D8.1 Review of ELSI issues in RWE approach

116020 - ROADMAP

Real world Outcomes across the AD spectrum for better care: Multi-modal data Access Platform

WP8 – Ethical, Legal, and Social Implications (ELSI)

Lead contributor	Brent Mittelstadt (1 – UOXF)
	brent.mittelstadt@oii.ox.ac.uk
Other contributors	Jenny Krutzinna (1 – UOXF)
	Andrew Turner (1 – UOXF)
	Zuzanna Angehrn (16 – Novartis)
	Clementine Nordon (16 – Novartis)
	Dianne Gove (8 – AE)

Due date	31/05/2017
Delivery date	23/06/2017
Deliverable type	R
Dissemination level	PU

Description of Work	**Version**	**Date**
	V1.0	27/10/2016

Table of contents

Document History

Version	Date	Description
V1.0	03/05/2017	First Draft
V1.4	20/05/2017	Comments from EAB (B. Prainsack), internal formal review (Jan Kors, Zuzanna Angehrn, Clementine Nordon)
V1.6	23/06/2017	Final version after Consortium Review

Definitions

- Partners of the ROADMAP Consortium are referred to herein according to the following codes:
  - **UOXF**. The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom) – **Coordinator**
  - **NICE**. National Institute for Health and Care Excellence (United Kingdom)
  - **EMC**. Erasmus University Rotterdam (Netherlands)
  - **UM**. Universiteit Maastricht (Netherlands)
  - **SYNAPSE**. Synapse Research Management Partners (Spain)
  - **IDIAP JORDI GOL**. Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (Spain)
  - **UCPH**. Københavns Universitet (Denmark)
  - **AE**. Alzheimer Europe (Luxembourg)
  - **UEDIN**. University of Edinburgh (United Kingdom)
  - **UGOT**. Goeteborgs Universitet (Sweden)
  - **AU**. Aarhus Universitet (Denmark)
  - **LSE**. London School of Economics and Political Science (United Kingdom)
  - **CBG/MEB**. Aagentschap College ter Beoordeling van Geneesmiddelen (Netherlands)
  - **IXICO**. IXICO Technologies Ltd (United Kingdom)
  - **RUG**. Rijksuniversiteit Groningen (Netherlands)
  - **Novartis**. Novartis Pharma AG (Switzerland) – **Project Leader**
  - **Eli Lilly**. Eli Lilly and Company Ltd (United Kingdom)
  - **BIOGEN**. Biogen Idec Limited (United Kingdom)
  - **ROCHE**. F. Hoffmann-La Roche Ltd (Switzerland)
  - **JPNV**. Janssen Pharmaceutica NV (Belgium)
  - **GE**. GE Healthcare Ltd (United Kingdom)
  - **AC Immune**. AC Immune SA (Switzerland)
- **Grant Agreement.** The agreement signed between the beneficiaries and the IMI JU for the undertaking of the ROADMAP project (116020).
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium.** The ROADMAP Consortium, comprising the above-mentioned legal entities.
- **Consortium Agreement.** Agreement concluded amongst ROADMAP participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.
- **GDPR.** General Data Protection Regulation, a European legal framework coming into force in May 2018.
- **AI.** Artificial intelligence.
- **RWE.** An approach to medical research and decision-making in clinical care and healthcare policy based on the aggregation of many types of evidence.

Publishable Summary

Ethical foresight may reduce the probability of regulatory and social 'whiplash' by informing public and policy debates. To contribute to this process for platforms enabling medical research based on 'real-world evidence' (RWE), a systematic and comprehensive review of academic literature on the ethical, social, and legal implications (ELSI) of medical data repositories was conducted to identify the issues of emerging importance for this novel form of data curation and analysis. This deliverable is a first step to defining the ELSI requirements for a RWE approach to AD research. Academic literature discussing ethical and social aspects of an RWE approach and medical data sharing was systematically surveyed to produce a narrative review of relevant concepts and issues. The review aimed to address one primary question:

> What ethical and social issues arise in a real-world evidence approach to medical research and healthcare decision-making?

In order to understand how these issues have already been identified and discussed in the context of RWE, a systematic survey of academic literature was conducted between November 2016 and March 2017. A total of 2484 non-unique sources were identified for review across the databases, with 81 sources fully reviewed. The results of the review are presented as a narrative overview, which highlights and comments upon key themes and topics in the literature. Through content analysis, seven major themes emerged from the literature: (1) informed consent; (2) autonomy and participation; (3) transparency; (4) ownership; (5) data provenance; (6) privacy; and (7) group harms and discrimination.

1. Introduction[1]

Europe now faces a substantial healthcare challenge due to an ageing population, increasing cost pressures, and more specialised treatments. Greater access for medical researchers and policy-makers to more and better quality data, enabled by biobanks and other medical data repositories, can help meet these challenges (Mostert et al. 2016). Approaches based on 'real world evidence' (RWE) are one promising area to deliver targeted and increasingly effective healthcare, while deriving additional value and knowledge from existing data sources. RWE refers to an approach to medical research and decision-making in clinical care and healthcare policy based on the aggregation of many types of evidence. RWE integrates available patient outcome data including routinely collected data, electronic health records (EHR) and registries, cohort and trial data, and patient reported outcomes. These data can be mined to elicit consensual and reliable indicators to provide an enriched clinical and health policy decision-making environment. Platforms that support RWE potentially better inform regulators (efficacy & safety), healthcare providers and payers (cost effectiveness and budget impact), industry (pricing & manufacturing), and scientists (mechanisms & pathways) to accelerate decision-making on re-purposing current treatments and developing new treatments.

RWE, and biomedical 'Big Data' more broadly, nonetheless face ethical, legal, and social issues (ELSI). RWE can be considered biomedical Big Data when it involves the aggregation and linkage of multiple datasets, often of different types and collected from different sources. Increasing interconnectivity between datasets gathered within and beyond traditional medical institutions can challenge accepted ethical, social, and legal norms and frameworks (B. Mittelstadt and Floridi 2016a). Practices centred on the mass curation and processing of personal data can quickly gain a negative connotation which, in a way similar to the public and regulatory whiplash over genetically modified organisms (cf. Devos et al. 2008), places potentially beneficial applications at risk through association with problematic applications. A 'whiplash effect' can occur, by which overly restrictive measures (especially legislation and policies) are proposed in reaction to perceived harms, which overreact in order to re-establish the primacy of threatened values, such as privacy. Such a situation nearly occurred for scientific research in Europe as reflected in the debate on the General Data Protection Regulation (Wellcome Trust 2014). Early drafts would have drastically restricted secondary uses of personal data for research by requiring specific and explicit informed consent for secondary processing of personal data (including pseudonymised data).

Ethical foresight may reduce the probability of regulatory and social 'whiplash' by informing public and policy debates. To contribute to this process for platforms enabling RWE, a systematic review of academic literature on the ethical, social, and legal implications of medical data repositories was conducted to identify the issues of emerging importance for this novel form of data curation and analysis. Section 2 provides a brief background on RWE. Section 3 describes the methodology of a systematic narrative review of academic literature discussing ELSI aspects of medical data repositories. The results of the review are presented thematically in Section 4. Shortcomings and further relevant issues not currently addressed sufficiently in the literature are then highlighted in Section 5, before concluding by reflecting on directions for further research in Section 6.

[1] The authors of this deliverable would like to thank the insightful feedback and comments from the reviewers, including members of the ROADMAP Ethics Advisory Board, which greatly improved the quality and scope of the deliverable.

2. Background

Alongside with 'gold standard' clinical trial data, routinely collected data, patient reported outcomes, national registries and electronic health records can be linked and aggregated to base research and decision-making as far as possible on evidence of the performance of interventions in the real world. Such 'real-world evidence' is varied and complex, consisting of data produced from a wide variety of sources, including "laboratory auto-analyzers, pharmacy systems, and clinical imaging systems...augmented by data from systems supporting health administrative functions such as patient demographics, insurance coverage, financial data, etc…clinical narrative information, captured electronically as structured data or transcribed 'free text'…electronic health records" to name but a few (Safran et al. 2006, 2). 'Non-medical' data, or data produced outside traditional medical institutions or for purposes not explicitly related to health management or care, can also be incorporated into RWE. Such data be generated explicitly or covertly via social media applications and health platforms (Lupton 2014, 858; Costa 2014), emerging 'personal health monitoring' technologies (B. Mittelstadt et al. 2011, 2013) including wearable devices (Boye 2012), home sensors (Niemeijer et al. 2010) and smart phone applications, and online forums and search queries (B. Mittelstadt et al. 2017; Costa 2014; Butler 2013).

RWE allows for longitudinal data describing large populations to be mined and support medical decision-making. Large model datasets can be assembled that resemble real populations in size and complexity (Collins 2016). Trial and cohort data can be treated as structured studies nested within a less well defined but wider population-based data-space. Data from national/regional EHR databases can be supplemented with individual patient and population outcomes, and overlaid with cohort and trials data to derive models which are mechanistically informative and show high predictive validity (Choudhury et al. 2014). Adherence to treatments often differs between clinical trials and real world usage, meaning individual-level adherence data (one type of 'real-world evidence') can provide a better basis to evaluate cost-effectiveness (Collins 2016). The effectiveness of interventions, such as drug interactions (Tene and Polonetsky 2013), can be measured according to real world usage. In line with the ideal of 'personalized medicine', "clinicians can select treatments that are most effective…[and] a dose that is tailored more to an individual, potentially making the healthcare system much more efficient" (Collins 2016).

RWE aims to make explicit the relationships and limitations of each data type, informing triangulation between highly focussed, question-specific evidence from trials, detailed etiologic inference from cohorts, self-reported data from patients, and population-based EHR data. From this, models of disease progression can be developed and hypotheses on efficacy, safety, cost and impact of therapeutic interventions can be tested (Collins 2016). RWE platforms can enable open, consensual and collaborative development of solutions to the healthcare challenges, including the involvement of additional stakeholders including regulators, health technologies authorities, and patient (advocacy) groups. RWE also provides the context for a new generation of trials and population studies that will attempt to integrate regulatory, payer and patient perspectives, reduce and spread risk among stakeholders, and reduce cost.

RWE is closely related but not identical to 'real-world data', defined by the Association of the British Pharmaceutical Industry as "data collected outside the controlled constraints of conventional randomised clinical trials to evaluate what is happening in routine clinical practice." According to this

definition, clinical trials alone are insufficient to guide clinical care. Real-world data, including routinely collected data, allows for individual-level assessment of treatment effectiveness and “insight into patterns of care, long-term drug safety, healthcare resource utilisation and disease epidemiology” (Collins 2016). The two concepts are thus closely related; real world evidence, however, further includes clinical trial data in its scope.

Aggregation and linkage of diverse data sources and types are a key element that enable seemingly limitless novel forms of data mining (Prainsack and Buyx 2013, 73; Laurie et al. 2010). An RWE approach thus emphasises secondary uses of existing medical and health-related data for research and healthcare decision-making. Secondary uses of data “allows the optimal use of already available resources, and reduces the costs of research activities.” In the related area of biobanking, networks have emerged for precisely this aim, “to enable meta-analysis of horizontally partitioned individual-level data” across their members (Tassé 2016).

RWE can be considered part of the broader ‘Big Data’ phenomenon (B. Mittelstadt and Floridi 2016a). The push towards ‘Big Data’ in healthcare and medical research aims to aggregate and link real-world datasets for research, and thus inform decision-making across healthcare (McDonald et al. 2016). The perceived value of such practices is variable, but stems inevitably from the ability to search for connections across vast datasets for a variety of research purposes (Floridi 2012). Methods based on mining and analysis of aggregated data nonetheless face important technical (e.g. interoperability, data harmonisation) and epistemological (e.g. selection bias, measurement bias) challenges that need to be acknowledged as potential limitations of their utility when seeking permission to link and aggregate datasets (Hoffman and Podgurski 2013).

‘Biomedical Big Data’ has recently gained significant attention due to a combination of two factors. On the one hand, there is the huge potential to advance the diagnosis, treatment, and prevention of diseases, to foster healthy habits and practices (Costa 2014), and to improve health policy-making. On the other hand, there is the inherent sensitivity of health-related data, the implicit vulnerability and needs of those potentially requiring treatments (Pellegrino and Thomasma 1993), and the informational risks posed to participants in research via secondary uses of health data (B. Mittelstadt and Floridi 2016a). RWE platforms likewise face many challenges due to aggregating and making individual and population level health data available for research and healthcare decision-making. Ethical, legal and social concerns can arise due to the sensitivity of data being manipulated and their seemingly limitless potential uses.

2.1. Legal context of RWE in Europe

This deliverable reviews ethical and social issues arising from an RWE approach in medical research. However, the legal context in which RWE and medical data sharing platforms in Europe operate must also be considered. A full review of current and forthcoming legal requirements for secondary usage of medical data processing will be undertaken in ROADMAP D8.2 (see: Section 3). Here, to provide additional context for the ethical and social concerns discussed, a brief overview of relevant legal requirements is offered.

An RWE approach has numerous legal implications. Chief among these is the requirement to obtain informed consent prior to participation in a research study, including studies involving secondary analysis of existing data. The duty for researchers to seek voluntary, informed consent to participate

in research was first established in the Nuremberg code, and later developed in the World Medical Association's Declaration of Helsinki (Kaye 2011; Boddington et al. 2011). As formulated in Article 9 of the latter, researchers have a positive duty to protect the confidentiality and well-being of participants in medical research beyond merely obtaining consent. Although the Declaration lacks legal force, it has proven influential on European national consent law and ethics review requirements for research (Boddington et al. 2011). Following this, RWE platforms, as gatekeepers of medical data, incur obligations related to consent. Without describing all possible models of ethical governance of medical data repositories (e.g. Knoppers et al. 2011b; Dove, Knoppers, and Zawati 2014), broadly speaking consent must be established prior to secondary processing of personally identifiable data. Precise consent requirements, in particular relating to the identifiability of data to be processed, will be explored in greater detail in D8.2.

Bodies acting as gatekeepers of personal medical data also incur obligations under data protection law. Currently, within the EU legal data protection requirements are set by Member States via national law and jurisprudence. These laws are primarily based on the EU's 1995 Data Protection Directive (95/46/EC). Broadly speaking, processing of personal data is legal when (1) it promotes vital interests of the subject or overriding national interests (e.g. disease surveillance; Nuffield Council on Bioethics 2015, 63), (2) explicit consent is given by the data subject or (3) processing occurs for purposes exempted from requiring consent. Concerning (3), legal exemptions facilitate certain types of socially beneficial processing (Kaplan 2014). According to Article 8 of the Directive, medical data can be processed without consent for "preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional…" (European Parliament 1995, Art. 8(3)).

In adapting the 1995 Directive into national law, many European member states have relaxed consent requirements for data-driven medical research. As an example, the UK's Data Protection Act 1998 allows processing of sensitive personal data for "medical purposes" undertaken by a "health professional" or person with an equivalent "duty of confidentiality" without explicit consent. Medical purposes here mirror those established in Article 8 of the Directive, but critically also include research (UK Parliament 1998, Schedule 3). Research data may additionally be "(i) used for alternative purposes; (ii) kept indefinitely; and (iii) exempted from a data subject's rights of access" in the UK (Boddington et al. 2011, note 26). The exemption of medical research from requiring *explicit* consent is a "significant diversion from the Directive" (Boddington et al. 2011, 498). Equivalent concessions have not been granted by all Member States. Such differences are barriers to international research collaboration and data sharing (Nuffield Council on Bioethics 2015; Knoppers et al. 2011b).

New data protection rules will, however, soon come into effect in Europe. From 26 May 2018, the General Data Protection Regulation (EU 2016/679), also known as the 'GDPR', will become directly applicable in each Member State. The GDPR aims to harmonise data protection standards across the EU. Despite this, the impact of the GDPR on medical research and secondary uses remains uncertain, the obligations the GDPR creates have been described as "onerous" (Allen & Overy 2017); although the same core rules as the Data Protection Directive are retained, there are some significant additional obligations introduced by the GDPR (Linklaters 2017).

An important change concerns the definition of "data concerning health," which is considered a "special category of personal data." It remains uncertain precisely which types of data 'concern health'. According to Article 4(15) of the GDPR, "data concerning health" is defined as "personal data

related to the physical or mental health of a natural person…which reveal information about his or her health status." This definition implies that the purpose of processing, rather than the data source, determines whether data can be considered 'health data' (Recital 35 GDPR). Data that do not directly describe health, but from which health-related inferences can be drawn (e.g. B. Mittelstadt and Floridi 2016a; B. Mittelstadt et al. 2017), appear to fall within the scope, and will thus be subject to greater restrictions and protections as a "special category of personal data" (Article 9 GDPR). Certain data sources are also explicitly highlighted as pertaining to health, including biometric and genetic data.

The full implications and meaning of key concepts of the GDPR will likely remain uncertain for several years following its enforcement in 2018. Clarification will be achieved through guidance issued by the European Data Protection Board (Article 68 GDPR), Member State law, and jurisprudence. Ethical governance of RWE platforms must take account of this legal context to ensure ethical and social requirements are harmonised with legal standards and desiderata. Despite remaining uncertainty over whether the GDPR will deliver a harmonised governance standard for medical data sharing across Europe (Mostert et al. 2016), such an approach is nonetheless urgently needed to eliminate unintended bureaucratic barriers facing cross-border medical data sharing and research (Kaye 2011).

As the discussion thus far indicates, medical data repositories are not a new phenomenon. Scholarship and governance of biobanks, disease registries, and other medical data repositories address many of the same issues that face RWE platforms. Nonetheless, the aggregation of data that addresses individuals and population, at different levels of identifiability, and which has been originally collected with different consent preferences and for different purposes, including in some cases for non-medical reasons (B. Mittelstadt et al. 2017), mean an RWE approach cannot merely be equated with prior practices. Recognising these subtle but important differences, it is necessary to investigate in more detail precisely the challenges and solutions relevant to RWE.

3. Objectives

As a first step to defining the ethical, legal, and social requirements for a RWE approach to AD research, this deliverable provides a narrative review maps these concerns as discussed in relation to medical data sharing more broadly. Academic literature discussing ethical and social aspects of an RWE approach and medical data sharing was systematically surveyed to produce a narrative review of relevant concepts and issues.

This deliverable presents a preliminary version of the review, which be expanded throughout the remainder of ROADMAP. Subsequent updates will be delivered in D8.2 and D8.5. The review aimed to address one primary question:

> What ethical and social issues arise in a real-world evidence approach to medical research and healthcare decision-making?

In order to understand how these issues have already been identified and discussed in the context of RWE, a systematic survey of academic literature was conducted between November 2016 and March 2017. Results of the survey are presented as a narrative review of the field. The review is systematic insofar as the search methodology used consistent keywords across multiple databases to identify an initial sample of literature (see: Table 1). However, the results are presented as a narrative overview, which intentionally does not assess the frequency of themes, theories, and concepts across the sample.

Systematic reviews and meta-analyses have long been used in evidence-based medical decision-making and other fields reliant upon empirical findings. Ethical decision-making is not an equivalent evidence-based procedure. In ethics, including medical ethics, systematic approaches have seen less uptake (Mertz, Kahrass, and Strech 2016; Stahl, Timmermans, and Mittelstadt 2016). This trend may not be an accident or sign of a lack of methodological rigour, but rather reflect a deliberate disciplinary choice. Ethical discourse is inherently argumentative. Normative positions rely upon argumentation and reasoning for support, as well as empirical evidence (Borry, Schotsmans, and Dierickx 2008). Normative positions are not normally taken to be 'right' based upon the number of people holding them; rather, it is the argumentation and reasoning supporting a position that makes it persuasive. Claims can and arguably should be grounded in empirical evidence; knowing the values and intuitions of physicians can, for example, reveal something of norms of 'good' practice in medicine (Pellegrino and Thomasma 1993; MacIntyre 2007). In short, the *frequency* with which an idea or concept appears in prior scholarship says little about the *persuasiveness* of the idea.

Defining ELSI requirements for an RWE approach to AD is a normative procedure. Consideration must be given to ethical values, legal requirements, concerns from relevant patient and carer populations, and prior best practice in similar platforms and organisations (e.g. Knoppers et al. 2011; Dove, Knoppers, and Zawati 2014). The first step in defining requirements is to understand the range of relevant normative positions, values, and concepts that will interact in the governance of an RWE platform. Defining the ELSI requirements for an RWE approach to AD can be considered a form of argumentation. Compared to a mapping of possible positions based on prior practice and academic discourse, systematic assessment and statistical analysis of the distribution of concepts across existing literature would contribute less to this argumentative process, which must decide which values, interests, and normative priorities *should* be reflected in the design and governance of an

RWE platform. The following narrative overview provides this type of ELSI mapping, beginning with ethical and social considerations.

4. Methodology

Four databases were searched (Web of Science, Scopus, PubMed and Google Scholar) to identify literature discussing ethical and social aspects of platforms supporting RWE, and medical data repositories more broadly. Legal considerations will explicitly be expanded via subsequent analysis reported on in D8.2 (see: Section 5); it is, however, recognised that literature addressing ethical and social issues will often also discuss relevant legal frameworks and contexts. The review thus addresses legal issues indirectly via a search methodology using keywords based on ethical and social issues. Keywords (with wildcards) were chosen to limit the review to articles addressing medical data sharing or secondary usage, and ethics. Keywords were chosen to identify sources explicitly addressing ethical or moral issues, or major areas of ethical concern in biomedical ethics and data ethics (i.e. privacy and consent; see (B. Mittelstadt and Floridi 2016a), in the context of medical data sharing repositories. A range of commonly used terms to describe medical data sharing were chosen. While a broader sample of keywords is possible reflecting a further range of ELSI concepts (e.g. autonomy, discrimination, equality), the selection was limited to keep the sample at a manageable size. The review could feasibly be expanded in future stages of ROADMAP to include a broader range of ELSI concepts. A breakdown of the search by database, keywords, and results returned can be found in Table 1.

Database	Search String	Returned
Web of Science	**TOPIC:** ((ethic* OR moral* OR privacy OR consent) AND (biobank* OR "real world evidence" OR "real world data" OR "secondary us*" OR "big data" OR "open science") AND (dementia OR "alzheimer's disease" OR medic*))	592
Scopus	**TITLE-ABS-KEY:** ((ethic* OR moral* OR privacy OR consent) AND (biobank* OR "real world evidence" OR "real world data" OR "secondary us*" OR "big data" OR "open science") AND (dementia OR "alzheimer's disease" OR medic*))	1084
PubMed	(ethic* OR moral* OR privacy OR consent) AND (biobank* OR "real world evidence" OR "real world data" OR “secondary us*” OR "big data" OR "open science") AND (dementia OR "alzheimer's disease" OR medic*)	608
Google Scholar	(ethic* OR moral* OR privacy OR consent) AND (biobank* OR "real world evidence" OR "real world data" OR “secondary us*” OR "big data" OR "open science") AND (dementia OR "alzheimer's disease" OR medic*)	3360 (first 200 reviewed)

Table 1 – Search Queries

The title and abstract of each returned article was reviewed by the author to determine relevance. Inclusion was based solely on the discussion of ethical issues in the article, with the goal of identifying themes in the literature. Limitations were not placed on the quality or length of the discussion, but rather on the mere presence of ethical concepts and issues. Further sources were also located through hand-searching and backtracking of citations provided within the reviewed articles. The

search was limited to English language articles appearing in conference proceedings or peer-reviewed academic journals. Date restrictions were not enforced.

4.1. Data Analysis

Each article was analysed and key passages highlighted for further interpretation and grouping into themes existing across multiple sources. These themes were allowed to emerge from the literature rather than starting from a pre-defined theoretical framework. To start, phrases and passages were highlighted that appeared to refer to ethical issues or concepts, understood as areas of ‘right’ and ‘wrong’ or the clash of competing values or normative interests among stakeholders. Highlighted segments were then coded to reflect the author’s interpretation of the text (cf. Gadamer 2004; Patterson and Williams 2002). Similar codes were then grouped and assigned to ethical themes. Once themes had emerged from the literature, a second systematic analysis was performed using the NVivo 10 software package. All sources were re-checked via text search for the presence of the themes that emerged.

5. Results

A total of 2484 non-unique sources were identified for review across the databases. Titles and/or abstracts were reviewed for each of the returned sources to determine initial relevance to the review. Following this initial sorting, 73 full-text sources were reviewed. Rejected sources were either off-topic or duplicates. An additional 8 sources were hand selected based on citation tracking and feedback from the reviewers, resulting in a total sample of 81 sources.

The results of the review are presented as a narrative overview, which highlights and comments upon key themes and topics in the literature. This overview is intended to address the aforementioned research question to inform the design of ethical governance of RWE platforms in the first instance. It is, however, acknowledged that the results may also have relevance for governance of other types of medical data repositories, and for 'Big Data' research projects more broadly.

Through content analysis (see: Section 4.1), seven major themes emerged from the literature: (1) informed consent; (2) autonomy and participation; (3) transparency; (4) ownership; (5) data provenance; (6) privacy; and (7) group harms and discrimination. While the themes are intended to be representative of the literature, the overview does not merely reflect the frequency of their discussion. Rather, the results discussed in the following sections were chosen for one of four reasons: (1) to draw attention to common interpretations of ethical, social and legal themes and concepts, (2) to emphasise individual cases and issues that reveal unique ELSI aspects of medical data repositories, (3) to highlight studies with an in-depth analysis of ELSI concepts and challenges, and (4) to identify gaps in the discussion in need of further research. The presentation of results therefore focuses on the authors' analysis and interpretation of the literature.

5.1. Informed Consent

Unsurprisingly given its historical prominence in medical ethics debate, informed consent emerged as a major theme in the reviewed literature. Several types of information are commonly supplied in informed consent procedures.[2] In a review of international guidelines for consent in biobanking, Hirschberg et al. (2014) identified four categories of information to be provided to participants: general information about the type and purpose of the study; conditions of participation including voluntariness, scope and consent; consequences of participation including risks and benefits; and information about data handling including confidentiality and third party access. In contrast, Beskow (2015) describe a greater range of information to be included specifically for biobank participants (see: Table 2).

Table 2: *Information to include in a biobank consent form; adapted from Beskow et al. 2015.*

- Purpose
- Procedures
    - Type of data collected
    - Access by (qualified) researchers
    - Recontact procedures for additional research participation (not secondary uses)
    - Large-scale data sharing (further sharing for research purposes)
- Duration (how long data will be retained)
- Risks (including possibility of unforeseen risks)
- Confidentiality protections
    - Coded form of data
    - Who will/will not have access
- Benefits and costs
    - Expected benefits from participation
- Voluntariness and alternatives
    - Refusal to participant won't impact care/benefits
- Discontinuing participation
- Contact point for questions
- Additional elements
    - Commercialization
    - Research results
        - Release of individual and aggregate results to participants (recommendation is to not release individual results)

Substantial attention was paid in the reviewed literature to the move away from traditional, single-study models of consent which are challenged by repositories aimed at enabling simple but responsible access to aggregated medical datasets for research. Models of specific, explicit informed consent (Angrist 2009) do not cleanly transfer to secondary research for a variety of reasons. Consent has traditionally been taken for participation in a single study to protect the autonomy and well-being of participants (Aicardi et al. 2016; Hirschberg, Kahrass, and Strech 2014). This model arose in

[2] For a full review of types of information to be mentioned in consent documentation for biobanks as recommended in international guidelines, see: Hirschberg, Kahrass, and Strech 2014.

response to concerns with medical research involving human participants (Aicardi et al. 2016). According to Johnsson and Eriksson (2016), “the main purpose of obtaining informed consent is to ensure that research participants are neither deceived nor coerced.” Ethics policies for medical research, in particular biobanks, “typically seek to protect and enhance participants’ rights to privacy, confidentiality, and self-determination” (Aicardi et al. 2016).

Explicit, single-study ‘specific consent’ does not normally cover secondary investigations resulting from sharing, aggregating, or even repurposing data within the wider research community (Choudhury et al. 2014, 4; Hirschberg, Kahrass, and Strech 2014). Secondary analyses of aggregated and re-purposed data can advance medical knowledge while minimising the burden on participants (Hirschberg, Kahrass, and Strech 2014). Obtaining specific and explicit consent is, however, often not feasible or practical for secondary analyses (Currie 2013). Many biobanks now seek to horizontally link their datasets to allow for meta-analysis of individual-level records. However, these biobanks, and in particular those collecting data for disease-specific uses, are unlikely to have included secondary uses outside of this disease in their consent processes (Tassé 2016). Alternative models of consent for data-based medical research are often used by repositories, which aim to minimise the burden placed on researchers in obtaining consent for secondary analyses of existing data.

Single-study consent can also be problematic because population-level analytics are often intended by design to reveal unforeseen connections between patients and records (Hirschberg, Kahrass, and Strech 2014). This means that both what the data reveals about the subject and its utility in future research present greater uncertainty than normal at the time of consent (Cato, Bockting, and Larson 2016; Beskow et al. 2015). For example, secondary effects of pharmaceuticals can be identified by comparing data not only from multiple clinical trials, but ‘informal sources’ as well, such as incidental self-reporting via social media and search engine queries. In this type of research, the connections that can be revealed through linking multiple data sets cannot be accurately predicted prior to carrying out the research. As a result, ‘consent’ cannot be ‘informed’ in the sense that data subjects cannot be told about future uses and consequences of their data, which are unknowable at the time the data is collected or aggregated. Inferring characteristics a patient did not choose or consent to disclose can violate medical ethical principles including respect for persons, beneficence, justice and patient privacy (Navarro 2008).

The uncertainty of future uses of data highlights a key aspect of secondary research, namely the desire for openness and creativity in identifying novel connections between data sets. For data collected explicitly for hosting in a medical data repository, the openness of the format does not create difficulties, although open data sharing may require a global type of consent, due to data travelling across the political and electronic borders of institutions and nations (Majumder 2005, 33).

The same cannot be said for historical data for which consent was granted for a specific purpose. By the ideal of explicit single-study consent such data should not be used without explicit consent for secondary uses by the person providing it. However, obtaining such consent years after a trial has been conducted can be very difficult, if not impossible (Wellcome Trust 2014; Clayton 2005). A tension therefore exists between the potential benefits of analyses of historical datasets and the interests of participants in contributing to medical research and advancement of scientific knowledge on the one hand, and the need for consent and the duty to protect participants from risks on the other.

Challenges to **single-study consent** are often discussed only in relation to data originally collected for research purposes. However, new issues are raised by the collection and analysis of data from

potentially 'unwilling' participants, for example data scraped from social media platforms, smart phone applications, or open web forums (Markowetz et al. 2014; B. Mittelstadt et al. 2017). As social scientific and other forms of research begin to utilise data collected from unwilling or unaware participants, the lack of an explicit *informed* consent mechanism in end-user agreements gives cause for concern, even when data are de-identified (Ioannidis 2013).

5.1.1. Alternatives to single-study consent

Recognising these challenges, calls to reform single-study consent based on the belief that it is a barrier to 'necessary' research and innovation can be found in the debate (Larson 2013). Numerous approaches to consent have been proposed to overcome such barriers in purely information-based research, which re-use existing datasets (e.g. Rothstein and Shoben 2013; Prainsack and Buyx 2013; Schadt 2012). Mere technical compromises are not obviously the solution - for instance, anonymising the data may be insufficient to eliminate the need for consent due to the possibility of re-identification (Mello et al. 2013; Terry 2014; Tene and Polonetsky 2013).

Broad consent mechanisms, which pre-authorise future secondary analyses, are often used in place of single-study consent (Aicardi et al. 2016; Clayton 2005; Ioannidis 2013; Chalmers et al. 2016; Hirschberg, Kahrass, and Strech 2014; Laurie et al. 2010). According to Smith et al. (2016), broad consent allows participants to agree in advance "to have their samples, genomic data, and/or health information retained for use in any future research deemed appropriate by the relevant oversight body." These samples, data or information will often be de-identified when broad consent is used. Following this, broad consent models are often justified on the basis that secondary research presents fewer risks and burdens for participants, at least in terms of potential bodily harm and time/physical commitments (Johnsson and Eriksson 2016), while allowing for less bureaucratic or more efficient research governance (Hofmann 2009). For this to be true, consent documents "need to explain the breadth of consent and other elements of the framework for future research such as, for example, cross-border use of biomaterials and/or data, property rights, commercial use, and data protection" (Hirschberg, Kahrass, and Strech 2014).

However, broad approaches to consent are not without challenges. The autonomy of data subjects can be limited when a choice is not offered over the parties accessing their data, and the purposes for which it is used (Master, Campo-Engelstein, and Caulfield 2014; Hofmann 2009). Patients do not always prefer broad consent to more interactive or specific forms of consent (Garrison et al. 2015). Privacy concerns and a lack of trust between the platform providers or governance bodies and participants have been linked to concerns with broad consent (Smith et al. 2016). Broad consent has been criticised for not sufficiently respecting the autonomy of participants who, by definition, lack the necessary information to make an informed choice about participation at the point consent is taken (Johnsson and Eriksson 2016). By definition, one-off consent that allows for more than one use of data cannot respect participants' autonomy to the same degree as consent as a continuous process. Periodically re-contacting participants with updates on uses of their data can make consent a process rather than one-off event (Knoppers et al. 2011a). Participants can thus object to proposed uses, assess the risks of participation, and ask questions over time as desired, without researchers needing to explicitly gain consent for every proposed secondary usage.

Other alternatives to single-study specific consent aim to address precisely these challenges. **Tiered consent** allows data subjects to permit specific, well-defined uses of their data (Beskow et al. 2015).

A participant can specify, for example, that their data only be used for cancer research, or in particular not for genomic research. Circumstances requiring re-consent can also be defined (Majumder 2005, 33). Exclusion clauses can be used for a 'line-veto' type of tiered consent, which can increase confidence in data subjects that custodians are actually respecting their beliefs and values as translated into prohibitions of specific re-uses (Master, Campo-Engelstein, and Caulfield 2014). Where such formats are used, governance mechanisms, such as review councils and committees, help distinguish *'bona fide'* and problematic requests for access to data, and can categorise studies for comparison to participants' pre-defined preferences.[3]

Differences in broad and tiered consent models can, however, cause problems when data are shared across institutional and national borders. Different categories or requirements may be used, for example, to define future research uses to which participants have consented. Convergence is required in legal requirements and governance standards to ensure participants' preferences are fully respected in external transfers.

Pragmatic solutions are also possible based on models of **'data donation'** and **open data sharing**. In the context of genomic sequencing, research is sometimes restricted to "information altruists" (Choudhury et al. 2014), or individuals willing to openly share their data (and sometimes, identity) on the basis that they possess the social status or economic resources to be sufficiently protected from future discrimination or harmful consequences. This itself can create sampling bias in studies using donated data, as by definition individuals protected from harm due to social or economic status will not be representative of broader populations. 'Radical honesty' models are similar, through which individuals volunteer de-identified genetic information for public sharing (Hayden 2012). Another approach is to establish "honest broker" and "stewardship" consent models by which impartial third parties mediate broad consent agreements to protect the interests of data subjects (Choudhury et al. 2014, 7). Emphasising professionalism or enacting punitive measures for misuse of data can shift some of the burden to researchers benefiting from access to the data and promote feelings of responsibility to data subjects (Fairfield and Shtein 2014). The hope here is that forbidding unacceptable forms or research, such as negligent re-identification of anonymised data without prior consent, will minimise potential negative impacts on data subjects (Hayden 2012, 314).

Opt-out approaches to consent have also been proposed. However, opt-out models (e.g. (Tene and Polonetsky 2013; Rothstein and Shoben 2013; Terry 2012; Hoffman 2014) should not be seen as ethically equivalent to informed consent. Opt-out consent models may take advantage of people in vulnerable moments (Hayden 2012), for example if consent is taken during a clinical encounter in which the data subject is seeking treatment (cf. Pellegrino and Thomasma 1993; MacIntyre 2007). However, the weaknesses of such approaches do not suggest that explicit consent for each instance of data use is the correct path either; rather, a revision of ethical standards which strikes a balance between the requirement for consent and the practical requirements of research may be appropriate. Tene and Polonetsky (2013, 262) suggest as much in calling for debate on the "merits of a given data use" as a broader societal issue, wherein distinctions can be drawn between 'types' of data uses requiring full informed, opt-in, opt-out or no consent at all.

Flexible, dynamic consent can address these challenges by allowing patients to consent to new and further uses of discoveries from their data (Cato, Bockting, and Larson 2016; Williams et al. 2015;

[3] See for example the UK Biobank Ethics and Governance Council: http://www.egcukbiobank.org.uk/.

Stein and Terry 2013). Supported by human-computer interfaces, dynamic consent systems can help meet emerging ethical and legal consent requirements, while accommodating "the fluidity of data flows in research networks" (Kaye et al. 2015). Dynamic consent is thus seen as a way to update single-instance consent to handle growing opportunities for data sharing, linkage, and aggregation. Users are enabled to access study information, make choices, and express study-specific consent preferences through a continuous process of repeated contact rather than a one-off event (Johnsson and Eriksson 2016). While broad consent can also involve repeat contact, dynamic consent is generally distinguished by the possibility of granular, study specific expression of preferences that can travel with the participant's data (Kaye et al. 2015). Dynamic consent models require consent preferences to remain attached to data as they are shared, aggregated and accessed by multiple parties. Metadata expressing consent preferences are one possibility to achieve this (Woolley 2016).

One potential challenge for dynamic consent is that greater information offered to participants can unintentionally lead to information overload and confusion over the most relevant or important details of a study brief. Greater involvement can also exacerbate the "therapeutic misconception" by suggesting to patients that studies are designed to benefit them directly, rather than passively via advances in medical knowledge. Dynamic consent can also subtly transfer ethical responsibility from research ethics committees and researchers to participants (Williams et al. 2015). The expectation that participants have the time, dedication and expertise required to meaningfully assess new study proposals may place "unreasonable demands" on participants, and "forces them to raise their guard rather than to place trust" in prospective studies (Johnsson and Eriksson 2016).

Ideally, all information that is or could be deemed relevant by the participant would be transmitted. This is, understandably, an unattainable ideal. As Johnsson and Eriksson (2016) note, "information can be discovered after consent has been given that could not have reasonably been known at the time consent is taken." This uncertainty owes both to inconsistencies between the assessment of the researcher or clinician and the participant of her needs and informational requirements, as well as to the uncertainty of future secondary uses of her data (or samples) that cannot be predicted when time consent is taken (Chadwick and Berg 2001; B. Mittelstadt and Floridi 2016a; Stein and Terry 2013). A hypothetical 'reasonable person' is often used as an objective standard to measure the validity of consent and adequacy of information being communicated (Beskow et al. 2015); however, this model is flawed insofar as it requires personally meaningless or unwanted information to be shared with the participant (Johnsson and Eriksson 2016).

Consent procedures that seek to 'empower' patients with more information and control over their data or samples should be aware of these challenges. While the desire to help participants make more informed choices is laudable, such autonomy-enhancing approaches can inadvertently serve as a technical measure to show more information was made available to participants than previously possible at the time a choice to participate was made (Johnsson and Eriksson 2016; Blasimme and Vayena 2016). The mere availability of information does not indicate its relevance or meaningfulness to the participant.

5.1.2. Alternatives to consent

It may be possible to reduce or eliminate the need for consent by focusing on the **concept of solidarity** (Chadwick and Berg 2001; Laurie et al. 2010) and the lack of physical risks to data subjects in data-based research (although informational risks, such as re-identification, must still be taken

seriously; Stein and Terry 2013). Prainsack and Buyx (2013) suggest a solidarity-based approach to biobank governance, focused on harm mitigation,[4] which recognises an empirically supported sentiment among the general public (in Europe) to want to participate in biobanking research (Kaye et al. 2012; Steinsbekk et al. 2013). The shift to solidarity is also said to free up the "significant resources" currently spent on (re-)consenting procedures for primary and secondary uses of data held in biobanks for research, innovation and infrastructural improvements including interoperability between repositories (Prainsack and Buyx 2013). This position rests on the assumption that significant resources are currently being spent on re-consent procedures in particular, which are a central concern for consent and Big Data (e.g. Wellcome Trust 2014), and that these resources would instead be spent on valuable research and structural improvements.

Rather than suggesting changes to consent processes, solidarity intends to re-define the relationship between biobanks and data subjects by emphasising the willingness to share data or assist others to support research and innovation (Prainsack and Buyx 2013). In contrast to autonomy-based consent approaches, biobanks would instead model consent on solidarity by providing data subjects with a 'mission statement', information on potential areas of research, future uses, risks and benefits, feedback procedures and the potential commercial value of the data, so as to establish a "contractual" rather than consent basis for the research relationship (Prainsack and Buyx 2013). Such an approach is claimed to be acceptable given the relatively low risks in genomic research. According to the authors, few examples of discrimination based on biobank-facilitated research exist and are incomparable in quality to the bodily harm possible in other types of medical research (Prainsack and Buyx 2013). The relative lack of reporting on harms stemming from abuses of biomedical data has been noted in a recent Nuffield Council report on the ethics of linking biomedical datasets for research (Nuffield Council on Bioethics 2015). The lack has been largely attributed to a lack of robust reporting mechanisms and empirical research on underreporting, with most cases coming from anecdotal accounts and notable media stories. As a result, a lack of evidence of harms should not be considered evidence for a lack of harms. For example, the authors dismiss one of the most commonly cited fears of insurance discrimination based on disease susceptibility due to the "limited predictive value" of genetic markers at the individual level.

Solidarity has already been shown to work as a governing concept in certain areas of biomedical practice. Vaccination, public health surveillance, and assumption of consent to organ donation after death all provide evidence that solidarity can be taken seriously as a motivation to take beneficial actions, even if consent cannot be confirmed. Despite these examples, it can nonetheless be ethically problematic to require individuals to give up consent rights in the name of scientific advancement and the social good of advances in medical knowledge (Schaefer, Emanuel, and Wertheimer 2009; Chadwick and Berg 2001). Recognising this point, Prainsack and Buyx (2011) argue that an initial point of 'opt-in' must be retained, meaning participants must be given a choice to give up consent rights, rather than them being forcefully taken away or disregarded in the name of solidarity (as opposed to public benefit or a collective good). Eliminating or reducing the need for informed consent is a decision that cannot be taken uncritically or without public debate of the potential benefits of data

[4] The shift to solidarity is also said to free up the "significant resources" currently spent on (re-)consenting procedures for primary and secondary uses of data held in biobanks for research, innovation and infrastructural improvements including interoperability between repositories (Prainsack and Buyx 2013, 80). This position rests on the assumption that significant resources are currently being spent on re-consent procedures in particular, which are a central concern for consent and Big Data (e.g. (Wellcome Trust 2014), and that these resources would instead be spent on valuable research and structural improvements.

sharing, weighed against the uncertainty of its impact and the harm to participants of being denied a choice (Chadwick and Berg 2001; Prainsack and Buyx 2011).

Another possibility not relying on such a problematic form of responsibility may be to emphasise the trustworthiness of researchers or bodies requesting data, wherein data is shared only between 'trusted' actors (cf. Hansson 2009). Koenig's (2014) notion of "consent to be governed" reflects the potential for consent to be thought of as an opt-in to trustworthy governance procedures, rather than a rational and fully informed choice between options. As Prainsack and Buyx (Prainsack and Buyx 2016)argue, collective responsibility via collective choice and oversight must be considered in tandem with individual choice and harm mitigation strategies to achieve meaningful participant empowerment via consent. In short, if patients can trust that a repository will act in their best interests and seek to prevent harms, the movement from individual or single-study consent to broad consent need not be seen as problematic. Rather, an informed choice can be made to enter a trusting relationship with a repository or governance body that will manage 'participation' across multiple studies, rather than choosing to participate in a single study.

Beyond solidarity, there are reasons to prefer alternative models of governance for data-based research. Single-study consent can inadvertently act as a barrier to data sharing that would be in accordance with the consent preferences of data subjects (Choudhury et al. 2014, 5). In doing so, researchers may be missing opportunities to derive valuable information and innovations from the samples and data offered by research participants. Requiring individual consent for each secondary usage of data or samples is often impossible and creates a significant financial and bureaucratic burden (Wellcome Trust 2014). Single-study consent can therefore prevent legitimate forms of research in addition to those rightly viewed as challenging. Such barriers prevent researchers from advancing scientific knowledge, from deriving beneficial applications, and more generally from fulfilling the moral obligation to data subjects that have volunteered their time, bodies, and data for research (Chadwick and Berg 2001; Laurie et al. 2010).

Data sharing to advance medical knowledge through research is increasingly seen as both an "ethical and scientific imperative" (Knoppers et al. 2011a).[5] Secondary usage can allow researchers to meet the moral obligation to maximise the value of data collected from research participants without the need for further data collection which places participants at risk (Currie 2013; Harris 2005; Mello et al. 2013). Further, when data are altruistically donated or shared, secondary usage respects the donator's intention to contribute to medical research and advancement of scientific knowledge. Failing to meet this implied duty accompanying donated data can be considered initially ethically problematic. With that said, due consideration must be given to other constraints, such as the duty to protect donators from risks that were not recognised or did not exist at the time of donation, and the duty to notify participants about studies being conducted with their data or samples. These latter notification duties may explicitly be required of repositories as data controllers under Articles 13-14 of the GDPR (see: Section 2.1).

[5] By some accounts moral obligations exist for medical research. As suggested by accounts of solidarity-based governance of biomedical Big Data (e.g. (Prainsack and Buyx 2013), patients may have a moral duty to participate in research due to the value generated through advances in medical knowledge and treatments (Harris 2005; Schaefer, Emanuel, and Wertheimer 2009; Kaye et al. 2015).

5.2. Autonomy and participation

According to the ideal of informed consent (see: Section 4.1), participants in medical data repositories should normally be given a choice in the collection and/or secondary processing of their data and samples. Governance structures define the conditions under which data can be accessed, by whom, and the relationship with and involvement of participants in the repository. Governance of medical data repositories depend upon trust between participants and the platform (Knoppers et al. 2011a; Meslin and Cho 2010). Informed consent is typically taken to demonstrate respect for participant autonomy (O'Neill 2003). Many formulations of a principled respect for autonomy are possible (O'Neill 2003), which set different requirements for informed consent and related protections downstream.

Much work in bioethics on informed consent is based on respect for autonomous *choice*, or the freedom of people to make decisions when their freedom, dignity, or other interests are at stake (Johnsson and Eriksson 2016). This approach is concerned with participants being adequately informed, in possession of the required knowledge or skills to process the information being provided, and not coerced into making a particular choice (Blasimme and Vayena 2016; Johnsson and Eriksson 2016; O'Neill 2003). The range of choices presented must be meaningful for the latter requirement to be met, with participants granted a realistic opportunity to make a choice that contravenes the preferences of the institution or researchers making the offer, including a choice in the type and quantity of information to be received (O'Neill 2003). Binary choices can nudge participants towards institutional preferences, rather than allowing for autonomous decision-making reflecting the participant's interests and values.

As Blasimme and Vayena (2016) argue, these criteria alone do not, however, ensure participants will have a *meaningful range of worthwhile options* to choose from. According to this stronger requirement, merely providing participants with more options, or more information, does not necessarily show greater respect for autonomy. Rather, the options or information provided should reflect the participant's "morally valuable interests."

Following this broader approach, other formulations of autonomy (or other complementary values) are possible that address a person's sense of belonging or community. Johnsson and Eriksson (2016) refer to this as respect for *authority* rather than *autonomy*: "to respect someone's authority is to respect their right to take responsibility for themselves, for their families and relationship to society." Authority is about a person's 'belongingness' to society or particular communities, as freedom, dignity and other interests often evolve through social interaction. In contrast to respect for autonomous choice as a type of "negative right to decline participation," respect for authority is conceived as a positive right that enables the individual to join and foster communities and relationships with other participants. As a positive right, authority imposes duties on researchers and data custodians beyond those imposes by respect for autonomy (e.g. non-interference or non-coercion of choice; Johnsson and Eriksson 2016).

Adequate provision of participatory options is particularly important to the broader conception of respect for autonomy as respect for authority, or meaningful choices. Such options include participation in governance decisions for sharing and usage of one's data, cohort dataset or repository. Participation demonstrates respect for autonomy by providing "the freedom to act…as members of a self-governing community that tries to attain what is of common interest" (Blasimme and Vayena 2016). However, the mere option to participate is unlikely to result in a community of

participants across a repository with common interests and agency (Blasimme and Vayena 2016; B. Mittelstadt 2017). Options to encourage partnership between participants are also required if a meaningful community is to be formed, wherein the community can mutually debate and form common interests and an ethos that plays a central role in the governance of the relevant dataset or repository. Blasimme and Vayena (2016) suggest that partnership and communitarian ethos as mechanisms to respect participant autonomy have largely been ignored thus far in ethical governance frameworks. With that said, as discussed above, discussions of solidarity and collective oversight of data sharing repositories emphasis both collective benefits and choice (see: Section 5.1.2).

Dynamic consent has been recognised as one such way to provide meaningful participation choices by allowing users to indicate per-project or per-type consent preferences. Dynamic consent systems can also be used to return individually relevant results and incidental findings, and to allow users to monitor relevant findings and health over their data or sample's lifetime. New data or samples can also be requested from existing participants, allowing for longitudinal research with interesting individuals or groups rather than one-off studies (Chalmers et al. 2016).

The potential of dynamic consent to allow for greater engagement between researchers and participants raises another question: under what conditions is further engagement or sharing of findings appropriate or imperative, including sharing of benefits? Setting appropriate requirements to share findings of secondary research is a much-discussed problem in bioethics and biobank governance (UK Biobank 2015). Factors often considered include the availability of treatments for detected conditions, limitations on re-identification and re-contacting participants, the consent preferences of participants concerning incidental findings, and the relative certainty of the finding or predictive correlation. Independent of these ethical concerns, opportunity benefits and participants' wishes must also be considered. Given an uptick in online services and sharing of personal information online, privacy concerns may hold less weight with participants than has historically been the case. Empirical studies of the attitudes of specific participant populations can assess the relative weight given to ethical concerns and participants' preference to share data openly or by default (cf. (B. Mittelstadt et al. 2017). Benefit-sharing mechanisms for contributing populations or individuals must also be considered in the broader focus on communities or collectives (Nicol and Critchley 2012; Collins 2016; Blasimme and Vayena 2016). The validity of a participant's choice to not be personally informed of findings must be respected in the first instance, while simultaneously accounting for case-specific valid external interests in disclosures.[6] Participants can reasonably choose not to be informed of population or individual level findings, particularly when therapeutic options are not available or offered by a study, or when a finding indicates a risk factor for genetically similar individuals, such as family members (cf. Pullman et al. 2012). New standards and policies of professional conduct for informing participants about individual or population-level research results generated via medical data repositories may be required (McGuire et al. 2008, 2012).[7]

As the link between trust and the principle to respect participant autonomy demonstrates, researchers can also be considered a part of the communities involved in the governance of medical data

[6] However, the right to 'not know' about incidental findings must also be balanced against the interests of related individuals or groups in disclosure. Work in genomics has unpacked the difficulties of balancing these interests in disclosure. See for example: (UK Biobank 2015; Joly et al. 2012).

[7] Regulatory action may be required, as Big Data creates new opportunities for "data aggregators and miners to…run around health care's domain-specific protections by creating medical profiles of individuals" not subject to existing legislation (Terry 2012, 386), as was the case with the Google Health platform which operated outside of HIPAA restrictions in the United States (Mora 2012, 373). This type of profiling based on personal data is, however, subject to the European Union's General Data Protection Regulation.

repositories. The relationship between medical researchers, participants and the general public is a longstanding issue in bioethics (cf. Chadwick and Berg 2001; Schaefer, Emanuel, and Wertheimer 2009; B. Mittelstadt et al. 2017). Reflecting the aforementioned communitarian concerns, a re-framing of this relationship has been proposed in various forms that further moves away from the protectionist stance afforded by the single-study informed consent model. A relationship based on reciprocity (Meslin and Cho 2010) or solidarity (Prainsack and Buyx 2013), for example, seeks a new balance between the respect for participant autonomy and the broader public and societal benefits of medical research.

At a practical level, movement away from a protectionist stance and single-study consent must re-establish the trust or protections for participants being lost. As the discussion above indicates, a closer, participatory relationship between researchers and participant communities in medical data repositories can fill the gap. Researchers must, in this new relationship, demonstrate trustworthiness and harm mitigation (Meslin and Cho 2010; Laurie et al. 2010), and assure the public that, “the goals of research projects are primarily to meet public needs, rather than self-interest” (Meslin and Cho 2010).

5.3. Transparency

Whether the aim is to respect autonomy, authority, or to enable a trusting relationship between researchers, individuals, communities and society, some level of transparency is required. Basic respect for persons suggests that individuals “should be offered appropriate ways to (not) consent to have their health records included in the central database” (Sterckx et al. 2016). Information about how participants’ data are being handled and used must be shared. Transparency grounds this exchange, requiring sufficient information to be shared with participants to enable informed decision-making about participating in a particular study, or allowing a third party to govern access to their data in line with the individual’s interests (Sterckx et al. 2016). A lack of transparency along these lines can damage participants’ trust in data sharing schemes (Sterckx et al. 2016; Blasimme and Vayena 2016).

In seeking transparency, different requirements can be set for the nature and frequency of information exchange with participants. Context-specific requirements will depend upon the purpose for which transparency is being sought, for example to demonstrate respect for the autonomy of participants, to protect their privacy, to encourage trust between researchers and participants, to demonstrate accountability in a governance or sharing structure, or to enable informed choices about future participation (Cato, Bockting, and Larson 2016, 217; Sterckx et al. 2016; Blasimme and Vayena 2016). Focusing on trust and informed choice, for example, can ground requirements that third parties gaining access to participant data must be clearly identified or categorised prior to seeking consent (Sterckx et al. 2016).

However, transparency does not come without risks. Opening data or access arrangements to the public can present risks for the privacy of participants. Data can also easily be misinterpreted by users lacking necessary resources or expertise: Collins (2016), for example, discusses misinterpretation of US Medicare cost data following public release.

5.4. Ownership

As an ethical concept distinct from its legal operationalisation in intellectual property rights, ownership can traditionally be distinguished in two senses: as an individual's right (1) to exclusive control of an entity (in this case, data), and (2) to benefit from its exploitation. These two senses of 'ownership' are evident in the reviewed literature: as a right to 'control' data, and as a right to 'benefit from' data.

When discussed as grounds for control over data, ownership overlaps substantially with issues of consent, privacy, and respect for autonomy. Distribution and modification of data can be restricted by the data owner for the sake of privacy, to maintain the data's integrity, or to limit its analysis by others (Laurie et al. 2010). These interests do not need to be protected by the owner directly. Third parties can be empowered to control data on behalf of data subjects; biobanks, for example, can be seen as custodians of the data and samples contributed by participants (Kinkorová 2016). To prevent secret databases or undesired uses, data subjects granted a right to control can be given mechanisms to track and verify creation, dissemination, and usage of their data (Tene and Polonetsky 2013; Choudhury et al. 2014). For biobanks, user-end control mechanisms are particularly important when third party access will be granted for commercial pursuits (e.g. (NHS England 2014) to which the participant may object if informed (Sterckx et al. 2016; Steinsbekk et al. 2013). Interests in human dignity can, for example, restrict commodification of medical data (Steinsbekk et al. 2013).

Ownership can also ground claims to share in benefits resulting from analysis of data that creates intellectual property, products or services. These claims extend beyond third party usage. As Tene and Polentsky (2013) argue, data controllers should offer data subjects "meaningful rights to access their data in a usable, machine-readable format" for personal analysis and innovation. Such mechanisms enable individuals subjects to derive benefits for themselves from the data they produce and communities (or aggregated datasets) in which it resides (Lupton 2014, 866).

Accessibility in both contexts is not without risks and necessary limitations. For instance, providing data subjects with unrestricted access to raw data may be harmful in the sense that it is practically useless or open to misinterpretation without the presence of a trained clinician or analyst to explain its significance (Watson, Kay, and Smith 2010). Revision rights can also undermine the integrity of datasets, introducing user-generated errors and inaccuracies. Gaps in coverage can also be introduced through revision, withdrawal, or erasure of a participant's data; however, these concerns with the quality of data must be balanced with the individual's right to withdraw from studies. In the case of anonymised records, revision, withdrawal, or erasure will often not be technically feasible.

5.5. Data provenance

Data integration can be conceptualised as a socio-technical (as opposed to purely technical) problem (Neff 2013). Data will further be generated or collected with different assumptions, standards and methods, all of which must be accounted for by the aggregating platform in order to be accurately interpreted and compared. Choices made in aggregating and making data accessible to researchers, concerning for instance levels of identifiability or analysis protocols (E. S. Dove, Knoppers, and Zawati 2014), affect the awareness of researchers to potential gaps, assumptions, or contextual factors in the records being accessed.

Context-sensitive understanding may be impossible in aggregated datasets. Context and meaning are often purposefully stripped from data representing a participant's behaviour or medical history, for example to protect identity. Methodological and individual context is lost through the aggregation of data unless extensive precautions are taken to preserve the 'assumptions' that helped generate the data. As Busch (2014) argues, "the categories to be used in collecting data, the procedures for handling missing data, the specific subjects of data collection, the nature of the sampling methods used, and the means by which to construct and aggregate the data" need to be consistently communicated to researchers requesting access. Aggregation obscures the complex methodological decisions and ontological assumptions that ground the research that produced the data to be aggregated.

Busch (2014) describes the following characteristics of aggregated datasets which contribute to a loss of context:

- **Lossiness** – Aggregation, case construction, standardisation and simplification of data to enable cross-sectional analysis may 'lose' certain aspects of the phenomena studied.
- **Drift** – Phenomena change over time, but the data representing them does not. The same can be said for the methods underlying the primary data collection and analysis.
- **Distancing** – Large datasets facilitate identification of patterns or 'clarity' by distancing oneself from the phenomenon.
- **Layering** – A 'realist' ontology is pre-supposed in that an assumption is made that the relations underlying the phenomena will remain over time as the data is aggregated and manipulated. The 'situatedness' or contextual meaning of each phenomenon must be removed for the (data representations of the) phenomena to be treated as sufficiently similar for aggregation. Context is lost by reducing the phenomena to a set of variables: "those aspects of things that are not amenable to numerical or statistical analysis - that situate particular phenomena - are systematically downgraded or removed from consideration" (Busch 2014).
- **Errors** – How are errors within the dataset identified and addressed?
- **Standards** – "The process of creating uniformity through standardization," or fitting data to discipline conventions or categories, "may obfuscate phenomena of considerable importance" (Busch 2014).
- **Disproportionality** – Outlying data may be deleted or treated as 'errors' to enable simplification and standardisation of the dataset.
- **Amplification/Reduction** – Aspects of phenomena amenable to quantified measurement are amplified in importance, while those that are not are reduced.
- **Narratives** – Large datasets can hide the role of interpretation in seeing data *as* something and obscure alternatives to the preferred interpretation.

Without explicit and consistent controls to communicate these assumptions and characteristics, aggregated data can mistakenly be treated as a complete and accurate representation of a particular population or condition (Crawford 2013; Bowker 2014). Weaknesses of data in terms of both coverage and loss of context need to be communicated to and acknowledged by researchers. As Bowker (2014) argues, datasets "cannot in principle contain the world in small…most slices of reality are not represented" (Bowker 2014, 1797). Increasing the scale and variety of data being aggregated and made available to researchers via repositories only increases the difficulty of identifying what contextual information and 'outliers' have been stripped from the dataset.

Lupton (2014) describes this phenomenon in terms of analytic metrics used to sort individuals and groups and highlight specific aspects or characteristics to 'understand' them. Metrics "make visible

aspects of individuals and groups that are not otherwise perceptible, because they are able to join-up a vast range of details derived from diverse sources" (Lupton 2014, 859). These metrics provide different ways of 'seeing' the groups and interpreting their behaviours; that a particular interpretation is correct or reflective of the meaning, identities or motivations given to acts by members of the group cannot be taken for granted.

With regards to data coverage, discrimination and benefits of secondary research may become localised around groups that present easy or interesting analysis opportunities. These 'blind spots' mean that analysis will tend to focus on data subjects and phenomena amenable to digitisation and measurement, meaning that the benefits and ethical burdens research enabled by medical data repositories will be placed, for better or worse, on specific social, cultural and economic groups (Majumder 2005; McGuire et al. 2008). For instance, analysis of social media datasets (B. Mittelstadt et al. 2017) will necessarily affect social media users and their underlying demographics in the first instance. Broad benefit sharing should be enabled as far as possible (Choudhury et al. 2014).

Localisation of burdens and benefits can be considered ethically problematic on the basis of inequality between groups. Knowledge and treatments will normally target populations for whom data is available by definition. Unrepresentative coverage has already been noted in relation to secondary uses of genomic and microbiomic data (Lewis et al. 2012). Existing gaps in medical practice and knowledge between "Euro-Americans of middle to upper socio-economic status" and others (Lewis et al. 2012, 2) can thus inadvertently be exacerbated by medical data repositories, unless explicit steps are taken to broaden the coverage of the data held. Bioethics work concerning global justice and pharmaceutical research ground these challenges. Even where studied populations are diverse, formal benefit sharing agreements may be required between participants and repositories to ensure data are not taken from one group purely to benefit another (Mathaiyan, Chandrasekaran, and Davis 2013; Chadwick and Berg 2001).

Data provenance also concerns the attachment of consent preferences to data as it moves between third parties (Woolley 2016). In part this challenge can be solved by requiring on-site access, or hosting data in a 'safe harbour' in which consent preferences can be assessed before access is ever granted (E. S. Dove, Knoppers, and Zawati 2014; Kaye et al. 2015). However, many emerging digital data sources accessible to researchers and private firms (e.g. the Internet of Things, wearable computing) provide longitudinal records of the behaviours, lifestyle, and histories of users (Collins 2016; B. Mittelstadt et al. 2014). These data can be collected via user agreements rather than institutionally approved informed consent processes (Collins 2016), which raises questions over the legitimacy of secondary usage by researchers. While concerning in itself, these gaps in knowledge about the study-specific consent preferences will not necessarily be communicated to researchers and firms involved in secondary uses. This presents a separate but important problem.

5.6. Privacy

In the reviewed literature, numerous concerns were described in terms of privacy, some of which relate to alternative concepts such as autonomy or freedom of information. It is worth noting that privacy can have very different theoretical grounding in international ethical and legal frameworks. In Europe, for example, privacy in relation to personal data is primarily protected through data protection law, which is partially grounded in the human right to privacy enshrined in the European Convention on Human Rights and European Charter of Fundamental Rights. In the reviewed ethics and social

literature, links were found with confidentiality, understood as “the duties that accompany the disclosure of non-public information within a fiduciary, professional or contractual relationship” (Majumder 2005, 33). Others discussed privacy in terms of the ‘invasiveness’ of secondary analyses. Invasiveness was connected in particular to analysis of combined data sets, particularly from geolocation and internet-based sources, even when such data is anonymised (e.g. (Moore et al. 2013; Markowetz et al. 2014).

The scope of data being collected can also be conceived of as a privacy issue. Traditionally, data collection has been limited by human perception and cognition. However, with automated and autonomous collection by information technologies, the scope of data, as can be seen over the past two decades, has grown exponentially. More personal and highly detailed data can be collected and analysed than at any other time in history. Furthermore, these data are designed to be stored in perpetuity, meaning that traditional limitations of memory no longer apply; data collected today may, in theory, be equally accessible and of the same quality in the future. Other issues are now emerging as important for the preservation of data, such as the obsolescence of software, the presence of malware, and the potential fragility of physical supports. While not a privacy issue *per se*, extending the lifespan of data describing phenomena that would otherwise be forgotten does increase the risks that privacy violations may occur.

Governance of ‘health’ data is currently facing significant challenges, introduced by the linkage and aggregation of disparate data sources (B. Mittelstadt and Floridi 2016a). Current data protection legislation in the USA and EU may not protect all medically-relevant or health-related data, or afford such data the protections granted to sensitive health data. Information about health can increasingly be derived from inferences about individuals and groups, enabled through linkage and aggregation of medical and non-medical datasets (Aicardi et al. 2016). Linkage, aggregation, and usage of these data are not always governed by traditional institutional ethical review systems and codes of conduct (Liyanage et al. 2014, 33). This situation is particularly concerning for privatized and internet-based health data sources, such as patient-driven databases (e.g. PatientsLikeMe; (Liyanage et al. 2014), which are likely to be subjected to less stringent requirements when compared to biobanks and repositories of clinical trial data, where restrictions can be enforced by governance bodies. Clarity is similarly required for non-commercial medical data repositories concerning the scope of data being held and accessed. Where medical and non-medical data are linked or aggregated, “finer distinctions between different types of biobanks and health databases according to their mission, practices, uses, and the commercial stakes involved in them” are required (Aicardi et al. 2016).

5.6.1. Anonymisation and identifiability

Anonymisation and privacy were closely linked in the literature. Privacy concerns are often thought to be mitigated by removing identifying information from a dataset. Anonymisation describes a set of techniques to remove identifying information from a record. ‘De-identification’ may be a more accurate term to describe these techniques, as re-identification is often still possible even when identifiers have been deleted and not stored in a separate coded record (i.e. ‘pseudonymisation’) (Ohm 2010). Nonetheless, anonymisation was frequently seen as the minimum requirement necessary to protect data subjects’ privacy in aggregating data.

For medical research, data is often de-identified to limit the necessity of explicit and specific consent from data subjects, to protect participant privacy, or to meet legal data protection requirements (Heurix

et al. 2016; Laurie et al. 2010). Knoppers and Saginur (2005) describe a “jumble of terminology” related to the identifiability of data. To clarify the vocabulary of identifiability, they propose that, distinct from ‘identified’ data, de-identified data should be described as either “coded (single or double) [or] anonymized.” Differences in language between national and institutional legal and ethical frameworks poses a challenge to interoperability, cross-border sharing (Tassé 2016; Knoppers and Saginur 2005), and participant understanding and appreciation of risks (Hull et al. 2008).

Further concerns relate to the actual protection offered by anonymization. Many examples have proven that anonymization cannot fully and beyond doubt prevent the re-identification of participants (Ohm 2010; Mostert et al. 2016). Linkage of new datasets and advances in techniques for reverse engineering identity suggest that anonymization must be considered temporary and potentially reversible. Re-identification is possible through cross-referencing with data concerning ethnic background, locational data, other metadata, health records or even small pieces of identified genetic data (Choudhury et al. 2014; Hayden 2012; Joly et al. 2012). According to Aicardi et al. (2016), “not only is the anonymity of data and material highly context-dependent, but data and material that are anonymized today may no longer be anonymous in the context of tomorrow’s technologies and data resources.” The categorisation of data as ‘anonymized’ must therefore be iteratively questioned and periodically re-visited. Concerns with re-identification must also likewise apply to anonymized data, particularly where exemptions from ethical and legal requirements governing data processing are being considered (Aicardi et al. 2016).

Further emphasising this need, Sterckx et al. (2016) note that, while de-identified data and samples often fall outside of the scope of data protection law and governance of human subjects research, ethical concerns can nonetheless arise. Controversy has, for example, surrounded the usage of de-identified newborn bloodspots in the USA, which “demonstrate the importance of providing information, transparency, guaranteeing individuals’ right not to participate, and preserving trust in the healthcare system” (Sterckx et al. 2016). De-identification can also affect the rights of participants, such as the right to withdraw (Knoppers and Saginur 2005) or to access outcomes or benefits of secondary uses.

Anonymisation imposes costs in terms of the value of the data for researchers that must be balanced with its privacy gains (Laurie et al. 2010). Removal of identifying information can prevent re-contacting of participants which can be privacy enhancing, but likewise prevent communication of incidental findings or follow-up requests for further information, such as the participant’s subsequent health status (Stein and Terry 2013). Recognising these costs, anonymisation cannot be straightforwardly considered a default option to protect participant privacy.

To address weaknesses of anonymization, alternative protections have been proposed including data audit trails (Ohm 2010) and digital identifier systems that allow “data tracing and prospectively limit the potential for malicious activities involving re-identification” (Knoppers et al. 2011a). Data can be hosted in ‘safe harbours’ within which data uses are screened and controlled (E. S. Dove, Knoppers, and Zawati 2014). Although these measures do not address the effects of secondary uses of de-identified data on groups, they are pragmatically responsive to possibilities of re-identification, while not further restricting movement of anonymised data.

5.7. Group harms and discrimination

The emphasis on identifiability in data protection and privacy law, as well as ethical frameworks for the governance of secondary uses of data, can give the impression that harms are only possible when individuals are identifiable from their data. This mindset ignores that potential group-level harms from analysis of aggregated data are clearly possible (B. Mittelstadt 2017; B. Mittelstadt and Floridi 2016b; Fairfield and Shtein 2014). When individuals are grouped according to geographical, socioeconomic, ethnic or other characteristics for analysis, anonymisation makes little difference (Choudhury et al. 2014, 6; B. Mittelstadt 2017). Discrimination and stigmatisation of affected groups are possible even from analysis of anonymised datasets (Docherty 2014). Such effects impact on all members of the community, not only those individuals who gave consent for data processing (Fairfield and Shtein 2014, 45). Biological material and data can likewise reveal traits of genetically similar individuals, including family members (Aicardi et al. 2016; Blasimme and Vayena 2016). Despite this, privacy and data protection law typically apply only to identifiable or identified individuals, but not groups (Knoppers and Saginur 2005).

Due to potential group-level harms, anonymisation can be criticised when presented as a 'silver bullet' that avoids, or at least minimises, the risk of being 'singled out' for discrimination or preferential treatment (McGuire et al. 2012). An 'ethics of care' approach may be appropriate when working with data aggregated from groups based on for example indigenous, demographic, ethnic or cultural features, to avoid possibilities of discrimination (Lewis et al. 2012, 3).[8]

Data analytics and data mining via unsupervised machine learning raise related concerns. Social discrimination can result from probabilistic correlations established from public or anonymised data that predict, for example, the likelihood of developing a future health condition. These correlations can have tangible impact (e.g. insurance eligibility) upon individuals possessing the relevant characteristics (e.g. biomarkers), independent of the correlation's accuracy or the individual's (lack of) participation in the initial analysis. How harms, benefits, rights and duties related to secondary usage apply to and are distributed across individuals and groups must be considered (Aicardi et al. 2016).

Reflecting these challenges, a "communitarian" (Blasimme and Vayena 2016) or "relational" (Jennings 2016; Edward S Dove et al. 2017) turn has been noted in bioethics in recent years, which challenges the notion that individuals are the only possible subject of moral rights or consideration. According to (Blasimme and Vayena 2016), "a whole constellation of values such as reciprocity, mutuality, solidarity, universality and citizenry has taken center stage" to highlight the need to give moral consideration to groups with "cultural, identity-shaping interests." This movement in bioethics argues that individual's interests, needs, and (autonomous) choices reflect relations with others, including members of their personal networks, communities, and cultures (Dove et al. 2017). As this movement away from individualistic ethics matures, formal protection of group-based or relational interests must be considered in platform governance and regulation of the future (B. Mittelstadt 2017).

[8] The applicability of theories on the ethics of care (e.g. Gilligan 1982; Noddings 2013; Slote 2007) to Big Data likely extend beyond discrimination against marginalised groups. For example, emphasising responsiveness and relationships between data subjects, custodians and analysts may provide avenues for development of new privacy protection mechanisms and group-level ethics which acknowledge the network ethical effects possible through Big Data. While a full account of this and related topics concerning ethics of care goes beyond the scope of this paper, existing work on the applicability of the ethics of care to public health (e.g. (Kass 2001) may provide a starting point for future enquiries.

6. Discussion

As originally proposed in the ROADMAP Description of Action, D8.1 was intended to cover ethical, social, as well as local issues arising from a RWE approach. This deliverable described the legal context of ROADMAP and RWE in Europe more broadly, before explicitly reviewing academic literature covering ethical and social issues with RW. The review was expanded to cover discussion of medical data sharing more broadly given the relative lack of literature returned in pilot database searches focused solely on ethical and social aspects of RWE. Legal issues were touched upon in this broader sample, but legal keywords and databases were not explicitly included in the search. The review thus covers ELSI in RWE, but will be expanded in D8.2 with an explicit review of legal databases to unpack uncertainties in the legal context in which ROADMAP will operate from 2018, following enforcement of the European General Data Protection Regulation. The subsequent review in D8.2 will also examine ethics and governance policies for European biobanks and other medical data sharing repositories, which reflect how Member State and Union law has previously been translated into governance requirements for medical data repositories. These latter sources will describe best practice for sharing, linkage and aggregation of sensitive medical data across all data types described in the ROADMAP Description of Action (i.e. trials, cohorts, patient reported outcomes, national registries and EHR databases). Finally, empirical studies of the attitudes of AD and dementia patients and carers towards medical data sharing and secondary uses of personal data for research will be surveyed in D8.2 to ground this process.

In subsequent months, the findings of this deliverable will be translated, along with findings of the aforementioned additional reviews, into initial recommendations for ELSI requirements for ROADMAP will be defined in D8.2, with final requirements following in D8.5. D8.2 and D8.5 will apply the concerns identified here and in the subsequent reviews to the specific case of RWE for AD research. Given these future plans, the work reported here can best be considered a review of ethical and social issues arising from ‘Biomedical Big Data’ and sharing of medical data for research, as opposed to the specific issues arising in sharing of medical data for AD research.

7. Conclusion and next steps

An overview of key ethical, legal, and social issues of medical data repositories has been offered. Biomedical applications of Big Data are particularly ethically challenging due to the sensitivity of health data and fiduciary nature of healthcare. Despite the wealth of literature available, challenges remain to translate contemporary discussion on ethical, legal, and social issues in medical data repositories into governance structures and best practice policies. As Chalmers (2016) argues, outstanding issues include "ongoing problems with the nature of consent particularly whether broad consent is ethically defensible; ensuring respectful and appropriate ongoing involvement and connection with participants; retaining public trust in an increasingly commercialised research environment; properly maintaining the physical space required to store tissue; keeping up to date with rapidly changing technology that may result in more accurate collection/storage/analysis; and, perhaps most problematically, sustainability in a constrained funding environment." The analysis undertaken here is intended to contribute to ethically responsible development, deployment and maintenance of novel datasets and practices in biomedicine concerning such challenges.

Resolving these challenges in a manner that enables cross-institution and cross-border data sharing will not be simple. According to Tassé (2016), ethical and legal frameworks governing secondary usage of data for medical research often contain discrepancies. Nonetheless, Tassé identifies nine criteria that can "form the basis for an internationally harmonized scheme for the secondary use of data in research" (see: Table 3). These criteria, along with the findings of this review and future analyses (see: Section 6), will ground an initial set of ELSI recommendations for an RWE approach to AD, to be reported in D8.2.

Table 3: Nine criteria for a harmonized ethical and legal framework governing secondary usage (Tassé 2016):

- Is obtaining a new consent impossible or impracticable?
- Is the waiver of consent granted by an REB or another authorized committee?
- Is access to identifiable personal data essential for the research?
- Are known preferences of the participants (or relevant groups) about the use of data taken into consideration?
- Does the proposed research use correspond to an important public interest?
- Does the waiver of consent adversely affect the rights and welfare of the subjects (balance of risks vs. benefits)?
- Does the research design involve no more than minimal risk?
- Are the privacy of research participants and the confidentiality of data protected?
- If warranted, is the waiver of consent consistent with international and domestic law?

Beyond the themes that emerged from the literature, a number of additional important issues received limited coverage. Commercialisation of participants' data and repositories as a whole, including access by private third parties, is an issue of growing significance given the development of new data sources in the private sector (Sterckx et al. 2016). Sharing of incidental findings with participants, while much discussed in relation to genomics and microbiomics, is also an important consideration for repositories producing individual and population-level results, particularly when actionable insights

are produced. The availability of therapeutic remedies must be considered in designing policy for sharing of incidental findings (Fischer et al. 2016). Following this, distinctions must also be drawn between research aimed at the development of diagnostic knowledge and methods on the one hand, and therapeutic interventions on the other (Hirschberg, Kahrass, and Strech 2014). While the relative absence of each of these issues may merely reflect the search method, they are nonetheless relevant to the governance of emerging medical data repositories.

8. References

Aicardi, Christine, Lorenzo Del Savio, Edward S. Dove, Federica Lucivero, Niccolò Tempini, and Barbara Prainsack. 2016. 'Emerging Ethical Issues Regarding Digital Health Data. On the World Medical Association Draft Declaration on Ethical Considerations Regarding Health Databases and Biobanks'. *Croatian Medical Journal* 57 (2): 207–13. doi:10.3325/cmj.2016.57.207.

Allen & Overy. 2017. 'The EU General Data Protection Regulation'. Accessed June 5. http://www.allenovery.com/SiteCollectionDocuments/Radical%20changes%20to%20European%20data%20protection%20legislation.pdf.

Angrist, M. 2009. 'Eyes Wide Open: The Personal Genome Project, Citizen Science and Veracity in Informed Consent'. *Personalized Medicine* 6: 691–99.

Beskow, Laura M., Carrie B. Dombeck, Cole P. Thompson, J. Kemp Watson-Ormond, and Kevin P. Weinfurt. 2015. 'Informed Consent for Biobanking: Consensus-Based Guidelines for Adequate Comprehension'. *Genetics in Medicine* 17 (3): 226–33. doi:10.1038/gim.2014.102.

Blasimme, Alessandro, and Effy Vayena. 2016. 'Becoming Partners, Retaining Autonomy: Ethical Considerations on the Development of Precision Medicine'. *BMC Medical Ethics* 17 (November). doi:10.1186/s12910-016-0149-6.

Boddington, Paula, Liam Curren, Karen Melham, Heather Gowans, Jane Kaye, Nadja Kanellopoulou, and Naomi Hawkins. 2011. 'Consent Forms in Genomics: The Difference between Law and Practice'. *European Journal of Health Law* 18 (5): 491–519. doi:10.1163/157180911X598744.

Borry, P., P. Schotsmans, and K. Dierickx. 2008. 'The Origin and Emergence of Empirical Ethics'. In *Empirical Ethics in Psychiatry*, 37–50. New York, NY: Oxford University Press.

Bowker, Geoffrey C. 2014. 'Big Data, Big Questions| The Theory/Data Thing'. *International Journal of Communication* 8 (0): 5.

Boye, Niels. 2012. 'Co-Production of Health Enabled by next Generation Personal Health Systems.' *Studies in Health Technology and Informatics* 177: 52–58.

Busch, Lawrence. 2014. 'Big Data, Big Questions| A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets'. *International Journal of Communication* 8 (0): 18.

Butler, Declan. 2013. 'When Google Got Flu Wrong'. *Nature* 494 (7436): 155–56. doi:10.1038/494155a.

Cato, Kenrick D., Walter Bockting, and Elaine Larson. 2016. 'Did I Tell You That? Ethical Issues Related to Using Computational Methods to Discover Non-Disclosed Patient Characteristics'. *Journal of Empirical Research on Human Research Ethics* 11 (3): 214–19. doi:10.1177/1556264616661611.

Chadwick, Ruth, and Kåre Berg. 2001. 'Solidarity and Equity: New Ethical Frameworks for Genetic Databases'. *Nature Reviews Genetics* 2 (4): 318–21. doi:10.1038/35066094.

Chalmers, Don, Dianne Nicol, Jane Kaye, Jessica Bell, Alastair V. Campbell, Calvin W. L. Ho, Kazuto Kato, et al. 2016. 'Has the Biobank Bubble Burst? Withstanding the Challenges for Sustainable Biobanking in the Digital Era'. *BMC Medical Ethics* 17 (July). doi:10.1186/s12910-016-0124-2.

Choudhury, Suparna, Jennifer R. Fishman, Michelle L. McGowan, and Eric T. Juengst. 2014. 'Big Data, Open Science and the Brain: Lessons Learned from Genomics'. *Frontiers in Human Neuroscience* 8: 239. doi:10.3389/fnhum.2014.00239.

Clayton, Ellen Wright. 2005. 'Informed Consent and Biobanks'. *Journal of Law, Medicine & Ethics* 33 (1): 15–21. doi:10.1111/j.1748-720X.2005.tb00206.x.

Collins, Brendan. 2016. 'Big Data and Health Economics: Strengths, Weaknesses, Opportunities and Threats'. *PharmacoEconomics* 34 (2): 101–6. doi:10.1007/s40273-015-0306-7.

Costa, Fabricio F. 2014. 'Big Data in Biomedicine'. *Drug Discovery Today* 19 (4): 433–40. doi:10.1016/j.drudis.2013.10.012.

Crawford, Kate. 2013. 'The Hidden Biases in Big Data'. *Harvard Business Review*. http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

Currie, J. 2013. '"Big Data" Versus "Big Brother": On the Appropriate Use of Large-Scale Data Collections in Pediatrics'. *PEDIATRICS* 131 (Supplement): S127–32. doi:10.1542/peds.2013-0252c.

Devos, Yann, Pieter Maeseele, Dirk Reheul, Linda Van Speybroeck, and Danny De Waele. 2008. 'Ethics in the Societal Debate on Genetically Modified Organisms: A (Re)Quest for Sense and Sensibility'. *Journal of Agricultural and Environmental Ethics* 21 (1): 29–61. doi:10.1007/s10806-007-9057-6.

Docherty, A. 2014. 'Big Data - Ethical Perspectives'. *Anaesthesia* 69 (4): 390–91. doi:10.1111/anae.12656.

Dove, E. S., B. M. Knoppers, and M. H. Zawati. 2014. 'Towards an Ethics Safe Harbor for Global Biomedical Research'. *Journal of Law and the Biosciences* 1 (1): 3–51. doi:10.1093/jlb/lst002.

Dove, Edward S, Susan E Kelly, Federica Lucivero, Mavis Machirori, Sandi Dheensa, and Barbara Prainsack. 2017. 'Beyond Individualism: Is There a Place for Relational Autonomy in Clinical Practice and Research?' *Clinical Ethics*, April, 1477750917704156. doi:10.1177/1477750917704156.

European Parliament. 1995. 'Data Protection Directive'.

Fairfield, Joshua, and Hannah Shtein. 2014. 'Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism'. *Journal of Mass Media Ethics* 29 (1): 38–51. doi:10.1080/08900523.2014.863126.

Fischer, T., K.B. Brothers, P. Erdmann, and M. Langanke. 2016. 'Clinical Decision-Making and Secondary Findings in Systems Medicine'. *BMC Medical Ethics* 17 (1). doi:10.1186/s12910-016-0113-5.

Floridi, Luciano. 2012. 'Big Data and Their Epistemological Challenge'. *Philosophy & Technology* 25 (4): 435–37. doi:10.1007/s13347-012-0093-4.

Gadamer, H.G. 2004. *Truth and Method*. Continuum International Publishing Group.

Garrison, Nanibaa' A., Nila A. Sathe, Armand H. Matheny Antommaria, Ingrid A. Holm, Saskia C. Sanderson, Maureen E. Smith, Melissa L. McPheeters, and Ellen W. Clayton. 2015. 'A Systematic Literature Review of Individuals' Perspectives on Broad Consent and Data Sharing in the United States'. *Genetics in Medicine*, November. doi:10.1038/gim.2015.138.

Gilligan, Carol. 1982. *In a Different Voice*. Harvard University Press.

Hansson, M G. 2009. 'Ethics and Biobanks'. *British Journal of Cancer* 100 (1): 8–12. doi:10.1038/sj.bjc.6604795.

Harris, John. 2005. 'Scientific Research Is a Moral Duty'. *Journal of Medical Ethics* 31 (4): 242–48. doi:10.1136/jme.2005.011973.

Hayden, Erika Check. 2012. *A Broken Contract*. NATURE PUBLISHING GROUP MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND. http://environmentportal.in/files/file/informed%20consent.pdf.

Heurix, Johannes, Stefan Fenz, Antonio Rella, and Thomas Neubauer. 2016. 'Recognition and Pseudonymisation of Medical Records for Secondary Use'. *Medical & Biological Engineering & Computing* 54 (2–3): 371–83. doi:10.1007/s11517-015-1322-7.

Hirschberg, Irene, Hannes Kahrass, and Daniel Strech. 2014. 'International Requirements for Consent in Biobank Research: Qualitative Review of Research Guidelines'. *Journal of Medical Genetics* 51 (12): 773–81. doi:10.1136/jmedgenet-2014-102692.

Hoffman, S. 2014. 'Citizen Science: The Law and Ethics of Public Access to Medical Big Data'. SSRN Scholarly Paper ID 2491054. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2491054.

Hoffman, S., and A. Podgurski. 2013. 'Big Bad Data: Law, Public Health, and Biomedical Databases'. *Journal of Law, Medicine and Ethics* 41 (SUPPL. 1): 56–60. doi:10.1111/jlme.12040.

Hofmann, B. 2009. 'Broadening Consent--and Diluting Ethics?' *Journal of Medical Ethics* 35 (2): 125–29. doi:10.1136/jme.2008.024851.

Hull, Sara Chandros, Richard R. Sharp, Jeffrey R. Botkin, Mark Brown, Mark Hughes, Jeremy Sugarman, Debra Schwinn, et al. 2008. 'Patients' Views on Identifiability of Samples and Informed Consent for Genetic Research'. *The American Journal of Bioethics : AJOB* 8 (10): 62–70. doi:10.1080/15265160802478404.

Ioannidis, John P. A. 2013. 'Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research'. *American Journal of Bioethics* 13 (4): 40–42. doi:10.1080/15265161.2013.768864.

Jennings, Bruce. 2016. 'Reconceptualizing Autonomy: A Relational Turn in Bioethics'. *Hastings Center Report* 46 (3): 11–16. doi:10.1002/hast.544.

Johnsson, Linus, and Stefan Eriksson. 2016. 'Autonomy Is a Right, Not a Feat: How Theoretical Misconceptions Have Muddled the Debate on Dynamic Consent to Biobank Research: Autonomy Is a Right, Not a Feat'. *Bioethics* 30 (7): 471–78. doi:10.1111/bioe.12254.

Joly, Yann, Edward S. Dove, Bartha M. Knoppers, Martin Bobrow, and Don Chalmers. 2012. 'Data Sharing in the Post-Genomic World: The Experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO)'. Edited by Philip E. Bourne. *PLoS Computational Biology* 8 (7): e1002549. doi:10.1371/journal.pcbi.1002549.

Kaplan, Bonnie. 2014. 'Patient Health Data Privacy'. *Yale University Institute for Social and Policy Studies Working Paper*, 14–028.

Kass, Nancy E. 2001. 'An Ethics Framework for Public Health'. *American Journal of Public Health* 91 (11): 1776–82. doi:10.2105/AJPH.91.11.1776.

Kaye, Jane. 2011. 'From Single Biobanks to International Networks: Developing e-Governance'. *Human Genetics* 130 (3): 377–82. doi:10.1007/s00439-011-1063-0.

Kaye, Jane, Liam Curren, Nick Anderson, Kelly Edwards, Stephanie M. Fullerton, Nadja Kanellopoulou, David Lund, et al. 2012. 'From Patients to Partners: Participant-Centric Initiatives in Biomedical Research'. *Nature Reviews Genetics* 13 (5): 371–76. doi:10.1038/nrg3218.

Kaye, Jane, Edgar A. Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. 2015. 'Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks'. *European Journal of Human Genetics* 23 (2): 141–46. doi:10.1038/ejhg.2014.71.

Kinkorová, Judita. 2016. 'Biobanks in the Era of Personalized Medicine: Objectives, Challenges, and Innovation'. *EPMA Journal* 7 (1): 4. doi:10.1186/s13167-016-0053-7.

Knoppers, Bartha M., Jennifer R. Harris, Anne M. Tassé, Isabelle Budin-Ljøsne, Jane Kaye, Mylène Deschênes, and Ma'n H. Zawati. 2011a. 'Towards a Data Sharing Code of Conduct for International Genomic Research'. *Genome Medicine* 3 (7): 46. doi:10.1186/gm262.

———. 2011b. ‘Towards a Data Sharing Code of Conduct for International Genomic Research’. *Genome Medicine* 3 (7): 46. doi:10.1186/gm262.

Knoppers, Bartha M., and Madelaine Saginur. 2005. ‘The Babel of Genetic Data Terminology’. *Nature Biotechnology* 23 (8): 925–27. doi:10.1038/nbt0805-925.

Koenig, Barbara A. 2014. ‘Have We Asked Too Much of Consent?’ *Hastings Center Report* 44 (4): 33–34. doi:10.1002/hast.329.

Larson, Eric B. 2013. ‘Building Trust in the Power of “Big Data” Research to Serve the Public Good’. *Jama-Journal of the American Medical Association* 309 (23): 2443–44. doi:10.1001/jama.2013.5914.

Laurie, G., P. Mallia, D. A. Frenkel, A. Krajewska, H. Moniz, S. Nordal, C. Pitz, and J. Sandor. 2010. ‘Managing Access to Biobanks: How Can We Reconcile Individual Privacy and Public Interests in Genetic Research?’ *Medical Law International* 10 (4): 315–37. doi:10.1177/096853321001000404.

Lewis, Cecil M., Alexandra Obregón-Tito, Raul Y. Tito, Morris W. Foster, and Paul G. Spicer. 2012. ‘The Human Microbiome Project: Lessons from Human Genomics’. *Trends in Microbiology* 20 (1): 1–4. doi:10.1016/j.tim.2011.10.004.

Linklaters. 2017. ‘Guide to the General Data Protection Regulation’. Accessed June 7. http://www.linklaters.com/Insights/Pages/General-Data-Protection-Regulation-survival-guide.aspx.

Liyanage, H., S. de Lusignan, S.-T. Liaw, C. E. Kuziemsky, F. Mold, P. Krause, D. Fleming, and S. Jones. 2014. ‘Big Data Usage Patterns in the Health Care Domain: A Use Case Driven Approach Applied to the Assessment of Vaccination Benefits and Risks. Contribution of the IMIA Primary Healthcare Working Group.’ *Yearbook of Medical Informatics* 9 (1): 27–35. doi:10.15265/IY-2014-0016.

Lupton, Deborah. 2014. ‘The Commodification of Patient Opinion: The Digital Patient Experience Economy in the Age of Big Data’. *Sociology of Health & Illness* 36 (6): 856–69. doi:10.1111/1467-9566.12109.

MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*. 3rd Revised edition. Gerald Duckworth & Co Ltd.

Majumder, M.A. 2005. ‘Cyberbanks and Other Virtual Research Repositories’. *Journal of Law, Medicine & Ethics* 33 (1): 31–39. doi:10.1111/j.1748-720X.2005.tb00208.x.

Markowetz, Alexander, Konrad Błaszkiewicz, Christian Montag, Christina Switala, and Thomas E. Schlaepfer. 2014. ‘Psycho-Informatics: Big Data Shaping Modern Psychometrics’. *Medical Hypotheses* 82 (4): 405–11. doi:10.1016/j.mehy.2013.11.030.

Master, Zubin, Lisa Campo-Engelstein, and Timothy Caulfield. 2014. ‘Scientists’ Perspectives on Consent in the Context of Biobanking Research’. *European Journal of Human Genetics*, July. doi:10.1038/ejhg.2014.143.

Mathaiyan, Jayanthi, Adithan Chandrasekaran, and Sanish Davis. 2013. ‘Ethics of Genomic Research’. *Perspectives in Clinical Research* 4 (1): 100. doi:10.4103/2229-3485.106405.

McDonald, Laura, Dimitra Lambrelli, Radek Wasiak, and Sreeram V. Ramagopalan. 2016. ‘Real-World Data in the United Kingdom: Opportunities and Challenges’. *BMC Medicine* 14: 97. doi:10.1186/s12916-016-0647-x.

McGuire, Amy L., Laura S. Achenbaum, Simon N. Whitney, Melody J. Slashinski, James Versalovic, Wendy A. Keitel, and Sheryl A. McCurdy. 2012. ‘Perspectives On Human Microbiome Research Ethics’. *Journal of Empirical Research on Human Research Ethics: An International Journal* 7 (3): 1–14. doi:10.1525/jer.2012.7.3.1.

McGuire, Amy L., James Colgrove, Simon N. Whitney, Christina M. Diaz, Daniel Bustillos, and James Versalovic. 2008. ‘Ethical, Legal, and Social Considerations in Conducting the

Human Microbiome Project'. *Genome Research* 18 (12): 1861–64. doi:10.1101/gr.081653.108.

Mello, Michelle M., Jeffrey K. Francer, Marc Wilenzick, Patricia Teden, Barbara E. Bierer, and Mark Barnes. 2013. 'Preparing for Responsible Sharing of Clinical Trial Data'. *New England Journal of Medicine* 369 (17): 1651–58. doi:10.1056/NEJMhle1309073.

Mertz, Marcel, Hannes Kahrass, and Daniel Strech. 2016. 'Current State of Ethics Literature Synthesis: A Systematic Review of Reviews'. *BMC Medicine* 14: 152. doi:10.1186/s12916-016-0688-1.

Meslin, Eric M., and Mildred K. Cho. 2010. 'Research Ethics in the Era of Personalized Medicine: Updating Science's Contract with Society'. *Public Health Genomics* 13 (6): 378–84. doi:10.1159/000319473.

Mittelstadt, Brent. 2017. 'From Individual to Group Privacy in Big Data Analytics'. *Philosophy & Technology*. doi:10.1007/s13347-017-0253-7.

Mittelstadt, Brent, Justus Benzler, Lukas Engelmann, Barbara Prainsack, and Effy Vayena. 2017. 'Is There a Duty to Participate in Digital Epidemiology?' *Life Sciences, Society and Policy*.

Mittelstadt, Brent, N. Ben Fairweather, Neil McBride, and Mark Shaw. 2011. 'Ethical Issues of Personal Health Monitoring: A Literature Review'. In *ETHICOMP 2011 Conference Proceedings*, 313–21. Sheffield, UK.

———. 2013. 'Privacy, Risk and Personal Health Monitoring'. In *ETHICOMP 2013 Conference Proceedings*, 340–51. Kolding, Denmark.

Mittelstadt, Brent, N. Ben Fairweather, Mark Shaw, and Neil McBride. 2014. 'The Ethical Implications of Personal Health Monitoring'. *International Journal of Technoethics* 5 (2): 37–60.

Mittelstadt, Brent, and Luciano Floridi. 2016a. 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts'. *Science and Engineering Ethics* 22 (2): 303–41. doi:10.1007/s11948-015-9652-2.

———. 2016b. 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts'. *Science and Engineering Ethics* 22 (2): 303–41. doi:10.1007/s11948-015-9652-2.

Moore, Philip, Fatos Xhafa, Leonard Barolli, and Andrew Thomas. 2013. 'Monitoring and Detection of Agitation in Dementia Towards Real-Time and Big-Data Solutions'. Edited by F. Xhafa, L. Barolli, D. Nace, S. Vinticinque, and A. Bui. *2013 Eighth International Conference on P2p, Parallel, Grid, Cloud and Internet Computing (3pgcic 2013)*, 128–35. doi:10.1109/3PGCIC.2013.26.

Mora, Fernando. 2012. 'The Demise of Google Health and the Future of Personal Health Records'. *International Journal of Healthcare Technology and Management* 13 (5): 363–377.

Mostert, Menno, Annelien L. Bredenoord, Monique C. I. H. Biesaart, and Johannes J. M. van Delden. 2016. 'Big Data in Medical Research and EU Data Protection Law: Challenges to the Consent or Anonymise Approach'. *European Journal of Human Genetics* 24 (7): 956–60. doi:10.1038/ejhg.2015.239.

Navarro, Robert. 2008. 'An Ethical Framework for Sharing Patient Data without Consent'. *Journal of Innovation in Health Informatics* 16 (4): 257–262.

Neff, Gina. 2013. 'Why Big Data Won't Cure Us'. *Big Data* 1 (3): 117–23.

NHS England. 2014. 'NHS England » The Care.Data Programme – Better Information Means Better Care'. http://www.england.nhs.uk/ourwork/tsd/care-data/.

Nicol, D., and C. Critchley. 2012. 'Benefit Sharing and Biobanking in Australia'. *Public Understanding of Science* 21 (5): 534–55. doi:10.1177/0963662511402425.

Niemeijer, A. R., B. J. Frederiks, I. I. Riphagen, J. Legemaate, J. A. Eefsting, and C. M. Hertogh. 2010. 'Ethical and Practical Concerns of Surveillance Technologies in Residential Care for

People with Dementia or Intellectual Disabilities: An Overview of the Literature'. *International Psychogeriatrics* 22: 1129–42.

Noddings, Nel. 2013. *Caring: A Relational Approach to Ethics and Moral Education*. Univ of California Press.

Nuffield Council on Bioethics. 2015. 'The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues'. Nuffield Council on Bioethics. http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf.

Ohm, Paul. 2010. 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization'. *UCLA Law Review* 57: 1701.

O'Neill, Onora. 2003. 'Some Limits of Informed Consent'. *Journal of Medical Ethics* 29 (1): 4–7.

Patterson, Michael E., and Daniel R. Williams. 2002. *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*. Vol. 9. Advances in Tourism Application Series. Champaign, IL: Sagamore Publishing, Inc. http://www.treesearch.fs.fed.us/pubs/29421.

Pellegrino, Edmund D, and David C Thomasma. 1993. *The Virtues in Medical Practice*. New York: Oxford University Press.

Prainsack, Barbara, and Alena Buyx. 2011. *Solidarity: Reflections on an Emerging Concept in Bioethics*. London: Nuffield Council on Bioethics.

———. 2013. 'A Solidarity-Based Approach to the Governance of Research Biobanks'. *Medical Law Review* 21 (1): 71–91. doi:10.1093/medlaw/fws040.

———. 2016. 'Thinking Ethical and Regulatory Frameworks in Medicine from the Perspective of Solidarity on Both Sides of the Atlantic'. *Theoretical Medicine and Bioethics* 37 (6): 489–501.

Pullman, Daryl, Holly Etchegary, Katherine Gallagher, Kathleen Hodgkinson, Montgomery Keough, David Morgan, and Catherine Street. 2012. 'Personal Privacy, Public Benefits, and Biobanks: A Conjoint Analysis of Policy Priorities and Public Perceptions'. *Genetics in Medicine* 14 (2): 229–35. doi:10.1038/gim.0b013e31822e578f.

Rothstein, Mark A., and Abigail B. Shoben. 2013. 'An Unbiased Response to the Open Peer Commentaries on "Does Consent Bias Research?"' *The American Journal of Bioethics* 13 (4): W1–4. doi:10.1080/15265161.2013.769824.

Safran, C., M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, and With input from the expert panel (see Appendix A). 2006. 'Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper'. *Journal of the American Medical Informatics Association* 14 (1): 1–9. doi:10.1197/jamia.M2273.

Schadt, Eric E. 2012. 'The Changing Privacy Landscape in the Era of Big Data'. *Molecular Systems Biology* 8 (September). doi:10.1038/msb.2012.47.

Schaefer, G. Owen, Ezekiel J. Emanuel, and Alan Wertheimer. 2009. 'The Obligation to Participate in Biomedical Research'. *Jama* 302 (1): 67–72.

Slote, Michael. 2007. *The Ethics of Care and Empathy*. New Ed edition. London ; New York: Routledge.

Smith, Maureen E., Saskia C. Sanderson, Kyle B. Brothers, Melanie F. Myers, Jennifer McCormick, Sharon Aufox, Martha J. Shrubsole, et al. 2016. 'Conducting a Large, Multi-Site Survey about Patients' Views on Broad Consent: Challenges and Solutions'. *BMC Medical Research Methodology* 16 (November). doi:10.1186/s12874-016-0263-7.

Stahl, Bernd Carsten, Job Timmermans, and Brent Mittelstadt. 2016. 'The Ethics of Computing: A Survey of the Computing-Oriented Literature'. *ACM Computing Surveys* 48 (4): 1–38. doi:10.1145/2871196.

Stein, Dorit T., and Sharon F. Terry. 2013. ‘Reforming Biobank Consent Policy: A Necessary Move Away from Broad Consent Toward Dynamic Consent’. *Genetic Testing and Molecular Biomarkers* 17 (12): 855–56. doi:10.1089/gtmb.2013.1550.

Steinsbekk, Kristin Solum, Lars Øystein Ursin, John-Arne Skolbekken, and Berge Solberg. 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): 151–62. doi:10.1007/s11019-011-9353-9.

Sterckx, Sigrid, Vojin Rakic, Julian Cockbain, and Pascal Borry. 2016. ‘“You Hoped We Would Sleep Walk into Accepting the Collection of Our Data”: Controversies Surrounding the UK Care.Data Scheme and Their Wider Relevance for Biomedical Research’. *Medicine, Health Care and Philosophy* 19 (2): 177–90. doi:10.1007/s11019-015-9661-6.

Tassé, Anne-Marie. 2016. ‘A Comparative Analysis of the Legal and Bioethical Frameworks Governing the Secondary Use of Data for Research Purposes’. *Biopreservation and Biobanking* 14 (3): 207–16. doi:10.1089/bio.2015.0121.

Tene, Omer, and Jules Polonetsky. 2013. ‘Big Data for All: Privacy and User Control in the Age of Analytics’. *Nw.J. Tech. & Intell. Prop.* http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11&section=20.

Terry, Nicolas. 2012. ‘Protecting Patient Privacy in the Age of Big Data’. *UMKC L. Rev.* 81: 385.

———. 2014. ‘Health Privacy Is Difficult but Not Impossible in a Post-Hipaa Data-Driven World’. *Chest* 146 (3): 835–40. doi:10.1378/chest.13-2909.

UK Biobank. 2015. ‘Incidental Findings Research’. http://www.ukbiobank.ac.uk/incidental-findings-research/.

UK Parliament. 1998. ‘Data Protection Act 1998’. Text. http://www.legislation.gov.uk/ukpga/1998/29/contents.

Watson, R. William G., Elaine W. Kay, and David Smith. 2010. ‘Integrating Biobanks: Addressing the Practical and Ethical Issues to Deliver a Valuable Tool for Cancer Research’. *Nature Reviews Cancer* 10 (9): 646–51. doi:10.1038/nrc2913.

Wellcome Trust. 2014. ‘Impact of the Draft European Data Protection Regulation and Proposed Amendments from the Rapporteur of the LIBE Committee on Scientific Research’. Wellcome Trust. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/WTP055584.pdf.

Williams, Hawys, Karen Spencer, Caroline Sanders, David Lund, Edgar A Whitley, Jane Kaye, and 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 Medical Informatics* 3 (1): e3. doi:10.2196/medinform.3525.

Woolley, J. Patrick. 2016. ‘How Data Are Transforming the Landscape of Biomedical Ethics: The Need for ELSI Metadata on Consent’. In *The Ethics of Biomedical Big Data*, edited by Brent Daniel Mittelstadt and Luciano Floridi, 171–97. Law, Governance and Technology Series 29. Springer International Publishing. doi:10.1007/978-3-319-33525-4_8.

ANNEXES

ANNEX I. ROADMAP Legal considerations and Data Protection Policy

Apart from ELSI recommendations made in the work of WP8, and in particular D8.2 and D8.5, the requirements for legal and ethical standards in the ROADMAP Project arise from two sources: legal requirements based on Member State data protection law and the forthcoming General Data Protection Regulation, and ethical guidelines provided by IMI in the Ethics Review Procedure Screening Report from 4 June 2016.

Concerning legal requirements, enabling access to RWE as pursued in ROADMAP has numerous legal implications. In addition to standard research ethics requirements and related legal rules, such as the need to obtain informed consent, acting as a gatekeeper to sensitive data (both personal and anonymized) creates legal obligations for a RWE platform as a data controller. Legal rules and regulations relating to the collecting, processing, linking and sharing of sensitive medical data need to be fully complied with in order to be on sound legal grounding.

At present, the legal requirements can be summarised as any data protection requirements set out within Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data) (Data Protection Directive). By the time ROADMAP will be set up, however, new rules will have come into effect. From 26 May 2018, the new General Data Protection Directive (GDPR) will become directly applicable in each member state and will lead to a greater degree of data protection harmonisation across EU countries. For ROADMAP, this means that any consideration of legal requirements should include consideration of changed and/or newly arising obligations from that Directive. It is therefore advisable to display foresight and consider these new – and stricter – rules to ensure data protection compliance within ROADMAP.

As a matter of due diligence, we contacted all ROADMAP partners to assess the data protection policies they currently have in place. In Annex 1, we provide an overview of applicable policies, which in some case are simple statements of compliance with national and European laws, as well as the appointed Data Protection Officer (DPO) where this is currently legally required. Under the new rules, most if not all partners will be required by law to have such a policy in place and to appoint a DPO.

The obligations the GDPR creates have been described as “onerous” (Allen & Overy 2017); although the same core rules as the Data Protection Directive are retained, there are some significant additional obligations introduced by the GDPR (Linklaters 2017). An important change concerns the inclusion of genetic and biometric data within sensitive personal data.

The extent to which the GDPR introduces new obligations partly depends on the current data protection laws in a particular EU member state. For instance, Germany already has numerous obligations imposed on data controllers (such as the duty to appoint a data protection officer), while this obligation will be entirely new in other member states. This is an important aspect to bear in mind in the context of ROADMAP: current levels of compliance with the GDPR will vary depending on the location of the partner.

For ROADMAP, the key focus with regard to the law is to ensure that all partners comply with the new rules, rather than the project in itself being subject to additional legal obligations. It might be prudential, however, for ROADMAP to request proof of individual data controllers of compliance with the new

GDPR rules, which should not pose an additional challenge as data controllers will already be placed under a legal obligation to demonstrate compliance with the GDPR principles ("Accountability"). To ensure compliance, it has been suggested that signing up to a Code of Practice or becoming certified might be sufficient for this purpose (Linklaters 2017) and this could be an option to be explored by partners.

Given the technical setup to be used for data access by ROADMAP (see: ROADMAP Description of Action), it would seem that the large majority of these requirements apply at the institutional level, i.e. the partner organisations which collect the data. Since only pseudonymised or anonymised data is fed into ROADMAP, it appears no additional obligations will arise from the GDPR. However, this initial opinion will need to be revisited if a) the process of data provision to ROADMAP is to change, or b) additional sources are to be added to ROADMAP. Legal requirements will also need to be assessed in ROADMAP Phase 2.

Concerning the IMI's Ethics Screening Procedure, one requirement imposed on ROADMAP was the provision of a Data Protection Policy (DPP) by the Consortium. DPP describes measures implemented which are undertaken in order to protect personal data during collection, storage and usage by institutions belonging to the Consortium. Even though all personal data shared as a consequence of the ROADMAP project will take place on a 1-to-1 basis and will be regulated by bilateral Data Sharing Agreements, all members of the Consortium can be viewed as prospective secondary users of the shared data in the project. In addition to that, all members might have access to personal data collected in the course of stakeholders' engagement activities and focus groups conducted within WP2 and WP8, respectively. In order to ensure that these potentially sensitive data are handled responsibly and fulfil the IMI requirements, ROADMAP created a portfolio of documentation on the DPPs and contact information of Data Protection Officers (DPOs) implemented at member institutions. Table 5 includes an overview of this portfolio. This portfolio will be reviewed at the next Ethics Advisory Board meeting in September 2017, prior to any newly or previously collected personal data being shared by consortium partners for research on ROADMAP.

Table 4. Description of processes assuring that ethical standards are met in the ROADMAP project (according to Description of Action)

Document name	From whom?	To whom?	When?	Special requirements		
Ethical Approval for all studies involving humans, issued by local Ethics Bodies	From Consortium Partners who collect/share data from studies involving humans within the ROADMAP project/Alba	IMI	prior to commencement of all studies involving humans	Will cover transfer of personal data generated in ROADMAP within the consortium, and to any third party	Will be managed by the project coordinator	Will be always accessible to EAB&IMI
Consent Form & Information Sheet	UOXF/WP8	All partners within the Consortium, who will collect data from patients	beginning of the project	will contain information on risks of participation	Will be managed by the project coordinator	Will be always accessible to EAB&IMI
Data Protection Policy of the Consortium	The Consortium	IMI	before any data sharing by third parties, before any data is collected from patients	Will follow best practices of EMIF's Ethical Code	Data Protection Officers of Consortium partners involved	Will be reviewed by the EAB (anonymisation measures)
Authorisations from data owners	From each data owner separately	To each receiving partner separately?	before any data is shared with the Consortium by any third party	Will be reviewed by the EAB		

Table 5. Details of the documentation on Data Protection provided by the ROADMAP partners

	Data Protection Officer	Data Protection Policy	Comments
Academic			
1. The Chancellor, Masters and Scholars of the University of Oxford (UOXF)	Giles Hamlin (giles.hamlin@ndphh.ox.ac.uk)	DPP Provided.	
3. Erasmus Universitair Medisch Centrum Rotterdam (EMC)	*pending*	*pending*	
4. Universiteit Maastricht (UM)	Bart van den Heuvel (Bart.vandenHeuvel@maastrichtuniversity.nl) (0031)(0)43 38 85526	Provided. Abstract of UM's 2013 Information Security Policy and measures.	Statement of compliance the relevant laws and regulations, in particular the Personal Data Protection Act (2001).
6. Fundació Institut Universitari per a la Recerca a l'Atenció Primària de	Leonardo Méndez Boo (lmendezboo@gencat.cat)	Provided DPP documentation (in Spanish language).	Contains statement of compliance with the requirements of the Organic Law 15/1999 on the Protection of Personal Data (Ley Orgánica 15/1999, Real Decreto 1720/2007).

Salut Jordi Gol i Gurina (IDIAP JORDI GOL)			
7. Københavns Universitet (UCPH)	Søren Brunak (soren.brunak@cpr.ku.dk)	ROADMAP Ethics and Data Integrity Document April 5, 2017 provided. General University level DPP not provided.	Refers to the internal ROADMAP processes rather than UCPH!
9. The University of Edinburgh (UEDIN)	Tracey Slaven (tracey.slaven@ed.ac.uk or recordsmanagement@ed.ac.uk)	Provided a DPP.	Statements compliance with The Data Protection Act 1998 (DPA)
10. Goeteborgs Universitet (UGOT)	*pending*	*pending*	
11. Aarhus Universitet (AU)	Lars A. Okkels (laok@clin.au.dk)	Provided a statement on compliance with The Danish act of processing personal data.	https://www.datatilsynet.dk/english/the-danish-data-protection-agency/introduction-to-the-danish-data-protection-agency/
12. London School of Economics and Political Science (LSE)	Jethro Perkins (J.A.Perkins@lse.ac.uk)	DPP provided.	Contains statement of compliance with The Data Protection Act (DPA).
15. Rijksuniversiteit	Raj Jagesar (r.r.jagesar@rug.nl)	Provided a ROADMAP-specific document about the data security. Provided links to University guidelines.	The UoG policy incorporates relevant guidelines and regulations from the VSNU Code of Conduct for Scientific

Groningen (RUG)		http://www.rug.nl/research/gelifes/data-management/ http://www.rug.nl/research/gelifes/research/data-management/rdmp-manual	Practice (2014), the CAO Nederlandse Universiteiten 2015-2016 and the VSNU SEP Guidelines & Requirements 2015-2021.
26. Academisch Ziekenhuis Leiden – Leids Universitair Centrum (LUMC)	*pending*	*pending*	**New partner**
EFPIA			
16. Novartis Pharma AG (Novartis)	Nicola Orlandi (nicola.orlandi@novartis.com)	Provided a DPP (global).	
17. Eli Lilly and Company Ltd (Eli Lilly)	*pending*	*pending*	
18. Biogen Idec Limited (BIOGEN)	Lee Parker (Lee.parker@biogen.com) 0041418820540	Provided a DPP.	
19. F. Hoffmann-La Roche Ltd (ROCHE)	Jason Hannon (jason.hannon@roche.com)	Provided a DPP. http://www.roche.com/dam/jcr:7e3782f2-7226-4b15-8064-ae2daab3b8fa/en/ch-ser-protectiondata.pdf	

20. Janssen Pharmaceutica NV (JPNV)	Henrik Olsson (holsson3@ITS.JNJ.com)	Provided Suplier Information Security Requirements, Clinical Research Policy, Data Safeguards Exhibits (all documents confidential!).	
21. GE Healthcare Ltd (GE)	Nestor Rivera (Nestor.Rivera@ge.com)	Provided Proprietary Information Classification Policy and Guidelines for Classification Handling.	
22. AC Immune SA (Ac Immune)	*pending*	*pending*	
23. Takeda Pharmaceuticals (TAKEDA)	*pending*	*pending*	**New partner**
24. H. Lundbeck A/S (HLU)	Lars-Peder Haahr (lph@lundbeck.com) 0045 3083 2534	Code of conduct provided, including a section on data privacy.	
Regulatory			
2. National Institute for Health and Care Excellence (NICE)	Julian Lewis (Julian.Lewis@nice.org.uk) 020 7045 2044	Provided a DPP.	Statement of compliance with The Data Protection Act 1998.
13. Agentschap College ter Beoordeling van	*pending*	*pending*	

Geneesmiddelen (CBG/MEB)			
Patient Organisations			
8. Alzheimer Europe (AE)	*No requirement for DPO*	Internal codes of conduct and guidelines for maintaining confidentiality.	Not currently required to have a DPO and DPP
SMEs			
14. IXICO Technologies Ltd (IXICO)	Kate McLeish (kate.mcleish@ixico.com)	Provided a Confidential Information Policy.	
5. Synapse Research Management Partners S.L (SYNAPSE)	Elena del Rey (edelrey@synapse-managers.com) 0034 93 3006061	Provided SOP on Data Protection and Data Storage.	Contains statement of compliance with the requirements of the Organic Law 15/1999 on the Protection of Personal Data (Ley Orgánica 15/1999, Real Decreto 1720/2007).