M31

Authors:
Indira Carr (U Surrey)

Deliverable lead beneficiaries:
University of Surrey

www.richfields.eu
#RICHFIELDS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Project	
Project acronym:	RICHFIELDS
Project full title:	Research Infrastructure on Consumer Health and Food Intake for E-science with Linked Data Sharing
Grant agreement no.:	654280
Project start date:	01.10.2015
Document:	
Title:	ETHICAL DESIGN At the Interface of Ethics for Big Data and the European Union's General Data Protection Regulation
Deliverable No.:	D 13.2
Authors:	Indira Carr
Reviewer:	Karin Zimmermann – Project Coordinator Pieter van 't Veer – Scientific Coordinator Harriet Teare – Project Advisory Board member
Start date:	01.03.2017
Delivery date:	17.04.2018
Due date of deliverable:	31.01.2018
Dissemination level:	PU
Status:	Final

Change history:		
Version	Notes	Date
001	Draft	25.01.2018
002	Final	17.04.2018

Karin Zimmermann
Project Coordinator

Prof Pieter van 't Veer
Scientific Coordinator

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Summary

Work Package (WP) 13.2 focuses on the ethical design of the RICHFIELDS research infrastructure that draws its data from diverse sources such as apps, social media, discussion lists, online forums, data from activity trackers, health platforms, store loyalty cards, and data provided by data brokers.

The use of big data that includes personal data for research purposes, whilst providing huge opportunities for researchers, raises ethical concerns. Section Two identifies these concerns as primarily relating to privacy of an individual as a result of the use of data containing personal data (namely identifiers such as name, telephone numbers, addresses, IP address, and biomarkers) and an individual's right of control over his or her data. Both privacy and right of control are matters addressed by the European Union's General Data Protection Regulation (GDPR). Section Three examines how this Regulation addresses these matters by imposing obligations on both the controller and the processor of data when it comes to processing personal data. Informed consent is key to legitimizing processing of personal data. However informed consent may not always be a feasible route where a research infrastructure obtains its data from a variety of sources. The GDPR does recognize the importance of data for research purposes and pseudonymisation of personal data is seen by the framers of the GDPR as reducing the risks to data subjects whilst enabling controllers and processors to meet their data-protection obligations. To facilitate legal compliance the GDPR requires the controller and the processor to designate a data protection officer (DPO). A DPO is required, for instance, where the core activities of the controller/processor consist of processing operations which by virtue of their nature, their scope and/or purposes, require regular and systematic monitoring of data subjects on a large scale. Lack of compliance with the GDPR will attract large fines. The GDPR also imparts a number of rights to the data subject which include withdrawal of consent, right to rectify and objection to processing. Trans-border data flows are subject to appropriate safeguards. These include legally binding and enforceable instrument between public bodies, binding corporate rules (BCR), codes of conduct and standard data protection contractual approved by the Commission.

Section Four in providing a framework for the design of the ethical and legal aspects of RICHFIELDS includes the following recommendations: (1) use of pseudonymisation with appropriate safeguards for unauthorised reversal of pseudonymisation; (2) use of appropriate technical and organisational measures to ensure GDPR compliance; (3) systems for dealing with queries and requests from data subjects; (4) appointment of a DPO; (5) mechanisms for handling freedom of information (FOI) requests; (6) use of suitable data protection clauses for trans-border data transfer; (7) obtaining insurance to cover liability in the event of data breaches; and (8) the establishment of an independent ethics committee with remit to monitor the activities of RICHFIELDS, its protocols on matters relating to security, transfer of data to third countries, assessing genuineness of requests from data users and procedures for dealing with ethically suspect requests, and procedures for handling requests from data subjects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Table of Contents

1. Introduction 5
2. Ethics for Big Data 6
2.1. Ethical Challenges 6
2.2. Ethics, Law and Governance 8
3. General Data Protection Regulation (GDPR) 9
3.1. Rationale for the GDPR 9
3.2. Legitimising processing through consent 11
3.3. Who does the GDPR protect? 13
3.4. What type of data does the GDPR apply to? 13
3.4.1. Prohibition in respect of processing of certain types of personal data 13
3.5. When does the GDPR apply? 14
3.5.1. Processing falling outside the scope of the GDPR (Anonymised data) 16
3.5.2. Processing pseudonymised data 17
3.6. Who has to comply with GDPR and their responsibilities 17
3.6.1. Controller's Responsibilities 18
3.6.2. Processor's Responsibilities 19
3.6.3. Security and Notification 20
3.6.4. Data Protection Officer 22
3.7. Rights of Data Subjects 23
3.8. Trans-border Data Transfer 25
4. Design of Ethical/Legal Aspects of RICHFIELDS 28
5. Conclusion and Potential for Future Developments 31
Reference 32

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

1. Introduction

Work Package (WP) 13.2 focuses on the ethical issues relevant to the RICHFIELDS project. Ethics for the purposes of this work package are restricted to the ethics of big data as understood through a review of academic literature. This report does not engage with ethics as propounded by philosophers such as Immanuel Kant and John Stuart Mill, or schools of moral thought such as utilitarianism, virtue ethics or pragmatism. Nor does it try to draw connections between these moral philosophies and ethics as conceived in the context of big data, though arguably some of the moral philosophies may have influenced our understanding of notions such as privacy and control over one's data (i.e. autonomy) that underpin the ethics of big data.[1]

Big data, consisting of huge volumes of data drawn from diverse sources such as apps, social media, discussion lists, online forums, Twitter feeds, status updates on Facebook, data from activity trackers, health platforms, home sensors and store loyalty cards and data provided by data brokers,[2] provide huge opportunities for researchers. Analysis of big data (that is discovering associations, trends and patterns in the data) has the potential to provide insight; for instance, in improving life style and the well-being of citizens and contribute to saving lives and improving health care management.[3] Whilst the value of big data for research can be fully appreciated, scant attention has been paid to the ethical consequences such as the individual's loss of privacy as a result of the use of data containing personal data (namely identifiers such as name, telephone numbers, addresses, IP address, biomarkers, religious affiliation, and trade union membership). The use of the software Beacon (which linked datasets) used by Facebook and the resulting backlash is an illustration of this. The software 'connected people's purchases to their Facebook account [which] advertised to their friends what a user had purchased, where they got it, and whether they got a discount. In one instance, a wife found out about a surprise Christmas gift of jewellery after her husband's purchase was broadcast to all his friends — including his wife … Others found their video rentals widely shared, raising concerns it might out people's sexual preferences and other details of their private life …'. [4] The out-of-court settlement, 'involved the establishment of a fund to better study privacy issues, an indication that progress was stepping well ahead of ethical considerations.' [5]

[1] See R Herschel & V M Miori (2017) 'Ethics & big data' *Technology in Society* 49: 33-36 who offer brief accounts of the different ethical theories and these theories can be applied to big data. They conclude that the 'collection and use of big data has little to recommend it from an ethical perspective … but it also opens the door to finding ways to mitigate any ethical shortcomings'(at 35).
[2] Data brokers typically collect and process data they source from social media, insurance claims, medical devices etc (N Terry (2014) 'Health privacy is difficult but not impossible in a post-HIIPAA driven world' *Chest* 146(3): 835-840).
[3] W Raghupathi & V Raghupathi (2014) 'Big data analytics in healthcare: promise and potential' *Health Information and Science System* 2:3, <https://doi.org/10.1186/2047-2501-2-3> accessed 1 September 2017.
[4] A Oboler, K Welsh & L Cruz (2012) 'The danger of big data: Social media as computational social science' *First Monday* 17:7 < http://firstmonday.org/article/view/3993/3269/ > accessed 5 September 2017.
[5] Ibid.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

This WP examines ethics as understood in the context of big data, the EU (European Union) legislation on data protection that embraces the ethical concerns, and the design of the ethical and legal aspects of the research infrastructure (RI) RICHFIELDS. Divided into five section, section Two (which follows this Introduction) provides an account of our understanding of ethics for the purposes of big data and also highlights the connection between the ethics of big data and law. Section Three provides an account of the salient features of the EU's General Data Protection Regulation[6] (GDPR) which comes into force on 25 May 2018, and highlights the provisions that are of particular importance along with the challenges they may pose for RICHFIELDS largely reliant on provision of data by third parties. It must be stressed at this juncture that this section is not a provision by provision commentary of the GDPR. Section Four addresses the ethical/legal aspects that RICHFIELDS needs to focus on in designing its RI. In doing so it highlights the areas that need to be addressed in the governance of RICHFIELDS. Section Five in concluding makes a few observations on the potential for future developments.

2. Ethics for Big Data

This section in the first part examines the ethical issues that confront the use of big data and in the second part delineates the connection between ethics, law and governance.

2.1. Ethical Challenges

Sophisticated communications technology, use of algorithms and storage capacity have greatly enhanced opportunities for researchers to analyse and use large data sets gathered from a variety of sources including apps, loyalty cards and activity trackers to understand links and trends and their impact on human beings. Insights thus arrived at have the potential to provide specifically targeted services to citizens in areas such as nutrition that would be of benefit to private sector organisations (such as those involved in the food industry), public sector organisations (such as hospitals), and also drive government policy in respect of healthcare. The biggest challenge to big data comes in the form of privacy. Human rights instruments, international and regional, protect an individual's right to privacy.[7] Against this context the pressing question is how should large data sets be handled for the purposes of research? How should we handle the ethics especially in the face of the potential for exposure as highlighted in the illustration given in the Introduction? Unlike medical research where there are some accepted standards such as the World Health Organisation's 'Standards and Operational Guidance for Ethics Review of Health-related Research with Human Participants',[8] and those from the UK Medical Research Council, the use of big data for purposes other than medical research have not as yet resulted in the formulation of common ethical standards to follow. It seems that attitudes to ethics of big data vary and there is much

[6] Regulation 2016/679.
[7] See for instance Article 12 Universal Declaration of Human Rights, Article 17 International Covenant on Civil and Political Rights, Article 8 European Convention on Human Rights.
[8] <http://apps.who.int/iris/bitstream/10665/44783/1/9789241502948_eng.pdf >accessed 1 October 2016.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

disagreement about the understanding of obtaining informed consent and use of ethics review boards.[9] Against this foggy backdrop of uncertainty regarding ethics it is important for any RI using big data to adopt ethical practices that do not undermine the rights enshrined in the human rights instruments and legitimises its standing as a respectable, reliable and accountable research tool for researchers.

A useful starting point for the ethical concerns in relation to big data is provided by Mittlestadt and Floridi.[10] Although their work relates to the bio-medical context, it is nevertheless of relevance to RICHFIELDS. Their extensive review of scholarly literature reveals that privacy, informed consent and ownership are the most frequently raised issues of ethical concern. Given the pervasive character of the human rights instruments and the recognition of autonomy it is no surprise that privacy was frequently raised as a principal ethical concern with many advocating anonymization and pseudonymisation as a means to preserve privacy.[11] Informed consent is not new to big data and is a mechanism widely used in research widely to legitimise the use of an individual's data for a particular purpose(s). It does not accommodate multiple data sets that are being used for purposes other than those for which the consent was sought. Mechanisms often utilised to enable re-purposing are blanket consent[12] for all potential uses, and tiered consent allowing individuals to permit specific uses.[13] Use of blanket consent (often used by app providers) is restrictive of individual autonomy.[14] Consent permitting specific uses also poses problems for research using big data since it is impossible to predict the uses of the data sets in the future. In order to balance the interests of both the individual and the researchers governance structures that recognise the rights of individuals to withdraw from participation and an independent ethics committee that review requests for access to data to determine whether the request would meet the ethical parameters without compromising privacy may be helpful. The independence of the ethics committee would contribute to the confidence of data subjects in the processing of their data for research purposes.[15]

[9] J Vitak, K Shilton & Z Ashktorab (2016) 'Beyond the Belmont principles: Ethics, practice and beliefs in the online data research community' *Proceedings of the 19th ACM Conference on Computer- Supported Co-operative Work & Social Computing* pp 941-953.
[10] BD Mitelstadt & L Floridi (2016) 'The ethics of big data; Current & foreseeable issues in biomedical context' *Sc Eng Ethics* 22: 301-341.
[11] A Markowetz, K Blaszkiewicz, C Mentag, C Switala, & TE Schlaepfer (2014) ''Psycho-informatics: Big data shaping modern psychometrics' *Medical* Hypotheses 82(4): 405-411; S Choudhury, J R Fishman, M L McGowan, & ET Joengst (2014) 'Big data, open science and the brain: lessons learned from genomics' *Frontiers of Human Neuroscience* 8: 239, doi: 10.3389/fnhum.2014.00239 accessed 1 September 2017.
[12] JPA Ioannidis (2013) 'Informed consent, big data and the oxymoron of research that is not research' *American Journal of Bioethics* 13(4): 40-42.
[13] MA Majumdar (2005) 'Cyberbanks and other virtual research repositories' *Journal of Law, Medicine & Ethics* 33(1); 31-39.
[14] Z Master, L Campo-Engelstein & T Caulfield (2015) 'Scientists' perspectives on consent in the context of bio-banking research' *European Journal of Human Genetics* i23(5): 569-574.
[15] See for instance the UK Biobank Ethics and Governance Council which acts as an independent guardian of the UK Biobank Ethics and Governance Framework and advises on its revisions, monitors and reports publicly on the conformity of the UK Biobank project with the framework and advises more generally on the interests of research participants and the general public in relation to UK Biobank <https://egcukbiobank.org.uk/Ethics-and-governance-framework.html > accessed 1 October 2017.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Ownership, the other feature identified,) is as Mittelstadt & Floridi acknowledge a complex concept. It could refer to the right to control data (namely, empowerment of the individual to control the means and ways in which his data is being utilised) and rights to benefit from the data (such as intellectual property rights that reside in the database and innovation from big data analysis which are examined in Work Package 13.1). Another related issue in respect to ownership is how an RI with vast quantities of data from different sources maintains data integrity.

There are three ethical principles that impact on research involving human subjects in biomedical and behavioural research. These succinctly stated in the Belmont Report[16] are also of general relevance. The principles enunciated are: respect for persons, beneficience, and justice. Respect for persons encapsulates two ethical concerns: the treatment of individuals as autonomous agents, and persons with diminished responsibility (vulnerable class) are protected. Beneficience is viewed as an obligation and finds expression in the duty not to do harm and to maximise possible benefits and minimise possible harms. Justice is understood as fairness in distribution and in the context of research refers to the duty to ensure that the selection process of research participants is directly related to the research study rather than on their easy availability due to their confinement in an institution, or vulnerability by virtue of belonging to a racial minority group. Where the research is publicly funded there is a duty to ensure wide distribution of the benefit of any advancement in therapeutic procedures or devices and that it be not restricted to those who can afford them.

2.2. Ethics, Law and Governance

The EU Charter of Fundamental Rights imparts everyone with the right to protection of personal data concerning him or her. Personal data and the rights of individuals are protected by the General Data Protection Regulation (GDPR). It focuses on privacy, informed consent, protection for certain types of data (e.g. racial origin, political affiliation), and ownership (in the form of control over data), identified as ethical concerns in 2,1 above. The ethical concerns elucidated through a legal instrument establishes the close relationship there is between the ethics of privacy in personal data and the law of privacy and personal data. At this juncture it would apt to point out that the right to privacy, covering the right to protection of person's identity, name, gender, appearance and dignity, is not the same as the right to private life. The right to private life is wider than the right to privacy and includes the 'right to establish and develop relationships with other human beings'.[17] The EU legislation

[16] The Belmont Report, available at <https://www.fda.gov/ohrms/dockets/ac/05/briefing/20054178b_09_02_Belmont%20Report.pdf> (accessed 21 January 2018). The other documents that impact ethics relating to research involving human subjects include the Nuremberg Code, and the Helsinki Declaration. See Bernard A Fischer Jr (2006) 'A summary of important documents in the field of research ethics' *Schizophrenia Bulletin* 32(1): 69-80.

[17] *Niemitz v Germany* (Application No 13710/88, Judgment of 16 December 1992. In Para. 29 stated:

> The Court does not consider it possible or necessary to attempt an exhaustive definition of the notion of "private life". However, it would be too restrictive to limit the notion to an "inner circle" in which the individual may live his own personal life as he chooses and to exclude therefrom entirely the outside world not encompassed within that circle.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

recognises, as will be seen in Section 3 below, the rights of individuals (data subjects) over their personal data and protects the individuals by imposing responsibilities and liabilities on those engaged in processing personal data of individuals. The GDPR also imposes obligations in respect of data and security, thus expecting those who process to have governance structures in place to ensure compliance with the GDPR. The governance framework of RICHFIELDS is enunciated in WP 13.3.

3. General Data Protection Regulation (GDPR)

This section provides an account of the salient features of the GDPR and examines the key provisions that are likely to be of relevance in designing the ethical and legal aspects of RICHFIELDS. By no means is this section a commentary on all the provisions of the GDPR.[18] There is no attempt to compare the GDPR with the Data Protection Directive, the reason being that the GDPR will come into effect in the EU Member States in May 2018. As for the UK, the Queen's Speech on 21 June 2017 confirmed that after its departure from membership of the EU the Government's intention is to bring the GDPR into UK law to ensure that UK's data protection framework is 'suitable for our new digital age, allowing citizens to better control their data'. [19] The plan is 'to implement the General Data Protection Regulation and the new Directive which applies to law enforcement data processing … helping to put the UK in the best position to maintain our ability to share data with other EU member states and internationally after we leave the EU'.[20]

3.1. Rationale for the GDPR

The GDPR replaces the data protection framework created by the EU Data Protection Directive 1995.[21] Being a Directive, Member States had some manoeuvrability in implementing the Directive resulting in divergent laws producing uncertainty and inconsistency. When data was transferred from one state to another, for instance from the UK to Spain, the data was subject to different national laws which undermined confidence in the digital economy. The GDPR changes this. In being a Regulation it is self-executing and will result in harmonisation in the laws relating to data protection amongst member states. The EU policymakers were also of the view that the GDPR would modernise the aging data protection framework of the Directive since the role and use of personal data by society and commerce had evolved since 1995 greatly making a noticeable impact on the economy.

Respect for private life must also comprise to a certain degree the right to establish and develop relationships with other human beings.

[18] For instance, this section does not consider provisions relating to supervisory authority, their competence and powers, or the establishment of the European Data Protection Board and their remit. Neither does it examine in any detail the provisions relating to fines and penalties.

[19] The Queen's speech and associated background briefing, on the occasion of the opening of parliament on Wednesday 21 June 2017.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/620838/Queens_speech_2017_background_notes.pdf > p 46, accessed 10 October 2017.

[20] Ibid.

[21] Directive 95/46/EC on the protection of individuals with regard to the protection of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

According to the EU Communication, 'the EU's 1995 Directive, legislative instrument for the protection of personal data in Europe, was a milestone in the history of data protection. Its objectives, to ensure a functioning Single Market and effective protection of the fundamental rights and freedoms of individuals, remain valid. However, it was adopted 17 years ago when the internet was in its infancy. In today's new, challenging digital environment, existing rules provide neither the degree of harmonisation required, nor the necessary efficiency to ensure the right to personal data protection. That is why the European Commission is proposing a fundamental reform of the EU's data protection framework.' [22] Protection of individuals' right to privacy is seen as pivotal whilst engaging with the digital economy. As the EU Communication notes, 'In this new digital environment, individuals have *the right to enjoy effective control over their personal information*.[23] Data protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union,[24] as well as in Article 16(1) of the Treaty on the Functioning of the European Union (TFEU)[25][26], and needs to be protected accordingly.'[27] The purpose of the GDPR is therefore to improve individuals' control over their personal data through consent and also enable through harmonisation the ease of transfer within the European Economic Area (amongst EU member states and Norway, Lichtenstein and Iceland). In better protecting an individual's control over his/her personal data the GDPR increases the responsibilities and level of compliance of most organisations that deal with such data. Consent, which is central to the legitimising of processing of individuals' data, and responsibilities and the compliance measures required by the GDPR are considered in the following sections.

Before proceeding it must be pointed out that there are similarities between some of the provisions of the GDPR and those of the earlier Data Protection Directive. National courts are likely to look at earlier European Court of Justice judgments relating to data protection for ease of interpretation. The courts which adopted a purposive approach when interpreting the Data Protection Directive[28] are also likely to apply the purposive approach by looking to the intention of the legislation and the specific provision. The recitals may also be useful for

---

[22] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century' COM(2012) 9 final, 25.1.2012<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0009&from=en >p,2, accessed 12 December 2016.

[23] Author's emphasis.

[24] Article 8 Charter of Fundamental Rights of the European Union ((2000/C 364/01) reads:

Protection of personal data

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collecte concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

[25] OJ C326/47, 26.10.2012. Article 16(1) reads:

Everyone has the right to the protection of personal data concerning them.

[27] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century' COM(2012) 9 final, 25.1.2012<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0009&from=en >p,1, accessed 12 December 2016.

[28] Case C-73/07 Satamedia [2008] ECR I-09831, paragraph 51.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

the purposes of interpretation. However they only are a means to explain by giving reasons for the legislation and should not be viewed as normative provisions.[29]

3.2. Legitimising processing through consent

As stated earlier, Article 8[30] of the Charter of Fundamental of the European Union provides every individual has the right to protection of personal data about him or her. To legitimise processing consent is pivotal. The GDPR emphasises consent and imposes significant responsibilities on entities that process data. For consent to be valid (1) it must be freely given; (2) a proper explanation of what the individual is consenting to must have been provided before the consent is obtained; (3) *separate consents must be given for separate purposes*;[31] (4) consent can be refused; and (most important of all) (5) consent can be withdrawn at any time.[32] The GDPR expects all consent (be they from a child or an adult) to meet the above conditions. However, in the case of children below sixteen they require the authority of the person with parental responsibility.[33]

The burden of proof that the data subject has consented lies with the controller. There is no requirement in the GDPR that it be in writing but for evidential purposes it is likely to be in writing. Silence on the part of the individual when agreement is sought would not meet the threshold imposed by the GDPR. For the individual to give consent freely he/she has to understand what is being sought by the controller. Clauses in consent agreements which are complex and lack clarity are unlikely to meet the consent threshold.

It is possible that there are existing consent agreements that have been obtained under legislation implementing the Data Protection Directive. These consent agreements need to meet the current standards set by the GDPR. If they do not then fresh consent needs to be obtained. This requirement is a challenge for RIs. Many of the third parties such as app providers collecting data on food purchase and food consumption who supply data to RICHFIELDS may have obtained general consent which does not indicate how the data could be used for other purposes or what those purposes are likely to be since it is difficult to predict hitherto unknown research uses for the data. These types of consent agreements are unlikely to meet the basic condition of separate consents given for separate purposes imposed by the GDPR. This creates a barrier to sharing data for research purposes. However Recital 33 recognises that, '[i]t is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection'. In these circumstances, 'data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research.

[29] European Union (2015) *Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union Legislation* Luxembourg: Publications of the European Union, Guideline10.
[30] See fn 20 for text of Article 8.
[31] Author's emphasis.
[32] Article 7.
[33] Article 8.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose'. While recognising the importance of data for research purposes this recital states that ethical standards for scientific research need to be met and the individual (i.e. data subject) is given the opportunity to give consent to specific areas of research or research projects. So RIs that collect and provide access to data for research purposes will have to ensure that ethical standards for research are met and provide opportunities to individuals to give their consent. Further processing which is for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes will not be considered to be incompatible with the initial purposes.[34] Ethical standards for scientific research may vary from state to state although instruments such as the Helsinki Declaration (which sets out the ethical principles for medical research on human subjects) and Code of Ethics produced by the International Sociological Association, may provide a workable framework for bringing about some degree of harmonisation in ethical standards for research purposes.[35]

In the context of informed consent a recent development may well prove to be an useful route for research purposes. Termed dynamic consent,[36] it is a personalised digital interface with the participant which enables continuous engagement between the participant and researcher. Being participant-centric it puts the participants in control of their data and their uses. It replicates the informed consent model common in medical research but uses a digital interface. Dynamic consent been used in a number of research projects such as RUDY (Rare UK Diseases of Bone Joints and Blood Vessel)[37] and CHRIS (Co-operative Health Research in South Tyrol).[38] Both these studies have limited number of participants, thus making it possible for continuous two-way communication. Adoption of this approach by RIs may pose difficulties especially where data is acquired from a variety of sources. The RI would need to make adaptations to its technical architecture and the procedures for acquiring information from third parties. The RI would also need to be clear about the role it plays: for instance, is the RI a broker between the researcher and the participants? If this is the case, it needs to reflect on questions such as the nature of its legal relationships with third parties who supply it with data and the researchers who access the RI.

---

[34] Article 5(1)(b).
[35] See also European Commission (2013) *Ethics for researchers: Facilitating research excellence in FP7 (*Brussels: European Commission).
[36] Haws Williams, Karen Spencer & William G Dixon (2015) 'Dynamic consent: A possible solution to improve patient confidence and trust in how electronic patient records are used in medical research' JMIR Med. Inrmar. 3(1) e 3, J Kaye, EA Whitley, D Lund et al (2015) 'Dynamic consent: a patient interface for twenty-first century research network' *Eur J Hum Genet* , 23(2) 141 – 146.
[37] See <https://www.ndorms.ox.ac.uk/research-groups/javaid-group-rare-bone-diseases-and-osteoporosis-epidemiology/projects/rudy> accessed 10 March 2018, M K Javaiad, L Forestier-Zhang et al (2016) 'The RUDY study platform – a novel approach to patient driven research in rare musculoskeletal diseases' *Orphanet Journal of Rare Diseases* (**11**):150 <https://doi.org/10.1186/s13023-016-0528-6> accessed 17 March 2018.
[38] For more details on this project http://www.eurac.edu/en/research/health/biomed/projects/Pages/default.aspx.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

3.3. Who does the GDPR protect?

As indicated in 2.1., the GDPR aims to give control to an individual over his/her data. It therefore lays down rules in relation to the protection of natural persons with regard to the processing of their personal data. The GDPR therefore does not protect data of entities such as companies, charitable organisations and associations.[39] The protection afforded by the GDPR is not linked to the nationality or place of residence of the natural persons.[40] So where an individual (that is a data subject) is present in the EU he/she is protected even in the absence of place of residence in the EU. To illustrate, if Joe (a US citizen with no place of residence in the EU) were to subscribe to a retail store (e.g. Boots in the UK) and provides personal data, Joe's data falls within the scope of the GDPR. The GDPR however does not apply to a citizen of an EU state (with a place of residence in the EU) were he/she to give her personal data whilst in a non-EU state. So if Jane whilst visiting the US were to give her data (such as her address, phone number, passport details) to Macy's in New York the personal data gathered will not be subject to the GDPR. The data given will be governed by US laws.

3.4. What type of data does the GDPR apply to?

The GDPR applies to personal data. The GDPR in Article 4(1) defines 'personal data' as any information relating to an identified or identifiable natural person ('data subject')'. So data such as height, weight, and address of John Smith would constitute personal data. Personal data also includes data of 'one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'. So if X can be identified on the basis of data such as the rich Russian who owns a football club and lives close to Oxshott, Surrey, then such data would classify as personal data because of indirect identification.

3.4.1. Prohibition in respect of processing of certain types of personal data

There are certain types of personal data the processing of which is prohibited by the GDPR.[41] Data 'revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation' fall within this category. There are however a number of exceptions to the general prohibition, which include 'processing necessary for the purposes of preventive or occupational medicine; for the provision of health or social care or treatment or the management of health or social care systems and services; for reasons of public interest

[39] Recital 14 states:
> This Regulation does not cover the processing of personal data which concerns legal persons and in particular undertakings established as legal persons, including the name and the form of the legal person and the contact details of the legal person.

[40] See Recital 14 which states:
> 'The protection afforded by this Regulation should apply to natural persons, whatever their nationality or place of residence, in relation to the processing of their personal data….'

[41] Art 9(1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

in the area of public health, such as protecting against serious cross-border threats'.[42] This prohibition also is not applicable where 'the data subject has given explicit consent to the processing of those personal data for one or more specified purposes' unless the laws of the EU or the Member State provide that the general prohibition referred to paragraph 1 may not be lifted by the data subject'.[43] The exceptions mean that processing of excluded data for the purposes of healthcare and preventive medicine is justifiable. Big data (with its potential to detect novel associations and enabling predictive links) is widely viewed as a vital source for medical innovations and health care. A report from McKinsey, for instance, highlights some of the benefits which include improving healthcare, well-being and improving the cost of care.[44] The contribution of big data is also recognised by politicians. For instance, in 2011, David Cameron (the then Prime Minister of the UK) said that every NHS (National Health Service) patient should be a 'research patient' and their data opened up to private healthcare firms.[45] He also went on to say this 'does not threaten privacy, it doesn't mean anyone can look at your health records, but it does mean using anonymous data to make new medical breakthroughs.'[46]

3.5. When does the GDPR apply?

The GDPR applies to both manual and automated processing of personal data. Processing is defined in Article 4(2) as 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.' The wide definition adopted indicates that manual filing systems also fall within the GDPR as long as it is structured to specific criteria.

The GDPR applies to the processing of personal data in the context of the activities of the establishment of a controller or processor in the European Union. The place of processing is immaterial. So where a controller/processor has a place of establishment in the EU but the processing takes place in China, for instance, the GDPR will apply. In order to determine whether the GDPR applies one has to look to the establishment of the controller/processor.[47] According to Recital 22 in determining establishment one has to consider where the activity is effectively carried out through stable arrangements. The legal form of the arrangements in the form of a branch or subsidiary with legal personality is not to be regarded as determining

[42] Art 9(2).
[43] Ibid.
[44] Center for US Health System Reform, Business Technology Office (2013) The 'big data' revolution in health care' McKinsey & Co, < http://achc.org.co/hospital360/contextos/Tecnologia_e_Informacion/Big_Data/Revolucion_de_la_Informacion.pdf> accessed 1 September 2017. See also D N Reshef, YA Reshef et al (2011) 'Detecting novel associations in large data sets' *Science* 334: 1518-1524.
[45] BBC News 'Everyone "to be a research patient", says David Cameron', 5 December 2011 http://www.bbc.co.uk/news/uk-16026827 accessed 1 September 2016.
[46] Ibid. The news inevitably attracted criticism in putting commercial interests(ahead of patient privacy. Patient Concern said there were real worries about the proposal to make patients' medical data available to private firms as the information would include postcodes and age profiles which would be possible to trace back to the individuals concerned.
[47] Art 3(1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

factor. To illustrate, an RI that collects data on consumers' in-store behaviour and shopping patterns with personnel in the Netherlands who decide on matters such as types of data to collect, mechanism for data collection, purposes of the data, types of access to researchers, will be subject to the GDPR even though the technology for storing/processing may be located in a non-EU state.

As stated earlier, the focus of the GDPR is to protect data subjects within the EU. In order to protect them in cases where their data is collected from outside the EU for instance through e-commerce, the GDPR extends its territorial scope. So data controller/data processors outside of the EU have to be GDPR compliant where they process personal dataof individuals who are in the EU (nationality, residence of data subject are irrelevant) and where the processing activities are related to the offering of goods or services to such data subjects in the EU, or the monitoring the behaviour of such data subjects.[48] In order to determine whether goods or services are being offered to data subjects in the Union intention is a core ingredient. It is important to establish whether the controller/processor intended to offer the goods to the data subject in the EU. Mere accessibility to the controller's website, contact details, email address, and the use of a language (that is generally used in the country of the controller's establishment) are insufficient for establishing the required intentions. Factors that may be relevant to establish intention include 'the use of a language or a currency generally used in one or more Member States with the possibility of ordering goods and services in that other language, or the mentioning of customers or users who are in the Union'.[49] To illustrate, if a data subject in the EU can order goods from the Metropolitan Museum of Art's shop in New York which provides shipping rates to destinations in the EU, it is apparent that it intends to sell goods to data subjects in the EU. This means that it needs to meet the standards set by the GDPR in the processing of personal data of the data subject.
As for the question whether the controller/processor's activity amounts to monitoring of data subjects in the EU, according to Recital 24 it should be ascertained whether natural persons are tracked on the Internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes'. This suggests that websites such as nutrition websites and recipe websites, that may collect information on searches for vegan diet, paleo diet and collect IP addresses and other details, would need to be GDPR compliant even if they are not located in the EU.
It is also possible for the GDPR to apply to processing as a result of the application of the rules of private international law. One such illustration provided by the GDPR is a controller not established in the Union, such as in a Member State's diplomatic mission or consular post.[50]

[48] Article 3(2). Author's emphasis of phrases 'offering of goods or services' and 'monitoring the behaviour' from Article 3(2).
[49] Recital 23.
[50] Recital 25.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

3.5.1. Processing falling outside the scope of the GDPR (Anonymised data)

De-identification (or anonymization)[51] of data is often offered as a panacea to avoid the rigid framework of data protection laws that act in the interests of data subjects. It is a process where by all the identifiers of a data subject are stripped (removed) which means that the subject cannot be identified, thus preserving the privacy of the data subject whilst allowing access to the data for research purposes. Anonymised personal data falls outside the scope of the GDPR as Recital 26 makes clear.[52] While this exclusion favours researchers there is a major problem with anonymisation as there are technologies currently available that can be used to re-identify individuals from anonymised data. According to Narayanan and Shmatikov, '[j]ust as medieval alchemists were convinced a (mythical) philosopher's stone can transmute lead into gold, today's privacy practitioners believe that records containing sensitive individual data can be de-identified by removing or modifying PII [personally identifiable information]'.[53] The distinction that is drawn between identifying and non-identifying attributes by existing privacy technologies for the purposes of anonymising[54] is 'increasingly meaningless as the amount and variety of publicly available information about individuals grows exponentially'.[55] Anonymization therefore is not a silver bullet. However anonymization should not be dismissed as a whole. There are codes of practice that may be helpful in reducing the risk of re-identification. The UK Information Commissioner's Office Code of Practice on anonymisation[56] is one such code of practice. It is helpful towards assessing the risk of re-identification even though it clearly admits 'that the risk of re-identification through data linkage is essentially unpredictable because it can never be assessed with certainty what data is already available or what data may be released in the future. It is also generally unfeasible to see data return (ie (i.e.) recalling data or removing it from a website) as a safeguard given the difficulty, or impossibility, of securing the deletion or removal of data once it has been published. That is why it is so important to take great care, and to carry out as thorough a risk analysis as is possible, at the initial stage of producing and disclosing anonymised data.'[57] It provides a list of issues to consider in the process of anonymization such as assessing the 'risk of piecing different bits of information together to

[51] The word 'de-identification' is used here to mean 'anonymisation'. It could also be used to refer to various methods which create a distance between the data and identities of subjects such as encryption, sharding (fragmentation of data) and pseudonymisation. (In this work package pseudonymisation is discussed separately since the GDPR does apply to peeudonymisation. For further on the various methods for distancing data from personal identities see W Kuan Hon, C Millard and I Walden (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 ,< https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577 > accessed 1 September 2016.

[52] It states, 'The principles of data protection should ... not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes'.

[53] A Narayanan & V Schmatikov (2010) 'Privacy and security: Myths and fallacies of "personally identifiable information"' *Communications of ACM* 53(8): 24-26.

[54] L. Sweeney (2002) 'Achieving k-anonymity privacy protection using generalization and suppression' International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems 10(5): 571-588.

[55] Narayanan (n 44) p 25.

[56] ICO *Anonymisation: Managing Data Protection Code of Practice*
< https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf > accessed 14 September 2016.

[57] Ibid. p 18.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

create a picture of someone .. whether the information have the same characteristics to facilitate data linkage [such as] the same code number … to refer to the same individual in different datasets, … what technical measures might be used to achieve re-identification'…. [and where] a penetration test has been carried out, what re- identification vulnerabilities did it reveal.'[58] In this context the ICO stresses the importance of governance and advises organisations anonymising personal data have an effective and comprehensive governance structure which also includes oversight of governance arrangements at a senior level.[59]

3.5.2. Processing pseudonymised data

The GDPR makes a distinction between anonymisation and pseudonymisation, with the latter falling within GDPR. Pseudonymisation, according to the GDPR, is 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information'. [60] The 'additional information is [to be] kept separately and … subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'.[61] Pseudonymisation of personal data is seen by the framers of the GDPR as 'reducing the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations.'[62] Pseudonymisation does not take the data processing outside the remit of the GDPR. It 'is not intended to preclude any other measures of data protection'.[63] Pseudonymisation is likely to prove an important tool for processing of large data sets for research purposes. So RICHFIELDS can utilise pseudonymisation for the purposes of research. However, the RI as controller[64] 'has to take technical and organisational measures necessary to ensure, for the processing concerned, that [GDPR] is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately. The controller processing the personal data should indicate the authorised persons within the same controller'.[65]

3.6. Who has to comply with GDPR and their responsibilities

The GDPR identifies two parties who are required to comply with the GDPR. They are the 'controller' and the 'processor'. 'Controller' is defined as '*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*' and 'processor' as '*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*'.[66] The definitions make clear that a controller or a processor can be an individual, companies, institutions such as associations, charitable organisations, educational

[58] Ibid. p 24.
[59] Ibid. p 39.
[60] Article 4(5).
[61] Ibid.
[62] Recital 28.
[63] Ibid.
[64] See 'Who has to comply with GDPR and their responsibilities' below for meaning of controller.
[65] Recital 20.
[66] Article 4.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

establishments, and public authorities including quasi-government bodies. An illustration may be helpful in understanding the difference between a controller and processor. An RI which does not, for instance, directly collect food consumption data through apps from data subjects but determines the means of processing personal data it receives from third parties would be a controller as opposed to an entity such as a data management company or educational establishment that stores or catalogues the data for the controller. The latter would be a processor. The GDPR does not say that a controller cannot be a processor as well. It may well be that an RI such as RICHFIELDS could take on both roles of controller and of processor. The GDPR imposes responsibilities on both controllers and the processors and both are equally subject to fines in the event of not meeting the obligations imposed by the GDPR.

3.6.1. Controller's Responsibilities

The controller is responsible for compliance with the GDPR and should also be able to demonstrate compliance with the principles relating to data processing of personal data.[67] These principles as set out in Article 5 are: lawfulness, fairness and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality; and accountability. To meet these responsibilities the controller must process personal data in relation to the data subject in a transparent manner, lawful and fair manner. This implies that the data subject is informed of why the data is being collected, how the data is to be processed and that the processing is legal. The data collected for specified and explicit purposes must not be processed in other contexts that is not compatible with the specified purposes. The data also must be adequate, relevant and limited to what is necessary for the purposes for which processed. The controller is also responsible for the data to be accurate and kept up to date. Where there are inaccuracies the controller is expected to take every reasonable step to erase or rectify the inaccuracies. The data kept in a form that permits identification of data subjects should not be kept for longer than necessary. And the data must be processed in a manner that ensures appropriate security against incidents such as unauthorised or unlawful processing and accident loss, damage or destruction through the use of technical and organisational measures. However Article 89(1) allows derogation from purpose limitation and storage limitation for archiving purposes in the public interest, scientific or historical research or statistical purposes subject to implementation of the appropriate technical and organisational measures to safeguard the rights and freedoms of the data subject.

The controller is responsible for the implementation of appropriate technical and organizational measures to ensure that processing is performed in accordance with the GDPR and must also be to demonstrate that processing is performed in accordance with the GDPR. The controller may demonstrate compliance through adherence to approved codes of

[67] Article 5(2).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

conduct[68] or certification mechanisms[69]. In implementing appropriate measures the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons have to be taken into account.[70] These measures have to be reviewed and updated where necessary.[71] Measures that the controller may take include allocation of responsibilities for data protection, a data protection impact assessment, risk mitigation plan, implementation of pseudonymization (the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information), and data minimization in order to meet the requirements of the GDPR and protect the rights of data subjects.[72]

Where there are several organizations that share the responsibility for the processing of personal data, the GDPR recognises the existence of joint controllers and they must determine their respective responsibilities by agreement and provide the content of this agreement to the data subjects, defining the means of communication with processors with a single point of contact.[73] According to Article 35 impact assessments will be required where the nature, scope, context and purposes of processing is likely to result in high risk to the rights and freedoms of natural persons. The GDPR itself does not give a methodology for assessing and managing risk but it is important that controllers and processors adopt a consistent approach.[74] The controller should have prior consultation with the supervisory authority[75] where the data protection impact assessment under Article 35 indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk.[76]

3.6.2. Processor's Responsibilities

Where a controller uses a processor to carry out processing on his behalf the controller must use processors who provide sufficient guarantees to implement appropriate technical and organizational measures in such a manner that the processing will meet the requirements of the GDPR and ensure the protection of the rights of the data subject.[77] So where a controller located in the EU uses a processor from a non-EU state it is imperative that the controller ensures that the processor complied with the standards imposed by the GDPR. Failure to comply will attract penalties and both the controller and the processor can be fined. In deciding on the amount of administrative fine the degree of responsibility of the controller or

[68] Article 40.
[69] Article 42.
[70] Article 24(1).
[71] Article 24(3).
[72] Article 25(1).
[73] Article 26 (1).
[74] The following documents may be useful for assessing privacy risk: CNIL *Methodology for Privacy Risk Management: How to Implement the Data Protection Act* https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf (accessed 10 October 2017); ICO *Conducting Privacy Impact Assessments Code of Practice* < https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf> (accessed 10 October 2017).
[75] Under Article 58 the supervising authority has both investigative and corrective powers.
[76] Article 36.
[77] Article 28.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

processor taking into account technical and organisational measures implemented by them[78] will be taken into account.

The GDPR mentions adherence to approved codes of conduct and also certifications as a means of demonstrating compliance. It is not clear whether codes of conduct and certification in the form of seals by an association or industry accredited body[79] will have pan-European effect. This is an area that requires further elucidation from the European Data Protection Body that is to be set up by the GDPR.[80]

At this juncture mention must be made of The International Organisation for Standardisation's (ISO) 27001 which may provide a partial framework for information security management though it does not meet all the requirements of the GDPR. There is also ISO 27018 which lists safeguards to increase the level of protection of PII in public clouds and these include: (1) rights of the customer to access and delete the data, (2) processing the data only for the purpose for which the customer has provided this data, (3) recording all the disclosures of personal data, (4) disclosing the information about all the sub-contractors used for processing the personal data, (5) notification to the customer in case of a data breach, (6) document management for cloud policies and procedures, (7) policy for return, transfer and disposal of personal data, (8) confidentiality agreements for individuals who can access personal data, (9) records of user access to the cloud, (10) specifying the minimum security controls in contracts with customers and subcontractors, (11) deletion of data in storage assigned to other customers, (12) disclosing to the cloud customer in which countries will the data be stored, and (13) ensuring the data reaches the destination.

It is difficult to say whether the adoption of the ISO standards would make an organisation or an institution fully GDPR compliant though certain aspects of ISO 27001 and 27018 might reflect best practices to follow especially where personal data are processed in the cloud. Hence they are useful documents to scrutinise closely within governance mechanisms of an RI.

3.6.3. Security and Notification

The GDPR requires controllers/processors to consider security measures. These include pseudonymisation, ensuring integrity, confidentiality, availability and resilience of processing systems, regular testing, assessing and evaluating security measures etc.[81] It should include not only day-to-day running but also the risks that such a system would face and what measures have been put in place to ensure (at least lower) the risk of security breaches. In

[78] Article 83 2(d).
[79] According to Article 40(2) '[A]ssociations and other bodies representing categories of controllers or processors may prepare codes of conduct, or amend or extend such codes, for the purpose of specifying the application of this Regulation …'.
[80] Article 68. See also Article 70 on the tasks of the Board.
[81] Article 32.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

the event of a security breach (however minor) affecting personal data it must be notified by the controller to the supervisory authority no later than 72 hours after becoming aware of the breach unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. If the controller is unable to notify within the time frame then the notification should be accompanied by reasons for the delay. [82] Where there is a high risk to the rights and freedoms of the data subjects then the notification also needs to be sent to data subjects.[83] As for the processor, he is under an obligation to notify the controller of the breach without undue delay.[84]

A notification to supervisory authority should describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records concerned; the likely consequences of the personal data breach, the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects; and communicate the name and contact details of the data protection officer or other contact point where more information can be obtained.[85] Article 33 also places the controller under an obligation to document any personal data breaches, comprising the facts relating to the personal data breach, its effects and the remedial action taken. This documentation is to enable the supervisory authority to verify compliance with Article 33.

Where a breach requires communication to a data subject it must express in clear and plain language the nature of the personal data breach and describe the likely consequences of the personal data breach, the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects, and provide the name and contact details of the data protection officer or other contact point where more information can be obtained.[86] However communication to the data subject is not required where '(a) the controller has implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the personal data breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption; the controller has taken subsequent measures which ensure that the high risk to the rights and freedoms of data subjects …is no longer likely to materialise; (c) it would involve disproportionate effort.' [87] In such a case, data subjects can be informed through public communication or a similar measure that is equally effective.

[82] Article 33.
[83] Article 34.
[84] Article (34(2).
[85] Article 33(3).
[86] Article 34(2).
[87] Article 34(3).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

While this Report does not look in any detail at the consequences of infringements of the Regulation for the controller and the processor Articles 82 gives a person[88] the right to receive compensation from the controller or processor for the damage (be it material or non-material damage) suffered as a result of an infringement. Administrative fines can be also imposed on the controller and the processor under Article 83. The decision about whether to impose a fine and the amount will be taken in light of factors such as the nature, gravity and duration of the infringement, the nature, scope or purpose of the processing concerned as well as the number of data subjects affected and the level of damage suffered by them, any action taken by the controller or processor to mitigate the damage suffered by data subjects and the degree of co-operation with the supervisory authority.[89]

The amount of fines will depend on the nature of the infringement such as the processes of obtaining consent or implementation of technical and organisational measures. For instance an infringement of Article 25[90] relating to implementation of data minimisation and pseudonymisation could attract 'administrative fines up to 10 000 000 EUR, or in the case of an undertaking, up to 2 % of the total worldwide annual turnover of the preceding financial year, whichever is higher'.[91] Whereas, infringements of the basic principles for processing,[92] including conditions for consent would attract fines of up to '20 000 000 EUR, or in the case of an undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher'.[93] The administrative fines are high thus reinforcing the seriousness with which the GDPR regards the rights of data subjects in respect of their personal data.

3.6.4. Data Protection Officer

The GDPR requires the controller and the processor to designate a data protection officer (DPO) where (1) processing is carried out by a public authority or body, or (2) where the core activities of the controller/processor consist of processing operations which by virtue of their nature, their scope and/or purposes, require regular and systematic monitoring of data subjects on a large scale; or (3) where the core activities of the controller/processor consist

[88] The GDPR does not use the term data subject in Article 82 but uses 'person' suggesting thereby it includes natural and legal persons.
[89] See Article 83(2) for a full list.
[90] Article 25 reads:
1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, 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.
2. The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.
3. An approved certification mechanism pursuant to Article 42 may be used as an element to demonstrate compliance with the requirements set out in paragraphs 1 and 2 of this Article.
[91] Article 83(4).
[92] See Articles 5,6,7 and 9.
[93] Article 83(5).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

of processing a large scale special categories pursuant to special categories of data or data relating to criminal convictions and offences.[94]

This means that an RI such as RICHFIELDS that is likely to monitor consumption behaviour of data subjects and purchase habits would require a DPO. According to the GDPR the DPO could be a member of staff of the controller or the processor or can be appointed through a service contract.[95] The appointment has to be on the basis of professional qualities and expert knowledge of data protection law and practices.[96] The DPO must be suitably qualified to perform the tasks as set out in the GDPR. In performing these he should have due regard to the risk associated with processing operations, against the context, nature, scope, context and purposes of processing.[97] The tasks are as listed below:

- to inform and advise the controller or the processor and the employees who carry out processing of their obligations pursuant to this Regulation and to other Union or Member State data protection provisions;
- to monitor compliance with this Regulation, with other Union or Member State data protection provisions and with the policies of the controller or processor in relation to the protection of personal data, including the assignment of responsibilities, awareness-raising and training of staff involved in processing operations, and the related audits;
- to provide advice where requested as regards the data protection impact assessment and monitor its performance pursuant to Article 35;
- to cooperate with the supervisory authority; and
- to act as the contact point for the supervisory authority on issues relating to processing, including the prior consultation referred to in Article 36, and to consult, where appropriate, with regard to any other matter.[98]

3.7. Rights of Data Subjects

As stated in Section Three individuals have the right to enjoy effective control over their personal information. The GDPR gives a number of rights to data subjects and these include the right to access,[99] right to rectification[100] and the right of erasure (also known as the right to be forgotten),[101] right to restriction of processing,[102] and the right to object.[103]

[94] Article 37(1).
[95] Article 37(6).
[96] Article 37(5).
[97] Article 39(2).
[98] Article 39(1) (a) – (e). See also 'Controller's responsibilities' above.
[99] Article 15.
[100] Article 16.
[101] Article 17.
[102] Article 19.
[103] Article 21.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Under the right to access, the data subject can ask the controller to provide information on whether they process any personal data about him or her and where this is the case require access to the held data. Data subjects can also seek information about matters such as the purposes of the processing, categories of personal data concerned, the recipients to whom the personal data have been or will be disclosed(in particular recipients in third countries or international organisations), and the period for which the personal data will be stored and the criteria for determining the period of storage. They can also seek information from the controller about rectification/restriction or erasure of personal data and to object to processing. The controller can also be requested information about the right to lodge a complaint with the supervisory authority. Since it is possible that the data was obtained from a third source and not personally collected from the data subject he/she can seek information from the controller of the source of the data. So RIs that are reliant on data from third parties should be prepared to inform the data subject of the source of the data supply when requested by the data subject.

The right to rectification of data is self-explanatory but when requested the controller must do so without undue delay. Similarly, where a data subject requests that information about him/her be erased these must be done without undue delay by the controller. The obligation to erase is dependent on the following:

- the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
- withdrawal of consent by the data subject;
- unlawful processing of personal data;
- objection to the processing on the basis that the data has been obtained for directing marketing purposes, including profiling.

In the event of the controller being obliged to erase but the data has been made public the controller whilst 'taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data'.[104]

The data subject also has the right to obtain from the controller a restriction of processing in situations listed below:

- the accuracy of the personal data is contested by the data subject, for a period enabling the controller to verify the accuracy of the personal data;

[104] Article 17(2).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

- the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead;
- the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims;
- the data subject has objected to processing pursuant to Article 21(1) pending the verification whether the legitimate grounds of the controller override those of the data subject.[105]

Under Article 21, the data subject has the right to object to processing. Where such a right has been raised the controller should not process the personal data unless the controller can demonstrate compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.[106] It is possible under Article 21(6) for the data subject to object, on grounds relating to his or her particular situation, where personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1)[107] unless the processing is necessary for the performance of a task carried out for reasons of public interest.

As can be seen from the above the GDPR gives the data subjects a lot of control over their personal data unless they are overridden by legitimate interests such as public interest and legal claims. Controllers therefore should have appropriate mechanisms in their organisation that can deal with the requests of data subjects and deal with them effectively and efficiently. An RI dealing with huge data sets will also need to have appropriate mechanisms. The appropriate form and its constitution will need to be addressed in the governance structure along with drawing up suitable protocols for complaints procedures to receive and deal with requests from data subjects.

3.8. Trans-border Data Transfer

As stated in the Introduction, the GDPR's intention is to bring about harmonisation in the data protection laws as between member states. So transfer of data between member states should pose no problems. However 'flows of personal data to and from countries outside the Union and international organisations are necessary for the expansion of international trade and international cooperation.'[108] The GDPR's aim is to ensure that the level of protection of natural persons imparted by it is not undermined as a result of onward transfer to third countries and onward transfer from that third country to controllers or processors in that country or another third country.[109] Transfers are made possible to third countries or

[105] Article 18(1).
[106] Article 21(1).
[107] See 'Rights of controller'.
[108] Recital 101.
[109] Article 44 reads:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

international organisation provided the safeguard mechanisms are satisfied. So transfer would be possible where the European Commission has decided that 'the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection.'[110] An adequate level of protection would essentially be equivalent to that ensured in the European Union.[111] The Commission assesses the adequacy of protection in the third state by considering a number of factors such as respect for human rights, rule of law, existence and effective function of independent supervisory authority with responsibility for ensuring and enforcing compliance with the data protection rules, in the third country or to which an international organisation is subject, and 'international commitments the third country or international organisation concerned has entered into, or other obligations arising from legally binding conventions or instruments as well as from its participation in multilateral or regional systems, in particular in relation to the protection of personal data.' [112]

In the absence of the adequacy safeguard, transfer of personal data by a controller or processor to a third country or an international organisation is possible where the controller or processor has provided appropriate safeguards and provided enforceable data subject rights, and effective legal remedies.[113] Appropriate safeguards include legally binding and enforceable instrument between public bodies, binding corporate rules (BCR), approved code of conduct or approved certification mechanisms along with binding commitments from the controller or processor in the third country to apply the appropriate safeguards in respect of rights of data subjects.[114] Standard data protection contractual clauses approved by the Commission as well as by the supervisory authorities can also be used and these would not need prior authorisation by the supervisory authority.[115] Standard clauses that are currently in use may be valid though their repeal is possible under the GDPR. The use of existing standard data protection clauses has an advantage in that they lower the administrative costs, such as legal costs in the drafting of suitable data protection clauses.

BCRs that have to approved by the competent supervisory authority should be 'legally binding and apply to and are enforced by every member concerned of the group of undertakings,[116]

Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or to an international organisation shall take place only if, subject to the other provisions of this Regulation, the conditions laid down in this Chapter are complied with by the controller and processor, including for onward transfers of personal data from the third country or an international organisation to another third country or to another international organisation. All provisions in this Chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.

[110] Article 45 (1).
[111] Recital 104.
[112] Article 45 (2) (a)-(c).
[113] Article 46(1).
[114] Article 46(2).
[115] Article 46 (2) (c) & (d).
[116] 'group of undertakings' is defined as 'a controlling undertaking and its controlled undertakings' Article 4 (19).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

or group of enterprises[117] engaged in a joint economic activity, including their employees' and 'must expressly confer enforceable rights in data subjects with regard to the processing of personal data'[118] The BCR must at the very least meet the specific requirements listed in Article 47(2) which include:

- structure and contact details of the group of undertakings, or group of enterprises engaged in a joint economic activity and of each of its members, the data transfers or set of transfers, including the categories of personal data, the type of processing and its purposes, the type of data subjects affected and the identification of the third country or countries in question; their legally binding nature, both internally and externally;
- the application of the general data protection principles, in particular purpose limitation, data minimisation, limited storage periods, data quality, data protection by design and by default, legal basis for processing, processing of special categories of personal data, measures to ensure data security, and the requirements in respect of onward transfers to bodies not bound by the binding corporate rules;
- the rights of data subjects in regard to processing and the means to exercise those rights, including the right not to be subject to decisions based solely on automated processing, including profiling, the right to lodge a complaint with the competent supervisory authority and before the competent courts of the Member States, and to obtain redress and, where appropriate, compensation for a breach of the binding corporate rules;
- the tasks of any data protection officer[119] designated in accordance with Article 37 or any other person or entity in charge of the monitoring compliance with the binding corporate rules within the group of undertakings, or group of enterprises engaged in a joint economic activity, as well as monitoring training and complaint-handling;
- complaints procedure; and
- the mechanisms for reporting to the competent supervisory authority any legal requirements to which a member of the group of undertakings, or group of enterprises engaged in a joint economic activity is subject in a third country which are likely to have a substantial adverse effect on the guarantees provided by the binding corporate rules; and, the appropriate data protection training to personnel having permanent or regular access to personal data.

BCRs are widely viewed as a favourable mechanism for the transfer of data since it substantially lowers the administrative burden of the controller and processor. Further a

[117] Enterprise is defined as 'a natural or legal person engaged in an economic activity, irrespective of its legal form, including partnerships or associations regularly engaged in an economic activity.' (Article 4 (18)).
[118] Article 47(1).
[119] See 'Data protection officer' above.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

group of undertaking or group of enterprises can use the same BCR for the purposes of transfer to non-EU states.

Where adequate safeguards or appropriate safeguards including BCRs cannot be met, Article 48(1) allows transfer or sets of transfers in specific circumstances. These, for instance, include transfers with explicit consent of the data subject to the proposed data transfer after having learnt of the possible risks in the absence of adequate and appropriate safeguards, transfers necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and another natural or legal person, transfers necessary on the public interest, and transfers necessary for establishing, exercising or in the defence of legal claims. However such transfers must not be repetitive and concerns only a limited number of data subjects and is necessary for the purposes of compelling legitimate interests pursued by the controller that do not override the data subjects' interests or their rights and freedoms. The controller is expected to have assessed all the circumstances surrounding the data transfer. This should guide the controller to suitable safeguards with regard to the protection of personal data. The supervisory authority and the data subject must be informed and the controller must provide compelling legitimate interests pursued for such a transfer. As for transfer for reasons of public interest it must be recognised in the law of the Union or the controller's Member State.

RIs established in Member States in the interests of furthering research are unlikely to limit accessibility only to those institutions and commercial enterprises operating within EU borders. The accessibility is likely to be global and many non-EU countries are yet to have strong data protection laws and where they do they are not as rigorous as the GDPR. In these circumstances, an RI needs to include within its governance structure mechanisms that ensure that the protection of data subjects is not compromised. The governance structures of the RI have to be sufficiently robust to assess requests, the third party's mechanisms and compliance with the GDPR and where appropriate negotiate suitable data protection clauses the do not compromise the data subjects' rights.

4. Design of Ethical/Legal Aspects of RICHFIELDS

As seen from Section Three the protection of personal data (or personally identifiable information) of data subjects is at the heart of the GDPR. With this focus it introduces a number of obligations on the controllers and processors for lawful processing which include informed consent from the data subject, purpose limitation (unless further data processing is compatible with the initial purpose for collection or for scientific or historical research purposes), use of pseudonymisation to reduce risks associated with processing of personal data, implementation of appropriate technical, organisational and security measures, trans-border data transfer, data protection officer, giving control to data subjects over their

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

personal data through mechanism such as right to rectification and withdrawal of consent. The ethical issues that were raised in Section II in fact are embedded in the EU data protection framework and finds legal expression in the GDPR. The end result is that a non-compliant organisation will attract stiff penalties. In designing RICHFIELDS attention must be paid to the obligations created by the GDPR.

RICHFIELDS will be collecting data from a variety of sources which are likely to include data from apps, social media and store loyalty cards. The consent obtained by these sources may not all meet the consent standards required by the GDPR where data is being processed for purposes other than originally envisaged. Close scrutiny of the consent forms and terms of use of the data suppliers will need to be examined to see whether data subjects have consented to further processing. If they are unsatisfactory in GDPR terms consent will have to be specifically obtained for processing by the RI. This is likely to prove administratively cumbersome and costly. It may however be possible for RICHFIELDS to use the data for scientific purposes under the GDPR. But, in the absence of any clear guidance (even though the Recitals of the GDPR recognises the importance of data for research) it is difficult to say with great conviction that re-purposing of the kind envisaged by RICHFIELDS would always pass scrutiny.

The new concept of pseudonymisation provides a useful route for RICHFIELDS. It is meant to reduce risks whilst maintaining the usefulness of the data. The GDPR actively encourages the controllers to pseudonymise the data they collect by introducing separation. In pseudonymisation the data cannot be attributed to a particular data subject without the use of additional information and this additional information has to be kept separately from the processed data. Where additional data are kept separately it may still be possible for there to be security breaches (for instance, obtaining of the key to the additional information) which enable the linking of the additional data to the pseudonymised data. To address this controllers need to implement appropriate safeguards for unauthorised reversal of pseudonymisation. Safeguards could include technical measures such as encryption and other organisational measures on the ways in which the de-identification key will be protected from access. Controllers may decide, taking into account the potential risks for reversal to delete the directly identifying data. In these circumstances the controller will have benefit of the exemption from the rights to access, rectification, erasure and data portability allowed to data subjects by the GDPR provided the inability of the controller to identify the data subject is demonstrated.

RICHFIELDS which will determine the purposes and means of the processing of personal data is a controller and will need to have appropriate technical and organisational measures that takes into account the nature, scope, context and the purposes of processing (without losing

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

sight of the risks for the rights and freedoms of data subjects) to ensure that the processing (even where it utilises pseudonymised data) is GDPR compliant. RICHFIELDS must consider the adoption of approved codes of conduct and certification mechanisms. There might not be a current off-the-shelf code of conduct that meets the compliance requirements of the GDPR and therefore RICHFIELDS should consider devising a code of conduct with the advice of lawyers, IT and security specialists and data protection practitioners.

The assumption here is that RICHFIELDS will be a controller but will outsource storage of data and processing to another organisation. In this event RICHFIELDS will have to ensure that the appointed processors are GDPR compliant.

Security measures are paramount and these need to be addressed by RICHFIELDS. Besides pseudonymisation, it should include integrity, resilience of processing systems and regular testing of the system for vulnerabilities and regular evaluation.

The GDPR, as indicated earlier, imposes stiff penalties in the event of a breach. RICHFIELDS therefore needs to have insurance to cover legal costs and for all the liabilities that arise under the GDPR.

There must also be systems in place that can respond to the data subjects' queries and their requests based on the rights under the GDPR such as the right to erase, rectify, object and restriction of processing. The body responsible for handling and dealing with such requests within RICHFIELDS could also take advice from the independent ethics committee (suggested below).

RICHFIELDS which is likely to monitor consumption behaviour of data subjects and purchase habits amongst others would need a DPO (data protection officer). It is not essential that the DPO is a member of the organisation where RICHFIELDS is located. A DPO with appropriate professional qualifications can be appointed on the basis of a service contract. The officer will be responsible for providing advice on compliance, monitor compliance, raise awareness and act as a contact point for the supervisory authority.

Freedom of formation is not specifically mentioned in the GDPR but RICHFIELDS may need to put in place a mechanism that enable such requests to be handled efficiently and effectively. One possibility is to ask the DPO to act also as an FOI officer.

Since RICHFIELDS is likely to be of global application there will be trans-border data transfer. Not all requests for research data are going to originate from states which are recognised by the EU as having adequate standards. RICHFIELDS should therefore consider using standard

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

data protection clauses or BCRs (Business Corporate Rules) which will require authorisation by the supervisory authority.

The establishment of an independent ethics committee drawing upon a range of stakeholders (research institutions, the legal profession, IT professionals, commercial entities, non-governmental organisations and consumer bodies) for monitoring the activities of RICHFIELDS, its protocols on matters relating to security, transfer of data to third countries, assessing the genuineness of request by data users and the rules of operation in the event of requests that may be ethically dubious or questionable, data subjects' requests, and complaints procedures.

The governance structure of RICHFIELDS therefore will have to incorporate within it the setting up of appropriate bodies and committees, allocation of responsibilities for various tasks such as appointment of personnel (e.g. the DPO), conducting risk assessment, establishing robust security systems, the processes for obtaining data from various data suppliers and their level of integrity, overseeing the contractual agreements with researchers, transfer of data to third countries, and reporting and monitoring processes for GDPR compliance. It should also consider the suitability of an independent ethics committee and how this should be structured.

5. Conclusion and Potential for Future Developments

This Report has outlined the ethical issues that arise in the context of big data and an account (albeit brief) of the provisions that are of interest to RICHFIELDS. Section 4 has listed a number of areas that RICHFIELDS should focus on to make it GDPR compliant. Given its purpose is to use the data sets in its repository for research, pseudonymisation is suggested as a means to process the data, provided appropriate safeguards are in place. In order to raise the integrity profile of RICHFIELDS externally the setting up of an independent ethics committee is also suggested. This would also help in bolstering the confidence of data in the utility of RIs such as RICHFIELDS as a research tool for promoting well-being and over time might usher in an era where data subjects in the spirit of altruism give their data for the sake of research and innovation.

At some point in the future there is a high likelihood that RICHFIELDS will link up with other RIs devoted to nutrition and healthcare thus creating remarkable opportunities for research and for devising innovative personalised diet and health care that would be of immense benefit to society. In the event of linking RIs ranging from food consumption and nutrition to health to form a super RI, compliance with the GDPR will still be of paramount importance subject to any derogations available in respect of research and health in the GDPR.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Reference

Advisory Council to Google on the Right to be Forgotten (2015) Report of the advisory council to google on the right to be forgotten Google , <https://static.googleusercontent.com/media/archive.google.com/en//advisorycouncil/advisement/advisory-report.pdf>.

Angrist M (2009) 'Eyes wide open: The personal genome project, citizen science and veracity in informed consent' Personalized Medicine, (6) 691–699.

BBC News 'Everyone "to be a research patient", says David Cameron', 5 December 2011 <http://www.bbc.co.uk/news/uk-16026827>.

Booch G (2014) 'The human and ethical aspects of big data' IEEE Software, 31(1): 20–22.

Bowker G C (2014) 'Big data, big questions the theory/data thing' International Journal of Communication, 8: 5.

Boyd D & Crawford, K (2012) 'Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon' Information Communication & Society 15(5): 662–679.

Center for US Health System Reform, Business Technology Office (2013) The 'big data' revolution in health care' McKinsey & Co, <http://achc.org.co/hospital360/contextos/Tecnologia_e_Informacion/Big_Data/Revolucion_de_la_Informacion.pdf>.

Choudhury S, Fishman JR et al (2014) 'Big data, open science and the brain: lessons learned from genomics' Frontiers of Human Neuroscience 8: 239, doi: 10.3389/fnhum.2014.00239.

Clayton E W (2005) 'Informed consent and biobanks' Journal of Law, Medicine & Ethics 33(1): 15–21.

CNIL Methodology for Privacy Risk Management: How to Implement the Data Protection Act <https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf>.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century' COM(2012) 9 final, 25.1.2012<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0009&from=en >.

Costa F F (2014) 'Big data in biomedicine' `Drug Discover Today 19(4): 433-440.

Fairfield J & Shtein H (2014) 'Big data, big problems: Emerging issues in the ethics of data science and journalism' Journal of Mass Media Ethics 29(1):38–51.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Fischer Jr BA (2006) 'A summary of important documents in the field of research ethics' Schizophrenia Bulletin 32(1): 69-80.

Floridi L (2012) 'Big data and their epistemological challenge' Philosophy & Technology 25(4):435–437.

Floridi L (2014) 'Open data, data protection, and group privacy' Philosophy & Technology 27(1): 1–3.

Herschel R & Miori V M (2017) 'Ethics & big data' Technology in Society 49: 33-36.

Hoffman S (2014). Citizen science: The law and ethics of public access to medical big data (SSRN Scholarly Paper) http://papers. ssrn.com/abstract=2491054.

ICO Anonymisation: Managing Data Protection Code of Practice < https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf >.

Ioannidis JPA (2013) 'Informed consent, big data and the oxymoron of research that is not research' American Journal of Bioethics 13(4): 40-42.

M K Javaiad, M K, Forestier-Zhang, L et al (2016) 'The RUDY study platform – a novel approach to patient driven research in rare musculoskeletal diseases' Orphanet Journal of Rare Diseases 11:150.

Karin M N, Wiford J C & Behrend T S (2015) 'Big data, little individual: Considering the Human side of big data' Organizational Psychology 8(4) 527-533.

Kaye J, Whitley E A, Lund D et al (2015) 'Dynamic consent: a patient interface for twenty-first century research network' Eur J Hum Genet , 23(2) 141 – 146.

Kuan Hon W, 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 ,< https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577 >.

Lupton D (2014) 'The commodification of patient opinion: The digital patient experience economy in the age of big data' Sociology of Health & Illness 36(6): 856–869.

Majumdar MA (2005) 'Cyberbanks and other virtual research repositories' Journal of Law, Medicine & Ethics 33(1); 31-39.

Markowetz A, Blaszkiewicz K et al (2014) ''Psycho-informatics: Big data shaping modern psychometrics' Medical Hypotheses 82(4): 405-411.

Master, Z, Campo-Engelstein A L & Caulfield T (2015) 'Scientists' perspectives on consent in the context of bio-banking research' European Journal of Human Genetics i23(5): 569-574.

Mittelstadt B D, Stahl B C & Fairweather N B (2015) 'How to shape a better future?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Epistemic difficulties for ethical assessment and anticipatory governance of emerging technologies' Ethical Theory and Moral Practice 1–21, doi:10.1007/s10677-015-9582-8.

Mitelstadt B D & Floridi L (2016) 'The ethics of big data; Current & foreseeable issues in biomedical context' Sc Eng Ethics 22: 301-341.

Narayanan A & Schmatikov V (2010) 'Privacy and security: Myths and fallacies of "personally identifiableInformation"' Communications of ACM 53(8): 24-26.

Nissenbaum H (2004) 'Privacy as contextual integrity' SSRN Scholarly Paper <http://papers.ssrn.com/abstract=534622>.

Oboler A, Welsh K & Cruz L (2012) 'The danger of big data: Social media as computational social science' *First Monday* 17:7 < http://firstmonday.org/article/view/3993/3269/ >.

Raghupath Wi V Raghupathi V (2014) 'Big data analytics in healthcare: promise and potential' *Health Information and Science System* 2:3, <https://doi.org/10.1186/2047-2501-2-3>.

Reshef DN, Reshef YA et al (2011) 'Detecting novel associations in large data sets' *Science* 334: 1518-1524.

Steinsbekk K S, Ursin LO, Skolbekken J A & Solberg B (2013) 'We're not in it for the money—lay people's moral intuitions on commercial use of 'their'' biobank' *Medicine, Health Care and Philosophy*, 16(2), 151–162.

Sweeney L (2002) 'Achieving k-anonymity privacy protection using generalization and suppression' International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems 10(5): 571-588.

Terry N (2014) 'Health privacy is difficult but not impossible in a post-HIIPAA driven world' *Chest* 146(3): 835-840

Vitak J, Shilton K & Ashktorab Z (2016) 'Beyond the Belmont principles: Ethics, practice and beliefs in the online data research community' *Proceedings of the 19th ACM Conference on Computer- Supported Co-operative Work & Social Computing* 941-953.

Williams H, Spencer J & Dixon W G (2015) 'Dynamic consent: A possible solution to improve patient confidence and trust in how electronic patient records are used in medical research' *JMIR Medical Informatics* 3(1) e 3, doi: 10.2196/medinform.3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.