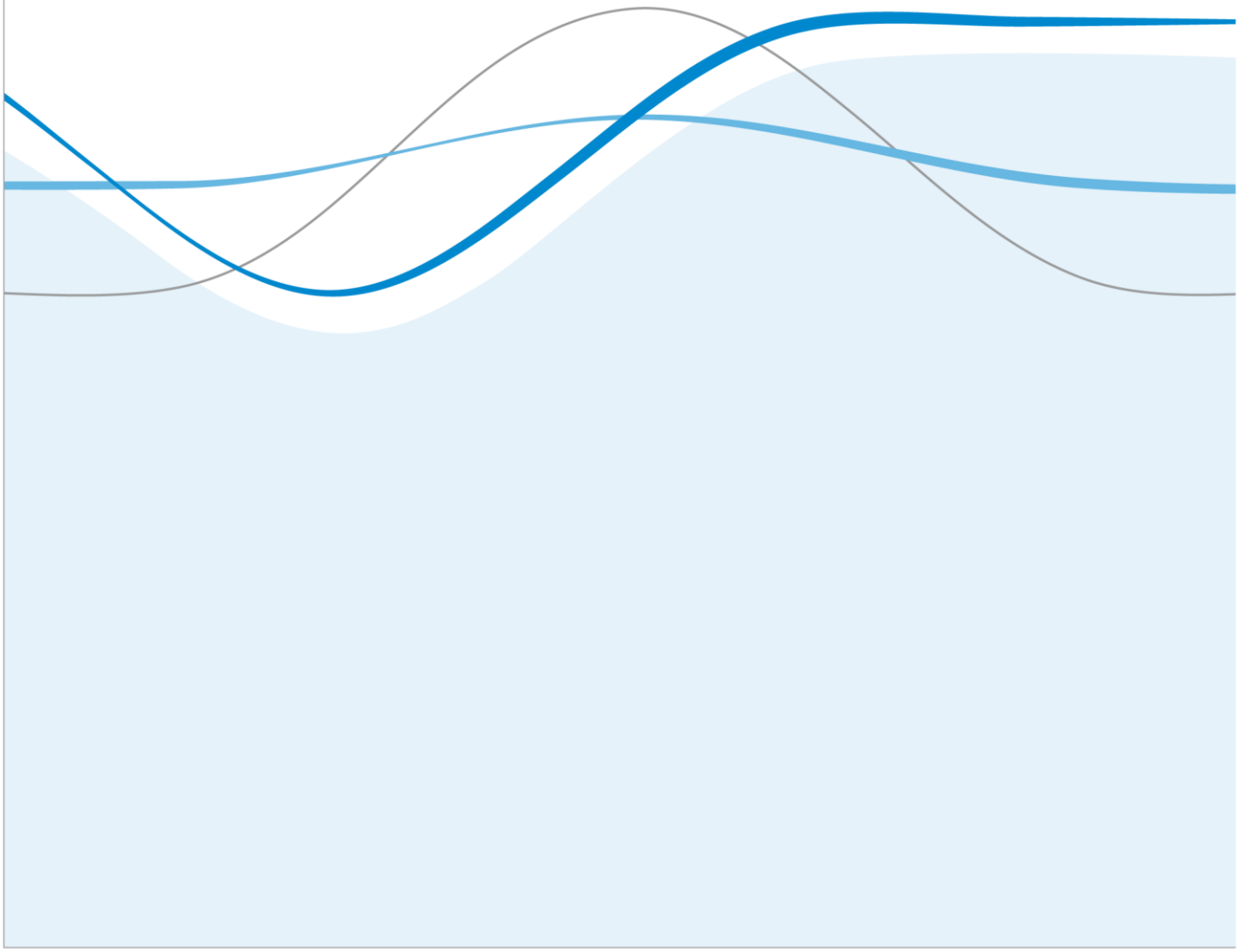
Office of Health Economics Consulting

Consulting Report

Data Governance Arrangements for Real-World Evidence

November 2015

Amanda Cole, Louis Garrison, Jorge Mestre-Ferrandiz, and Adrian Towse



Data Governance Arrangements for Real-World Evidence

Amanda Cole[a], Louis Garrison[b], Jorge Mestre-Ferrandiz[a] and Adrian Towse[a]

[a] Office of Health Economics, London [b] University of Washington

November 2015

Consulting Report prepared for Lilly

Submitted by:
OHE Consulting
Office of Health Economics
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom

For further information:
Jorge Mestre-Ferrandiz
Director of Consulting
Tel: +44(0)207 747 8860
Or: +44(0)7920 496 833
jmestre-ferrandiz@ohe.org

About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies OHE Consulting deems to be of interest to a wide audience, however, are made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in peer reviewed publications.

Studies published as OHE Consulting Reports are subject to internal quality assurance but do not go through the OHE Editorial Board peer review process. Publication is at the client's discretion.

Acknowledgements

We would like to thank Lilly for commissioning and funding this project and for their permission to publish the report. We would also like to thank Kai Yeung and Preeti Bajaj for their important contribution to this report, Carol Pratt for her feedback, to members of the Steering Group for their comments, and to our interviewees. We alone are responsible for any errors and omissions.

TABLE OF CONTENTS

Executive summary

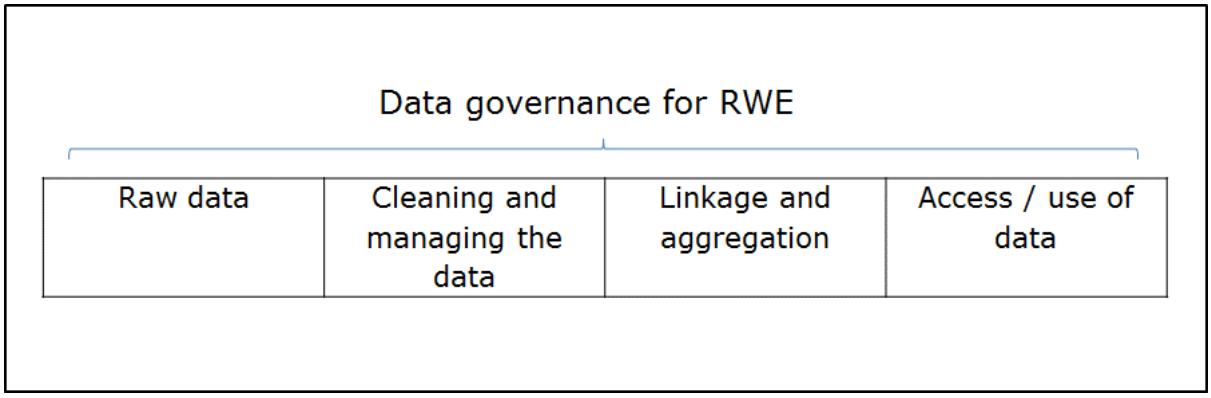
Introduction and Objectives

Lilly's objective in commissioning the research presented in this OHE Consulting Report was to understand and develop a view on the core principles that should govern how Real-World Data (RWD) is *accessed or generated*, and *used* credibly to produce or generate Real-World Evidence (RWE), thereby working toward a set of "international standards".

In this report, we firstly outline models of data governance in eight key markets of interest that have been selected by Lilly: the UK, France, Italy, Sweden, Germany, the Netherlands, Australia and the U.S. With the insights generated from these case studies, we then develop an illustrative framework of a top-performing governance model. By defining this model, we provide recommendations of what constitutes a good governance framework, which will support the favourable environment for the creation of RWE.

The process we use to illustrate the different steps in the data-to-evidence chain is illustrated in Figure ES1.

Figure ES1 Framework: Key elements of data governance for RWE



Recommendations for an ideal governance framework for RWD

National policies for the collection and use of health care data differ country to country, and often the legal framework is not completely prescriptive. We propose an aspirational governance framework that could guide the management of data access and use, as well as the processes that would facilitate constructive interactions among the relevant stakeholders, whilst maintaining accountability and public trust. By setting out the relevant stakeholders and their key roles in relation to the various elements of the governance framework, we illustrate the shared responsibilities between stakeholders and recognition of their shared values. Our recommendations fit under the dimensions "Raw data", "Cleaning and managing the data", "Linkage and aggregation" and "Access/use of data", as shown in Figure ES1.

"Raw data"

RWD can take various forms. Routinely collected data is that which is already collected for other purposes, such as electronic health records, and health care utilisation datasets (generally used for administrative or payment purposes). 'De novo' data describes the collection of further datasets (or further data fields in existing databases or registries) created for the purposes of a specific project. Both have the potential for data to be included with patient identifying information.

Data protection legislation outlines the lawful means by which personal data can be obtained and processed. Data ceases to be personal once it has been anonymised. Data which is anonymised but carries the risk of re-identification or de-anonymisation may still be treated as personal data.

There is often a line drawn between 'audit'/ 'service evaluation'/'quality assurance' activities on the one hand and 'research' on the other: the former generally does not invoke the need for ethical approval for data collection and/or use but the latter does. When ascertaining the differences between these two data collection purposes the following themes are generally referenced:

(1) Intent. Primary research aims to achieve generalisable results, whereas audit / service evaluation measures standards of care. In other words, research is to find out what you should be doing, whereas audit is to investigate planned activity.

(2) Clinical support. In audit / service quality evaluation, treatments have a firm basis of support in the clinical community.

(3) Allocation of treatment. Audit / evaluation does not involve allocation of treatment by protocol. If randomisation is used, it is research.

We found a line drawn in practice between the rules for processing personal data for audit purposes, versus those for research purposes, with the latter more strictly controlled and subject to more stringent data consent requirements. This is because it is generally considered that use for an audit exercise is 'in line' with the implied consent given by patients. The use of data for research, on the other hand, is a re-purposing exercise.

The ideal framework for "raw data" should include the following:

- ***Data protection legislation***: Clear data protection requirements that recognise the legitimacy of health care data utilisation beyond the direct care of patients.
- ***Data quality assurance****.* Requirements that records are accurate, and up-to-date. Patient identifiers which conform to national standards should be used and stored with the record.
- ***Patient consent****.* Where patient consent is not feasible, the collection of data for purposes beyond direct care can be supported with relevant legislation. Requirements that new legislation be passed for each new dataset poses prohibitive restraints on legitimate and worthwhile data collection activities. Greater flexibility can be administered through a legislative framework that grants statutory exemption for the requirement of consent where this would be too burdensome and where the purpose of the exemption is in the interests of the public. This should be decided after careful assessment by an ethical review board. This kind of regulation can be government-sanctioned but privately administered, it doesn't have to be administered by a government entity.
- Where data collection is to be collected on a routine basis across a large patient cohort, an ***opt-out, rather than an opt-in, system of patient consent*** may serve to maximise coverage and allow patients to contribute data more easily.
- **Patient information:** There must be clear communication to data subjects of potential future uses of their data.
- ***Approval of data collection activities to be based on intended use***. This relates to de novo data collection. The requirements for new data collection activities should be cognizant of the future intended use of the data. For example data collection activities that often form part of managed entry agreements (MEAs) or risk-sharing

arrangements between payers and manufacturers should be recognised as essential to the appropriate and optimal treatment of patients. Clear and transparent roles for the various actors in the collecting and eventual sharing of data should be well set out, which will enable access to data without harm or impact on privacy and public interest positions.

- ***Clear and transparent criteria****.* The criteria of Ethics Committees for data collection projects ('de novo' data) should be clear, transparent, and replicable. For national projects, there should ideally be a central ethical review board whose decision is accepted by the relevant national and local parties; this would reduce duplication of effort and promote consistent coverage.
- ***Data ownership****.* Responsibility (to be distinguished from 'ownership') for the data after collection passes to the data controller, who must act in the interest of patients and the public as specified by law.

"Cleaning and managing the data"

Data controllers are the organisations responsible for collecting, managing, and linking patient data. In order for the public to have trust in a system that collects and manages patient data, that system and those organisations that work within it must demonstrate strong and robust processes and meet quality criteria that give the public and data users confidence in the quality and security of the data held.

The ideal framework for "cleaning and managing the data" should include the following:

- ***Recognised data stewardship entities***. Data stewardship entities manage the acquisition, storage, aggregation, and de-identification of data. The interests of those entities must be aligned with those individuals whose data is being collected.
- ***De-identification of data****.* Where appropriate, data can be de-identified by removing any personally identifiable information and replacing the unique patient identifier (which in some countries is used across different sectors of the economy and therefore highly sensitive) with a pseudonym. Where data is not managed by one single entity, care should be taken that the algorithm for the pseudonymisation process is replicable for other datasets so that they may be linked, or else that the pseudonymisation process be reversible when desirable.
- ***Data quality***. Organisations processing patient data must ensure that the quality and integrity of the data is maintained.
- ***Security arrangements***. Security arrangements for the protection of confidential patient data should be assured through sound security processes, ranging from physical and technical computing protections and to the legal, security, and confidentiality training of staff involved in processing the data.
- ***How long data are kept***. In many countries, it is specified through data protection legislation that data should be kept 'no longer than necessary'. This is difficult to define, but the importance of rich longitudinal data that follows a patient over time through the care pathway and its benefits for research should be considered.

"Linkage and aggregation"

The ability to link data across datasets is incredibly important for research for which we need a common identifier. But therein lies the potential risk for re-identification (as we need a common link to match data sets). The ability for central linkage of datasets may

be impeded in countries where there are multiple data custodians each managing distinct datasets.

The ideal framework for "linkage and aggregation" should include the following:

- ***Develop a clear set of nationally agreed and implemented standard rules to optimize interoperability of health record systems***, for datasets to be compatible.
- ***Data linkage by trusted third party***. Common organisational and technical barriers to data linkage arise when there is no single group or organisation that has the responsibility or technical expertise required to manage the linking process. This could be minimised if linkage is undertaken by a single trusted third party. Whilst pseudonymisation helps to reduce the potential identifiability of data, there will always remain some residual risk of jig-saw re-identification. Therefore, it is still appropriate for requests for non-aggregated data to be examined by information governance panels (often through ethics committees) which consider the balance between the risk to patient confidentiality and the public interest in the research.

"Access / use of data"

Permissions and processes for access to health data vary substantially among countries. Use of data for research can indirectly benefit the population as a whole. For this reason, all countries have processes (some simpler than others) in place whereby the consent requirement can be waived in certain specific circumstances, including where collecting consent it not practicable.

The ideal framework for "access/use of data" should include the following:

***Forms of data access*.** Different access arrangements may be employed to achieve the needed balance between protection of private information and informing real-world research:

- An often used model involves the potential data user applying for access and following privacy review and contracting, from the data provider. In this scheme, the data provider may offer information at varying levels of detail and scrutiny:
    - When data is provided at the aggregate level and either (i), the data provider has capacity for in-house analyses which could be shared with the applicant, and/or (ii) data are provided at the level of individual patients but with most or all individually identifying elements removed, these situations involve minimal risk to privacy.
    - Data provided at the level of individual patients with most or all individually identifying elements intact, requires the highest level of scrutiny, as this level of information carries greater risk to privacy. However, this may be justifiable in some cases when investigators specifically need the patient identifiers to link the dataset to other data sources for research.
    - Data at the individual patient level could alternatively be provided to researchers in a physical space, which allows for direct control and monitoring of data use in cases where those data are highly sensitive.
- Another model which is able to allow access to individual patient data, data linkage across data providers, while protecting individual privacy, is the distributed network model. This could help to overcome the difficulties that can arise when there are multiple data custodians.

- ***Approval panels / ethical review.*** Ethical review boards (also called institutional review boards) grant access to health care data, so must be assured that the interest to society of the research project significantly outweighs the risk of violation of personal integrity of the individual that the processing may involve. A 'consent or anonymise' approach is too polarised and not a proportionate system. This risk of re-identification can be minimised with requirements for security procedures, training of staff that will process the data, and carefully written confidentiality agreements which assure correct use and reporting of data and which carry with it sanctions for inappropriate use. Approval panels should be composed of representatives with a broad range of relevant expertise and standpoints. The criteria used by committees to grant access to data should be clear, consistent, and transparent.
- The onus should be on data custodians to communicate how information is being shared and with whom in order to ensure public trust and transparency.
- ***Data use agreements and confidentiality requirements.*** Permission for data access should be granted with contractual requirements around the protection of confidentiality. The agreement should clearly define the scope and define duration of use.
- ***Affiliation of the data user.*** The type of organisation requesting access to data may influence the potential risk associated with its distribution (both realised and perceived). However, whilst the organisation's remit may influence their motivation for requesting access, this should not be the only consideration by data providers. Where the appropriate safeguards are in place, authorisation should be based on careful consideration of the motivation for and outputs of the research facilitated, rather than on the basis of the organisation's status. This is particularly important where manufacturers are tasked by HTA agencies or regulators with assessing the evidence for their products in routine practice.

Access costs. Arrangements for the cost of data access will vary according to the nature of the data controller. For many datasets collected and held on a national basis, data charges are based only to recover the costs of data extraction and cleaning. Cost of access should be fair and not excessive, but in recognition of the need for the sustainability of the system.

Conclusion

The evidence that is used to support decision-making in health care is becoming increasingly diverse, reflecting the increased complexity of the regulatory and reimbursement processes. Increasingly, the importance of understanding the impact of health care interventions in real-world settings is being recognised.

Appropriate and facilitative governance arrangements for RWE are imperative to facilitate evidence collection to meet the demands of regulators and HTA bodies, and to make the most of health care information and the role it can play in improving patient care. Problems arise due to the fact RWD is being used for purposes beyond those for which it was originally collected – to directly manage the care of the patient. As a result, legal frameworks are playing catch-up in order to accommodate these new secondary uses of data which clearly benefit patients and society but in a different way. With the general progressive move toward evidence-confirmatory pathways for the regulation and

HTA of medical products, legislation that permits the utilisation of RWD for activities, such as monitoring care quality and research to generate RWE, is becoming ever more important. This is evident through the increased reliance on and appetite for managed entry agreements, whose primary goals are usually one or more of: matching performance with payment, managing use, or to generate RWE. Research scientists and others, such as the companies tasked with providing the data as part of these arrangements, should be given every opportunity to support these goals. We have provided recommendations for an ideal governance framework that could lead to a more facilitative environment for the transformation of RWD into RWE.

1. Data governance arrangements for RWE

1.1. Introduction, objectives and context

Lilly's objective in commissioning the research presented in this OHE Consulting report wish to understand the core principles that should govern how Real-World Data (RWD) is *accessed or generated*, and *used* credibly to produce or generate Real-World Evidence (RWE), thereby working toward a set of "international standards".

In this report we firstly outline models of data governance in eight key markets of interest selected by Lilly: the UK, France, Italy, Sweden, Germany, the Netherlands, Australia and the U.S. With the insight generated from these case studies, we then develop an illustrative framework of a top-performing governance model. We then provide recommendations on what constitutes a good governance framework, and in particular one which will create a favourable environment for the generation of RWE.

Before outlining current approaches to governance, it is worth considering the central aim of governance arrangements for the collection and use of data in health care, and by doing so highlight the process by which RWD (the raw data) is transformed into RWE (the insight); the framework to support this flow of data to evidence will be developed in section 12.

RWD and RWE

The evidence that is used to support decision-making in health care is becoming increasingly diverse, in recognition of the need to understand the impact of health care interventions in the real-world. Decision-makers are interested more and more in the relative or comparative *effectiveness* of new treatments, which can sometimes deviate from the relative or comparative *efficacy* results obtained through randomised controlled trials (RCTs). Evidence of effectiveness collected in real-world settings is by its nature more generalisable, though this higher external validity must be balanced with likely less internal validity and potential biases (Luce et al., 2010). Another factor leading to our increased reliance on the collection of data in real-world settings is a gradual shift in the timing of drug appraisals, which are being conducted progressively earlier in the lifecycle of a product. Notably, models of earlier access to treatments, adaptive licensing, and managed entry agreements (MEAs) in general employ a system of provisional approval which is subject to re-assessment at a later time, at which point data collected alongside use of a product can help inform subsequent decisions.

RWD can be in various forms – but two of its key characteristics are that it is collected outside a clinical trial and is used for health care decision making (Garrison et al., 2007). Broadly, it could consist of either data that are already routinely collected in a health care system (electronic medical/health care records, administrative reimbursement databases, pharmacy data used to fill prescriptions, etc.) or data that is collected specifically for the purposes of a project (e.g. new patient registries for a disease or clinical procedure or pragmatic clinical trials)[1]. Other uses of RWD can include achieving appropriate levels of access and reimbursement, improving safety surveillance and risk

[1] While important to consider other sources of healthcare data, such as digital devices and social media, these are outside the scope of our remit.

management, supporting better outcome measurement, and informing drug development decisions throughout the product lifecycle. These represent important and increasing *applications* of RWD. In order for this data to be used credibly to inform clinical practice, an appropriate governance framework is needed for its access and use.

Governance

Governance has been described as covering: *"…the processes, roles, standards and metrics that ensure the effective and efficient use of data and information in enabling an organisation to achieve its goals"* (Gartner, 2014). The ultimate goal of the information that is collected around health care is to provide evidence for health care interventions and thereby to influence clinical practice and the treatment of patients. The process we use to illustrate the different steps in the data-to-evidence chain is illustrated in Figure 1.

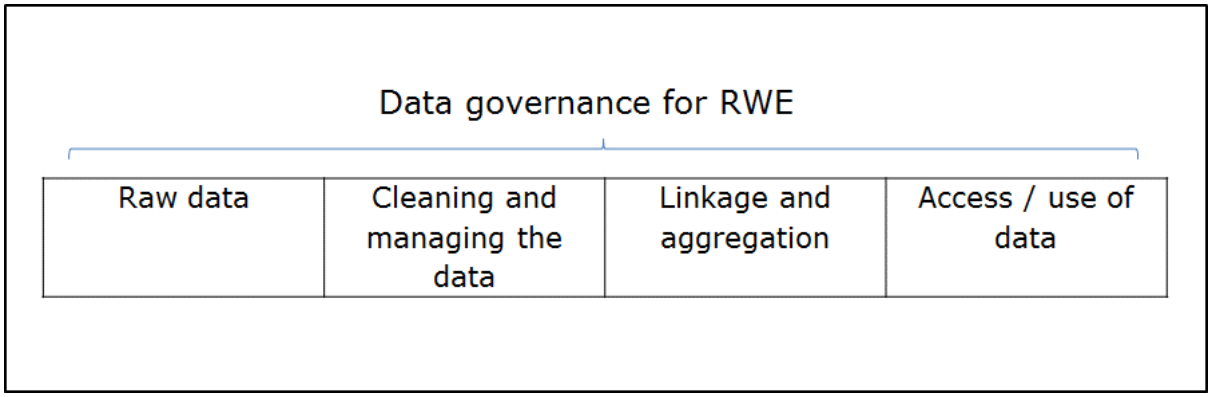


Figure 1 Key elements of data governance for RWE

1.2. Current state of data governance for RWE

This report has three stages: [stage 1] evaluate existing data governance models/frameworks in the eight selected countries; [stage 2] identify strengths and weaknesses with a view to outline an ideal governance framework; and [stage 3] develop policy recommendations that would support the 'ideal' of a favourable RWE governance environment.

Data extraction for each country followed a systematic approach by adhering to a structured pro-forma, thereby ensuring completeness of information for each country. We sought expert input advice from Lilly colleagues to finalise this pro-forma. We provide two 'deep dive' case studies – the UK and the U.S. – which demonstrate in detail the systems in place that govern data collection and data access. This is followed by the six further case studies, which also outline the governance models in place, but provide less detail of the specific agencies involved and datasets available. See Appendix (1) for the pro-formas utilised. Governance arrangements are separated according to their relevance for: (i) routinely collected data and (ii) de novo studies collecting new patient data (whilst accepting that for some countries the processes / governing bodies will overlap). We also describe the capacity, rules, and governance arrangements for linking datasets, accessing data, and using that data, as well as any changes (recent or imminent) to the data governance environment.

1.3. RWE governance - International perspective

All stakeholders involved in the development, regulation, assessment, provision, and use of health care have a vested interest in how the health care market will respond to the increased need for and availability of RWD. National responses to calls for more and better data, whilst maintaining ethical standards for its processing, are varied. This presents difficulties for organisations with an international perspective who must work within these heterogeneous frameworks which vary from country to country.

Governance arrangements for health care data, the national approaches to which we summarise in this section, include the core principles and legislation in place around how patient data can be generated and accessed, including rules around: transparency, patient confidentiality, ethics, and how data can be utilised credibly. According to a study by the OECD on health system performance, country variation is linked primarily with risk management in: granting exemption to patient consent, sharing identifiable data, and in granting access to data (Oderkirk et al., 2013). The authors highlight that actions are needed to address the current heterogeneity in data protection practices. By outlining the governance arrangements in eight key markets below, we will characterise this heterogeneity and in so doing set the scene for the best model framework and recommendations to follow.

2. The EU: unifying framework for data protection

An important overarching framework to consider for European countries before assessing individual Member States' national frameworks is the EU legislation for the protection of patient data. This data protection framework is essential for projects that involve multiple European countries, as well as affecting the national legislative arrangements for the protection of patient data. Below, we describe a selection of multi-national data initiatives in Europe, followed by a description of the EU framework for data protection and how this may change in the near future.

European data projects

The IMI GetReal project represents an important current EU initiative which focuses on methods of RWE collection and synthesis, particularly for the earlier adoption of pharmaceuticals; this will be particularly pertinent for part two of this report: RWE for managed entry agreements. There are many examples of specific cross-border EU health data projects, many of which have been pilot projects. One such project was the EU adverse drug reactions (EU-ADR) project (the final report for which was published in 2012). The aim was to supplement the current spontaneous reporting system, by making use of clinical data from 30 million patients' electronic health records from The Netherlands, Denmark, the UK and Italy (EU-ADR, 2015). By integrative mining of clinical records and biomedical knowledge, 'signals' are detected by linking combinations of drugs and suspected adverse events that warrant further investigation. The initiative was funded by the European Commission, and demonstrates the possibilities of joint working between countries on real world data projects where the unified data protection framework allows it. Another project, which ended in June 2014, was the epSOS project (European Patients Smart Open Services) which combined e-health records among many Member States epsos (epSOS, 2015). The project was undertaken in compliance with the EU regulatory framework and did not require changes to national legislation, though was conceived as a pilot project.

The Vaccine Adverse Event Surveillance and Communication (VAESCO) project provided a research infrastructure for vaccine safety data across eight European countries: Italy, the UK, Norway, Denmark, the Netherlands, Sweden, Finland and Spain, covering nearly 27 million patients. In a 'Lessons learnt' exercise conducted at the end of the project, the project team highlighted the critical limiting factor for data linkage studies to be the data protection laws and regulations in each country. In particular, they highlight the distinction for projects between public health functions (e.g. surveillance and response) versus research, with protections being more stringent for projects classified as research (Destefano and Vellozzi, 2012). This finding is supported by the findings in this report for each individual country.

Another important EU data platform is the European Medical Information Framework (EMIF), which is supported by EFPIA, FP7 and IMI, whose aim is to provide an environment which allows the efficient re-use of existing health data, in particular for Alzheimer's disease and metabolic diseases (EMIF, 2015). A similar platform is the Electronic Health Records for Clinical Research (ehr4cr), also supported by IMI, which is a 5 years project ending in 2015. A governance framework for that platform is being worked on and will be published soon. For cross-border data collection initiatives, it is clear that a cohesive legislative pathway would be beneficial. Below we describe the EU framework for data protection and how this might be changing.

European data protection

The right to data protection is currently underwritten by Article 8 of the EU Charter of fundamental rights, according to which data must be processed fairly, for specified purposes, and on the basis of either patient consent or some other legitimate basis laid down by national law. In addition, everyone has the right to access the data collected on them, and the right to have it rectified if incorrect (Working Party, 2012).The current *EU Data Protection Directive* (Directive 95/46/EC) provides a unifying framework for national policies around data protection, which extend to the protection of health care data. The Directive has been in place since 1995. However it is widely acknowledged that the legislation leaves considerable room for interpretation, with individual countries implementing national data protection policies which differ with respect to how confidentiality is maintained and how sensitive data is processed and managed. A report by the OECD outlines some of the commonalities and differences between EU national policies (Oderkirk et al., 2013). Some countries have national data protection offices to grant the use of data which has been collected without patient consent, but this is difficult without authorising legislation (Belgium, Italy, Cyprus). According to experts in Germany, data is only shared where patient consent has been obtained, or when authorised by a law or regulation. The authors of the OECD report found a similar situation in Portugal, where data linkage is illegal in the absence of authorising legislation, and in Poland where national data linkages were illegal. Policies in France, Sweden, Denmark, Finland and the UK were noted to be more permissive, with data protection legislation outlining a framework for sharing identifiable data without consent, where data custodians or national approval bodies make such decisions on a case-by-case basis, trading off the risk to privacy and the benefits of research that is in the public interest (Oderkirk et al., 2013).

Given the room for interpretation in the current directive, the European Commission is currently looking to revise the current arrangements, with the objective of harmonising data protection and privacy across the EU, as well as to respond to the changing technological environment. The matter first arose through a 'Communication' on a *"comprehensive approach on personal data protection in the European Union"* (European Commission, 2010). The Communication acknowledged that the right to privacy is not an absolute right, by suggesting that data protection and privacy should not unnecessarily limit other fundamental rights, such as the right to health and health care. The Communication also acknowledged the divergent implementation of the 1995 Data Protection Directive, and resultant lack of harmonization amongst Member States' legislation on the various aspects of data protection.

Two years after the 2010 'Communication', the EU Commission enacted a *"Proposal for regulation of the European Parliament and of the Council on protection of individuals with regard to the processing of personal data and on the free movements of such data"* (European Commission, 2012), where a legislative proposal for a 'General Data Protection Regulation' (DPR) was put forward. An important clarification that the proposal makes is to define the 'public interests' that serve to justify exemptions to the general prohibition of processing data. Public health and scientific research interests are recognised to serve the good of society, thus requiring protection in order to guarantee other fundamental rights, including the right to health care (Di Iorio et al., 2014). This recognition is an important step forward. However, the European Parliament has

produced a draft report which proposes several amendments to the European Commission's General Data Protection Regulation (European Parliament, 2012). In a paper by Di Iorio and colleagues (2014), the authors propose that, if the amendments stand as written, *"the right to privacy is likely to override the right to health and health care in Europe"* (Di Iorio et al., 2014).

Most significantly, the Draft Report eliminates from the text proposed by the Commission which refers to the possibility to deviate from the general prohibition of processing sensitive data for scientific research or statistical purposes, when this is done by a law with the balance of public interest in mind. These public interest grounds are currently the basis upon which most EU countries that have permissive data governance allow for the processing of data to facilitate research and evaluation activities. Therefore, this change would have a profound impact on information governance and management processes for individual Member States.

Whereas the proposal would have allowed the processing of personal data concerning health "which is necessary for historical, statistical or scientific research purposes" the amendment adds that the processing for these purposes "*shall be permitted only with the consent of the data subject*" (European Parliament, 2012). This would lead to a situation where any patient data that is potentially identifiable cannot be accessed or linked without explicit consent of individual patients; this is in contrast to the justification provided for processing sensitive data for "management of health care services", for which grounds are provided. Whilst the amendments allow Member States to provide exemptions for research which is of "exceptionally high public interest", these are not defined. There is considerable concern that this would have a strong negative impact on medical research (NHS European Office, 2014). This impact is summarised by Di Iorio and colleagues: *"…the amendments would make difficult or render impossible research and statistics involving the linkage and analysis of the wealth of data from clinical, administrative, insurance and survey sources, which has contributed to improving health outcomes, to reducing unsafe practices and to improving health systems performance and governance"* (Di Iorio et al., 2014, p.490).

The risk to current national governance arrangements for the collection and processing of RWD that the new regulation would engender is clear. Moreover, by implementing a Regulation to replace the current Directive would leave less flexibility for countries to implement legislation to suit their own contexts (potentially both a benefit and a disadvantage). Despite the fact that the Regulation was first officially proposed by the European Parliament in 2012, timelines are still unclear and it has and continues to undergo amendments; it has been predicted that the final text may not be confirmed until Spring 2016 (Hunton & Williams, 2015).

3. The United Kingdom

1. **Brief overview** of the health system and collection / management of patient data

Health care in the UK is provided predominantly on a public basis through the National Health Service (NHS), which accounts for around 83% of total health care spending (Payne.C, 2013). This means that there is a large volume of data in health care records collected and maintained by the NHS, and the opportunity to link information sets across NHS services is high. A useful summary and comparison of the health care systems across the four countries of the UK is provided by The King's Fund (Ham et al., 2013). Whilst the primary function of health care records is to maintain and share information on a patient's medical history with health care professionals involved in their care, an important secondary function is for data to be used for research and to monitor and improve services. It is this secondary function to which much of the legislation and governance arrangements are directed, details of which are summarised below.

2. Routinely collected patient data

Routinely collected patient data in the UK is extensive, and encompasses records of health care activity and interactions with the NHS in both primary and secondary care, as well as through disease-specific national registries and audits.

a. **Core legislation governing the collection / use of routinely collected patient data.** Key documentation outlining principles of governance and data protection.

The legal framework for using personal confidential health data in the UK is complex, and includes various legislative frameworks, including: the Data Protection Act 1998, the Human Rights Act 1998, the Health and Social Care Act 2012, and the NHS Act 2006.

As patient consent is not obtained for patient data that is routinely collected, the governing principles that surround its management and use are very important. The key piece of legislation that underlies the governance arrangements for the obtaining, holding and using or disclosing this data is the Data Protection Act 1998 (DPA) (Government, 1998). The DPA was developed in response to the EU directive (Directive 95/46/EC) which came into place in 1995, to which member states were required to respond with legislation by October 1998.

Under the DPA, organisations that handle and store information that is identifiable to patients are legally obliged to adhere to 8 data protection principles, which cover:

- only collecting information that is needed for a specific purpose,
- keeping the information secure,
- ensuring the data is relevant and up to date,
- only holding as much as is needed and only for as long as is needed, and
- allowing subjects of the information to see it on request (ICO, 2014).

In addition, The Human Rights Act (1998) underpins the right for patients to keep their health records confidential (NHS Choices, 2013).

It is under these guiding principles that information centres in the UK which collect and hold information are governed. The Health and Social Care Information Centre (HSCIC)

in England, which is the national provider of information, data and IT systems for health and social care, outlines its information governance and standards to which it is held on its website (HSCIC, 2014g).

Another key piece of legislation is the Health and Social Care Act 2012 in England, which for example re-established the HSCIC (as well as NICE) in primary legislation, the intention of which was to give the organisation greater autonomy and clearer powers to make information more open and transparent (Government, 2012). This was intended to make the HSCIC the focal point for information collection in England to improve quality and minimise information burdens. In addition, the Act gave HSCIC power to advise organisations on how to handle confidential information securely, which is summarised in a report 'Guide to confidentiality in health and social care' (HSCIC, 2013a). According to the HSCIC the legal obligations under which patient information is held go over and above those specified in the DPA (HSCIC, 2014f).

The primary basis which underpins the lawful processing of confidential information used for secondary purposes is that an organisation must have either:

- Obtained informed consent from the data subject (i.e. the patient), or
- Been granted a statutory basis for no consent.

Statutory exemption is generally through Section 251 of the NHS Act 2006 (Government, 2006) (this was a re-enactment of section 60 of the Health and Social care Act 2001). Section 251 *'allows the Secretary of State for Health to make regulations to set aside the common law of duty of confidentiality for defined medical purposes'* (NHS, 2014). To attain section 251 support and access confidential (patient identifiable) information, the purposes of the information recipient must be related to improving patient care, and must be in the public's interest, and will only be granted where it is either not possible or too expensive / technically difficult to get consent from every patient; applications are considered by the Confidentiality Advisory Group (CAG) (HSCIC, 2014k). In 2011 'The NHS Care Record Guarantee' – a document produced by the National Information Governance Board for patients – this provision is described in the following terms: *'We will not share health information that identifies you for any reason other than providing your care, unless […] we have special permission because the public good is thought to be of greater importance than your confidentiality'* (NIGB, 2011).

If neither informed patient consent has been obtained nor section 251 support granted, the transfer of secondary data must be anonymised. Anonymised data is defined by the Medical Research Council (MRC) as *'data prepared from personal information, but from which the person cannot be identified by the recipient of the information. The term is used […] when referring to robustly pseudonymised / linked anonymised data or unlinked anonymised data'* (MRC, 2014).

An important report that is commonly cited is the Caldicott report, which was published in 1997 and provided guidance on the use of *patient identifiable information*, commissioned by the Chief Medical Officer of England in response to increasing concern about how information was used in NHS England and Wales (Department of Health, 1997). In order to set out an updated view on information governance, with a view to ensure there is an '*appropriate balance between the protection of patient information and the use and sharing of information to improve patient care'*, Dame Fiona Caldicott was asked to lead a review of that report which was conducted by the Independent Information Governance Oversight Panel (IIGOP), entitled *The Information Governance*

Review (or 'Caldicott2') (IIGOP, 2013). Of particular importance, the report outlines a third 'grey area' which sits between (a) completely anonymised data and (b) non-anonymised data for which there is a legal basis for processing. This grey area includes data that has been de-identified by use of pseudonyms / coded references, but could potentially be re-identified when combined with other data. Linked data which may have this property is of particular interest to researchers and possibly to companies interested in longitudinal patient data; accredited safe havens (e.g. HSCIC) are tasked with managing the risks and benefits associated with dissemination of data of this sort. HSCIC have developed a new anonymisation standard is response to this (ISB, 2013). On 13th November 2014 Dame Fiona Caldicott was appointed the (new) role of 'National Data Guardian'. The datasets available in the UK as well as the organisations involved in their upkeep and access are outlined below.

b. Datasets. Overview of what data is collected, and from what parts of the health service

In a study of health information infrastructure across OECD countries, the UK was noted to have the most comprehensive suite of national datasets (OECD, 2013). However coverage is generally limited to an individual UK member country, and arrangements for their access are also organised at a national (or sub-national) level. There has been a call to make the health care data collected across the four UK countries more systematic and comparable, thereby facilitating comparative analyses of performance and outcomes across the countries (Bevan et al., 2014).

As noted, there is a large focus in the UK on routinely collected patient data, which covers primary care services, secondary care, social care, prescribing, patient experience, public health, as well as numerous national audits for specific disease areas. The data sets are too numerous to list here – please see the website of the information centres listed below for a comprehensive list of the data managed by the respective centres. These include mandatory routine submission of data which summarise a patient's interactions with primary care and hospital episodes.

c. Information providers. Who 'holds' the data (likely to be a mix of public and private organisations), what data do they hold, how is it collected, and what are the core governing principles in handling the data?

As the players involved in the management of health care data are different across the four countries of the UK, the main (national-level) information providers for each country are summarised below. In each of the four countries there is a strong emphasis placed by the organisations involved on the principles of information governance and protecting sensitive information.

England

In England, The *Health and Social Care Information Centre (HSCIC)* is the national provider of information and data for health and social care, and is an executive non-departmental public body sponsored by the Department of Health. A list of the data sets and services that the HSCIC can provide has been published on their website (HSCIC, 2014c). A very important resource held by the HSCIC is the Hospital Episode Statistics (HES) dataset, which contains over 1 billion records of patient attending Accident and Emergency units, outpatient clinics and NHS hospitals in England. HES is a records-based system which was designed for secondary use, disclosure of which is managed by

HSCIC. HES data is comprehensive, as hospital payments are based on the information submitted to HES. The core governing principles under which the HSCIC operate were outlined above. As described, the Health and Social Care Act 2012 gave HSCIC further powers to provide guidance on information governance in the UK.

HSCIC is working towards a national 'Diabetes data set', which will bring together and combine four national diabetes work streams: National Diabetes Audit, the diabetes chapter of the GMS QoF, the diabetes e-performance management tool, and the diabetes indicators for the better metrics performance indicator project (HSCIC, 2014e). In addition, Office of National Statistics (ONS) cancer data is available through the HSCIC; approval for granting access to that data lies with the ONS.

The Clinical Practice Research Datalink (***CPRD***) is the English observational and interventional research service which is jointly funded by the National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA). CPRD collects and hosts primary care data from General Practices (GPs) in the UK, which covers patient registration information and all care events that GPs record as part of usual medical practice (including diagnoses, referrals, prescriptions, lifestyle information, etc.) (CPRD, 2014d). There are four main providers of software for GP practices in the UK: EMIS, Vision, TPP and SystmOne. At present, CPRD collects data from Vision practices only, though there is an initiative underway to collect from EMIS practices which would expand coverage[2].

The data held by CPRD is all 'anonymised' by the HSCIC before being received by CPRD. However, in recognition that data of this sort could potentially identify individual patients, this anonymisation is not relied upon by CPRD on its own, and the organisation also operate under specific Charters to protect confidentiality. Privacy Enhancing Technologies are used, which include security systems, passphrases, and computer certificates which mean that CPRD is aware of where and by whom the data is being used (CPRD, 2014b). In addition, CPRD enters legal agreements with all individuals working with the data, has Standard Operating Procedures (SOPs) which mandate how each stage of data release is undertaken, and has the right of audit for any researcher or organisation using CPRD data to ensure compliance. An independent panel – Independent Scientific Advisory Committee (ISAC) – reviews data requests.

A major private organisation working in this space is IMS Health, who in England have filled a gap that exists across the UK: unlike primary care, there is no central NHS collation of information on medicines issued and used in NHS hospitals. IMS Health collect data on the use of pharmaceuticals, at an aggregate level only for primary care but at a patient level for secondary care usage through their 'Hospital Pharmacy Audit' Index (HPA/HPAI). This allows data to be linked with other health care data (details below), for which section 251 exemption has been attained by IMS Health to link with HSCIC HES data. HPA data is said to cover 99% of all drugs prescribed in Hospitals, but known limitations include drugs dispensed from a ward trolley and drugs prescribed in hospital but dispensed in the community (HSCIC, 2013b). The specific governance framework and legal obligations of IMS Health in collecting and managing NHS pharmacy data are not set out on their website. However the Information Commissioner's Office, from whom organisations working with health data must receive approval, outlines a code of practice for all organisations that handle patient data (ICO, 2012).

[2] Information obtained from conversation with a research scientist at CPRD, October 2014

Another important source of real world data for oncology in England is data collected through the National Cancer Intelligence Network (NCIN). NCIN, which is operated by Public Health England, is a unified cancer registration service for England, which aims to provide a near-real time comprehensive data collection and quality assurance system over the entire cancer pathway for all patients treated in England. The data collected (which is submitted to the National Cancer Registration Service [NCRS]) is comprehensive, with the cancer outcomes and services dataset (COSD) collecting data across diagnoses, demographics, referral, staging, imaging, treatment, surgery, recurrence, etc. (NCIN, 2014a). The NCRS collects and integrates information collected through COSD with related datasets such as cancer waiting times, national clinical audits, ONS data, radiotherapy dataset, as well as the Systemic Anti-Cancer Therapy Dataset (SACT). The SACT dataset collects data on all drug treatments with an anti-cancer effect, including traditional chemotherapy as well as newer agents (NCIN, 2014b); submission of data to SACT has been mandated in England since May 1st 2014. Cancer registries receive identifiable data under a regulation made under section 60 of the Health and Social Care Act 2001 with continuing effect under Section 251 of the NHS Act 2006, indicating that the perceived potential benefits from the assessment of this data outweigh the associated risks.

Wales

In Wales, a central access point for routinely-collected national health data is the Secure Anonymised Information Linkage databank (SAIL). SAIL is managed by the Health Information Research Unit (HIRU) at the Swansea University School of Medicine, and describes itself as a 'research resource focused on improving health, well-being and services'. SAIL receives core funding from the Welsh Assembly Government in their commitment to the UK Clinical Research Collaboration, through the National Institute of Social Care and Health Research (NISCHR) (Ford et al., 2009). As well as linking health data, SAIL is able to link other population-based data such as school outcomes and demographic data. Data providers to SAIL include: NHS Wales (Emergency department dataset, National Community Child Health Database, Outpatient Data set, Patient Episode Database for Wales (PEDW), Welsh Demographic Service), Public Health Wales (Bowel Screening Wales, Breast Test Wales, Cervical Screening Wales), Office for National Statistics (ONS – birth and death extract), Welsh Cancer Intelligence and Surveillance Unit (WCISU) and individual, Congenital Anomaly Register & Information Service, and GP practices (SAIL, 2014c). The coverage of GP practice data held by SAIL is 70%[3]; diagnoses and drug prescriptions are collected through these records. Hospital data is collected from the Patient Episode Database (PED – equivalent to HES in England), but hospital prescriptions are still collected mainly in paper format, and therefore not routinely captured. The information governance standards to which SAIL is held are summarised on their website (SAIL, 2014b) and in the Data Management Policy document (SAIL, 2013).

Scotland

In Scotland the Information Services Division (ISD) collects a wide range of health-related administrative data on behalf of NHS National Services Scotland. The electronic Data Research and Innovation Service (eDRIS) provides a single entry point for research approvals and data access to Scottish health data that is routinely collected. The data accessible through eDRIS can be viewed on the ISD website (ISD, 2014b). Until March

[3] Information obtained from research analyst at SAIL, October 2014.

2014 The Scottish Health Informatics Programme (SHIP) – a work stream undertaken by a consortium of Scottish Universities supported by a grant from the Wellcome Trust among others – offered a Scotland-wide research platform for the collation, management, dissemination and analysis of electronic patient records. The group provided useful guidance and commentary on information governance for health data in Scotland, and recommendations for appropriate governance frameworks (Laurie and Sethi, 2011; Laurie and Sethi, 2012).

Building on the progress of SHIP is the Farr Institute, a health informatics research UK-wide initiative which has a branch in Scotland. eDRIS is now located at the Farr Institute, representing a move towards a joint informatics centre. Information governance in Scotland is outlined on the web pages of 'eHealth' – part of the Health Finance and Information Directorate – which is part of the Scottish Government Health & Social Care Department (Scottish Government, 2014). A useful document on this website is the open response to the second Caldicott review by the Chief Medical Officer for Scotland, Sir Harry Burns, to Dame Fiona Caldicott. The response highlights the implications of the key recommendations from that report, and how they apply or otherwise to the situation in Scotland.

Northern Ireland

In Northern Ireland, Health and Social Care Northern Ireland (HSCNI) is the information hub for health services in Northern Ireland. Data is collected through 'regional data warehouses', access to which is managed through the Honest Broker Service. For a list of data available see the HSC Business Services Organisation (BSO) website (HSC BSO, 2014). In comparison with health data held and managed in the other UK countries, N. Ireland appears to have more limited data availability.

UK Audits and Registries

The Healthcare Quality Improvement Partnership (HQIP) maintains a list of clinical databases and registers across the UK, details of which can be found on their website (HQIP, 2014a).

d. Data linking. To what extent can / are patient data linked across databases – how and by whom? Who are the major organisations involved?

The ability to link individual patient data across datasets is facilitated by the provision of a unique patient identifier, the use of which are mandated in NHS IT systems across the UK. For England and Wales these identifying numbers are the same – 'NHS numbers' – which are codes containing 10 digits and are used to maintain electronic health records for all health care users. Scotland use a Community Health Index (CHI) and Northern Ireland a Health and Care Number (HCN); both have the same format as an NHS Number, but with the 10 digits over a separate range (to avoid overlap) (NHS, 2008). These numbers are used exclusively for the provision of health services, and are not related to any other public service or taxation.

Across the UK, sensitive patient data is linked by a 'Trusted Third Party' or 'Accredited Safe Haven'. In England this is the HSCIC, which de-identifies patient data before this is sent to the relevant organisations that work with or provide access to this data. In addition, it provides a linking service for outward-facing organisations such as CPRD and IMS Health.

CPRD is able to provide data that is linked across datasets by individual patient identifiers. Whilst GP data coverage is around 9% of the total UK population, most of this data is held for GP practices based in England, and it is only these English practices from which permission has been sought by CPRD to link data. CPRD is able to link data with various datasets: secondary care HES data, ONS data and Index of Multiple Deprivation (IMD) scores. CPRD 'host' all of these datasets. CPRD is also able to link with NCIN cancer registry data and the Myocardial Ischemia National Audit Project (MINAP). However for these two datasets the owners remain the custodians of the data, which are not held in-house by CPRD. This means that each have their own additional independent panels to approve access to data. All data linking is undertaken by the HSCIC.

IMS Health is also able to provide linked data through their 'Hospital Treatment Insights' (HTI) dataset. This data set links hospital pharmacy audit data with HES data, facilitating analyses of drug use and patient treatment / outcomes. Permission to link this data has been sought by IMS on an individual trust-level basis, and coverage is around 25% of English hospitals - around 3.3 million patients[4]. Section 251 exemption allows IMS to be able to receive this data. The data linkage process for this dataset is managed and performed by HSCIC. 'HTI-CPRD GOLD' is a further development of this dataset, in which HTI data is linked with CPRD data and thereby able to reflect the full patient journey. As this requires patients to overlap all three datasets, coverage is lower: around 317,000 individual patients.

In Wales, datasets linked by SAIL were described above; this is facilitated using the unique identifiers of patients as in the other UK countries, generally with NHS Number (or probabilistic matching on personal identifiers e.g. postcode, gender, etc.). Organisations that provide SAIL with data do so via the NHS Wales Informatics Service (NWIS), which acts as the trusted third party in Wales for anonymisation and encryption.

In Scotland, health data can be linked and assessed by eDRIS. The Scottish Government in 2012 released proposal for a 'Joined up data for better decisions', which proposed a 'Data Sharing and Linking Service' (DSLS) which could link health and non-health data (Scottish Government, 2012); a technical consultation for the service was completed in 2013 (eDRIS, 2013). In this scheme the National Records of Scotland (NRS) would provide the trusted third party 'Indexing' service to handle the personal identifying data and create a linked dataset, access to which would be through an Analysis Safe Haven provided primarily through the NHS National Services Scotland (NSS). It is not clear to what extent the DSLS has progressed since this consultation, but there is some indication that his may be co-located with the Administrative Data Research Centre for Scotland (ADRC), with eDRIS acting as the first point of contact and the Farr Institute housing the health research function (Scottish Government, 2013). Another recent development is the Health and Social Care Data Integration Project which will come into play from April 2015, for which the Information Services Division (ISD) has been commissioned to link health and social care data on an individual level and create a nationally agreed core dataset (ISD, 2014a). Few details are available on the mechanism of data linking in Northern Ireland.

e. Data access. To what extent is data shared, with whom is it shared, how does permitted access differ according to organisation (i.e. access by

[4] Information obtained from Manager at IMS Health, October 2014.

pharmaceutical companies versus access by public bodies / academic institutions), what are the processes involved in being granted permission to access data, what are the costs involved in data access (where available), and in what form is data access granted (e.g. provision of raw data / ability to use in-house data analysis services?

The processes and methods of data sharing differ by country and by organisation. As discussed, accessing patient data by those health care professionals directly involved in their care is relatively straightforward. Secondary uses of health care data include uses by commissioners. It should be noted that HSCIC host a data warehouse called 'Secondary Uses Service' (SUS), which is the single comprehensive repository for health care data in England and whose purpose is to provide NHS providers and commissioners with data for use other than primary clinical care (i.e. for health care planning, commissioning, payment by results, improving services, etc.) (HSCIC, 2014l). Access to identifiable data in SUS by *commissioning organisations* used to be supported by Section 251 exemption. However as of the end of October 2014 this has ended (HSCIC, 2014d). These organisations (including clinical commissioning groups and commissioning support units) must ensure there is a data sharing agreement in place with HSCIC, and will only be allowed to receive *pseudonymised* data from 1 November 2014. For commissioning organisations to receive 'weakly pseudonymised' data, they must now become an 'Accredited Safe Haven' (ASH).

For data requests that are not for the direct care of patients and not for commissioners, access to health data through any of the organisations begins with an application process. These are generally well set out by the respective organisations, and require the applicant to complete and submit an application. For the HSCIC's the data access request process applicants are asked to provide details of the purpose of the data being requested, demonstrating that the request is being made to support the provision of health and social care and the promotion of health. The applicant must also provide evidence of the approvals required (e.g. patient consent or Section 251 Support),details of the security arrangements in place (e.g. ISO 27001 and NHS Information Governance Toolkit), as well as the type and amount of data requested (HSCIC, 2014a). The application is then considered by the Data Access Advisory Group (DAAG), which makes a recommendation to the HSCIC Senior Information Risk Officer who has responsibility for releasing data. Similarly, provision of patient level and linked CPRD data is for "publicly-benefiting medical research" only, and applications are considered by the Independent Scientific Advisory Committee (ISAC) (CPRD, 2014c). It should be noted that ISAC approval is not required for the use of patient level data to generate aggregated data, not destined for publication. For Welsh health data, applicants submit a SAIL Data Acquisition Request form (SAIL, 2014a), which is considered by the SAIL Data Management Committee. The process involves review by the Information Governance Review Panel (IGRP) which ensures proposals are "appropriate and in the public interest", and the committee consists of representatives from the British Medical Association (BMA), the National Research Ethics Service (NRES), Public Health Wales, NHS Wales Informatics Service (NWIS) and a Consumer Panel. For access to UK audit data, the Healthcare Quality Improvement Partnership (HQIP) (including the Diabetes audit), requests must be for a clinically appropriate use of the data and applicants must have in place the necessary storage accreditation and information governance policies (HQIP, 2014b).

The form in which data is accessed differs according to the requirements of the data applicant, and also by the processes in place in the respective information centres. Little detail is provided publicly on the precise form of data access, as this will be considered on a case-by-case basis. Most organisations offer and show a preference for data to be analysed 'in-house' according to a research question posed by the client, and the results shared with the applicant (e.g. eDRIS in Scotland). On the other hand, CPRD offers licences for use of primary care data on an annual basis. Depending on the requirements of clients, SAIL can either conduct analysis in-house or provide a limited (anonymised) dataset via a secure laptop / online server ('Gateway'), which would be in the form of raw data extracted specifically for the project. In recognition of the demand for researchers to interrogate the data themselves, many organisations have in place or are proposing mechanisms to share data in a 'safe haven': a physical research space that can be managed in a secure and monitored setting. For example, a 'Customer Survey' was sent out in September 2014 by the HSCIC to members of the pharmaceutical industry to collect views on a model of "on-site" access to HSCIC data, in response to a recommendation from their board that "…the HSCIC actively pursues a technical solution to allow access to data, without the need to release data out of the HSCIC to external organisations" [email correspondence, September 2014].

Similarly, Public Health England (PHE) is currently developing a model for the sharing of pseudonymised linked data in a secure environment: The Cancer Data Research Centre. In recognition of the fact that record-level data will always have the potential risk to be identifiable, they propose to minimise that risk through data sharing in a controlled environment where users can be contractually bound, under defined protocols, with a controlled dataset, keeping a history of release and physical controls (e.g. no USB sticks, cameras, etc). It is proposed that under these strict conditions the data could be regarded as anonymous and therefore not subject to the Data Protection Act. To mitigate reputational risk for Public Health England, the research centre would act as an "ethical intermediary" – a third sector social enterprise. In a meeting hosted by Public Health England in September 2014 Jem Rashbass, National Director for Disease Registration for Public Health England, said that he expected the service to be running by April 2015. Access will be under a 'membership scheme' whereby a desk might be rented on an annual basis. The secure office environment will be located in Cambridge and act as a proof of concept. If successful, expansion to other PHE locations will be considered, before exploration of a virtual space.

In Scotland the NSS National Safe Haven is a secure environment where data can be linked and accessed, with an access point in the form of a physically secure area with no external devices which allows trusted and authorised researchers to analyse linked individual data (eDRIS, 2014). Secure access points are located in the Farr Institute Scotland (Edinburgh); remote access via a VPN is considered in some cases. The concerns that arose from the consultation for DSLS included geographic access difficulties, as well as not being able to employ specific platforms and tools which might otherwise be available to them (ISD, 2013). There are examples of innovative solutions to this problem. For example, PHE propose to create a secure record-level server with all relevant cancer data fields but with spoof data, so users can develop and test software outside of the safe environment which they can then bring in and work on the data with.

Most organisations are not explicit in who can access data, but all discuss the requirement that data and research requests should be in the public interest; HSCIC say that data cannot be released "solely for commercial purposes" (HSCIC, 2014a). In

Scotland the Farr Institute express that "industry does not have access to patients' data in Scotland, and they cannot buy datasets from NHS Scotland" (Farr, 2014). They go on to explain that if a research project proposed by industry is in the public interest and a partnership with an academic or NHS institution is attained, then permission may be granted. However the pharmaceutical company would not be given access to the data or given permission to undertake any of the analyses. This appears to be different from the model in England, for example where CPRD offer yearly licences to industry for (anonymised) patient datasets (subject to the qualifying criteria listed). IMS Health is clearly a commercial-facing organisation, and industry represents their main client base. In Northern Ireland the BSO was asked to set up the 'Honest Broker Service' to manage access to health care data. Customers to data have been colleagues from within the HSCNI and academic researchers. Allowing pharmaceutical companies access to data may be some way off[5].

The cost of accessing health data is generally not explicit on the websites of the relevant organisations. This is mainly because charges will be dependent on the type of data request. Many information centres that work with public data (SAIL, CPRD and HSCIC) specify that they do not "charge" for the data, but rather act on a cost recovery basis, specifying a price to cover the costs associated with administering, preparing and supplying the data. According to the HSCIC the money received is put back into the HSCIC to support the running of the organisation (HSCIC, 2014i). HSCIC provide indicative prices for data charges on their website (HSCIC, 2014b). An annual licence for access to primary care data through CPRD costs approximately £255,000 per year (CPRD, 2014a); there are further charges for more than two nominated users. IMS Health does not publish prices for their services on their website, and prices would have to be obtained by contacting them with a specific enquiry. When asked how data from the HTI-CPRD GOLD database should be accessed, a manager and IMS Health indicated that this could either be through IMS *or* CPRD, and that prices would be equivalent (as they were mutually agreed).

3. Collecting de novo patient data

a. **Governance arrangements for research to collect new data.** Key documentation outlining research ethics and governance for the collection of new patient data (i.e. setting up registries, pragmatic clinical trials, etc.).

Governance arrangements for collecting de novo patient data for the purposes of evaluation or research vary depending on the primary purpose of the data collection. The governance arrangements for 'research' are rigorous, whereas data collection activities for audit or service evaluation do not require approval by an ethics review committee.

The generally accepted definition of research is an *"attempt to derive generalizable knowledge by addressing clearly defined questions with systematic and rigorous methods"* (IRAS, 2014a). According to the same source: *"Although some research projects include evaluation, where a project is considered to be solely audit or service/therapy evaluation, it will not be managed as research within the NHS or social care. Such projects do not require ethical review by a NHS or Social Care Research Ethics Committee or management permission through the NHS R&D office. There is no need to submit applications in IRAS either to the REC or R&D office*."

[5] Information obtained from conversation with HSCNI, October 2014.

Given the understandable ambiguity or grey areas for some projects, there are many sources which provide guidance and decision tools to help applicants decide (HQIP, 2014c; HRA, 2014a; HRA and MRC, 2014). Whereas research aims to find out what you should be doing, audits are conducted to find out whether planned activities are being undertaken and to assess whether they are working. For programs whose categorisation may be ambiguous it is not entirely clear under which category they would fit in; if the health intervention is disseminated widely and its access not based on specific restrictive criteria, yet future commissioning decisions are based on their result, then it may unclear under which category these data collection activities should lie.

For research projects undertaken within the NHS, ethical approval must be sought. Links to the key documentation for the overarching research governance frameworks for England, Scotland, Wales and Northern Ireland are provided online by the NHS Health Research Authority (HRA) (HRA, 2014b). This framework is outlined by the Department of Health Research and Development Directorate in England, the National Institute for Social Care and Health Research in Wales, the Chief Scientist Officer is Scotland, and the R&D Division, Public Health Agency in Northern Ireland. Informed patient consent is at the heart of ethical research, and the ethical review process places a strong emphasis on this (Department of Health, 2005). Exception is granted in limited circumstances, such as cases where section 251 support has been granted where identifiable patient information can be collected for research without consent.

b. **Research application process.** Process by which application for new data collection is considered, and governing principles of the committees that grant approval.

If it has been decided that a project is 'research', there are various types of approvals required from various bodies. As a minimum, NHS management permission ('R&D approval') is needed for each research site, and ethical approval must be considered by a Research Ethics Committee (REC). The requirement for ethical review are set out in a UK-wide 'harmonised' edition of the Governance Arrangements for Research Ethics Committees (GAfREC) (DH et al., 2011).

Application for ethical review is now centralised in the UK into the Integrated Research Application System (IRAS). This represents a single system for applying for the permissions and approvals required to undertake health research in the UK, and to meet regulatory and governance requirements. In order avoid duplication of effort the system captures the information needed for the relevant approvals for all relevant bodies (IRAS, 2014b):

- Administration of Radioactive Substances Advisory Committee (ARSAC)
- Gene Therapy Advisory Committee (GTAC)
- Medicines and Healthcare products Regulatory Agency (MHRA)
- NHS / HSC R&D offices
- NRES/ NHS / HSC Research Ethics Committees
- Confidentiality Advisory Group (CAG), formerly the National Information Governance Board (NIGB)
- National Offender Management Service (NOMS)
- Social Care Research Ethics Committee

4. **Data use.** What are the rules governing the use of RWE?

There appears to be less information available in the public domain around the governance issues relating to the *use of* (rather than access to) real world data. As indicated, in general, access to data that is provided by public organisations may not be granted for solely commercial purposes. With data provided by HSCIC this is regulated through a data sharing agreement, which specifies the purpose to which the recipient can put the data and restricts how they store, share, use and eventually destroy the data. The CPRD data access licence (CPRD, 2014a) specifies that all nominated users of the data must undergo training, and lists the permitted use and restrictions, which can be viewed in Appendix (2) of this document. Specifically, it specifies that the use of the data must be restricted to Medical and Health Research Purposes on a non-profit making basis, which "*shall not prevent the Licensee from: […] recovering a profit from any application of the results of the Licensee's research provided that such profit is solely attributable to the value added by the Licensee in its analyses or interpretation of the Data*". The CPRD also specify that in reporting the findings of research using CPRD data, users must include the ISAC protocol in journal submissions and include details of the ISAC approval in the manuscript. Users should also submit copies of all peer-reviewed publications based on CPRD data to the ISAC Secretariat. In addition, in disseminating results users must not report any cells containing fewer than 5 events, in order to reduce the possibility of unintentional deductive disclosure (CPRD, 2014c).

There is also an emphasis on transparency for the use of NHS patient data; the HSCIC publishes a list of approved data releases which contains information on who the data was shared with and for what purpose (HSCIC, 2014j). On this register the HSCIC discloses details of the form of the data provided (identifiable, pseudonymised, anonymised or aggregated-anonymised) as well as the legal basis for provision of the data (i.e. informed patient consent or the relevant act and clause such as Section 251 support). No pharmaceutical companies appear on the list of organisations to which data has been released directly by HSCIC in 2014. However various consultancy firms do appear on this list, e.g. IMS Health, PricewaterhouseCoopers LLP and McKinsey and Co.; for all of these data releases the 'Health and Social Care Act 2012' is listed in the description of the legal basis for provision of data. In the glossary that the HSCIC provides for the register, the description of this legal basis is: "*The Act of Parliament that set up the HSCIC. This is cited as a legal basis for some instances of sharing pseudonymised data because the statute set the legal framework for the collection and dissemination of health and social care data"*.

With regards the use of non-randomised or non-controlled evidence for health technology assessment (HTA) (i.e. observational data, of the kind that we are interested in for this project), there is little detail on the acceptability or methods to be employed in its use in the NICE Technology Appraisal Process and Methods guide (NICE, 2013). It is simply stated that for observational studies, inferences are necessarily more circumspect, and that potential biases should be identified and ideally quantified and adjusted for. This is in contrast to the methods guide for the medical technology evaluation programme that assesses medical devices, where is it accepted that non-RCT data will provide important input into the evidence base (NICE, 2011). However, in considering changes to their technology appraisal methods for drugs, NICE has acknowledged that there should be more productive risk sharing between companies and the NHS, and have proposed that value should progressively reflect value of treatments as our knowledge of what they can offer patients increases (NICE, 2014). They offer the example of NHS England's 'Commissioning through Evaluation' (CtE) process which could be used for this. To date, only interventional procedures have been managed through

CtE (the first of which began at the end of 2013), and there are no clear guidelines to their conduct. We have yet to see how the information gathered (largely through registries) will be assessed by NHS England or by NICE. Additionally, it is unclear whether the activity should be considered as 'research' or 'service evaluation', for which the ethical requirements are distinct.

5. **Suggested principles or guidance for data governance, and the adapting environment for such.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

The Nuffield Trust offers an interesting thought-piece on the social values associated with information governance in health, and outlines the interpretations of the legal frameworks that regulate the use of patient information, the most dominant of which in the UK has been the idea of 'consent or anonymise' (Clark and Weale, 2011). The authors describe justifications in the law for the use of patient data for medical research based on public interest, but assert that these are inadequately defined and therefore researchers are reluctant to use them. They describe a more open 'free access' position and an intermediate 'controlled access' approach, interpretations which could all be supported depending on interpretations of the legal and ethical provisions in place.

A paper that draws from the research of the Scottish Health Informatics Platform (SHIP) initiative (funded by the Wellcome Trust in partnership with NHS Scotland) outlines information governance challenges in health-related research and advance a template for 'good governance' (Laurie & Sethi, 2012). They assert that the current governance model is disproportionate relative to the benefits of health-related research, and encourage a more facilitative environment.

Despite the call by many research professionals for a less bureaucratic environment for access to and analysis of health data in the UK, public trust is of paramount importance in consideration of these issues, and the public voice against recent proposals to improve data linkage in the UK has had an important impact on this landscape. In order to improve the utility of health data that is already collected, NHS England along with HSCIC and other stakeholders worked together to design a modern data service known as 'care.data' (NHS England & HSCIC, 2013). The purpose of this data service was to bring together patient-level information across all health care settings, in order to provide a joined-up view of patient care. This would have seen the transformation of HES data into Care Episode Service (CES), to include: hospital dataset, mental health data, GP data, community health services data, social care data, clinical audit data and disease registry data; management of the dataset would be through the HSCIC. Whilst this would represent a big step forward in improving the value of health data, progress was halted due to a strong public reaction against the plans and fears that health data and entire medical records would be "on sale" to drug companies among others (Ramesh, 2014). As a result, widespread roll-out of the program has stalled. However NHS England are pursuing a staged roll-out through an initial "pathfinder" stage starting with 265 GP practices (NHS England, 2014a). The immense public scrutiny of plans to modify arrangements for the management of health data demonstrates the importance of robust and transparent governance arrangements and the need for clarity in communications with the public (HSCIC, 2015).

The Health and Social Care Act 2012 introduced various legal and organisation changes to the NHS, which has given rise to various evolving challenges which must be addressed. As a result, an Information Governance Transition Programme was introduced to develop a range of short and medium term measures to address the issues (NHS England, 2014b). The solutions and guidance developed by this Transition Programme disseminated and updated through a monthly *information governance bulletin*, which is directed at those using data for secondary uses (e.g. commissioners, data analytics providers, clinical audit teams, researchers, managers etc.) (NHS England, 2014c).

We have demonstrated that sources of guidance for information governance are disparate, and it is therefore challenging to present a coherent picture of the frameworks and processes for data governance in place. In response to this, there is a work programme underway to establish an 'Information Governance Alliance' (IGA) through 2014/2015. The IGA was set up in July 2014 in response to a request from the Independent Information Governance Oversight Panel (IIGOP), chaired by Dame Fiona Caldicott (HSCIC, 2014h). On its board are members from the core organisations that fund the IGA: Department of Health, NHS England, HSCIC and Public Health England. The intention is for IGA to act as a single authoritative resource of information and guidance on governance in the health sector. Another 'Information Governance' work stream in development is being facilitated through the Farr institute, and will bring together governance leads from each of the four Farr Institutes in the UK to conduct governance methodology research and develop and share best practice governance standards for data use.

As indicated, the legislative changes that may occur in the wider European environment could have a profound impact on the collection and use of patient data for research. In the UK the potential impact of this is summarised by the NHS European Office (NHS European Office, 2014).

SUMMARY

The legal framework in the UK and its four nations for handling patient health data attempts to balance the benefits and risks of holding personal information. As a result, there is room for interpretation and there are few hard and fast rules which dictate exactly how organisations should operate. Therefore there is an element of organisations 'feeling their way'. This is evidenced by the fact that many of these organisations are currently issuing consultations on how best to share data (e.g. Scottish Government, HSCIC, PHE). This move towards extending access to data rather than simply providing in-house data analysis services requires that organisations manage the risks associated with various 'degrees' on anonymisation. There is a recognition that the law does not give absolute value to privacy, and that a balance must be struck by those holding the data between proportionate mitigation of risk and the potential benefits for research.

The question of how to ensure this balance is correct is central to a clear and transparent governance framework. Measures to address this so far have included ensuring that patient and public representatives are present on the governance and advisory boards of information centres, providing information publicly on all data releases, and incorporating sanctions into data use agreements. This is of particular importance for organisations that manage or provide access to patient data which is linked across

datasets. In recognition of the sensitivities involved in data linking, in the UK this is generally performed through Trusted Third Parties.

In theory there is a clear basis for data governance in terms of accessing patient health data: obtain patient consent or anonymise the data. However there are various characteristics / provisions in law which in reality make this more complex. The first is that there are various degrees of anonymisation, and depending on the use to which data is put, re-identification of pseudonymised data could be a risk. This means that monitoring of data release and follow-up falls within the remit of information centres. A second complicating factor is the provision in law for statutory exemption through section 251 support of the NHS Act 2006, which bypasses the need for consent where it is in the public's interest and where patient consent is not feasible.

Whereas in Wales access to health data is managed by a central point of contact — SAIL—in England there are many access points for routinely collected health data depending on the type of information needed. This can create some confusion; for example, for access to the linked primary care and HES-linked pharmacy audit data, either the CPRD *or* IMS Health can be approached (supposedly for the same price).

In principle, it would be helpful to observe a set of guidelines for governance around (1) accessing data, (2) use to which data is put and (3) collecting new patient data. Whilst we have tried to characterise the relevant information for each of these of the UK, there is most information around the first of these.

Once an organisation has access to health care data, there are various technical and security procedures and functions that must be in place which are often specified in the contractual agreements between organisations. However when it comes to using that data to provide evidence and inform best practice, there are few guidelines available. In an article on *'Building trust in the collection and use of real world health data'*, the research director of Deloitte UK Centre for Health Solutions highlights the need for guidance around Good Evidence Practice, in order to build trust across stakeholder groups in the form of a mutually accepted governance process for the use of health data to generate real world evidence, in the same way that Good Clinical Practice (GCP) guidance exists to provide confidence to clinicians and researchers in the collection of patient data (Taylor, 2014).

With regards to collecting de novo patient data, the line between audit and research may prove problematic in the collection of real world data to inform decision-making. Whilst much of the burden may be on clinicians / auditors rather than patients themselves, the governance arrangements for setting up new data collection initiatives are not clear. Moreover, data collection initiatives to support earlier and more iterative assessments of health technologies that are supported by commissioners, such as Commissioning through Evaluation, are welcome, but there should be clearer guidance around how data is to be assessed, and the governance arrangements for data collection and management.

4. The United States

1. Brief overview of the health system and collection / management of patient data

The U.S. health care system is distinguished from that of other countries due to the existence of multiple payers, including both public and private payers. Approximately 30-40% of the U.S. population have health insurance coverage through the public system (Rice et al., 2013; US Census Bureau, 2014). The largest public health care payers include Medicare and Medicaid. Medicare is a federal program that provides coverage for individuals 65 years of age and older and certain disabled persons. Medicaid is administered at both the state and national levels and covers low income individuals and those with limited assets.

Private health care plans are numerous, and are often supplied by individuals' employers, but may also be directly purchased by individuals. Most Americans receive coverage through private health insurance, and approximately 80-90% of Americans with private insurance have employer-sponsored coverage; however, the proportion of the population that purchases insurance individually has likely increased in 2014 with the passage of the Affordable Care Act (Rice et al., 2013; US Census Bureau, 2014). Private insurance can be categorized as using one of preferred provider organizations (PPOs), health maintenance organizations (HMOs), or high deductible plans.

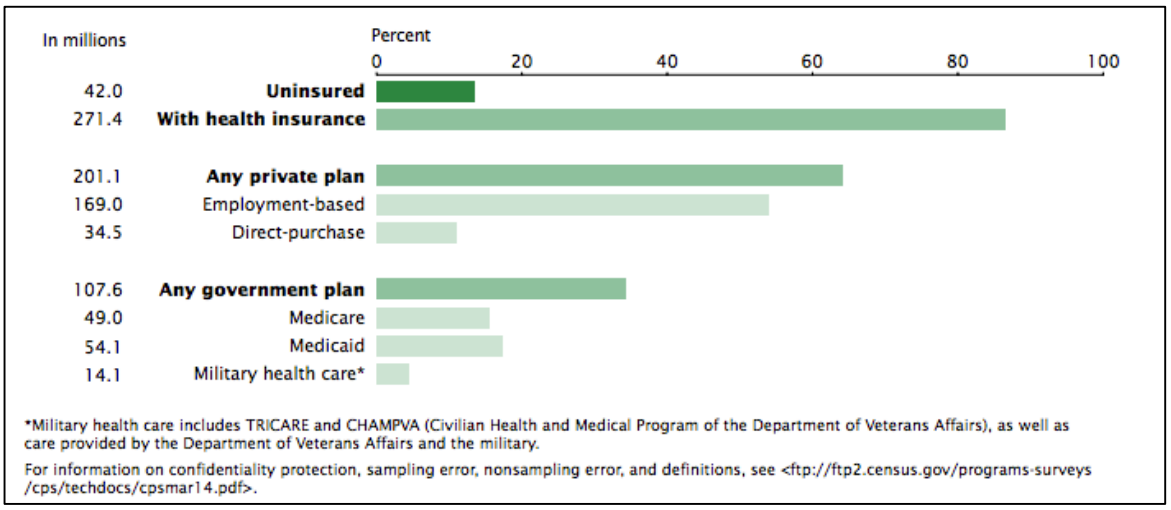


Figure 2 Number and percentage of people by health insurance status in the US, 2013 (population as of March the following year)

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economics Supplement.

Both public and private payers purchase health care services from providers subject to regulations imposed by federal, state, and local governments as well as by private regulatory organizations.

Figure 3 provides an overview of the structure of the U.S. health care system.

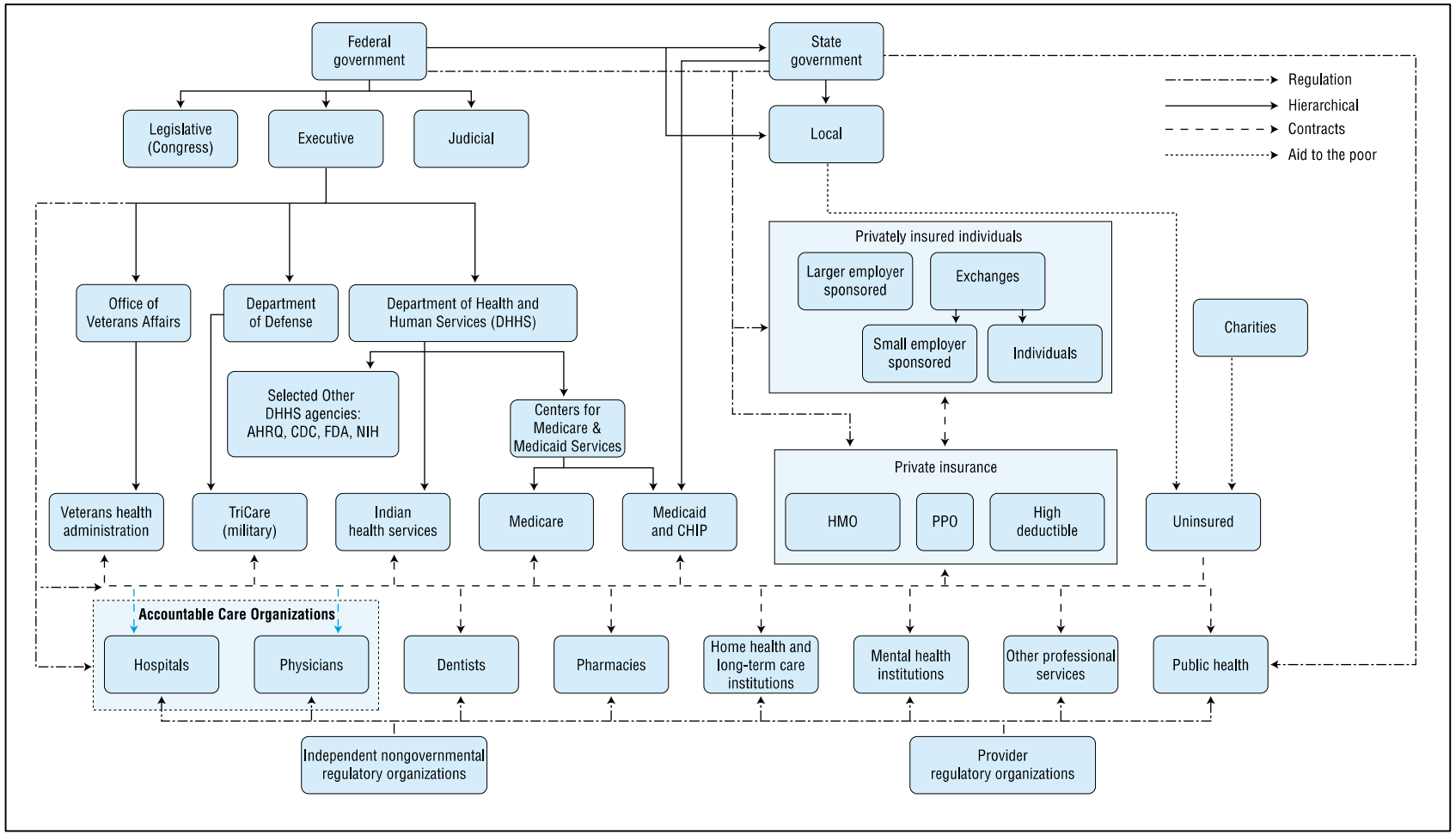


Figure 3 Organization of the U.S. Health System after Implementation of the Affordable Care Act

Reproduced with permission by the European Observatory on Health Systems and Policies. Source: Rice T, Rosenau P, Unruh LY, Barnes AJ, Saltman RB, van Ginneken E. United States of America: Health system review. Health Systems in Transition, 2013; 15(3): 1– 431.

2. Routinely collected patient data

a. **Core legislation governing the collection / use of routinely collected patient data.** Key documentation outlining principles of governance and data protection.

HIPAA-Introduction

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is the most extensive federal legislation regarding data governance and protection in the United States (OCR, 1996). Title II of HIPAA, known as the Administrative Simplification provisions requires the Department of Health and Human Services (HHS) to draft rules aimed at creating standards for the access and use of health care information. In response, the HHS has promulgated the Privacy Rule, which protects the privacy of individually identifiable health information and the Security Rule, which sets national standards for the security of electronic protected health information. The Privacy Rule (updated in 2013 based on the Health Information Technology for Economic and Clinical Health Act) regulates the use and disclosure of Protected Health Information (PHI) held by covered entities (generally, health care clearinghouses, employer sponsored health plans, health insurers, and medical service providers that engage in certain transactions) and business associates (DHHS, 1996). PHI is any information in the medical record that can be used to identify an individual and that was created, used, or disclosed in the course of providing a health care service such as diagnosis or treatment. A covered entity may only disclose PHI without a patient's express written authorization to facilitate treatment, payment, or health care operations. Any other disclosures of PHI require the covered entity to obtain written consent from the individual for the disclosure. The HHS has also established enforcement rules which delineate the procedures for investigation and hearing for HIPAA violations as well as civil and criminal penalties for infractions.

HIPAA-Implications for RWE Research

HIPAA regulations cover any research which uses, creates, or discloses PHI. Research is generally defined as a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge. Research may also be defined as any clinical investigation that involves an item regulated by the FDA, and involves one or more humans subjects and involves either any use of a drug in living persons other than use of an approved drug in the course of medical practice or the use of a device in living persons that evaluates the safety or effectiveness of the device or the use of a test article regulated by the FDA to obtain data that is intended to be eventually submitted to the FDA. Hence HIPAA regulations can apply to both retrospective and prospective RWD collection which involve medical record review or the creation of new medical records. The Privacy Rule permits the use or disclosure of PHI for research under the following most common conditions:

- If the subject of the PHI has granted specific written permission through an Individual authorization
- If the Institutional Review Board (IRB) has granted a waiver of the authorization requirement which in turn require all three of the following criterion are met:
    - The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals;
    - the waiver or alteration will not adversely affect the rights and welfare of the subjects; and

    - the research could not practicably be conducted without the waiver or alteration.
- If the information is released in the form of a limited data set, with specified direct identifiers removed, and with a data use agreement between the researcher and the covered entity which limits who can use or receive the data and which prohibits the researcher from contacting the individual
- If the PHI has been de-identified in accordance with the standards set by HIPAA

The Privacy Rule makes two methods available for de-identifying health information such that it is no longer considered PHI. The safe harbour method requires the removal of 18 unique identifying characteristics (Table 1). The Expert Determination Method requires that a "person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable", applying these statistical and scientific principles and methods, determines that the risk of identification by an anticipated recipient is no more than "very small."

Table 1 Summary of 18 patient identifiers under the HIPAA Privacy Rule

Names	Account numbers
Geographic subdivisions smaller than a state	Certificate/license numbers
All elements of dates (except year) for dates directly related to an individual	Vehicle identifiers and serial numbers, including license plate numbers
Telephone numbers	Device identifiers and serial numbers
Fax numbers	Web Universal Resource Locators (URLs)
Electronic mail addresses	Internet Protocol (IP) address numbers
Social security numbers	Biometric identifiers, including finger and voice prints
Medical record numbers	Full face photographic images and any comparable images
Health plan beneficiary numbers	Any other unique identifying number, characteristic, or code

HIPAA imposes both scientific and administrative constraints on RWE Research. Scientifically, HIPAA may reduce sample size and introduce selection bias due to an inability to directly contact potential participants. A study which attempted to recruit patients for an Acute Coronary Syndrome registry observed a 62% decrease in the number of patients who were willing to enrol after the implementation of HIPAA (Armstrong et al., 2005). Further, enrolees in the post-HIPAA period were found to be significantly different from non-enrolees in terms of age, marital status and mortality. Administratively, HIPAA may also add burden and costs due to increased efforts required for documentation and recruitment. Another study found that HIPAA implementation was associated with 73% decrease in patient accrual, a tripling of time and costs for spent recruiting patients (Wolf and Bennett, 2006).

The Common Rule

The Federal Policy for the Protection of Human Subjects or the "Common Rule" was published in 1991 and codified in separate but identical regulations by 15 Federal departments and agencies (published by the HHS in Title 45 Code of Federal Regulations Part 46, subpart A) and applies to research on human subjects conducted, supported or otherwise subject to regulation by these departments and agencies. The HHS

regulations, includes subparts B, C and D which extend additional protections to pregnant women, human fetuses, and neonates; prisoners; and children respectively (DHHS, 2009). Title 45 Code of Federal Regulations Part 46 also establishes IRBs which are administrative committees designed to review human subject research conducted under the auspices of the institution with which it is affiliated. The IRB has the authority to approve, require modifications in, or disapprove all research activities that fall within its jurisdiction as specified by both the federal regulations and local institutional policy. IRBs judge human subject research for approval based on the following seven criteria: (1) risk minimization, (2) risk/benefit comparison, (3) equitable subject selection, (4) informed consent, (5) data monitoring to ensure safety, (6) privacy protection and confidentiality, and (7) protection of vulnerable subjects. IRBs are themselves regulated by the Office for Human Research Protections within the HHS.

The Privacy Act

The Privacy Act, enacted 1974, is a federal law which establishes a code of fair information practice that governs the collection, maintenance, use, and dissemination of personally identifiable information about individuals that is maintained in systems of records by federal agencies (US Congress, 1974). The Privacy Act prohibits the disclosure of information from a system of records without the written consent of the individual, unless the disclosure is pursuant to one of twelve statutory exceptions. A system of records consists of any item, collection, or grouping of information about an individual, where those records can be retrieved by the name of the individual or by some other type of identifier unique to the individual. The Act also provides individuals with a means by which to seek access to an amendment of their records, and sets forth various agency record-keeping requirements.

The National Research Act

The National Research act, also enacted in 1974, is a federal law which established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (which was succeeded by the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research) to develop guidelines for human subject research and to oversee and regulate the use of human experimentation in medicine (US Congress, 1917). The National Commission published the Belmont Report in 1979 which established the unifying ethical principles for the federal protection of human subject protection and which has informed subsequent policies such as HIPAA and the Common Rule. The three ethical principles laid out in the Belmont report are:

- Respect for persons: research must protect the autonomy of all people and allow for informed consent
- Beneficence: research must "Do no harm" and maximize the benefits and minimize risks to the research subjects
- Justice: research must be reasonable, non-exploitative, and distribute the costs and benefits to potential research participants fairly and equally

b. Datasets. Overview of what data is collected, and from what parts of the health service

Health care data are owned by individual health care providers, payers who adjudicate claims and process payments, and registries that are established for the purpose of tracking patients with certain conditions and their outcomes.

Claims data. Claims data are data maintained by payers, and are typically managed at the organizational level. Claims data include billing codes that physicians, pharmacies, hospitals, and other health care providers submit to payers. These data follow a relatively consistent format across payers and use standard, pre-established codes that describe specific diagnoses, procedures, and drugs. Additionally, claims are generated for almost all interactions between a payer and health system, leading to a fairly comprehensive and standardized source of patient information: claims data provide a holistic view of the patient's interactions with the health care system (Wilson and Bock, 2012). However, given the fluid nature of the market with patients changing employers and therefore health plans, and individuals potentially entering and exiting the Medicare and Medicaid systems based on their eligibility, the data typically do not provide long-term information for individual patients.

Patient records. Patient medical records are a mechanism for physicians and other caregivers to record the details of medical care administered to patients. Medical records contain rich patient-level data and are often the only source of such detailed information. Records include information such as physician notes, laboratory reports, surgical dictations, copies of correspondence, appointment schedules, imaging records, and so forth. Non-digitized versions of patients' records are included in paper charts, but are often from one practice and not shared across practices. As organizations move towards electronic records, patient records will become more tenable to research efforts. Electronic medical records (EMRs) contain standard medical and clinical data gathered in one provider's office, whereas electronic health records (EHRs) are intended to go beyond the data collected in the provider's office and include a more comprehensive patient history (DHHS, 2014). EHRs are designed to contain and share information from all providers involved in a patient's care. According to the Institute of Medicine, the four components of EHRs are defined as: clinical documentation, results reporting, physician order entry, and clinical decision support (IOM, 2013). EHR data are intended to be created, managed, and consulted by authorized providers and staff from across more than one health care organization. However, until electronic records are universally adopted by all providers, information for a given patient may be incomplete within this data source.

c. Information providers. Who 'holds' the data (likely to be a mix of public and private organisations), what data do they hold, how is it collected, and what are the core governing principles in handling the data?

As described above, ownership of health care data in the U.S. is very much siloed, and, therefore, access to data requires liaising with individual companies to establish data access agreements and obtain data. Below, we have summarized examples of data or databases that can be accessed in partnership with various organizations.

PRIVATE SECTOR
Truven Health Analytics: MarketScan (Truven Health Analytics, 2014)
Patient Population: 200 million unique patients from commercial, Medicare supplement and Medicaid populations.
Type of Data: Inpatient, outpatient, drug, lab, health and productivity management, health risk assessment, dental, and benefit design.

Restrictions: Cannot attempt to identify specific individuals contained in the data, cannot attempt to link MarketScan commercial claims and encounters and the MarketScan Medicare Supplemental Database to local, regional or state level data, cannot report results at a three digit zip code level, must obtain review and consent by Truven Health before reporting or publishing results for geographic areas larger than the three-digit zip code level (such as whole states).
Cost: Approximately $25,000 for federal/non-profit studies, and $50,000 for for-profit funded studies.

HealthCore Integrated Research Database (HealthCore, 2014)
Patient Population: 43 million lives in commercial plans across 14 states.
Type of Data: Medical claims, pharmacy claims, laboratory result data; includes access to electronic medical records, medical charts, and an array of prospective information from patients, physicians, and other health care professionals.

Humana: Comprehensive Health Insights (Humana, 2014)
Patient Population: 11.3 million total lives, 2.6 million commercial members, 1.8 million fully insured commercial members with integrated data.
Type of Data: Fully insured commercial members – medical and pharmacy claims; Medicare Advantage members – medical and pharmacy claims; Medicare Prescription Drug Plan Members – pharmacy claims. Data include: medication utilization and adherence, hospital admissions and procedures, physician office visits and procedures, diagnosis codes, laboratory procedures, laboratory results, disease state information.

United Health care: Optum Database (Optum, 2014)
Patient Population: 114 million commercial and Medicare lives.
Type of Data:
Affiliated single-payer administrative data (123.1 million lives)

- Commercial members with medical and pharmacy benefits: 57.2 million, 2.7 average years of enrolment per member, 81% with one or more medical claims, 72% with one or more pharmacy claims, 44% with a lab procedure also have a result.
- Other commercial members with medical and pharmacy benefits: 8.0 million, 2.2 average years of enrolment per member, 74% with one or more medical claims, 70% with one or more pharmacy claims, 41% with a lab procedure also have a result.
- Commercial members with medical benefits only: 39.2 million, 3.0 average years of enrolment per member, 72% with one or more medical claims, 41% with a lab procedure also have a result.
- Medicare members with medical and pharmacy benefits: 4.7 million, 2.7 average years of enrolment per member, 92% with one or more medical claim, 81% with one or more pharmacy claim, 40% with a lab procedure also have a result.
- Other Medicare members with medical and pharmacy benefits: 1.9 million, 4.0 average years of enrolment per member, 88% with one or more medical claims, 83% with one or more pharmacy claims, 43% with a lab procedure that also have a result.
- Medicare PDP members with pharmacy benefits only: 12.1 million, 3.2 average years of enrolment, 89% with one or more pharmacy claims.

Non-affiliated multi-payer administrative data (34.6 million cumulative lives)

- Members with medical and pharmacy benefits: 27.5 million lives, 3.2 average years of enrolment per member, 82% with one or more medical claims, 70% with one or more pharmacy claims, 20% with a lab procedure also have a result.
- Members with medical benefits only: 7.1 million, 2.6 average years of enrolment per member, 71% with one or more medical benefit, 11% with a lab procedure also have a result.

Non-affiliated multi-employer administrative data (7.2 million cumulative lives)

- Commercial members with medical and pharmacy benefits: 7.2 million, 2.7 average years of enrolment per member, 88% with one or more medical claim, 82% with one or more pharmacy claim, 17% with a lab procedure also have a result, 9% with attendance at work have STD, LTD or WC claim.

Clinical EMR/EHR data (24.5 million cumulative lives)

- All clinical lives: 24.5 million, 21% with a drug administration, 60% with a lab result, 81% eligible for physician notes, 89% with a DX, 61% in an IDN, 69% with vitals.
- Clinical lives linked to administrative data: 3.1 million, 13% with asthma/COPD, 32% with hypertension, 2% with prostate cancer, 14% with diabetes, 5% with CHF, 16% with MI/stroke.

Additional linkable data:

- Health risk assessment data: 3.4 million, 99% height, weight, BMI, BSA, 92% tobacco use, 90% alcohol use, 84% physical/emotional problems, 99% health status, 97% stress, 88% pain, 79% sleep.
- Mortality data: 1.9 million (date of death).
- Pharmacy benefit design data: 53.9 million, tier levels, types, and cost-sharing, premiums, deductibles, copays, coinsurance, limits, thresholds, mail order benefits.
- Oncology management data: 85,000, 81% with stage at diagnosis, 84% with histology, 84% with current clinical status.
- Socioeconomic data: 44.5 million, 100% education attainment, 92% net worth, 68% household income, 99% language, 96% race/ethnicity, 78% home ownership.
- Primary data collection: 4.1 million, direct to patient survey, medical chart review, device information, NDI link, 66% with home phone number, 7% opted in to email contact.

Restrictions: Data cannot be merged with other data sources. Neither patients nor providers can be re-identified.

PUBLIC SECTOR

The U.S. Food and Drug Administration-Sentinel Initiative

In the fall of 2007, Congress passed the FDA Amendments Act (FDAAA), mandating FDA to establish an active surveillance system for monitoring drugs, vaccines, other biologics, and medical devices using electronic data from health care information holders (FDA, 2007). The Sentinel Initiative is the FDA's response to that mandate. Its goal is to build and implement a new active surveillance system that will eventually be used to monitor all FDA-regulated products. This system will link existing automated health care data from multiple sources to actively monitor the safety of medical products continuously and in real-time.

The Mini-Sentinel is a pilot project used to inform and facilitate development of full active surveillance system. Mini-Sentinel is comprised of routinely collected

administrative claims data, outpatient and inpatient electronic health records (EHRs), demographic information, outpatient pharmacy dispensings, and registry data from 18 participating data-partners, including some of the largest private health plans in the United States. These data are created or acquired through the normal business activities Through Mini-Sentinel, FDA has the capability to better understand the safety outcomes using electronic health care data of approximately 178 million covered lives (FDA, 2014). This accumulation of data represents 358 million person-years of observation time and 4 billion of prescription dispensings.

The Mini-Sentinel Collaborating Institutions include both Data and Academic Partners that provide access to health care data and scientific, technical, methodologic, and organizational expertise as needed to meet the requirements of the project. Representatives of the Collaborating Institutions participate in various capacities, including as members of the Planning Board, the Safety Science Committee, the Project Operations Committee, the Data, Methods, and Protocol Cores, and workgroups engaged in specific projects and other Mini-Sentinel activities.

Data Partners retain stewardship and possession of both original source data and data transformed into Common Data Model format. The Common Data format is a data structure that standardizes administrative and clinical information across Data Partners. It relies on existing standardized coding schema (e.g., ICD-9-CM, HCPCS/CPT and NDC) to minimize the need for ontologic mapping and enable interoperability with appropriate evolving health care coding standards and is compatible with other common data models using the same data types. Data Partners manage and store the data in accordance with their own institutional policies.

Table 2 Mini-Sentinel Collaborating Institutions as of January 1, 2014 * Indicates data partners

Aetna: Aetna Informatics*	Kaiser Permanente Center for Effectiveness and Safety Research • Kaiser Permanente Colorado* • Kaiser Permanente Hawaii* • Kaiser Permanente Mid-Atlantic* • Kaiser Permanente Northern California* Kaiser Permanente Northwest*
America's Health Insurance Plans: Clinical Affairs Department	OptumInsight, Inc.*
Brigham and Women's Hospital: Division of Pharmacoepidemiology & Pharmacoeconomics in the Department of Medicine	Outcome Sciences, Inc., a Quintiles company
Cincinnati Children's Hospital Medical Center: James M Anderson Center for Health Systems Excellence	Rutgers University: Center for Health Services Research on Pharmacotherapy, Chronic Disease Management and Outcomes at the Institute for Health, Health Care Policy and Aging Research
Columbia University: Department of Statistics	University of Alabama at Birmingham: Center for Outcomes and Effectiveness Research and Education
Critical Path Institute	University of Illinois at Chicago Medical Center: Departments of Pharmacy Administration, Pharmacy Practice, General Internal Medicine, and Biostatistics
Duke Clinical Research Institute	University of Iowa: Department of Epidemiology in the College of Public Health
HealthCore, Inc.*	University of Pennsylvania School of Medicine: Center for Clinical Epidemiology and

	Biostatistics and Department of Biostatistics and Epidemiology
HMO Research Network • Group Health Research Institute* • Harvard Pilgrim Health Care Institute* • HealthPartners Institute for Education and Research* • Henry Ford Health System: Public Health Sciences Department* • Marshfield Clinic Research Foundation* • Meyers Primary Care Institute*	Vanderbilt University Medical Center*
Humana Comprehensive Health Insights, Inc.*	Weill Cornell Medical College: Department of Health care Policy and Research

The National Patient-Centered Clinical Research Network

Since 2013, the Patient-Centered Outcomes Research Institute (PCORI) has committed over $100 million to the development of PCORnet: The National Patient-Centered Clinical Research Network which was designed to be a large nationally representative network for conducting CER in "real time" and in "real-world" settings (PCORI, 2014b). Electronic health record data, claims data, and other patient-generated data will be collected and stored in a standardized, interoperable Common Data Model under rigorous security protocols, and data sharing across the network will be accomplished using a variety of methods that ensure confidentiality by preventing patient identification (PCORI, 2014a). Access, use and data privacy policies are still currently being developed. PCORnet is currently comprised of 29 health data networks and a Coordinating Center.

The NIH Collaboratory Distributed Research Network

The NIH Collaboratory Distributed Research Network is a network developed to improve the conduct of clinical trials, particularly pragmatic clinical trials (Richesson et al., 2013). The network is composed of a registry of data partners who list detailed information about their health system, data sources, and preferences for collaboration. Investigators who wish to collaborate identify and contact the data partners using the registry listing. If data partners agree to collaborate, then the organizations holding data can allow secure distributed querying of their research datasets by individuals whom they authorize on a case-by-case basis. Results that are returned are often aggregate results, without confidential or proprietary data. The level of data sharing is determined in advance as part of the collaboration agreement. Features of the network include:

- Data partners are able to maintain possession of, and analyze, their own data;
- Data partners are able to provide results, not data, to their external collaborators;
- Data partners have complete control over both the individuals or organizations with whom they collaborate, including those from whom they accept queries, as well as the queries they accept.

The Surveillance, Epidemiology, and End Results

The Surveillance, Epidemiology, and End Results (SEER) program funded by the National Cancer Institute (NCI) is a network of 18 population-based cancer registries covering approximately 28 percent of the U.S. population (NCI, 2014). This registry system, which began as in1973 collects data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status. The registries identify cases prospectively at the site of clinical care and follow

up either actively through continued clinical care or passively through data linkages to vital statistics databases such as the national death index or the social security administration. The registries then de-identify these cases and send the data to NCI for compilation, quality checking, and assembly into public use data sets. Data generated from the SEER program is freely available for analysis, however the program requires users to comply with a data use agreement. Key features of the agreement are summarized below:

- All research results must be presented or published in a manner that ensures that no individual can be identified
- There must be no attempt either to identify individuals from any computer file or to link with a computer file containing patient identifiers
- The data provided by SEER must not be shared with any person except those who have signed the data use agreement

Health care Cost and Utilization Project

The Health care Cost and Utilization Project (HCUP) is a collection of databases developed through a Federal-State-Industry partnership and funded by the Agency for Health care Research and Quality (AHRQ, 2014). HCUP aggregates data collected by state and Federal-level data organizations since 1988 to create a uniformly formatted national information resource of discharge-level health care data. HCUP contains administrative data and contain encounter-level, clinical and nonclinical information including diagnoses and procedures, discharge status, patient demographics, and charges for all patients, regardless of payer. Information provided in HCUP data sets are consistent with the definition of "limited data sets" under the HIPAA Privacy Rule and contain no direct patient identifier. The collection of databases includes the National Inpatient Sample, Kids' Inpatient Database, Nationwide Emergency Department Sample, State Inpatient Databases, State Ambulatory Surgery and Services Databases, and State Emergency Department Databases.

Centers for Medicare & Medicaid Services

The Centers for Medicare & Medicaid Services (CMS) provides a collection of over 100 datasets collected on the population served by the CMS (ResDAC, 2014). These datasets include patient level institutional, non-institutional and medication administrative claims, beneficiary surveys, quality of care indicators, patient demographics and enrollment, and provider, and facility characteristics. Notably, many datasets are available at three privacy levels: Public Use Files (PUFs), Limited Data Sets (LDSs), and Research Identifiable Files (RIFs). PUFs have been edited to contain only completely de-identified beneficiary information. As such, PUFs contain only aggregate level data on Medicare beneficiary or provider utilization. PUFs are freely available from the CMS website, may be freely shared, and do not require a data use agreement. LDSs do contain patient-level protected health information, but particular variables have been removed or edited. For example the LDSs only provide geographic locations at the state or county level, not at the zip code level. Beneficiary ages are provided in 5-year ranges and no physician identifiers are provided; although institutional facilities can be identified. In order to access LDSs, the data requestors must submit an application, pay a fee and establish a Data Use Agreement. RIF data contain beneficiary level protected health information. Although no direct identifiers (social security numbers or Medicare numbers) are provided, enough information is provided that beneficiaries could potentially be identified for example, by date of birth, zip code, exact dates of service. RIF data can only be

requested for research purposes and not for commercial purposes. Applications require CMS Privacy Board review and approval prior to being released. Files contain a unique, encrypted beneficiary ID that allows linkage across files and across years. LDS data contain less beneficiary information than RIF data and have encrypted professional provider IDs. RIF data may be linked to the Health and Retirement Survey a longitudinal study of health, retirement, and aging. This linkage requires approval from both CMS and Health and Retirement Survey privacy boards(ResDAC, 2013).

Table 3 Summary of Privacy Level Differences. Adapted from the Research Data Assistance Center

	Public Use Files	Limited Data Sets	Research Identifiable Files
Beneficiary level data?	No	Yes	Yes
Data files customizable to specific cohort?	No	No	Yes
Linkable to non-CMS data at the beneficiary level?	No	No	Yes
Require Privacy Board Review?	No	No	Yes
Require a DUA	No	Yes	Yes

d. Data linking. To what extent can / are patient data linked across databases – how and by whom? Who are the major organisations involved?

PRIVATE SECTOR

Data-linking most commonly occurs by the information providers themselves; these information providers often have restrictions around the linkage of data information across different sites.

PUBLIC SECTOR

The U.S. Food and Drug Administration-Sentinel Initiative

In order to best protect patient privacy, the data from each partner is maintained behind each individual health plan fire-wall (Table 2) (FDA, 2014). This "distributed data" approach allows a single coordinating center to distribute FDA safety questions in the form of "queries," to each of the participating data partners to be run against their own data. Further, Data Partners do not share direct patient identifiers with the coordinating center and adhere to the HIPAA minimum necessary standard. Data are provided by Data Partners in summary form, unless there is a specific need for person-level information for example, information (stripped of direct patient identifiers) regarding individuals who received specific vaccines on specific dates when such information is required to respond to a particular FDA query. Data Partners execute the standardized data queries and then share the output of these queries, with the coordinating center for final analysis.

The NIH Collaboratory Distributed Research Network

For this network, the level of data sharing is determined in advance as part of the collaboration agreement. Features of the network include:

- Data partners are able to maintain possession of, and analyze, their own data
- Data partners are able to provide results, not data, to their external collaborators
- Data partners have complete control over both the individuals or organizations with whom they collaborate, including those from whom they accept queries, as well as the queries they accept

Centers for Medicare & Medicaid Services

RIF may also be linked to the SEER dataset to compare claims and enrollment in cancer patients. Access to RIF data requires a formal data request to CMS. A request, if approved, will establish require a Data Use Agreement with CMS which is applicable on a per study basis.

e. Data access. To what extent is data shared, with whom is it shared, how does permitted access differ according to organisation (i.e. access by pharmaceutical companies versus access by public bodies / academic institutions), what are the processes involved in being granted permission to access data, what are the costs involved in data access (where available), and in what form is data access granted (e.g. raw data / in-house data analysis services only?

PRIVATE SECTOR

Data access is generally permitted on a case-by-case basis and requires payment to the organization that owns the data. Costs of accessing the data generally differ for research conducted by non-profit versus for-profit organizations. Additionally, companies may require pre-approval of research findings prior to publication.

PUBLIC SECTOR

The U.S. Food and Drug Administration-Sentinel Initiative

The FDA obtains unlimited rights to access and to use all Mini-Sentinel data in the possession of the Operations Center. Access to the non-summarized data is limited to authorized individuals within the coordinating center. Data transfer between Data Partners and the Operations Center and between the Operations Center and the FDA is done by means of a secure web-based file sharing system. The Operations Center complies with the standards established by the HIPAA and the Federal Information Security Management Act of 2002 (OCR, 1996; US Congress, 2002).

3. Collecting de novo patient data

a. Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data (i.e. setting up registries, pragmatic clinical trials, etc.).

PUBLIC SECTOR

As indicated for routinely collected data, collecting de novo patient data for human subjects research conducted, supported or otherwise subject to regulation by Federal departments and agencies is covered under the Common Rule. The collection of new data may involve greater than minimal risk therefore more through assessments of the

risk/benefit criteria in the Common Rule may be required. Research is considered to have greater than minimal risk when the probability and magnitude of harm or discomfort anticipated in the proposed research are greater than those ordinarily encountered in daily life, or during the performance of routine physical or psychological examinations or tests. Risks of harms may be physical, psychological, social, or economic. In research presenting more than minimal risk, potential subjects must be informed of the availability of medical treatment and compensation in the case of research-related injury, including who will pay for the treatment, and the availability of other financial compensation.

> **b. Research application process.** Process by which application for new data collection is considered, and governing principles of the committees that grant approval.

PUBLIC SECTOR

Applications for human subjects research require approval by the IRB institution with which they are affiliated .The Common Rule requires that IRBs carefully consider whether the risks of research are reasonable in relation to anticipated benefits. In research involving an intervention expected to provide direct benefit to the subjects, a certain amount of risk is justifiable. In research where no direct benefit is anticipated, the IRB must evaluate whether the risks presented by procedures performed solely to obtain generalizable knowledge are ethically acceptable.

4. Data use. What are the rules governing the use of RWE?

PRIVATE SECTOR

The rules around data use vary by companies that own the data, but generally include restrictions around trying to identify individuals from de-identified data and linking disparate data sources. Rules and restrictions may exist around publication of results from data analysis of data from these sources – e.g., review and approval by the data owner prior to publication.

PUBLIC SECTOR

The US Food and Drug Administration-Sentinel Initiative

Mini-Sentinel Collaborators, both institutions and individuals, retain all rights and privileges, including those of patent and copy, to all data and materials they owned prior to engagement in the Mini-Sentinel pilot. Collaborators are allowed to use any non-confidential and non-proprietary summarized Mini-Sentinel data in presentations and publications. It is important to point out that the Mini-Sentinel activities are considered public health practice and not research. Therefore the Common Rule does not apply and it is therefore not necessary for the Collaborating Institutions to obtain approval from their respective IRBs or Privacy Boards, or to obtain waivers of authorization under HIPAA, to participate in Mini-Sentinel (DHHS, 2009; OCR, 1996). This approach is in contrast to other RWD distributed networks which require IRB oversight (HMO, 2011).

Centers for Medicare & Medicaid Services

Major elements of the LDS and RIF dataset data use agreement include the following:

- CMS retains all ownership rights to the data;

- Data may only be used and retained up to the agreed date;
- Users must establish appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access;
- Users must not contact any individuals in the datasets;
- Users must grant access to the data to CMS representatives for the purpose of inspecting to confirm compliance with the terms of this agreement;
- Users must not to disclose any information which may result in an individual's identity being revealed;
- All published findings are limited to patient de-identified data that conform with the HIPAA Privacy Rule.

Health care Cost and Utilization Project

Data generated from HCUP is available for purchase, however the project requires users to comply with a data use agreement. Key features of the agreement are summarized below:

- Do not attempt to learn the identity of individuals and prohibit others from doing so;
- Do not publish or report on the identities of individual hospitals or health institutions;
- HCUP raw data may be shared only with immediate research group only if each member of immediate research group has signed the data use agreement;
- Linkages to external datasets to enhance analyses are allowed but must not be result in identification of individuals;
- Data may not be used for commercial or competitive purposes involving those individual establishments reported in HCUP.

5. **Suggested principles or guidance for data governance, and the adapting environment for such.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

Institute of Medicine

The Institute of Medicine (IOM) is an independent, non-profit organization established in 1970 as the health arm of the National Academy of Sciences. The IOM works to provide unbiased and authoritative advice to decision makers and the public. In cooperation with the Office of the National Coordinator for Health Information Technology, the IOM convened a roundtable released a report summarizing the role of the digital health data systems a learning health system (Grossman and McGinnis, 2011). The report points out the need of complying with patients' expectations for data de-identification to avoid embarrassment or economic loss. This report proposes that the protection of health data should include: techniques for de-identification; a defined process for determining trustworthiness of data recipients; and a physical security protocol for the location in which the data will reside.

The IOM has also produced an interim report on data sharing in the clinical trial setting (IOM, 2014). The report proposes several data sharing models in which the level of data

access is inversely related to the confidentiality of the information: (1) open access (2) controlled access to one or more organizations' data (3) closed consortium.

The open access model allows the data to be broadly available to the public through an open-access website and no data use agreement will be necessary. This data will be provided as summarized or de-identified individual patient data.

The controlled access data model also allows access to members of the public via a website. This data is provided as summarized, de-identified individual patient data or limited datasets. However, in order to access this information, the user must sign a data use agreement, provide and adhere to a publication plan, must not use data for commercial purposes, must not attempt to re-identify data, inform the data generator about any safety concerns identified, and agree that the data generator retains exclusive rights to inventions or other intellectual property generated by the data recipient. The data user must also be registered, demonstrate that they have the needed expertise to analyse the data, disclose funding sources, provide a purpose for the data request that meets specified criteria, or provide an analysis plan.

Information shared within the closed consortium data model is only available to members of the consortium. The shared data may be summarized, de-identified or identifiable individual patient data. The members of the consortium (either individuals or organizations) who wish may use the data may be required sign a partnership/consortium agreement, an agreement governing rights to inventions or other intellectual property generated by the data sharing activity, or an agreement dictating whether and how future publications will be undertaken.

Indiana Health Information Exchange

Indiana Health Information Exchange (IHIE) is a non-profit organization which operates the nation's largest health information exchange, providing a secure and robust statewide health information technology network that connects over 90 hospitals, long-term care facilities, rehabilitation centers, community health clinics and other health care providers in Indiana. IHIE espouses 6 key attributes of data governance summarized in Table 4.

Table 4 Adapted from IHIE

Attribute	Description
Availability	The data must be available to the applications of all relevant users when needed
Accessibility	The data must be accessible regardless of the applications used
Interoperability	The data must be both semantically and syntactically interoperable across systems
Auditability	There must be a trail of the data from its source to its destination
Quality	The data must be accurate and complete
Security	The data must be kept secure

Adapted from IHIE (IHIE, 2012)

American Medical Informatics Association

The American Medical Informatics Association (AMIA) is a professional scientific association that aims to transform health care "through trusted science, education, and the practice of informatics" (AMAI, 2014). In 2012, AMIA held it's 7th annual Health Policy Meeting to discuss key challenges for data use, and consider topics such as data stewardship principles and effective approaches to reduce or eliminate data silos and protect patient privacy. The key principles of health data use generated in this meeting were (Hripcsak et al., 2013):

1. Access to and use of health data should be viewed as a public good. Data should be available and 'fit-for-use', with proper security, for appropriate purposes beyond direct patient care.
2. Health data must be consistent, comparable, timely, accurate, accessible, complete, and reliable as possible. Users must be able to track the degree to which the data have attained these attributes. Understanding the context and the provenance of the data is also critical in determining their 'fitness for use'.
3. Integration and sharing of health data that currently reside in silos are necessarily for the optimal use of the data.
4. The rights and responsibilities of everyone (including patients, families, providers, researchers, payers, and organizations) involved in collecting and using health data must be understood and respected.
5. Data uses must be transparent to all, including patients and their agents.
6. The potential benefits of data use must be weighed against the potential risks and costs of loss or inappropriate disclosure of personal health information.
7. Data stewards (those who collect, maintain, aggregate, analyse, and use health data) must demonstrate that they understand and are willing to assume the responsibilities of effective stewardship in order to earn and retain the support of patients and the public. Data stewards must demonstrate that they use data appropriately and in accordance with applicable laws and regulations.
8. Data use policies should not be so binding that they restrict or prevent uses of data from emerging technologies or impede as yet unknown data sources or technologies.
9. All health care system stakeholders must continue to study the benefits and risks of new data sources and uses and to refine data use principles as needed.

Other concepts that were generated form the meeting include:

- Continuous use of data: "To the extent possible, data should be collected once and used continuously. Ensuring data quality and certainty, data collectors, data stewards, and data aggregators must help assure that data are available continuously for appropriate querying and uses."
- Understand the risks of data use: "A coordinated strategy and action plan should address intellectual property, ethical proprietary, and commercial issues such as organizations' reluctance to share data, and concerns around the sale of data. It should also tackle emerging public policy complexities in the area of data use arising from widespread adoption of technology-based advances such as EHRs, exchange of data via the cloud, and data becoming available beyond the point of care (e.g., mobile devices, biomedical sensors, genomic data, social media)."

Office of the National Coordinator Health IT Governance Workgroup (Tang, 2010)

In 2010, the Health IT Policy Committee charged a Governance Workgroup to make recommendations regarding the mandate in the Health Information Technology for Economic and Clinical Health (HITECH) Act that the Office of the National Coordinator (ONC) establish a governance mechanism for the nationwide health information network (NW-HIN). An important strategic goal of the Office of the National Coordinator (ONC) is to enable a wide range of innovative and complementary approaches that will allow secure and meaningful information exchange within and across states, grounded in a common foundation of standards, technical specifications, and policies.

The workgroup made the following recommendations related to principles for NW-HIM governance:

1. Transparency and openness
2. Inclusive participation and adequate representation
3. Effectiveness and efficiency
4. Accountability
5. Federal governance and devolution
6. Clarity of mission and consistency of actions
7. Fairness and due process
8. Promote and support innovation
9. Evaluation, learning and continuous improvement

IMS Health (Busalacchi and Moyer, 2012)

IMS is a provider of information, services, and technology for the health care industry. In 2012, leadership at IMS published an article on preparing for big data and suggested the following as key components of data governance:

- Setting standards around data definitions and taxonomy, metrics and measures, technology and tools, and reference data.
- Determining policies and processes related to data definitions, monitoring, measurement, change management, and access and delivery.
- Establishing organizational readiness to include defining roles and responsibilities, identifying training requirements, and applying change management techniques.

National Committee on Vital and Health Statistics (NCHVS, 2009)

The National Committee on Vital and Health Statistics (NCVHS) is the Department of Health and Human Services' "public advisory body on health data, statistics, and national health information policy". In its primer on health data stewardship, the committee suggests the following practices related to data stewardship: "transparency about use; identification of the purpose for data use; participation of individuals; security safeguards and controls; de-identification (when relevant); data quality, including integrity, accuracy, timeliness, and completeness; limits on use, disclosure, and retention; oversight of data uses; accountability; and enforcement and remedies."

Summary

The use of patient information in the U.S. is governed by a number of Federal Acts, principally HIPAA. These acts define the process for use of data for research, including patient consent, instutional review and de-identification. A central factor of how the regulations affect RWD collection and use is the determination of whether the RWD is conducted by a covered entity to facilitate treatment, payment, or health care operations. Any other disclosures of PHI (such as for research) require the covered entity to obtain written consent from the individual for the disclosure, IRB approval and/or a process for de-identification.

Despite these constraints, the U.S. is a prolific producer of RWD, both in the public and private sector. RWD is available as pharmacy and medical claims, electronic health records, disease registries, and hospital records. Consistent with the diversity of the U.S. health care system, access to the data varies greatly based on the characteristics of the data producer and the data user. For example a number of private data providers will grant access to their de-identified data to any user willing to pay for access, but the fees may vary based whether the use is by academic, industry or government organizations. In contrast other data providers will only grant access to members of the data-sharing consortium with explicit agreements amongst the members governing rights to intellectual property and publications.

The production of RWD is expected to increase with the passage of the Patient Protection and Affordable Care Act which increased the number of Americans with health insurance coverage and therefore the expected use of health services. A number of organizations advocate for greater translation of RWD into RWE through more standardization of data definitions, more timely access and better data quality.

5. France

1. **Brief overview** of the health system and collection / management of patient data

Health care in France is managed by a national programme of statutory health insurance, which is a branch of the wider social security system (sécurité sociale). Health insurance is publicly financed through employee and employer payroll contributions and taxes. Reimbursement by statutory health insurance takes place after direct payment has been made by the patient, with schemes for low-income groups guaranteed universal access to care. There is voluntary health insurance (VHI) to cover treatments beyond those provided by the national scheme and to cover most out-of-pocket payments; approximately 90% of the population have access to a voluntary health insurance plan (Green et al., 2013).

Health professionals are required to apply official rates for their service that are set out in agreements, though some doctors have the right to exceed these official charges, for example those who have opted for the so-called 'second sector' which has variable fees (Coudert, 2008). On the other hand, doctors working in hospitals are state employees whose condition of employment is similar to civil servants. Hospitals can be public, private non-profit, or private for-profit.

All residents of France are entitled to national health insurance, and all residents are issued with a 'carte vitale' which is a plastic card which indicates national insurance

rights and enables the government to credit patients immediately. The carte vitale is embedded with a chip containing address, social security details, etc.

The French scheme for electronic health records was set in place by law in 2004, but uptake / roll-out has been slow. The system was formally launched in 2011 after a first pilot phase in 2006; coverage is still very low – less than 1% (De Lusignan and Seroussi, 2013; Lantieri and Pelsy, 2014). The Agence des Systèmes d'Information Partegés de Santé (National Agency of Health Shared Information Systems: ASIP Santé) since 2009 has responsibility for setting operability standards for EHRs and agreements with data custodians (OECD, 2013); the committee is composed of representatives from industry, patients, health professionals, and people with legal expertise.

The National Health Authority, Haute Autorité de Santé (HAS), was established in 2004 to coordinate and undertake a number of activities to improve patient care and ensure equity in health care. HAS assesses medicines, medical devices and procedures, produces guidelines, and accredits health care organisations and doctors (Green et al., 2013).

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. Routinely collected patient data.

Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

Due to the insurance-based system for health care in France, routinely collected data generated by claims can provide useful sources of information. For health care activity in the ambulatory and primary care setting (and of independent practitioners in private hospitals), the Système National d'Information Inter-Régime de l'Assurance Maladie (SNIIRAM) is the national health insurance database of claims data paid by the Social Security System (HDN, 2015). This contains data on consultations (but not their content), procedures, dispensed drugs, diagnostic tests (but not their results), medical devices, and personal patient data (age, birth date, gender, region of residence, long-term / chronic disease diagnoses, low income indicator and date of death). Data is only available for to the past three years.

For the secondary care setting, the Programme de Médicalisation des Système d'Information (PMSI) is the national hospital discharge database containing details of hospital activity and claims data. This is reflective of the activity-based payment system in France for claims paid by the Social Security System for all public and private hospitals. Data include discharge data and diagnoses (ICD-10 codes), medical procedures, length of stay, and medicines and medical devices (Fagot, 2012). However, neither SNIIRAM nor PMSI contain data of over-the-counter drugs, results of clinical exams or laboratory tests, drug use in a hospital stay apart from the most costly and necessary drugs (from a specific list), and cause of death. An important health survey in France is the Enquête Santé et Protection Sociale (ESPS). The survey collects information on health status, health care coverage and access, social determinants, as well as health care needs and demands (HDN, 2015). There is a collaboration between

ASIP Santé and the National Institute for Cancer to build a database of shared oncology records (OECD, 2013).

The Commission Nationale de l'Informatique et des Libertés (CNIL) is the French national data protection authority, an independent administrative authority which also assesses and approves access to projects requiring access to personal health data. CNIL is authorised in this role by the Loi Informatique et Libertés Act N°78-17, 1978 (Data Protection Act) (CNIL, 1978). The Data Protection Act has since been amended by the Act of 6 August 2004 relative to the protection of individuals with regard to the processing of personal data, and the Act of 13 May 2009 relative to the simplification and clarification of law and lighter procedures. There are two Chapters in the Act which are particularly relevant for our purposes:

- Chapter IX: Processing of Personal Data for the Purpose of Medical Research
- Chapter X: Processing of Personal Medical Data for the Purposes of Evaluation or Analysis of Care and Prevention Practices or Activities

Chapter IX (Articles 53 – 61) describe the rules around the use of data for medical research, and do *not* apply to the direct care of patients, nor do they apply to studies conducted by and for staff involved directly in the care of the patient[6]. Authorisation for research using personal data must be granted by CNIL, and any research using data that allows identification of individuals must be codified before transmission, *unless:* the study is for pharmacovigilance, is part of a research agreement in the context of a national or international co-operative study, or if there is some distinctive feature of the research that requires it. If granted, the person in charge of the research must be compliant with security arrangements for the data, and is bound by a duty of confidentiality (sanctions apply). Whilst the duty of confidentiality is exempt for health care data (i.e. requirement to obtain patient consent), patients have the right to object to this. In addition, according to Article 57, all individuals from whom medical data is collected must be informed individually of the nature of data transmitted, the purpose, details of the entities receiving the data, the right of access, and the right to object to data being processed without obtaining their consent.

Chapter X (Articles 62 – 66) describes the provision on law for using data from medical files for the purposes of statistics of evaluation or analysis of practices and activities of care and prevention, but only in aggregate form or individual form but where data subjects cannot be identified. However, the chapter does not apply to the processing of personal data for the purpose of reimbursement by health insurance schemes. According to Article 64, personal data should not be attached with any names *or* their social security number, 'NIR'. As mentioned below, in France, all citizens have a Numéro de'intification au répertoire (NIR), which is their social security number.

The Health Insurance Reform Act (Act 2004-810 of 13 August 2004) specified the parliamentary control over the health care system, as well as clarifying the respective roles of the state and the health insurance system (European Observatory on Health Systems and Policies, 2015). Of note, it put into place the EHR scheme for France which, as described, is still in its early phases. The law stipulates that if a patient has an EHR,

[6] "The processing of personal data for the purpose of therapeutic or individual medical follow-up of patients shall not be subject to the provisions of this Chapter. The same shall apply to processing that allows the carrying out of studies based on the data obtained if these studies are carried out by the staff responsible for the follow-up of patients and are intended for the exclusive use of the staff" (CNIL, 1978: Article 53)

the health care professional associated with their care *must* refer to it and complete it (OECD, 2013). In addition, the same law mandates health care providers to use an international clinical terminology standard, SNOMED 3.5 vf, and to adopt CDA HL7/CDA R2 interoperability standards (OECD, 2013). The structure format of this data entry should be facilitative for good quality health care monitoring or research. It is difficult to know whether there are any particular quality issues for EHRs as it is early in their implementation. The Agence Nationale de la Sécurité de Systèmes d'Information (ANSSI) conducts security audits, whilst data confidentiality audits are conducted by CNIL.

There is no 'minimum dataset' in France collected as part of the EHR. Patients must provide explicit consent for an EHR to be kept for the patient, and the patient has access to this record, as well as being able to monitor which health professionals have accessed it (Artmann et al., 2015) and to specify the elements of their record that can be shared (OECD, 2013). This is facilitated through the requirement that physicians use thier 'Carte de Professionel se Santé' (health professional's card) to establish a connection with the care record which permits control over access (Lantieri & Pelsy, 2014). However, there are legal provisions to allow physician access in cases of emergency where the patient is incapable of consenting (OECD, 2013). In addition, free text may be added by the patient to their own record, labelled "patient's personal expression" (OECD, 2013). Article 34 of the Health Insurance Act specifies that ePrescriptions can be transferred electronically to a pharmacist. A modification to the Act in 2007 permits a pharmaceutical care record for all beneficiaries of social health insurance, provided the patient has consented (Artmann et al., 2015). France is not yet building datasets from routinely collected electronic health records, but plans to do so in the next phase of the national EHR strategy (OECD, 2013).

Another relevant piece of legislation in France is the Public Health Code (Code de la Santé Publique), which was reengineered significantly in 2002 to clarify all aspects of medical law, including the rights of patients, the obligations of physicians and other health care professionals, as well as fully integrating the Medical Code of Ethics (Coudert, 2008) L.161-36-4-2 (part of this code) entrusts the implementation of the pharmaceutical record 'Dissier pharmacetique' to the College of Pharmacists.

In 2011, a public health scandal in France brought the issue of patient data collection and assessment to the fore. The medicine 'Mediator', prescribed for diabetic care as well as weight loss, was found to be associated with cardiovascular deaths. This highlighted the importance of using routinely collected data to monitor health consequences of prescription medicines and also the problems associated with lack of a unique patient identifier. In reaction, a decree was published in December 2011 which facilitates access to national primary care data (SNIIRAM) for public health organisations as well as research centres working in the field of public health (OECD, 2013).

b. Collecting de novo patient data.

Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

In France there are 39 regional Ethics Committees: 'Comités de Protection des Presonnes' (CPP) distributed amongst seven regions (EUREC, 2015a). In accordance with

the Law of Public Health (Loi de Santé Publique, 2004) and the Bioethics Law (Loi de Bioethique, 2004-2011), these Ethics Committees are responsible for the approval of *interventional* studies, standard of care studies, medical and other health products and further research areas such as genetics or physiology. Committees are composed of 28 members which represent a variety of disciplines or backgrounds including biomedicine, ethics, law, social sciences, medical and patients. On a country-wide level there is the National Consultative Ethics Committee which considers controversial issues arising from the progress of life sciences, and issues public statements and recommendations (EUREC, 2015a).

The Agence Nationale de Sécurité du Médicament et des Produits de Santé (ANSM) is the regulatory authority for health in France, to which all applications for interventional studies must be submitted, and a document completed for the national register of authorized clinical trials (ANSM, 2015). All other studies, including observational studies, must be registered with ANSM via the website (Roggemans, 2012).

According to Roggemans (2012) registry and observational studies should not be submitted to the regional Ethics Committees (CPPs) but instead to the Comité Consultatif sur le traitement de l'Information en matière de recherche dans le domaine de la santé (CCTIRS – the advisory committee on the handling of health information for research). The CCTIRS considers proposals for research prior to their submission to the CNIL, and assesses the research methodology and the relevance of the registered personal data for the purpose of the research proposed (CCTIRS, 2015). The committee meets monthly, and returns a decision to the applicant within one month from data of receipt (in 90% of cases).

3. **Data linking.** To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

Linking requires a unique patient identifying number. In France, all citizens have a Numéro de'intification au répertoire (NIR), which is a social security number and is used by medical authorities for the issuance of a "carte vitale" for insurance purposes. However, these are not used by hospitals as they were considered to be too sensitive to be used for electronic medical records (OECD, 2013). Identifying numbers used by hospitals still vary, which can act as a barrier to data linkage projects as well as impede the sharing of information between hospitals. In 2007 the development of national identifying numbers for medical records was approved by law, as a way to de-risk the use of highly sensitive NIR numbers (CNIL, 2007; OECD, 2013). This is called the 'Identifiant National de Santé' (INS), and is created through a hashing algorithm which uses the patient's name, birthdate and NIR to come up with a new number (Lantieri & Pelsy, 2014). To enable the matching of people with their medical record, it is being suggested that either a third party holds the key to match health insurance record (with an anonymised NIR) with medical records with the new INS health identifying number, or to have the insurance system adopt the new same INS identifying number (OECD, 2013). A report published in 2012 by High Council for Public Health (Haut Conseil de la Santé Publique – HCSP) suggests a change in line with the latter, which would impede matching of records with non-health databases (Haut Conseil de la Santé Publique, 2012).

A report by the OECD suggests that in France projects involving linking several databases are undertaken on a regular basis, in particular linking primary care data (SNIIRAM) to data on in-patient hospitalisations (PMSI) and survey data (ESPS); however, lack of access seems to be an important issue. Health care quality indicators in France are in development, which may involve the regular linking of databases in the future (OECD, 2013).

With regards data linking across national borders, under French law CNIL may approve projects involving sharing personal data with other EU countries, under the premise that similar protections for data security and the protection of privacy exist. Where a project involves sharing data between France and non-EU countries, the non-EU country must demonstrate an equivalent level of data security (OECD, 2013). Such projects have been passed, for example through a safe harbour agreement that was reached with a researcher in the U.S. which confirms that U.S laws offer similar protection of security and data privacy as those of EU countries.

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

Despite comprehensive claims data, research output in France, as measured by peer reviewed journals, is low compared with other countries. This is likely to be due mainly to the restrictive governance arrangements and data access models for third parties. There is a strong emphasis in France of protecting privacy of personal health data, and access to research purposes in limited (OECD, 2013).

The CNIL, the data protection authority, acts as a central decision-making body for approval of research projects that require personal health data (OECD, 2013). Its committee is composed of both medical experts as well as research experts, who are able to advise on the scientific merits of the proposed projects. The committee considers the legality of the request, the security measures and confidentiality protections that will be put in place to protect the data, as well as the legitimacy of the researcher; this includes an assessment of affiliation and whether this is with a 'credible' organisation (OECD, 2013). Non-government researchers, as well as seeking approval from CNIL must also be approved by the Conseil National de l'information statistique (CNIS) – the national council for statistical information.

As described, the rich information from the administrative insurance financing databases are able to provide strategic information on patient care pathways and their cost implications, especially when linked to other databases. However, the data are *not for public and research use* (HDN, 2015). Raw data is apparently complex and administrative-based, requiring a high level of expertise in the French social security payment and reimbursement system. However, according to IMS Health, access to PMSI data for commercial organisations has recently been relaxed, and examples of the databases' use within research projects can be seen through sponsored research studies such as PROSPERE and CONSTANCE (Hughes and Kessler, 2013). According to the open data commission, access has improved in recent years through the efforts of the CNAMTS (the French national health insurance fund) and the ATIH (the Agence Technique de l'Information sue l'Hospitalisation) (Touraine, 2014). According to the ATIH website, application to the national information system on hospitalisation is accessible only to: health facilities, regional health agencies and national organisations (public structures, hospital associations, and social security) (ATIH, 2015).

A report by the High Council for Public Health refers to two main barriers to the utilisation of health data for surveillance and research. The first is the judicial framework, which is very complicated and which depends on the nature of the organisation receiving data: health agencies, public statistics services, or not-for-profit private organisations (note the omission of 'private for-profit' organisations from this list). The second is the organisational and technical obstacles, namely that there is no single or group of organisations who have the responsibility and technical expertise required to manage data identified using the NIR (HCSP, 2012).

Although access to data is described by most commentators in France as difficult, the demand for RWD is strong, particularly given the French HTA process for medicines by HAS, which increasingly involves a value assessment for conditional reimbursement which is re-evaluated after three to five years based on further data collection in the real world.

5. Data use. What, if any, are the rules governing the use of RWD, including arrangements between data suppliers and recipients and rules around use for HTA

For most countries, guidance around the use of RWD has been scarce, save for the contractual arrangements between data providers and data recipients, which cover the required security measures and handling processes to ensure the confidentiality and integrity of the data that is transferred. As described, HAS—which is tasked with evaluating the medical/economic and public health benefits of medical treatments—often requests the gathering of additional data through post-registration studies in order to minimise uncertainty and facilitate re-evaluations. The responsibility for implementing these studies lies with the manufacturers of the medicine of device, and failure to perform these studies may result in regulatory or financial penalties.

In recognition of the challenges associated with these requirements, HAS has produced a document providing practical points of reference on the methodological aspects of post-registration studies (HAS, 2011). One of the major aims of these studies is to capture data on the effectiveness of a product in real-life conditions of use. Between 2004 and 2010 there have been 346 post-registration studies, which were largely epidemiological observational studies. Prior to implementation, the manufacturer must set up a scientific committee, and submit a protocol which must be evaluated by the National Committee for the Evaluation of Healthcare Devices and Healthcare Technologies (CNEDiMTS) for devices, or to the Transparency Committee (TC) for medicines. Study protocols are assessed against:

- 'The recommendations on professional ethics and good practice in epidemiology' by the Association of French language epidemiologists (ADELF, 2007);
- Recommendations on improving the quality of observational study reports in epidemiology (Strengthening the Reporting of Observational studies in Epidemiology – STROBE initiative) (Elm et al., 2007); and
- The principles of high quality research on comparative effectiveness (Good ReseArch for Comparative Effectiveness – GRACE Initiative) (GRACE Initiative, 2015).

The document defines two types of study to look at a medicine's conditions of use in real life situations: Cross-sectional studies without patient follow-up, or Prospective studies with patient follow-up. HAS refers to the use of routinely collected administrative claims

data (SNIIRAM and PMSI), but emphasises some draw-backs in these, namely that clinical information is not currently registered, and their use statistically is complex due to the administrative purpose nature of their collection (HAS, 2011). HAS does not discuss the governance issues or barriers for companies in *accessing* the data.

For assessing the impact of a healthcare product on morbidity/mortality, HAS describes four types of studies: Pragmatic trials, Observational Studies, Other types of epidemiological studies, and Modelling, as well as studies based on databases such as disease or intervention registries. Again, there is no discussion of the process of research application or data access; the focus, rather, is on the methodological aspects of the post-registration studies, for which it provides ann overview of its expectations.

See also (Barron et al., 2014) for a regulatory governance perspective on HTA in France.

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

In France the call for health care data and its management to be more open and transparent is strong. For example, a report by HCSP outlines the current obstacles for utilising administrative health data for public health and research purposes, noting in particular the impracticable requirement that any access to data containing the unique identifier NIR requires a decree from the Council of State. The report also highlights another risk on the horizon for health care data usage, which is the deployment of the 'INS' (Identifiant national de santé), which will be the only identifier used for health applications, which would make linking this data with other type of data sources impossible (HCSP, 2012). In addition, various initiatives both formal (e.g. (Touraine, 2014)) and through social media (Twitter, 2015)) have been calling for greater transparency of data.

A report written by the Inspector General of Social Affairs in France, Pierre-Louis Bras describes, on the governance and use of health data was published in 2013. In it he argues that [SI – amalgamation of PMSRI and SRIINAM, i.e. routinely collected data on payments] has huge potential, and that such a 'bien public' (public good) cannot be under the ownership of a single actor in the system, but should be administered in the public's interest by a legitimate authority that engages in dialogue with the appropriate parties (Bras and Loth, 2013). He goes on to recommend that where data present a risk of re-identification access should be restricted, but where there is no risk then they should be communicated widely or made public.

In order to put into practice the principles outlined in the report, Mr Bras proposes that data that are fully anonymous should be distinguished as far as possible from data that could indirectly be identifiable to patients. This classification should be made through a public assessment of re-identification risk, with those that can be shared or made public without risk defined as such by CNIL. Access to anonymous data should be opened up and made available free of charge (specific extractions or analyses should be chargeable). He proposes that access to data that could directly or indirectly identify patients should be limited, on the basis on the public's interest, the quality of the protocol, the need to access those data, the security in place for handling data, and the quality of the applicant. Permanent access should only be granted by the Ministry of Health under the control of CNIL, only to public organisations, and after careful consideration of the risks (re-identification) and benefits (health monitoring, health-

economic intelligence, etc). Health monitoring organisations (such as ANSM, INVS and HAS) should be able to make full use of those routinely collected data (SRIINAM and PMSI). He goes on to recommend that one-off requests for access be granted through a unique system which will ensure that the envisaged 'collective benefit' of the research justifies the risks of misuse (Bras & Loth, 2013).

Importantly, it is recommended that the data linkages that are authorised through this proposed process should not require a decree from the Conseil d'Etat (the French Council of State). The data should be hosted in a secure environment with strong security procedures for the access (Bras & Loth, 2013). Mr Bras states that if this advice is upheld legislative modifications would be required, in particular: the arrangements relating to SNIIRAM and PMSI in the social security and health (legal) codes, and the measures under the 'Loi Informatique et Libertés' (Law on Information Technology, Data Files and Civil Liberties) which prevents the use of NIR (personal identifying numbers) in health research and evaluation of health care. It also requires, in order to exploit the exceptional potential of the data and to expand its usage, an expansion of the platform of services for users. In order to cover the costs for this extra capacity, data extractions and processing costs would be charged to the organisation responsible for the study. This would also apply to pharmaceutical companies (Bras & Loth, 2013). Three institutional alternatives were proposed for the service: (1) an autonomous structure, (2) a structure connected with the CNAMTS but with its own dedicated resources, or (3) a structure tied with the statistical management of the Ministry of Social Affairs: the DREES.

SUMMARY

In France there is a strong demand for RWE to support periodic re-evaluations by HAS of the medicines and medical technologies on which it issues recommendations.

Routine data through electronic health care records is emerging in France and might in the future be a useful source of clinical information on the French population, particularly as regulations around the structure and terminology in data collection are strong. However, uptake is still low, which is perhaps reflective of the system by which patients opt-in to the programme. Although a personal identifying number for health, the INS, is being rolled out, there needs to be consistency between the identifiers used for health insurance purposes and electronic records in order to be able to link these two types of information source (OECD, 2013). It is unclear to what extent the deployment of the INS (as a substitute for the NIR social security number) has been made across the health insurance claims databases, which represent the richest source of routinely collected health care information in France.

There is a clear need for a governance structure to support data use to translate the abundant RWD into RWE, as its application is low to date. On the other hand, post-registration studies which monitor treatment impact in real-life settings, which are demanded by HAS to support periodic re-evaluations of health care products, are common in France.

6. Italy

1. **Brief overview** of the health system and collection / management of patient data

In Italy the National Health Service (Servizio Sanitario Nazionale – SSN) was established in 1978, replacing a system of health insurance funds, with the objective of providing uniform and comprehensive care to all under a system financed by general taxation (France et al., 2005); the SSN was modelled on the UK's NHS. In 2012 health spending accounted for 9.2% of GDP in Italy, 77% of which was funded by public sources (OECD, 2014a). A reduction in overall expenditure on health in recent years is partly attributable to lower spending on pharmaceuticals - down 14% in real terms between 2008 and 2012. The public health care system is financed by a mixture of national and regional taxes and patient co-payments. Only 15% of Italians have private health insurance (France et al., 2005).

According to the Italian constitution, responsibility for health care is shared between the State – which has the power to set the 'essential levels' of care for all citizens – and the 20 regions, which have almost exclusive responsibility for organising and administering publicly financed health care. According to France et al (2005) regions differ substantially in terms of their demography and economic development, as well as their health care infrastructure and health expenditure, with a clear north – south divide. Regions have strong autonomy, which has been exploited in the. A catalogue of health services that must be provided is issued, with any services beyond these needing to be financed by a region's own-source revenues. Likewise, a 'negative' list is provided for services that are ineffective or fall outside the remit of SNN, and therefore should not be provided. At a regional level, providers compete for contracts based on a prospective payment system (PPS) (fee for service) according to activity based on diagnosis-related groups (DRGs), a system which was introduced in 1994.

The Italian Medicines Agency (the competent authority for drugs), AIFA, has been in existence for 10 years, and is relatively unique in that it acts as both the regulatory agency as well as making decisions on drug reimbursement. This means that it is able to combine licensing and negotiating activities, which puts it in a strong position to implement conditional reimbursement strategies such as MEAs (Moroni, 2014). AIFA is a public body which operates autonomously under the direction of the Ministry of Health and under the vigilance of the Ministry of Economics, and cooperates with the Regional Authorities (AIFA, 2015b). With regards its pricing and reimbursement activities, pricing is set through a negotiation between AIFA and the pharmaceutical companies, in accordance with Law n.326 of 2003 and the Interministerial Committee for Economic Planning Resolution of 2001.

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. Routinely collected patient data.
Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

There are various registries in Italy that routinely capture patient treatment and outcomes information, which often feed into the set-up of and follow-up for conditional reimbursement decisions. In order to support their regulatory, HTA and MEA activities, AIFA manage Drug-monitoring Registers, which have operated since 2005; products are entered into the registers immediately after their authorisation. As of December 2011 these covered 78 therapeutic indications, corresponding to 66 active compounds (Ferrario and Kanavos, 2013). The main motivation of the registers is to ensure drugs are being prescribed correctly and in accordance with guidance and to collect the clinical data needed to support conditional reimbursement decisions However, according to AIFA in the "near future" the registers may be useful to evaluate the effectiveness of drugs and assess their economic impact (AIFA, 2015a). In addition, the Italian Association of Cancer Registries (AIRTUM) which was established in 1997, acts as a portal for the registration of cancer cases in Italy and aims to make incidence, mortality, survival and prevalence data for tumours in Italy available to health service bodies and the scientific community (AIRTUM, 2015).

Whilst birth and death registries exist at a national level, constructing disease registries at a national level remains a big challenge, as this involves the consolidation of regional data, which would require its own authorising legislation. According to the OECD (2013) regions are becoming reluctant to participate in research studies for fear that the Data Protection Agency will revoke approval.

When the European Data Protection Directive of 1995 was introduced, the ease with which health research could make use of identifiable personal health data was reduced. The first Data Protection Act to be put in place after this, in 1997 (no. 675/1996) stated that personal health data should be de-identified. Only when de-identification was impossible should identifiable data be processed (OECD, 2013). The most relevant piece of legislation in Italy is the Data Protection Code, which is supplemented by numerous 'Authorisations' which have been published over the years since.

In 2004 the Data Protection Code no.196/2003 was introduced which brought together various laws and regulations relating to data protection. This included a special chapter on the processing of health care data (sections 75-94) (Garante, 2015). It specified categories for processing personal identifiable health data that could be considered as being in the substantial public interest (OECD, 2013). The code permits processing identifiable data either if (a) consent is attained (which must be provided in writing) or (b) law authorises it. An important exception to this rule may be accounted for as a 'special circumstance' by Section 41 of the Code, which enables the 'Garante' (the Data Protection Authority) to authorise studies to be undertaken without consent, where obtaining consent entails an effort that is manifestly disproportionate in relation in particular to the number of individuals involved (Garante, 2003; Garante, 2013[Section 6 'Authorisation Requests']). The criteria for this exemption are laid out in section 4 ('Impossibility to Inform Data Subjects') in an authorisation published in March 2012: General Authorisation to Process Personal Data for Scientific Research Purposes [1884019] (Garante, 2012).

Patients that wish to object to their data being processed may do so under section 7(4) letter a) of the Data Protection Code; however this does not refer to administrative records (Garante, 2003).

Further clarification was offered by the Data Protection Authority — 'Garante' (the 'Privacy Guarantor') — in 2013 with an updated Authorisation no. 2/2013 Concerning

Processing of Data Suitable for Disclosing Health or Sex Life [2941268] (Garante, 2013). This allows health care professionals and 'public health care bodies' (including universities acting as such) to access health data where consent has not been attained, when the purpose of said data and processing operations is to protect patients from harm or protect the health of a third party or the community as a whole. Whereas under Section 26(1) of the Data Protection Code private bodies (and profit-seeking public bodies) may only process sensitive data upon authorisation of the Garante, this 2013 'Authorisation' specifies that private entities (including private research entities) may process health data without the requirement for specific authorisation by the Garante, but still only when patient consent has been obtained (as per section 106 107 and 110 of the Code) and if "*the availability of exclusively anonymous data concerning population samples does not allow achieving the purposes of said research'*. In addition, data '*shall be processed in such a way as to prevent data subjects from being identified even indirectly, unless matching of the research data with the data subjects' identification data is performed on a temporary basis, is fundamental for the research purposes, and is accounted for in writing. Research findings may only be disclosed in anonymous form*" (Garante, 2013). If a patient cannot offer his/her consent, then consent should be obtained from a legal representative, next of kin, a family member, a person cohabiting with the data subject, or failing these the manager of the institution where the data subject is resident (Garante, 2013). Where the above conditions are met, data controllers need not lodge an authorisation request to the Garante.

A common barrier to the use of health data for research is that data protection legislation generally specifies that data is not to be used for purposes different from those for which it was collected. Helpfully, the Italian Data Protection Code sets out in section 99(1) that the "*Processing of personal data for historical, scientific or statistical purposes shall be considered to be compatible with the different purposes for which the data had been previously collected or processed*". In addition, it is specified that the processing of data for those purposes may be carried out upon expiry of the period that was necessary for achieving the different purposes for which that data had previously been collected (Garante, 2003). In addition, Section 110(2) of the Code specifies that personal health data may be processed for medical research purposes if the research programme has obtained a reasoned favourable opinion from the geographically relevant competent Ethics Committee, along with the Garante's authorisation partly in pursuance of section 40 of the Code (Garante, 2003).

With regards to data collected on drugs that are paid for (even in part) by the National Health Service, the code specifies that prescriptions are to be written to allow the data subject's identity to be established "*only if this is necessary in order to check that the prescription is correct or else with a view to administrative controls or for epidemiological and research purposes, in compliance with the applicable rules of conduct"* (Garante, 2003).

In 2009 Guidelines were issued by the Garante on the Electronic Health Record and the Health File (Garante, 2009). These allow patients to freely decide whether to have an EHR and whether it should include all or just part of their medical information. Patients should have the right to hide any piece of information in the electronic record, and should have access to it. It is specified that future use of EHRs for further purposes related to scientific, epidemiological or statistical research is not ruled out per se, but any such use should be compliant with the sector-specific legislation; this includes data that is kept regionally.

b. Collecting de novo patient data.
Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

In 2002 the Ministerial Memorandum no.6 2nd September 2002 (issued by the Ministry of Health) set forth the regulatory requirements for observational (i.e. non-interventional) studies[7] for drugs in Italy. However it is considered that the regulation was insufficiently clear, especially with regards to the criteria of the Ethics Committees (Baccetti, 2007). Additionally, regions have in the past enacted their own regionally applicable guidelines for the evaluation of protocols for observational studies. A good demonstration of this heterogeneity can be viewed in a paper by Santarlasci and colleagues where the authors did two things: first, they traced back publications of observational studies that were conducted in a particular hospital in Italy to the list of registries notified to the hospital's Ethics Committee. None of the studies had been notified. Second, they investigated the opinions expressed by 28 different Ethics Committees (at different hospitals) on the same multi-centre clinical protocol for an oncologic observational study, and found that evaluations were inconsistent and ranged dramatically, from unconditional approval by some committees to rejections by others (Santarlasci et al., 2005).

In recognition of the need for greater clarity, the Italian Ministry of Health together with the Italian Drug Agency, AIFA, produced a set of Guidelines for the classification and conduct of observational studies on drugs (AIFA, 2007). In it they note the particular importance of observational studies for assessing the safety profile of drugs under normal conditions of use, for further evidence of effectiveness, and for pharmacoeconomic evaluation. The guidelines include details on when a study is to be considered an observational study, requirements of the study protocol, reporting procedures for adverse drug reactions, and requirements to make the results of the study publicly available. It also includes a list of documents to be presented to the Ethics Committee. Of particular importance is the distinction that is drawn between types of observational study, and its implication for ethical review:

- *Prospective cohort study.* A study where patients are included based on their taking a certain drug, and then followed up over time to evaluate outcomes. For this type of study ethical approval must be attained from each of the relevant ethical committees (i.e. from the ethical committees of each participating hospital).
- *Other observational studies of etiological nature [retrospective cohort studies, case control studies, case cross over studies, transversal studies, ecological studies] and descriptive studies.* Approval need not be attained, but Ethics Committees must be notified.

Therefore, approval must only be attained for prospective cohort studies, whereas for retrospective studies notification is sufficient. Among the documentation to be submitted to the Ethics Committees, Patient Information sheets and Consent forms should be included, but *only for studies where there is a direct relationship with the patient*, i.e. if

[7] Studies in which drugs are prescribed according to its approved indication, as part of normal clinical practice, where the decision to prescribe is independent of the decision to include the patient in the study.

the information gathered for the study goes beyond current data collection activities. Additionally, the applicant must provide details of the procedures in place to ensure confidentiality of information.

3. **Data linking.** To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

In Italy, data linkage is facilitated by the 'TS' number, which has nearly universal population coverage and is used for both health and tax purposes. The system is managed through a publicly owned private company SOGEI, which acts as a trusted third party (OECD, 2013).

There is a relatively strong Italian infrastructure for health data, due to the fact there is universal health coverage through a national health system, broad collection of data in public data files, unique patient identifying number which allows datasets to be linked, a large academic community, and established data flows (OECD, 2013). However, one of the biggest challenges for research in Italy is the fragmented nature of health service administration, which makes data sharing to facilitate data linkage very difficult. An example is the National Outcomes project which links hospital and death records. Despite the assistance of the National Agency for Regional Services (AGENAS), linkage is still only occurring in a few regions. In some regions this is apparently due to technical problems, but in many it is due to the uncertainty in whether they can legally share data for a national project (OECD, 2013). This is because "strict" interpretation of the privacy legislation would only allow local authorities to link data for the purposes of direct patient care. However there appears to be some discordance here with the legislative environment described under [1b], and highlights the difficulty in observing a coherent interpretation of permissions across Italian regions. This places a restrictive pressure on research in Italy.

According to the OECD (2013) report, there are no routine or standardised ways to request linkage of a researcher's own cohort data with governmental databases.

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

As noted in the description of the relevant legislation, informed patient consent is required for third party access to identifiable data; the 'general authorisation' specified by the Garante permits that this be undertaken without the need to request authorisation form the Garante if this condition is met. Applications to conduct such studies are to be approved by the relevant regional Ethics Committees.

There are 20 distinct regions in Italy, each with its own local health authority which processes personal health data for their area. Access to routinely collected and registry data for research purposes is considered by the relevant Ethics Committees. Little information is available on the criteria for access to that data or data that is collected through AIFA's drug monitoring registers; according to feedback received from Lilly, access for pharmaceutical companies is not permitted.

The landscape for research that makes use of regional data appears difficult, due in large part to the difficulty to engage with the regional data controllers; the processes of and requirements for applying for access are not well defined. Criteria used to evaluate proposals are not set out clearly, and sharing data between regions, even for official institutions, is therefore very difficult (OECD, 2013). However, in order to have greater confidence in project approval, many researchers seek funding or collaboration with public authorities. Some helpful criteria for the ethical and approval requirements for different types of observational studies are provided by AIFA (AIFA, 2007; Baccetti, 2007).

As described in a position paper form the Italian Society of Medical Statistics and Clinical Epidemiology Working group on Observational Studies, Regions are effectively both the producers of health care utilisation datasets as well as the bodies responsible for handling the sensitive data they contain, thus both having ownership of the datasets as well as regulating their use (Corrao, 2014). Attempts are being made in some regions to set out criteria for access to the regional DataWarehouses by external bodies more explicitly, for example in the Lombardy Region. They specify that data may be provided to external accredited bodies, provided that those data are used to fulfil the Region's planning needs, which will fall into a set of research areas that the region will establish yearly (Corrao, 2014). To receive accreditation the external body must already have conducted and published models for integrating use of health care utilisation databases for the purposes of monitoring in the past.

The picture is therefore patchy, but it seems that access to RWD in Italy from the perspective of a pharmaceutical company is severely restricted.

5. **Data use.** What, if any, are the rules governing the use of RWD including cover contract arrangements between data suppliers and recipients and rules around use for HTA

As discussed, there is a lack of clear criteria for gaining access to health care data, and this poses a major barrier for research in Italy. The IMS Health report on the RWE market impact on medicines describes that, like in France, there is significant demand for RWE but limited supply (Hughes & Kessler, 2013). The requirement for the collection and use of RWE in Italy is largely driven by AIFA's reimbursement requirements. There are various conditional reimbursement options: no reimbursement, unconditional reimbursement or reimbursement in the frame of an MEA (Ferrario & Kanavos, 2013). These MEAs could include price-volume agreements, cost-sharing, budget cap, monitoring registries, payment by results, risk-sharing and 'AIFA' notes. The monitoring registries ensure that prescriptions are being carried out correctly according to these agreements, and are mainly used for high-cost drugs (Navarria et al., 2015). Noting the complexities of the current system for introducing medicines into the National Health Service, an alternative strategy for performance-based agreements has been proposed: the "success fee" (Navarria et al., 2015). This would consist of an ex-post payment to the manufacturer, which would only be applied to patients that receive benefit from the therapy. In other words, rather than a system of refund for non-responders, the company and the National Health Service would agree on a threshold for effectiveness (i.e. criteria to categorise as a 'responder'), and payment would only be exchanged on the basis of number of responders. This system has been applied for the first time in

Italy to a novel drug indicated for idiopathic pulmonary fibrosis: pirfenidone (Esbriet). Coverage with evidence development schemes represent the most significant use of RWE in Italy. Other potential uses of RWE are guideline development through the National Guideline System (SNLG), though little emphasis is placed on RWE in guideline development (Hughes & Kessler, 2013).

The guidelines issued by the Ministry of Health and AIFA specify that, for observational studies, there must be a written commitment to summarise the results of the study and to put these in the public domain (AIFA, 2007).

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

In the report by the OECD on strengthening health information infrastructure for health care quality governance, respondents to the survey from Italy noted that it is getting *harder* to use personal health data to monitor health and health care quality. There is apparently a view by some in Italy that patient identifiers should be removed completely from patient health data; this would make data linkage impossible both regionally and nationally (OECD, 2013). Environment for research would improve if clear guidelines from public authorities were issued on how health research projects are approved and best practices for processing and linking data: There is currently no office at the national level to fulfil this role (OECD, 2013). This highlights the impact of a de-centralised and fragmented system of data collection and management. In recognition of this issue, some have proposed a single regional evaluation committee to consider proposals for data use, which would provide a consultative role for the regional health care authorities; in addition this could identify training needs for regional research groups and thereby enhance the quality of research undertaken using the data (Corrao, 2014). However we have found no indication that these proposals are being taken forward.

An eCard scheme in Italy is currently being rolled out, and in the regions of Lombardy, Tuscany and Fruili Venezia Guilia, data is already being collected through this system (Hughes & Kessler, 2013). Additionally, there is a major effort to drive the adoption of ePrescribing, and a tool to capture data to support, monitor and oversee health care services is currently being implemented: the National Heath Information System (NSIS). These developments may serve to improve the supply of RWD in Italy.

SUMMARY

In Italy, health care professionals and public health care bodies may process personal data without consent where the purpose is to protect the individual from harm or to protect the health of a third party or the community as a whole; this seems to pave the way for research to be undertaken with this data, with provisions, of course, for making any results that are disclosed anonymous. Since the general authorisation of 2013 concerning data processing, private organisations may process data without specific application and authorisation by the Garante, but patient consent must always have been obtained. However, implementation of these national legislative criteria seems to be fragmented across the Italian regions, leading to disparate environments and

processes for research across Italy. Contributing to this is the fragmented nature of data collection activities, and inadequate mechanisms for data sharing across territories.

However, there is strong demand for RWE in Italy, due to the widespread use of pay-for-outcomes and Coverage with Evidence development (CED) by AIFA; however, the ways to meet those requirements with the data available are unclear, other than by AIFA itself collecting and analysing the data.

7. Sweden

1. **Brief overview** of the health system and collection / management of patient data

Health care in Sweden is tax-funded under the premise of equal access to all. Whilst central government establish principles and guidelines and set the political agenda for health and medical care at a national level, provision is de-centralised and responsibility for delivering health care is devolved to county councils or municipal governments (Swedish Institute, 2015). There are 21 regional county councils that govern health care, to which the government reaches out to implement health care guidelines (OECD, 2013); the 290 municipalities, at a local level, are accountable for social services and elderly care. County councils may choose to deliver health care themselves, or using private companies, cooperatives or non-profit organisations (Doupi et al., 2010b). Health care expenditure in 2012 accounted for 9.6% GDP, with 81.3% of total spending on health being public expenditure (OECD, 2014b).

The National Board of Health and Welfare in Sweden — the *Socialstyrelsen*—is a government agency under the Ministry of Health and Social Affairs (Socialstyrelsen, 2015g) which compiles, analyses and passes on information with the aim of improving evidence-based practice and evaluating the effects on the population of political decisions concerning health and social care. The Socialstyrelsen also develops standards, issues regulations and guidelines, and is responsible for the development and management of the national information infrastructure: eHealth (Socialstyrelsen, 2015a), which was initiated by the Ministry of Health and Social Affairs in 2006 (Socialstyrelsen, 2011). This includes setting terminology standards such as statistical classifications and coding systems, including the Swedish translation of the clinical terminology SNOMED CT. For more information on the National Information Structure Strategy see (Eftimovska, 2014).

Whilst strategic responsibility for eHealth and the national coordination of electronic health record (EHR) implementation falls with the (national) Socialstyrelsen, county councils are responsible for their own implementation of EHR (OECD, 2013). County councils are represented by the Swedish Association of Local Authorities and Regions (SALAR), which is responsible for all health care providers, pharmacies and suppliers. The Centre for eHealth in Sweden is governed by the SALAR.

In a report by IMS Health on the RWE Market Impact on Medicines, Sweden is ranked second after only the UK in terms of supply of and demand for RWE to inform decisions (Hughes & Kessler, 2013). The authors note that the country has good medical records as well as registries, but that access to data can be problematic; whilst datasets in

Sweden are richer and more integrated than those in the UK, their use to generate insight and inform decision-making is lower.

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. Routinely collected patient data.

Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

Routine collection of in-patient data from hospitals began in Sweden in the 1960s through the 'National Patient Register' (NPR) (sometimes called the National Inpatient Register or Hospital Discharge Register). In 1984 the Ministry of Health and Welfare along with the Federation of county councils decided to make participation mandatory, and since 1987 NPR includes data for all in-patient care in Sweden (Socialstyrelsen, 2015h). Since 2001 the register also includes information on outpatient visits including day surgery, provided both privately and publically. Primary care is not included in the NPR, and therefore primary care data in Sweden – although captured through some specific disease registries — is not as complete as the information available for secondary care. However the use of EHRs in Sweden is widespread across both secondary and primary care.

The Socialstyrelsen is responsible for the NPR, which captures several data: fields relating to the patient such as personal registration number, sex and age, geographical data, administrative data relating to the date and length of admission, and medical data capturing diagnoses (according to ICD-10-SE classifications) and procedures (Socialstyrelsen, 2015h). The NPR is updated once a year, and captures the whole Swedish population. As well as hospital discharge information captured through NPR, the Socialstyrelsen is responsible for the following additional mandatory registers which also cover the whole population:

- The Swedish cancer registry, which was founded in 1958 and captures personal identifying data, medical and diagnostic data, and follow-up (Socialstyrelsen, 2015f).
- The Swedish medical birth register which began in 1973 and captures information on all deliveries (Socialstyrelsen, 2015i).
- The prescribed drug register, which contains information on all prescription medicines dispensed at pharmacies since 2005. Drugs delivered in hospitals are not included. Data is available on a monthly basis but annual overviews are published once a year (Abrahamsson, 2015b).

In addition to these mandatory registers that cover the whole Swedish population, there are several "national quality registers" in Sweden (104 funded registers as of 2014) which enable monitoring of care quality and focus on various areas of care or specific treatments (Eftimovska, 2014). The Swedish National Diabetes Register is one of the largest diabetes registers globally (Hallgren Elfgren et al., 2013). Participation by hospitals in collecting data to contribute to these registers is voluntary, but some councils encourage participation by offering financial incentives (Mattsson, 2014; OECD, 2013). A recent review of the coverage of these registries found that around 60% of

quality registers covered more than 80% of the target population (Emilsson et al., 2015).T

In Sweden, the collection and use of personal data is regulated through the Personal Data Act 1998 (*Personuppgiftslag 1998:204*) (PDA), which implements the EU Directive 95/46/EC and contains provisions to protect privacy of individuals from the processing of personal data (Hager and Mirsch, 2014). This applies broadly to the processing of all types of personal data, and is regulated through the Swedish Data Inspection Board (*Datainspektionen*). In addition, there are various sectoral laws which regulate the use of personal data; in the health care sector, the Patient Data Act (*Patientdatalagen* 2008:355) introduces legislation to increase the protection of patient's privacy in the processing of personal data.

The Personal Data Act (1998) dictates that personal data (i.e. data that is identifiable to a specific person) can only be collected for explicit and legitimate purposes, and that data gathered for a particular purpose cannot later be processed for a different purpose or in a different manner unless new legal grounds have been established (Hager & Mirsch, 2014). As in other countries, it is this point from which questions arise around the secondary use of health care data for research.

The main rule under the Personal Data Act is that personal data can only be processed if consent has been attained from the data subject. However there are various exceptions to this rule, the most relevant of which for health care are likely to be:

- Protect the data subject's vital interest
- Perform a work task of public interest
- Satisfy a purpose that concerns a legitimate interest of the data controller, or of a party to whom personal data are provided, if this interest is of greater weight than the prevention of the possible violation of the data subject's personal privacy

The Personal Data Act also requires that data is correct and if necessary up to date, and not kept for longer than necessary. Whilst as a general rule the data controller must provide information to a data subject at the point of collection, this may not be necessary if (a) it is impossible or (b) would involve a disproportionate effort. Exemption from the rule of consent also explicitly exists in the case use of personal health data in the care or treatment of a patient (Hager & Mirsch, 2014).

The Patient Data Act, introduced in 2008, replaced the Patient Record Act (1985:562) and the Health Care Register Act (1998:544), and allows health care providers to have electronic access to information held by other health care providers (Doupi et al., 2010a). Whilst the law enables care professionals to share information within health and medical care services, it also empowers patients by giving them access to their medical care record, allowing them to see what personnel has accessed their record, and giving them the right to block certain data from being shared (Doupi et al., 2010b). However, some suggest that whilst the law permits patients to access their own health information, the IT infrastructure to support this on a national basis is not yet widely available (Eftimovska, 2014; Gray et al., 2011). Other relevant legislation is: The Public Access to Information and Secrecy Act (Offentlighets- och Sekretesslag), Public Access to Information and Secrecy Ordinance (Offentlighets- och Sekretessförordning) and Tryckfrihetsförordningen (regarding freedom of press) (CODEX, 2014).

In order to facilitate a national register of prescribed pharmaceuticals, legislation has been introduced to facilitate the collection and processing of this information. The

Swedish eHealth Agency is a government agency which was formed in 2014 and took over the activities of the state-owned company Apotekens Service AB. The eHealth Agency stores and transfers all electronic prescriptions issued in Sweden (around 90% of all prescriptions are e-prescriptions), and is responsible for national drug statistics. The processing of personal data is governed by the Personal Data Act, as well as the Pharmaceutical Register Act (2005:258) (Nordqvist, 2014). Under this act, a patient cannot require that their data be excluded from the register; however, the registered information can only be accessed [it is not made clear by whom] following the patient's explicit consent, and information must be deleted from the register after a period of 15 months. If for some reason consent cannot be attained, as an exception to the main rule the information may be accessed by a prescriber to ensure accurate treatment (Ashjari and Strom, 2005). Extracts from the register of the personal information held can be requested by patients free of charge once per calendar year (Nordqvist, 2014).

The Patient Data Act of 2008 addresses national quality registers specifically in Chapter 7, and stipulates that patients control their data in any quality registry, and that they may opt out at any time and request that their data be removed from the national registry (Eftimovska, 2014). That is, whilst written informed consent is not required, all patients must be informed of their participation in a register (in accordance with the Patient Data Act, Chapter 8, Section 6) and have the right to withdraw their participation (Henriksson et al., 2014; Mattsson, 2014)[8]. This therefore represents an 'opt-out' system. The nationally aggregated data may be used for three purposes: statistics, analysis of health care quality, and research - but only with permission from the Ethical Review Board.

The arrangements described above appear to differ from those for the mandatory government-administered registries like the National Patient Register, where every patient is registered by default without the possibility of opting out (Emilsson et al., 2015).

b. Collecting de novo patient data.
Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

To the extent that research involving the collection of non-routinely collected data requires that personal data be processed, the Personal Data Act and the Patient Data Act apply equally to the governance for the collection of new patient data. In Section 19 of the Personal Data Act (1998), sensitive personal data may be processed for research and statistics purposes provided "*…the interest of society in the research or statistics project within which the processing is included is manifestly greater than the risk of improper violation of the personal integrity of the individual that the processing may involve*" (Swedish Government, 1998). The judgement that the benefit from the research

[8] This is a notably difficult issue for patient populations that permanently lack decision-making capacity, such as the severely disabled or elderly. In 2011 it was declared that for patients who could not understand the information provided for registration, if they did not have a legal representative who could opt out on their behalf then no data could be collected for that patient. In October 2014 a new regulation came into force in the Patient Data Act (Chapter 7 section 2a) whereby so long as there is no reason to believe that the patient would have opposed the registration of data if he/she had been mentally competent, then data could be collected (Mattsson, 2014).

outweighs the risk to privacy will be deemed to have been met if the research has been approved by a research Ethics Committee: a "*special body for consideration of research ethics issues that has representatives for both the public and the research and that is linked to a university or a university college or to some other instance that to a very substantial extent funds research*" (Swedish Government, 1998). Informed consent should be attained wherever possible, though where certain criteria are met and where consent would be impossible or impracticable to obtain, a waiver can be issued (e.g. for quality registers). In general, research collecting de novo information from patients will require patient consent and must meet certain criteria and openness for such.

The sensitivity of registers, through which data is collected for unspecified future research, is sensitive and therefore government and parliament maintain control over their creation (CODEX, 2015). Otherwise, data collection must be for a specific and particular purpose.

The Swedish Ethical Review Act (2003:460) stipulates that research ethics review is mandatory in research involving any intervention in humans, including biological material as well as *personal data* obtained from registers or questionnaires / interviews (EUREC, 2015b). There is one central ethical vetting board in Sweden – EPN – and six regional boards which are centred around Universities. An English translation of the application for ethical review of research involving humans is provided on the EPN website (EPN, 2015). The fees vary from SEK 5,000 up to SEK 16,000, and regional boards should normally make a decision within sixty days of receiving a fully completed application (EPN, 2015).

3. **Data linking.** To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

In Sweden, the personnummer (Personal Identity Number) is used for all official purposes (tax, social welfare, health care, education, income etc.). Swedish health care and all national health registers depend on this identifier, which is comprised of date of birth, a three-digit birth number and a check digit (Eftimovska, 2014). This systematic recording of national ID numbers across health care settings makes linkage across datasets feasible, though ethics approval for this process for purposes outside of national policy review can be long (Hughes & Kessler, 2013; OECD, 2013).

According to Oderkirk et al (2013), internationally, Sweden performs amongst the highest in terms of regular data linkage projects, though their health information infrastructure for data linkage is reportedly stronger sub-nationally than nationally, particularly where quality register coverage is better in some areas than others (OECD, 2013).

Routine use of RWD in Sweden can be seen through Quality and Efficiency assessment of clinical guidelines, which cover stroke care, care for four types of cancer, dental care, diabetes care and mental health care (OECD, 2013). This involves data linkage activity to review and prioritise health care quality indicators, and involves linking quality registers with the mandatory routinely collected datasets which capture mortality, prescribed drugs, and cancer care. For example, for cardiac and stroke care, individual patients are linked to health care encounters to see how processes of care as well as health have changed since the introduction of a guideline. With this information, the

Socialstyrelsen and the SALAR jointly publish the Quality and Efficiency in Health Care reports, which provide regional comparisons and are updated on an annual basis (Socialstyrelsen, 2013). These analyses then feed into national guideline updates (OECD, 2013).

There are many examples of data linkage projects outside of national policy review, which demonstrate the use of Sweden's expansive data collection programme. For example, a project assessing diabetes care by linking databases in primary and secondary care, the pharmaceutical registry, quality registers for diabetes and ongoing population-based health surveys (Rolandsson et al., 2012). Drug cost-effectiveness research and post market surveillance can be enabled by linking the relevant disease registry with the Quality Registry Drug Follow Up, which contains information on drug treatments classified by drug product and diagnosis, as well as certain outcome measures (Eftimovska, 2014).

A specific unit within the Socialstyrelsen is permitted access to identifiable data, and performs the database linkage activity in Sweden. Data is de-identified by removing names and other identifying information such as address and birth date, and by replacing personal identity numbers with unique serial numbers (Eftimovska, 2014; OECD, 2013). This means that data analysts within government, or external researchers with approved projects, are only provided with de-identified datasets, and therefore never see identifiable information.

There are various measures in place to ensure that identifiable data from linked datasets is protected. Linked data are locked in a secure building and protected from unauthorised access, and are not stored on computers that are connected to a network. Data use is tracked by a security officer and special confidentiality rules apply for example where information has been contributed by a hospital with very few patients. All staff are undergo legal, security and confidentiality training (OECD, 2013).

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

A recent summary of all the ongoing quality registers is provided in a recent review by Emilsson and colleagues (2015), which provides registry name, disease area, year of initiation, coverage, participating health care units, volume per year and completeness. Further detail can be found on the Socialstyrelsen website, including a list of all of the variables collected within each register (Socialstyrelsen, 2015c).

As is the case with collecting de novo patient data, *access* to patient data for a specific research project is subject to the approval from an ethics review board.

On their website, the Socialstyrelsen describe the process by which data or statistics can be ordered (Socialstyrelsen, 2015b). Statistics, which are described to be aggregated tables where no data can be attributed to any individual, are available directly from the statistics page of the Socialstyrelsen, but may be ordered if alternative statistics are required. Processing time is described to be at least 3 weeks, with a charge of 1,100 SEK per hour (excluding VAT) for the work associated with the order (most orders are said to take between 2 and 20 hours). Disclosure of information through this channel is tested against the rules of confidentiality, and any identifiable information is completely removed (Socialstyrelsen, 2015b).

For requests relating to patient-level data for research purposes, a similar scenario applies whereby the client is charged on a per hour basis for the time associated with extracting and processing the information (1,100 SEK per hour). Processing times for this type of request is estimated to be between 2 and 6 months (or longer if other agencies are to be included), and most orders are said to take between 10 and 40 hours (Socialstyrelsen, 2015d). The process involved in making such request follows a path set out by the Socialstyrelsen:

- Order forms are sent directly to the Socialstyrelsen for registry data hosted by that organisation. If data linkage is required across datasets, for example matching the hospital discharge or pharmaceutical register with a relevant quality register, then requests for disclosure must be made to each of the relevant authorities directly. Whilst the disclosure authorities will work together, each will make its own 'secrecy examination'.
- A specification and confirmation of order is then agreed, and an approximate price and time is provided.
- Documentation of the application and approval of the project from an Ethical Review Board must be submitted with the application for the Socialstyrelsen to consider the request.
- A privacy examination is then undertaken by the Socialstyrelsen to make a decision on whether to disclose the information. Disclosure of information will always be de-identified, unless it is clear that the research cannot be conducted without the personal identifiers (and this has been approved by an Ethics Committee). The purpose of the request must be for research.
- If permission is granted, a confidentiality agreement must be signed by the recipient which governs how the researcher may use the released data (Socialstyrelsen, 2015e).
- Data is made available in the form of anonymised microdata, with aggregated tables available on request. Data usually transferred through SAS or Excel files, but other programmes can be supported (Abrahamsson, 2015a; Abrahamsson, 2015b).

Sweden does not rule out access to data by commercial companies. According to an OECD report, in Sweden there is a perceived difficulty sometimes in ascertaining whether research requests from pharmaceutical companies are really in the public interest or solely for commercial purposes (in which case access is denied) (OECD, 2013). Sweden is therefore considering the introduction of new legislation which sets out the conditions for personal data access for research and analysis more clearly.

5. Data use. What, if any, are the rules governing the use of RWD, including arrangements between data suppliers and recipients and rules around use for HTA,]

All staff of the Socialstyrelsen that work with patient data are trained in confidentiality requirements and data security, thus minimising the risk that data is analysed or used inappropriately (OECD, 2013). For use of data outside of this government agency, as mentioned, approval for access to health care data is contingent upon a confidentiality agreement between the data supplier and the data recipient, in order to ensure that data integrity and confidentiality is maintained. There is little public documentation which details the rules around use of data once it has been received. However, compared with

many other countries, the role for observational data in HTA decisions in Sweden is pronounced, due to the adoption of coverage with evidence development.

The Dental and Pharmaceutical Benefits Agency (TLV) – the HTA body in Sweden – has for a long time employed 'coverage with evidence development' decisions to permit access to drugs in the presence of uncertainty, with the ability to revise decisions in the future after the collection of observational data (Willis et al., 2010). Reimbursement is thus sped-up, by using ex-ante value based pricing (VBP) as a form of risk sharing. This means that to secure reimbursement, manufacturers are often required to commit to the collection of evidence in real world settings (in actual clinical practice) (Persson et al., 2010). Guidelines around the rules governing this appear to be lacking.

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

The Government in Sweden is considering introducing new legislation around data access and data linkage, specifically addressing the issue of commercial interests in relation to personal health data (OECD, 2013). There is concern in particular about health insurance companies using such data to determine whether to approve or deny insurance coverage. However, there is a worry that new legislation to tackle this may tighten restrictions in such a way that there would be a negative impact (tighten restrictions) on research that *is* in the public interest (OECD, 2013).

There are various emerging initiatives which will increase the need for a shared understanding of appropriate governance for the collection and use of real world data. One of these is the emergence of 'three party agreements': agreements between pharmaceutical companies, TLV and county councils to collect efficacy and outcome data. In addition, the emergence of Managed Introductions (MEAs) as a new national process is in its pilot phase, currently with eight drugs (Janusinfo, 2015).

SUMMARY

In Sweden there are various mandatory registers collecting real world data on a routine basis: the National Patient Register (hospital discharge), cancer registry, medical birth register and the prescribed drug register. In addition, there are several quality registers relating to specific disease areas or procedures which can be linked to other datasets to draw a longitudinal picture of clinical practice and outcomes in Sweden.

The Personal Data Act, which was written into legislation in 1998 in response the EU Directive (95/46/EC) states that personal data cannot be processed without consent from the data subject. However, information on patient health is noted to be a special case. In order to clarify the lawful processing of health care data, various other legislative arrangements have been introduced to specify the lawful uses of health care data, in particular the Patient's Data Act in 2008 which outlined the lawful transfer of data between health care practitioners, and the Pharmaceutical Register Act in 2005 which permitted the collection of prescription data. Under the Patient's Data Act, patients must be informed of their participation in quality registers and given the opportunity to withdraw, but no explicit written consent is required. This is a similar "opt-out" system

as that is employed in the UK. Potential changes in the EU-wide legislation may change this and no longer allow this to be the default position.

Ethical review for collection of and access to patient data balances the interest to society of the research and risk of improper violation of personal integrity. Whilst informed consent is optimal and the most straightforward way to collect and access data, this can be waved where ethical review has deemed the research or data collection initiative to be worthwhile and where obtaining consent would be impossible or impractical. It is for this reason that consent is not required for the mandatory registers, from which patients cannot abstain. This could be perceived as a similar system to the UK and the distinction their between audit and research.

8. Germany

1. **Brief overview** of the health system and collection / management of patient data

In Germany there is a statutory health insurance system, which is part of social insurance that also covers among other things pensions and unemployment benefits; these are regulated by the Social Code Book (SGB) (HDN, 2013). Private health insurance is also prevalent (covering around 11% of the population) in Germany, which can offer substitutive or supplementary health care, though the service providers are the same. Only those earning above a certain level of income can purchase private insurance. Since 2009 all residents of Germany are legally required to have health insurance (HDN, 2013). A small proportion of the population (4%) are covered by free government health care, which applies to civil servants, soldiers, the police force, welfare recipients and asylum seekers. Ambulatory care is mainly provided by for-profit private providers, whereas acute, long-term, and hospital care are provided by a mix of public and private, for-profit and not-for-profit organisations.

Contributions to the statutory health insurance, also known as "sickness funds", are shared 50:50 by employees and employers, and people can select the sickness fund that insures them (the funds are obliged to accept any applicant) (Holtorf et al., 2009). At the federal level, the Federal Joint Committee (G-BA) issues directives for the 'benefit catalogue' of the sickness funds, thereby specifying which services are to be reimbursed; the G-BA is a public legal entity which was established in 2004 as a result of the Health care Modernisation Act and comprises the leading organisations of the self-governing German health care system (G-BA, 2015). The G-BA is under the statutory supervision of the Federal Ministry of Health, and is supported by an independent Institute for Quality and Efficiency in Health care: IQWiG, which evaluates the costs and benefits of medical interventions.

In Germany 'Gematik' is an organisation of representatives from the statutory health insurance system and health care providers; this group has responsibility for the establishment of a national telematics infrastructure for health care (OECD, 2013). Whilst Gematik provides guidance on interoperability of documentation systems, there is no national organisation to set clinical terminology standards at a national level (OECD, 2013). There is no law in Germany to mandate health care providers to adopt EHRs or to adhere to particular standards.

In 2006 the distribution of smart cards – elektronische Gesundheitskarte (eGK) – was planned for all 71 million German legal health insurance customers, with the ambition of facilitating an EHR with the capacity to enable authentication, authorization and secure data storage (Smart Card Alliance, 2006). The health minister Philipp Rösler suspended the project because of concerns around security and confidentiality, and proposed restricting the functionality (Hoeksma, 2010). However roll-out seems to be ongoing.

EHRs are widespread in primary care, but in secondary care are more limited – covering around 38% of hospitals (Hughes & Kessler, 2013).

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. Routinely collected patient data.

Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

Sources of aggregate health data in Germany include the Federal Bureau of Statistics (Destasis, 2015) which produces aggregate demographic data on birth and fertility rates, mortality, life expectancy, causes of death as well as hospital statistics and expenditure. The 'GBE de Bundes' (Federal Health Reporting System) collects statistics on diseases and health problems (GBE des Bundes, 2015a).

The Robert-Koch Institute (RKI) in Berlin is the central federal institution for disease control and prevention, and cooperates with the Public Health Service, the Health care Sector, as well as the scientific research community. After the passing of the Law for the Prevention of Infection (Infektionsschutzgesetz: IfSG), the RKI has become the federal epidemiological centres for infectious diseases, and maintains epidemiological registers at the federal level (RKI, 2004). These include disease-specific registers such as for Diabetes (GBE des Bundes, 2015b). The German Centre for Cancer Registry Data (ZfKD) is also located within the RKI, and is responsible for pooling and quality-checking data from population-based cancer registries from each German federal state; it is noted by the organisation that data is heterogeneous in quality and content, due to the varying history and initiation data of each of the submitting registries. Data is pooled and analysed, with national statistics published every two years. The "pooled data set" of the German Epidemiological Cancer Registries is available for research through a 'scientific use file' through application to the Institute (RKI, 2015a).

Germany has rich electronic data captured through widespread use of EHR in primary care. There are also large claims datasets held by payers, as all settlements must be processed electronically for payment (Hughes & Kessler, 2013). There is no central hospital discharge database, which for a lot of other countries provides important information for research. However, the Institute for Hospital Remuneration System (Institut für das Entgeltsystem im Krankenhaus) is responsible for the DRG-payment system. In Germany there is no 'minimum dataset' that is captured for all electronic patient health records. Rather, datasets are defined and managed by organisations of health care professionals, and are only used mainly for the purposes of direct care (OECD, 2013). However there are some similarities across datasets in terms of patient identification and the reporting to diagnoses and medication. For data on drug consumption, IMS Health operate in Germany, as well as a company called *Pharmafakt*

which sells reimbursement drug consumption data; on its website it describes the measures taken by the organisation to meet the relevant data protection legislation in Germany (Pharmafakt, 2015).

In Germany the main legal source of data protection is the Federal Data Protection Act: Bundesdatenschutzgesetz (BDSG) which was introduced in 2003 (most recently revised in 2009) to implement the EU Directive 95/46/EC (BMJV, 2009; Jansen and Hinzpeter, 2014). Under the general requirements of the BDSG, the collection and processing of personal data is only allowed if the data subject has expressly offered their (written) consent (section 4a of the BDSG), or if it has been expressly permitted or ordered by law. In order to offer this consent data subjects must be informed of the purpose of the processing and the identity of all recipients (Jansen & Hinzpeter, 2014). There is no specific rule around consent of minors, but it is generally accepted that an age of 12 to 14 years may be regarded as a general threshold for the child to be capable of understanding the extent and meaning of the declaration. According to section 4f of the BDSG, public and private bodies which process personal data must appoint in writing a 'data protection officer' who must comply with various duties as set out in the Act (BMJV, 2009). The rights of the data subject to access (sections 19 and 34), and to correct, delete or block data (section 20 and 35) must not be excluded or restricted by any legal transaction.

In the absence of consent, the data controller must rely on a statutory provision that allows the processing of that data. However, special rules apply whereby sensitive data may be processed "*to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving consent*" or for the purpose of scientific research where "*the scientific interest in carrying out the research project significantly outweighs the data subject's interest in excluding collection"* and where "*the purpose of the research cannot be achieved in any other way or would otherwise necessitate disproportionate effort"* (BMJV, 2009, Section 13). Data subjects have rights to access their data and details of how it has been used, and in general this information must be provided free of charge (Jansen & Hinzpeter, 2014). The data subject may object to the processing of his/her data. In addition, there are various security requirements. Health data is considered in the BDSG as a special category of personal data (section 3(9)) and is subject to stricter rules.

The sectoral law which covers the processing of health and other personal data in connection with the provision of medical and social security services is the Social Security Code (SGB) I, II; IV, V and X. Section V relates to the reimbursement of patient treatment by statutory health insurance companies. Section 305a of SGB V limits the disclosure of prescribing information, as it is argued that if the data is provided to a service provider or pharmaceutical company then this may lead to conclusions about the physician's prescribing behaviour (Jansen & Hinzpeter, 2014).

In regard to data collected through personal electronic health cards, the only mandatory applications of data for which consent are not required are: provision of administrative data, provision of information about private co-payments, transmission of electronic prescriptions, and provision of data required by EU regulations for having access to treatment in member states of the EU (HIQA, 2010).

Data protection is governed partly at a federal level and partly at a state level (OECD, 2013). The Federal Data Protection Commissioner is responsible for public sector entities and providers in the social security administration as well as private-level entities that

fall under public law. At a local level there are State Data Protection Commissioners, of which there are 16 in total.

Submission of state cancer registry data on an annual basis to the Centre for Cancer Registry Data (ZfKD) is carried out in accordance with the 2009 Federal Cancer Registry Data Act (BKRG); data is transferred in an anonymised format (RKI, 2015b). In accordance with section 5 para.3 of the Act, the ZfKD may make the verified dataset available for use by third parties, provided a justified, scientific interest can be credibly demonstrated by the applicant.

b. Collecting de novo patient data.
Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

In Germany there are a total of 53 research Ethics Committees, 33 of which are attached to Faculties of Medicine in Universities, 17 of which are attached to Medical Associations, and 3 of which are attached to State Governments (EUREC, 2015d). For studies that are carried out by an investigator attached to a University, the research committee of that University is entitled to assess the application. For studies carried out by an investigator outside a University, the Ethics Committee of the regional Medical Association is legally competent (EUREC, 2015d). At a national level, the National Council of Ethics may give recommendations which are not binding.

The German Society for Epidemiology (DGEpi) have produced guidelines and recommendations for Good Epidemiologic Practice (DGEpi, 2004). It is specified in this guidance that approval should be obtained from an Ethics Commission before an observational study is conducted. Explicit and operationalisable research questions must be formulated and a protocol submitted, with details of how data protection will be ensured.

3. Data linking. To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

The Public Health Insurance Act of 2003 introduced the electronic health card, and through this modernisation bill a new universal health identifier was introduced: the Krankenversichertennummer (HIQA, 2010). The format of the number is a 10 digit alphanumeric sequence, generated from the social security number using a one-way algorithm. Therefore there is potential for data linkages, and once the electronic health card (eGK) is introduced across the country, these numbers could be used in all aspects of care provision (OECD, 2013). According to the OECD, data linkage projects are undertaken at the state rather than the national level, and only when authorised by law. Examples include sickness fund data linkages in the state of Hessen, the development of a mortality index in Bremen state, and linkage of population-level health surveys in Essen and Augsburg.

According to Anderson and Storm (2013), linkage of cancer data works through an encryption system. Cancer registration is divided into two separate offices, a notification

office where the personal ID is known, but which is encrypted (pseudonymised) before sending to the registration office. For three months it is possible to link back to the individual data, but after this period, linkage and data analysis can only be carried out on the encrypted data. Therefore new data can only be linked using the same pseudonymisation key. All German states are able to use this same national pseudonymisation algorithm, so it is possible to merge de-identified datasets at the Centre for Cancer Registry Data. However, it means that errors in the original data cannot be identified or corrected as the link to the actual individual is lost. In addition, as the anonymisation algorithm is based on gender, data of birth, residence, etc. changes in these as well as misspellings or errors may result in missing linkages or duplicate entries. It is thought that this trusted third party system of pseudonymisation constitutes 30-50% of the total cost of cancer registries in Germany (Andersen and Storm, 2013).

Data is collected at a state level in general. Any data linkage, e.g. amalgamating data from different states for research projects, or linking cancer registry data with other data sources, requires authority from each individual state to proceed (OECD, 2013). There are legal provisions to allow data from the statutory health insurance 'morbidity-orientated risk adjustment scheme' which is conducted at the state-level to be analysed at federal level in order to facilitate health services research (OECD, 2013). Only *de-identified* data is shared / provided to researchers (OECD, 2013).

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

As described, under German data protection laws, personal medical data that is collected is only allowed to be used for the purpose for which is originally collected (the medical care of patients) unless explicitly allowed by law (OECD, 2013). However some datasets can be extracted, within the constraints of the Data Protection Act, such as the monitoring of health care quality from within health care organisations (OECD, 2013). Those organisations that manage health care data routinely, such as the RKI, must conform to strong physical and technical security standards.

Access to data for researchers is to anonymised data only. Access to payer claims data is limited, and to date payers have often ignored regulations requiring them to share datasets with IQWiG (Hughes & Kessler, 2013). Access to individual data by third parties is only allowed if the patient has given their consent, and strict confidentiality procedures must be adhered to. Whilst reimbursement data is often used for health economic analyses, access by private organisations must be through an academic collaboration, which according to feedback received by Lilly is becoming more complex.

Access to cancer registry data for third parties, as discussed, is permitted through the Federal Cancer Registry Data Act. Applicants must detail what variables are required and must demonstrate the "justified scientific interest" of the project. Applications are considered by an Advisory Committee, and are generally processed within three months (RKI, 2015b). A written agreement is drawn up which regulates the scope of data usage and publication rights. Use of the released data is solely for the purposes described in the application, and in particular specifies the **exclusion of a commercial use** of the dataset. In addition, data protection requirements must be met, no attempts to de-anonymise the data can be made, and individual cases cannot be linked to other data sources. Data users must conform to 'Good Practice in Secondary Data Analysis" (AGNES

et al., 2008); these guidelines specify the requirements as set in the federal data protection law [BDSG section 3a] to anonymise / pseudo-anonymise data (see Appendix 3). Also, a qualified person in the research team must have responsibility for complying with data protection standards.

5. **Data use.** What, if any, are the rules governing the use of RWD, including arrangements between data suppliers and recipients, and rules around use for HTA?

Access to RWD is subject to the provisions of the Data Protection Act in Germany, which will always be covered through the contract arrangements between data suppliers and data users.

Output of RWD in Germany as measured by peer reviewed research that uses RWD is very low, likely reflecting the restrictive access arrangements in Germany (Hughes & Kessler, 2013). In the IMS Health report on RWE market impact, Germany was noted to be the lowest among the countries assessed in terms of public use of RWD. Despite high electronic data capture, there is strong conservatism relating to privacy and use of data, and also scepticism around data quality. Therefore decisions around medicines at a public level are informed exclusively by RCTs. However, some payers use RWE to inform decisions in activities such as disease management programmes (Hughes & Kessler, 2013).

Legislative changes by AMNOG created a theoretical role for RWE in decision-making around medicines, allowing for the use of RWE to supplement the AMNOG benefit assessment for market access and pricing (Hughes & Kessler, 2013). However, the assessment is undertaken rapidly and does not allow time for this information to feed into the process. However, some progress has been made in using RWE outside of AMNOG, for example for price negotiations with IQWiG, where RWE has been accepted by sickness funds with short-acting insulin analogues (Hughes & Kessler, 2013).

According to the same report there is some evidence that payers and clinicians are using RWE and collaborating with other stakeholders, including industry, to create disease registries to collect data on safety, usage, adherence and health outcomes (Hughes & Kessler, 2013). However, very little information is in the public domain, and evidence of its application to decision-making is extremely limited.

An example of RWE use in practice by a pharmaceutical company in the area of diabetes is a Sanofi-run study looking to confirm the effectiveness and safety of Lantus, which restored access and premium pricing in Germany (Hughes & Kessler, 2013).

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

Conservativeness in German data protection can be seen through the long and protracted issuance of electronic health insurance cards for all residents, which has been ongoing for many years and halted at various times for confidentiality concerns. In a document for patients, residents are reassured that *"each insured person can decide for himself or herself whether and how data is stored, who is able to access what data, and what information can be shared"*. Without patient consent, no data can stored and access is not possible, and only administrative data *must* be stored on the card. Even

access to data by physicians treating a patient can only be granted with the patient's consent (by the patient entering a PIN). The patient can also view the last 50 data accesses to their data (Gematik, 2012). The benefits of the new electronic cards are discussed largely in terms of reduced administrative burdens, rather than facilitating further data use utility.

SUMMARY

According to the Data Protection Act (Bundesdatenschutzgesetz) personal data may only be accessed if explicit consent is obtained. Provision in the law for use of data for scientific research purposes is through a special rule which describes, as in other countries, permitted use where obtaining consent is impossible or would require disproportionate effort, and where the scientific interest of the research significantly outweighs the risk to privacy. Data subjects always have the right to object to the processing of their data, and health data in particular is subject to strict protection rules.

Few data linkage projects are undertaken at a national level. When the issuance of electronic health cards is fully implemented, the environment for data collection and linkage may improve as all interactions should be attached to a unique health ID. Where they are undertaken currently, data linkage is carried out with a pseudonymisation process, allowing linkages to be undertaken with a common pseudonym by a trusted third party. However this pseudonymisation process is irreversible, so mistakes in the information that feed into it can lead to missing data or duplicates. Whilst cancer data registration is advanced in Germany, the quality of data linkage based on pseudonyms and other limited identifiers used is questionable (OECD, 2013). It is unclear whether the changes that would be necessary to facilitate deterministic linkages would be acceptable to the German population.

9. The Netherlands

1. **Brief overview** of the health system and collection / management of patient data

Total health care spending in the Netherlands is around 12% of GDP, which is above the OECD average (Statistics Netherlands, 2013). A recent report published by The Commonwealth Fund provides an excellent and succinct overview of the Dutch health care insurance and delivery system (Wammes et al., 2015). Figure 4 is the authors' summary diagram.

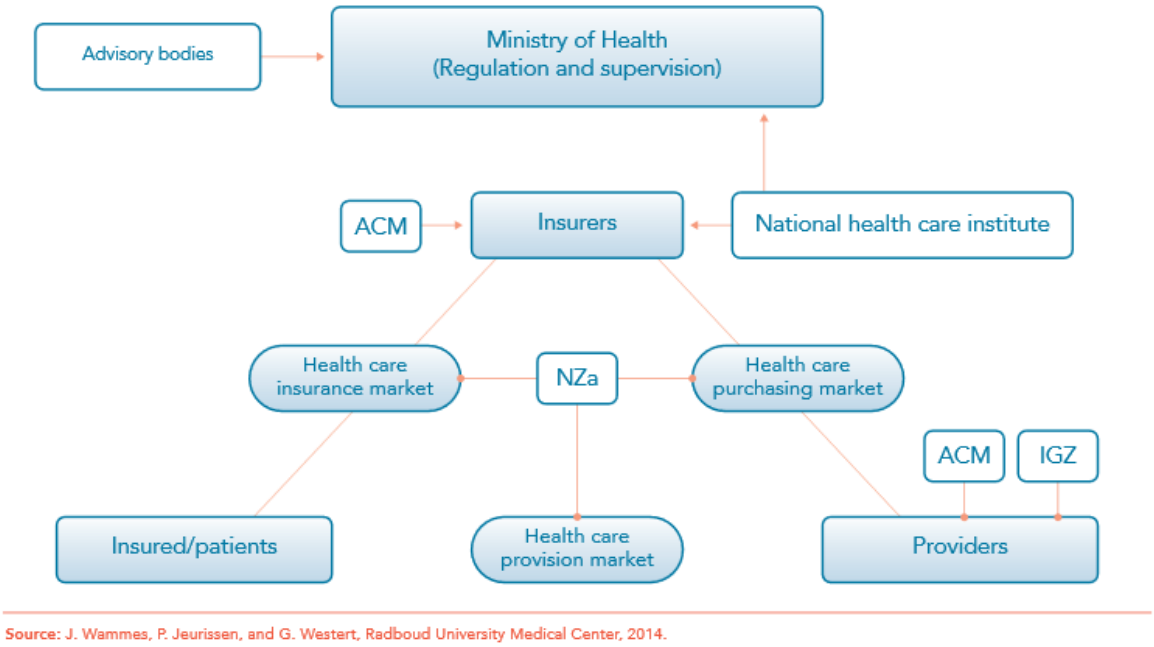


Figure 4 Organisation of the health system in the Netherlands

Source: Wammes, J., Jeurissen, P., and Westert, G., 2015. 'The Dutch Health care System, 2014' in: International Profiles of Health Care Systems. The Commonwealth Fund

Notes- NZ: Dutch Health care Authority (*Nederlandse Zorg autoriteit)* responsible for the supervision and regulation of the Dutch health care system; IGZ: Health care Inspectorate (*Inspectie voor de Gezondheidszorg*) monitors and controls the quality of health care services, prevention measures, and medical products; ACM: Dutch Competition Authority (Autoriteit Consument en Markt)

The Ministry of Health has overall responsibility for the health care system, including setting health care priorities and monitoring access, quality and costs. In accordance with the Health care Insurance Act 2006 (Zorgverzekeringswet: ZVW), all residents are legally obliged to purchase statutory health insurance, which is provided by private insurers but regulated under public law. The Government dictates the cover that must be provided by the standard package of health care insurance companies, who must accept anyone who applies for the standard package and must charge all policyholders the same premium (Government of the Netherlands, 2015a). People may buy additional private insurance to cover further services. Whilst premiums are fixed and the same for everyone, income-related contributions are made as laid out in the ZVW. The Government pays for the costs of insuring children under the age of 18.

The Ministry of Health relies on advice from the National Health Care Institute in defining the standard benefits package. Care is provided by private providers who then bill their services to insurers based on a DRG-type system called Diagnose Behandeling Combinatie (DBC), of which there are several hundred. Citizens generally buy their health insurance from one of four large insurers (1 non-profit; 3 for-profit). Contributions are pooled centrally and paid to insurers using a risk-adjustment formula based on age, gender, labour force status, and "health risk" (based on past hospital and

drug utilization). Insurers compete for enrolees through their purchasing and contracting with providers.

The Drug Reimbursement System in the Netherlands is based on classification into groups of 'interchangeable' drugs, where a fixed refund price is set based on the average list price. When a new drug cannot be clustered, it cannot be reimbursed unless there is a clinical benefit and it is cost-effective; the requirements for submission of health economic data for HTA are similar to those for the UK, Australia and Canada (ISPOR, 2007).

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. Routinely collected patient data.
Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.

The Dutch health care system generates two major types of routinely collected patient data: 1) electronic health records (EHR), and 2) insurance claims data. The EHR capabilities and claims data systems are impressive in their breadth, but to date have not generated as much RWE as smaller more focused databases.

Nearly all practitioners in the Netherlands (97%) utilise EHRs in their practice. Patient records are maintained at the practitioner or regional level. Collection of data into a central database – the Dutch National Health care Information Hub (LSP) – was planned, through an *opt-out* system of patient consent, to be undertaken by the National Information and Communication Technology Institute for Health care (NICTIZ). This would have connected remote information hubs into a national, searchable database (Goldstein and Rein, 2010). The system has been called "health care Google", where regional exchanges would be linked and records found via "*a searchable database accessible to eligible practitioners throughout the country (i.e., those who meet a set of minimum security and functionality requirements)*" (Goldstein & Rein, 2010). Patients would have the option of segmenting data based on provider, care delivery setting, and data type, as well as being able to opt-out of the exchange entirely. However, despite advanced progress against this national exchange point for electronic patient information, it was halted in 2011 when the Senate voted unanimously against the law that would have been required to support it. However, there are plans to re-launch the initiative on an *opt-in* basis (OECD, 2013).

In terms of insurance claims, each of the four major insurers must process and pay claims from providers, such as GPs and hospitals. All such care is delivered under the (DRG-type) DBC system via defined Diagnosis-Treatment Combinations (DTCs). Each insurer has its own set of tariffs for each DTC. As described in a recent report by The RAND Corporation, the Achmea Health Database (previously called the AGIS Health Database) has been constructed from claims data by a private insurer in the Netherlands (Miani et al., 2015). It has data on 1.2 million patients but lacks clinical information (e.g., test outcomes and information on adherence) and contains only demographic patient data and recent diagnoses. It includes data on providers and services provides as well as prescription drug dosage and costs. This large cohort dataset provides opportunities for epidemiological research, though its main purpose is to provide

information on health care consumption and the interaction between services provided in primary care, secondary care and public health (UMC Utrecht, 2015). The data can also support health care management and quality evaluation. According to the RAND report, the data has high reliability due the economic motivation in its collection. Smeets and colleagues (2011) provide an overview of the potential and limitations of the database.

There are many other datasets of routinely collected data in the Netherlands, an important source for which is 'PHARMO', an independent scientific research organisation which is *"dedicated to the study of epidemiology, drug utilisation, drug safety, health outcomes and utilisation of healthcare resources"* (PHARMO, 2015). Datasets include: a GP database (a longitudinal observational dataset containing computer-based records from collaborating practices covering 1.5 million patients); Out-patient pharmacy data (covering 20% of the Dutch population); Clinical Laboratory Register; In-patient Pharmacy (drug database containing 1.5 million patients and data on drug, dose, duration, diagnosis and length of stay); The Dutch Medical Registry (data on all hospital admissions in the Netherlands containing all treatments and diagnoses); Mortality Register; the Eindhoben Cancer Registry (containing detailed diagnostic information); Perinatal Registry; the Dutch National Pathology Registry; and the Thrombosis Register.

The Netherlands implemented the EU Data Protection Directive 95/46/EC on 1 September 2001 via the Dutch Personal Data Protection Act. Enforcement is through the Dutch Data Protection Authority: "College Bescherming Persoonsgegevens". According to Article 21 of the Act, the prohibition of the use of personal data concerning a person's health does not apply where the processing is being carried out by (a) medical professionals or institutions where access is necessary for proper treatment and care or administration; (b) insurance companies (as specified in the Insurance Supervision Act 1993) provided that this is necessary for assessing risk (provided the data subject has not objected) or for the performance of the insurance agreement; (c) schools where special arrangements in relation to a child's health are required; (d) institutions of child protection; (e) Minister of Justice where this is necessary in connection with implementation of a prison sentence, and; (f) administrative bodies, where access its necessary for the implementation of laws, pensions, or reintegration of workers (Upper House of the Dutch Parliament, 2012). According to Article 23, prohibition of processing personal data does not apply if either express consent from the data subject has been obtained, or if it is necessary with a view to important public interest, in which case this is *provided for by law or else the Data Protection Commission has granted exemption.* In addition, the prohibition of processing personal data (as set out in Article 16) does *not* apply for scientific research or statistics purposes where: the research serves a public interest, the processing is necessary for the research or statistics concerned, it appears impossible or would involve disproportionate effort to obtain express consent, and where there are sufficient guarantees to ensure that the processing of data "*will not adversely affect individual privacy to a disproportionate extent*" (Upper House of the Dutch Parliament, 2012).

b. Collecting de novo patient data.
Governance arrangements for research to collect new data.
Kdocumentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

In the Netherlands, The Central Committee on Research Involving Human Subjects (CCMO in Dutch: Centrale Commissie Mensgebonden Onderzoek) is the body responsible for implementing the Medical Research Involving Human Subjects Act (WMO). The CCMO has a broad range of tasks, which include accrediting Research Ethics Committees, acting as the competent authority for clinical research with medicinal products, reviewing protocols for medical research involving human subjects, registering protocols, and acting as the administrative body for appeals and objections around the implementation and application of the WMO (CCMO, 2015).

There are 24 accredited Research Ethics Committees in addition to the CCMO. Research must be submitted to an accredited Ethics Committee for approval before being conducted, which will review protocols in relation to Dutch law, in particular the WMO. The Committees consider all investigational trials as well as non-therapeutic observational studies (EUREC, 2015c).

3. **Data linking.** To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

In the Netherlands there is no unique identifying number for patients for healthcare specifically, and therefore other variables have been used for research requiring data linkage (OECD, 2013). However, since 2009, all care providers and health insurers must refer to the 'citizen service number' when exchanging information about patients and in electronic patient records. This could in principle facilitate direct linkages across providers and between EHR and claims data: however, the latter has apparently been rare. The citizen service number is allocated to all residents and is the number used also for passports, driving licenses and identity cards; it also replaces the social security and tax number (Government of the Netherlands, 2015b).

As described above, the planned Dutch National Healthcare Information Hub would have acted as an exchange point for electronic patient information on a national basis, thereby creating a central database of routinely collected data which would have operated through an opt-out system of consent (Goldstein & Rein, 2010). Progress has halted, and an opt-in model is being considered (OECD, 2013).

There are several individual data linkage programmes in the Netherlands which can provide an access point for linked health care data to facilitate pharmaceutical research. One such project is the 'Mondriaan', which started in 2007 with the goal of creating a national network containing health care and research databases, which arose from a partnership between GlaxoSmithKline, Sanofi, University of Utrecht and University of Groningen (Mondriaan, 2011a; TIPharma, 2015). This links GP data (from certain providers / networks), pharmacy data, claims data (from the Achmea Health database), and research cohorts. In order to comply with legal regulations, data processing is subject to various 'Privacy Enhancing Techniques and Procedures' (PETs) that protect privacy both in the anonymisation of data and in preventing that data being re-identifiable. An example of such PETs is in their use of a trusted third party to conduct encryption and linkage between datasets. This means that encrypted research data can be separated from the identifying information such as names and addresses (Mondriaan, 2011a).

Another source of linked data is the PHARMO Record Linkage System, which links six major sources of data that covers 2 million Dutch residents (NCI, 2013). The linked dataset provides information on outpatient and inpatient drug prescriptions, hospital admissions, diagnoses, treatment procedures, and GP data. It was established at the Utrecht and Rotterdam Universities in the early 1990s to provide insights into the effectiveness, safety and value of prescription drugs used in daily practice (NCI, 2013). The Pharmo Institute emerged originally from a linkage project which predated any national patient identifiers between community pharmacy data and the National Dutch Hospital Registration; linkages were established via semi-deterministic means based on a combination of birth date, gender and GP identification number; this evolved into Bayesian-probabilistic models. In 1999 the PHARMO Institute was established as an independent organisation responsible for the governance, maintenance, collection, linkage, privacy protection and analysis of data (PHARMO, 2015).

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

For access to the Mondriaan database, consideration is given to the application with the condition that the research must have a scientific aim. In addition, the source dataset organisations must all agree with the research proposal for data to be shared (presumably with each having its own process for review). Depending on the nature of the data request, Mondriaan may request a review of the proposal by a Medical Ethics Review Committee, or through the Scientific Advisory Council of the University Medical Center Utrecht (Mondriaan, 2011b).

In contrast to the Mondriaan project which provides data to research organisations but does not conduct in-house data analysis services, the PHARMO Institute is a research organisation which itself conducts analyses of patient data to derive real-life insights into the performance of medicines. It is not clear from the PHARMO website to what extent data is available or how data access or in-house services are facilitated. However, it is mentioned that pharmacoepidemiological studies have been conducted with a range of stakeholders including the pharmaceutical industry.

5. **Data use.** What, if any, are the rules governing the use of RWD, including arrangements between data suppliers and recipients and rules around use for HTA?

The IMS report on RWE reported that the Netherlands produces relatively high-quality data as well as providing explicit guidance on how to use of RWE for cost modelling, outcomes research, and cost-effectiveness. Furthermore, RWE has been used for a long time, for example, by PHARMO in drug studies (Hughes & Kessler, 2013). The claims data generated by the insurance system has not been used to any great extent for health services research or cost studies. However, there is apparently increasing interest in constructing a longitudinal claims database at the National Health Care Institute, which could be made available to academic researchers. The IMS report (Hughes & Kessler, 2013) placed the Netherlands behind Sweden and the UK in terms the application of RWE application due to the lack of impact. This is consistent with other assessments of the impact of their efforts at RWE generation (Garrison et al., 2013).

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

In the report by the OECD on strengthening health information infrastructure for health care quality governance, respondents to the survey from the Netherlands noted that it was '*very unlikely'* that electronic health records will be used to monitor national health care quality over the next five years. This is in part due to the legal barriers which need to be addressed, and may also depend on decisions around an opt-in or opt-out model of patient consent for these activities.

SUMMARY

Like other countries, data protection legislation in the Netherlands restricts the processing of personal data but makes special exemptions for health data where these are used for the direct purposes of care. Processing for research is permitted if either consent has been obtained from the data subject or if it has been permitted by law or granted exemption by the Data Protection Commission. This is facilitated by the wording in the legislation regarding the use of data for scientific research where it is in the interest of the public and where obtaining consent would require disproportionate effort.

An important step toward a centralised platform for data linkage would have been achieved by the 'Dutch National Healthcare Information Hub', which was planning to operate with an opt-out model of patient consent. However, this was opposed by the Senate on privacy grounds; there are plans to re-launch the initiative on an *opt-in* basis. This could be compared with the care.data programme in the UK, which has also met with opposition. However, there are several organisations in the Netherlands which do link a variety of datasets which can provide anonymised individual-level data for the purposes of research.

Although EHR and claims data systems are generating a lot of "big data" in the Netherlands, there is little to support the notion that health authorities have been able to leverage this information to improve clinical decision-making or overall health system outcomes. PHARMO, an NGO research enterprise, has, however, been able to use data on medicines use to conduct numerous excellent pharmacoepidemiological studies, that have had international impact on medical practice. Little work has been done to link insurance claims data to EHR to study costs and cost-effectiveness.

10. Australia

1. **Brief overview** of the health system and collection / management of patient data

Health care in Australia is provided by a combination of public and private institutions (AIHW, 2015). Public sector health services are provided by all levels of government: local, state, territory and the Australian Government. Private sector health service providers include private hospitals, medical practices and pharmacies. In 2011-12, health expenditure in Australia was estimated at 9.5% of gross domestic product (GDP). Almost 70% of total health expenditure during 2011-12 was funded, with the Australian

Government contributing 42.4% and state publicly and territory governments 27.3%. The remaining 30.3% ($42.4 billion) was paid for by patients (17%), private health insurers (8%) and accident compensation schemes (5%)

The Australian Government's funding contributions include a universal public health insurance scheme known as Medicare. Medicare was introduced in 1984 to provide free or subsidized treatment by health professionals such as doctors, specialists and optometrists. The Medicare system has three parts: hospital, medical and pharmaceutical. The major elements of Medicare include free treatment for public patients in public hospitals, the payment of benefits or rebates for professional health services listed on the Medicare Benefits Schedule, and subsidization of the costs of a wide range of prescription medicines under the Pharmaceutical Benefits Scheme. Individuals can have Medicare coverage only, or a combination of Medicare and private health insurance coverage.

Three parties are involved in coverage and reimbursement policy for pharmaceuticals: the Pharmaceutical Benefits Advisory Committee (PBAC), the Pharmaceutical Benefits Pricing Authority (PBPA), and the Minister of Health and Ageing. PBAC is responsible for assessing the comparative clinical and cost-effectiveness of interventions as well as overall budget impact and advises the PBPA regarding the value of an intervention. The PBPA in turn, negotiates with the pharmaceutical manufacturer to establish a reimbursement price for the product subject to approval by the Minister of Health. Performance-based arrangements are used as a tool to improve the value of interventions. The process for establishing these arrangements is publically documented (Department of Health, 2014). In 2010, the Australian Government and the local pharmaceutical industry agreed to implement a Market Access Scheme program (Wonder et al., 2012). This program introduces a "mechanism whereby the PBAC may recommend PBS coverage at a price justified by the existing evidence, pending submission of more conclusive evidence of cost-effectiveness to support listing of the drug at a higher price." Further the "PBAC will provide advice in relation to sources of uncertainty and specific evidence required to support a subsequent application." These submissions will be restricted to those where there is an agreed clinical need, PBAC would not otherwise recommend listing at the proposed price, and there is a program of evidence generation related the identified uncertainty due to report within a reasonable timeframe (Department of Health, 2010).

2. Core legislation and governance arrangements for the collection and/or use of patient data

a. **Routinely collected patient data.**
   Core legislation governing the collection / use of routinely collected patient data. Key documentation outlining principles of governance and data protection.
b. **Collecting de novo patient data.**
   Governance arrangements for research to collect new data. Key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

The Privacy Act 1988

The Privacy Act 1988 is the key Australian national legislation which governs the protection of personal information (Australian Government, 2015). All organizations that provide a health service are covered by the Privacy Act. Under the Privacy Act a 'health service' includes any activity that involves:

- Assessing, recording, maintaining or improving a person's health; or
- Diagnosing or treating a person's illness or disability; or
- Dispensing a prescription drug or medicinal preparation by a pharmacist.

Medical Research

The Privacy Act permits the handling of health information for health and medical research purposes, without individuals' consent in certain circumstances. The National Health and Medical Research Council (NHMRC) has issued two sets of legally binding guidelines for handling health information for research purposes without individuals' consent (Anderson, 2014; NHMRC, 2014; Pilgrim, 2014). The guidelines also assist Human Research Ethics Committees (HRECs) in deciding whether to approve research applications. The guidelines are produced under sections 95 and 95A of the Privacy Act. Section 95 of the Privacy Act sets out procedures that HRECs and researchers must follow when personal information is disclosed from a Commonwealth agency for medical research purposes. Section 95A provides a framework for HRECs to assess proposals to handle health information for health and medical research without individuals' consent (medical research includes epidemiological research). The guiding principle is to ensure that the public interest in the research activities substantially outweighs the public interest in the protection of privacy.

The Privacy Amendment (Enhancing Privacy Protection) Act 2012 amended Section 14 of the Privacy Act with the Australian Privacy Principles (APPs) that govern the conduct of Commonwealth agencies in their collection, management and use of data containing personal information (Australian Government, 2014). The APPs do not permit agencies to use or disclose identifiable records of personal information for research and statistical purposes, unless specifically authorized or required by another law, or the individual has consented to the use or disclosure.

Other legislation and regulations

In addition to the Privacy Act, there are also some regulations at State and Territory level, either in the form of legislation related to privacy generally, or administrative codes of practice, that may have a bearing on either access to personal information to be used in research or the way in which proposed research must be conducted. Some jurisdictions have included stricter limitation on the handling of personal information as part of the administrative structure of health departments and agencies.

Table 5 National, State and Territorial legislation and regulations related to privacy or access to personal information from the Australian Commission on Quality and Safety in Health care

National	The Privacy Act 1988 (Section 95) including Information Privacy Principles (applicable to Commonwealth agencies and the ACT, not applicable to other States and Territories)
Private Health Sector	The Privacy Act 1988 (Section 95A) including National Privacy Principles (applicable to all health service providers in the private health sector)
Australian Capital	Privacy Act 1988

Territory	Health Records (Privacy and Access) Act 1997
New South Wales	Health Records and Information Privacy Act 2002
Northern Territory	Information Act 2002
Queensland	Information Privacy Act 2009 Health and Hospitals Network Act 2011 Private Health Facilities Act 1999 Public Health Act 2005
South Australia	Cabinet Administrative Instruction 1/89: Information Privacy Principles 1, 2 & 3; Code of Fair Information Practice
Tasmania	Personal Information Protection Act 2004
Victoria	Health Records Act 2001 Health Services Act 1988 Mental Health Act 1986
Western Australia	Hospital and Health Services Act 1927

Source: (ACSQHC, 2014)

Databases and Research Institutes

George Institute for Global Health

In 1999, with the support of the University of Sydney Medical School, The George Institute (TGI) was established in Australia with the aim of creating independent medical research institute for global health. TGI has adopted the Australian Privacy Principles set out in the Australian Privacy Act as the minimum standard across all of their offices worldwide. TGI also complies with the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guidelines for Good Clinical Practice with respect to the use, protection and security of health information collected, as well as guidelines issued by the NHMRC with respect to health information that may be accessed in the conduct of research.

TGI (or an approved third-party operating on their behalf) will collect personal information and health information (and at times, other sensitive information) from individuals who participate in human clinical trials and observational post-launch registries undertaken by TGI. Such information collected may include:

- Gender, nationality, heritage, and date of birth;
- Medical history and treatments;
- Medicare number (or similar) and private health insurance information;
- Current medications and treatments;
- Health services and treatments;
- Symptoms, test results and hospital care; and
- Consequential health factors.

TGI may also collect personal information of health practitioners and health providers who are involved in the care of study participants (e.g. general practitioners, physiotherapists, other health care service providers). Such information collected may include name, address, contact details, professional qualifications, experience, and interaction records with TGI (as part of the particular research study or trial). This information is collected for the purpose of administration, management and operation of TGI and the particular research study or trial.

TGI may also collect the personal information on medical experts, researchers and other professionals advising on, overseeing, or assisting in the conduct of a particular research study or trial. Such information collected may include name, address, contact details,

professional qualifications and experience, and registration information. TGI may collate statistical data from study/trial results that have been collected over years for the purposes of future research, or advising on health care policy to Governments and decision-makers.

Skin & Cancer Foundation

The Skin & Cancer Foundation (SCF) is a not-for-profit organization established in 1987 which provides specialist treatment, education and research for a wide variety of skin disorders, skin cancers and melanomas. The Foundation houses the Australasian Psoriasis Registry (APR) which is a national database of patients in Australia with moderate-to-severe psoriasis being treated with biological disease modifying drugs, as well as a suitably large comparison cohort of patients with moderate-to-severe psoriasis not receiving biologic therapy. By collecting long-term information about people's psoriasis management, their health status and quality of life, the APR to provides outcome data to Australian doctors, consumers, policy makers, drug development companies and approval agencies. The APR focuses on the long-term safety and efficacy of established and new generation 'biologic' drugs, and the impact of living with psoriasis.

Australian Rheumatology Association Database

The Australian Rheumatology Association Database (ARAD) is a national database which collects health information from individuals with inflammatory arthritis. The aim of ARAD is to determine effectiveness and safety of new biological drugs used to treat inflammatory arthritis condition, such as Enbrel, Remicade, Humira, Kineret, MabThera and Orencia. ARAD collects information from patients every six months via questionnaires. Questions about medical history, medication history, responses to medication, physical functioning and quality of life are included. ARAD is currently funded through unrestricted educational grants from AbbVie Pty Ltd, Pfizer Australia, Astrazeneca, Bristol-Myers Squibb Australia Pty Ltd.

The ARA owns ARAD and controls access to the data and its release. The management structure of ARAD is shown in Figure 5. ARAD has both a Steering Committee and a Management Committee. The Management Committee comprises the principal investigators. The ARAD Steering Committee conforms to the ARA committee structure and the Operating Principles and Technical Standards for Clinical Quality Registries. The Steering Committee comprises an ARA representative from each state, the Chair of the ARA Therapeutics, Quality Assurance, Scientific Committees, a member of the ARA Executive, a consumer representative from Arthritis Australia and members of the Management Committee (ex-officio). The Steering Committee reports to the ARA Executive.

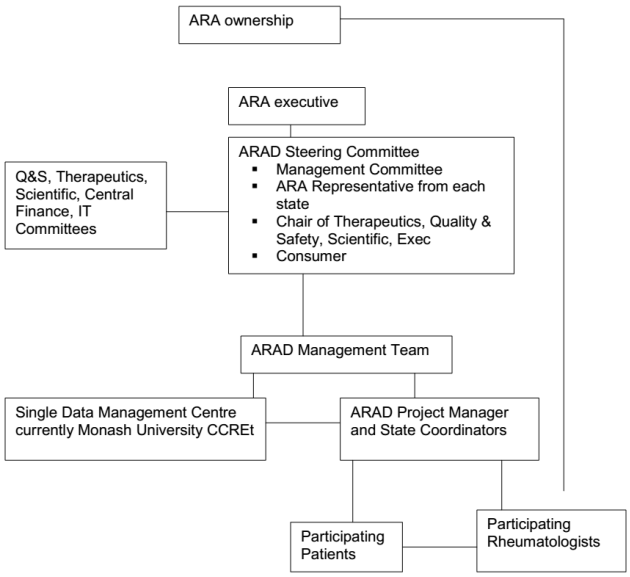


Figure 5 The management structure of ARAD from the ARAD governance document

Source: ARAD Governance Document. Australian Rheumatology Association Database. Available at: https://arad.org.au/Documents/ARADgovernanceMar30_2011.pdf [Accessed 29 January 2015]. (ARAD, 2011)

Personally Controlled Electronic Health Records (PCEHR) system

The Personally Controlled Electronic Health Records (PCEHR) system is a national system of shared electronic health records which can be viewed by patients and their authorized health care providers (Department fo Health, 2014). Its objectives are to provide access to people's health information to help overcome the fragmentation of health information, improve the availability and quality of health information, and improve the coordination and quality of health care provided to patients by different health care providers. The PCEHR can include information on medications, allergies, Medicare benefit and pharmaceutical benefit claims data, organ donation status, location of advance care directives, emergency contacts, and for children – immunizations and early development. The Personally Controlled Electronic Health Records Act 2012 (PCEHR Act) established the legal framework for PCEHR and enrollment in the PCEHR began on July 2012 (Parliament of Australia, 2012). PCEHR Act also authorized the PCEHR system operator to prepare and provide de-identified data for research and other public health purposes. A framework will be developed to ensure that appropriate protections are put in place around the preparation and disclosure of de-identified data.

Population Health Research Network

Population Health Research Network (PHRN) is a national data linkage network comprising of a Program Office located in Perth, Western Australia, a Centre for Data Linkage located at Curtin University in Western Australia, a remote Access Laboratory located at the Sax Institute in New South Wales and a network of Project Participants and Data Linkage Units located in each Australian state/territory. The PHRN links routinely collected data from hospitals, state and territory health departments, and Births, Deaths and Marriages registries as well as de novo collected data such as on those 45 and up to understand healthy aging or Aboriginal health survey (PHRN, 2015).

3. **Data linking.** To what extent can patient data be linked across datasets? Who are the organisations involved, and what are the core governing principles under which they operate?

Australian Rheumatology Association Database

ARAD performs data linkage with various registries including the Australian Institute of Health and Welfare database to monitor morbidity and mortality. International data linkage may also be performed (ARAD, 2011).

Population Health Research Network

To allow data about the same person to be linked across different data collections, Data Linkage Unit (DLU) staff (employees or associates of a government agency) create unique Linkage IDs. To do this, the data owners provide the personal information portion plus the local record ID of each record in their data collections to the DLU. The data owner requires approval from an HREC before providing the data. The other portion of the record containing the health, education or other data remains with the data owner, meaning that the data linkers never have access to this data. Upon receiving the personal information and Record IDs at the DLU, the DLU staff assign a Linkage ID to each person. These Linkage IDs are stored on secure computer servers and can only be accessed by authorized DLU staff. Data owners provide regular updates of the personal information and Record IDs to the DLU.

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

George Institute for Global Health

TGI may disclose personal information to staff, related parties, and approved third-parties (e.g. agents, service providers, collaborators and research partners) who are working on the study or research program for which your personal information was collected; but only to such persons who need to know. TGI staff must comply with privacy and confidentiality terms as part of their employment. To be an approved third-party of TGI, that party must be subject to similar privacy and confidentiality laws, or have a professional and/or contractual obligation of confidence.

Skin & Cancer Foundation

Information will only be used or disclosed for the primary purpose for which it was collected. Personal information about an individual will not be used or disclosed for a secondary purpose unless:

- The purpose is closely related to the primary purpose and the individual would reasonably expect the information to be used in that way; or
- The information is health information and its use is necessary for records or statistical analysis relevant to public health; or
- The individual has consented (recognizing the competence to consent); or
- The Skin & Cancer Foundation has a legal obligation to disclose personal information which overrides the provisions of the primary legislation.
- The Skin & Cancer Foundation will not sell or exchange or release personal information about an individual for commercial gain.

Sensitive information about an individual will not be collected without that individual's consent; or the information is necessary for research relevant to public health, compilation or analysis of public health statistics, or the management or monitoring of a health service and that purpose cannot be served by collection of non-identified information and it is impracticable to seek the individual's consent.

Australian Rheumatology Association Database

ARAD complies with the Australian Commission on Safety and Quality in Health Care Operating Principles for Clinical Quality Registries and Commonwealth and State privacy laws. All ARAD research staff are trained appropriately and sign confidentiality agreements. Only de-identified patient data will be made available to third parties. No ARAD patient personal information will be released to third parties without explicit patient consent. An ARAD manual describing ARAD data collection and management processes is continually being updated. All policies concerning consent, requests for involvement, follow up and data management and privacy are available upon request. The ARA executive accepts legal liability for ARAD as covered under the ARA insurance policy.

Population Health Research Network

Only researchers who have approval from a Human Research Ethics Committee (HREC), who have signed confidentiality agreements with data custodians and who have sanctioned data security plans in place are allowed to access PHRN data. Researchers are only permitted to use the data for the particular project and must only use it in the precise way that has been approved. Approval is evaluated according to a number of criteria:

- Appropriate purpose:
    - To facilitate research which may contribute to the promotion, protection and maintenance of the health of the public;
    - To facilitate the planning, evaluation and delivery of health services; and
    - To contribute to knowledge regarding research methodologies relating to health data collection, linkage of health-related data and compilation and use of health related statistics generally.
- Eligible researchers:
    - Researchers with the appropriate experience, qualifications, facilities and funding to conduct the proposed research;
    - Students and early career data users who are part of a research team with appropriate experience and qualifications; and
    - International collaborators, depending on the nature of the project and the form of the data requested.

Access to PHRN data is generally provided on a first come first served basis but may involve prioritization based on a number of criteria:

- Data availability;
- Complexity of project/technical feasibility;
- Public interest;
- Resource availability e.g. funding
- National Health Priority Areas determined by the Australian Health Ministers' Conference; and

- Strategic priorities.

At the end of the approved access period researchers must dispose of the data in accordance with the data destruction plan contained in their project application and agreed to by the relevant HREC/data custodians and provide notification that this has been done. Both the HREC and the data custodians have the right to audit/monitor/check that the researchers are adhering to the agreed retention and disposal plan.

5. Data use. What, if any, are the rules governing the use of RWD, including arrangements between data suppliers and recipients and rules around use for HTA

Australian Rheumatology Association Database

Research proposals requesting access to identifiable data will require ethical approval from the researcher's institution as well as ethical approval from one of the following:- the Cabrini Hospital Human Research Ethics Committee, Melbourne; Royal North Shore Hospital Human Research Ethics Committee, Sydney; and South Eastern Health Human Research Ethics Committee Southern Section, St George Hospital.

The ARAD Steering Committee may grant access to de-identified data only (ARAD, 2015a). ARAD participants may be invited to participate in external research projects if approval has been granted by both the ARAD Steering Committee and relevant HRECs. Initial contact for those studies will be made by ARAD personnel. Identifying information of participants will not be available for any purpose other than reports of health benefits and harms to their treating rheumatologist or other recipient nominated in writing by the participating patient and rheumatologist. No ARAD identification details will be released to third parties without consent (written or electronic) of ARAD participants. All third parties must sign ARAD and any other relevant confidentiality agreements. Industry partners have no rights to directly access the data. They may receive reports of grouped non-identifiable data upon request. The process for external projects:-

1. Researchers write to ARAD Steering Committee with proposal;
2. Researchers can either wait till they have ethics approval and funding before submission or be processing approvals concurrently but the program will not proceed until ethics clearance and funding are available;
3. Proposal also contingent on ARAD staff having time to provide the service;
4. ARA Executive to be informed of all projects;
5. ARAD newsletter to participants and updates in ARA e-bulletin to notify members and patients of ongoing projects;
6. Rheumatologists will be notified of proposed studies;
7. Steering committee to make recommendation about need for individual patient approval from the ARAD participant's treating rheumatologist on a case by case basis;
8. Steering committee, in collaboration with ARAD PIs, to make a recommendation about level of funding required for the project to proceed;
9. Authorship to be discussed a priori between researchers, ARAD Management and ARAD Steering Committee and decided on a case-by-case basis in accordance with international guidelines.

When access to ARAD data is granted to external researchers, the researchers are required to enter into a confidentiality agreement (ARAD, 2015b). Elements of this agreement include:

- Restrict use and disclosure to that which was pre-specified
- Make no attempt to attempt to identify or make unauthorized contact with any individual
- Not to make any unauthorized linkages to other datasets
- Notify ARAD staff regarding any breach

6. **Governance ideals and changes to the environment.** Key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

The National Health and Hospitals Reform Commission (NHHRC) was established by the Australian government in 2008 to develop a long-term health reform plan for Australia. In 2009, the NHHRC prepared the report "A healthier future for all Australians" with over 100 short and long term recommendations to transform the Australian health system (NHHRC, 2009). Among these were a number of data governance recommendations.

General recommendations for use of data

- Data should enhance decision making, drive improvements in clinical practice, guide how resources are marshalled and deployed and provide the basis for feedback loops to promote improvements in access to and quality and efficiency of care:
  - Develop a credible and well-resourced national health data system for monitoring and comparing performance in both private and public settings;
  - Compare, analyze and report data back to clinicians, health services and consumers in a user-friendly format ;
  - Use data to understand extent of a clinical problems, how we should target improvement efforts for best effect, and metrics for success.

Specific Recommendations to accomplish the General Recommendations

- Introduce unique personal identifiers.
- Develop a clear set of nationally agreed and implemented standard rules to optimize interoperability of health record systems.
- Legislate to ensure the confidentiality and privacy of a person's electronic health data, while enabling secure access to that data.
- Develop sound patient outcomes data for primary health care
- Collection and linkage of public and private hospital episode data to the Medicare Benefits Schedule and Pharmaceutical Benefits Scheme using a patient's Medicare card number.

SUMMARY

Nationally, the collection and use of personal information in Australia is governed by the Privacy Act. This act permits the use of health information for research without individuals' consent if public interest in the research activities substantially outweighs the public interest in the protection of privacy.

Two key sources of RWD in Australia are the Population Health Research Network (PHRN) and the Personally Controlled Electronic Health Records (PCEHR) system. The

PHRN is a national data linkage network of routinely collected data from hospitals, state and territory health departments, and Births, Deaths and Marriages registries as well as de novo collected data. Linkage is performed by specific data linkage staff. Access to this data for research requires approval from a Human Research Ethics Committee (to ensure privacy protection) and from PHRN (to ensure scientifically appropriate research). The PCEHR is a national system of shared electronic health records which can be viewed by patients and their authorized health care providers. The PCEHR system may provide de-identified data for research and other public health purposes but a framework for linkage, access and use has yet to be developed. More generally, the National Health and Hospitals Reform Commission recommends the use of unique personal identifiers, standardized health record systems, legislation to balance confidentiality and privacy with access to data.

11. Country comparison

In order to move from our detailed country case studies to our proposed ideal framework for governance, it is useful first to summarise and compare the main characteristics of the governance arrangements for RWD in our eight markets of interest. In Table 6 we describe these main characteristics according to: the main features of data protection, data linkage, access, and governance ideals and changes in the environment. This is followed by a high-level summary of the issues arising.

Table 6 Data governance country comparison

	Data Protection – Health [Patient consent & Exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
The United Kingdom	Data Protection Act 1998. No consent required for service evaluation/audit. To process confidential information for research purposes organisation must have either: - Obtained informed consent from data subject - Been granted statutory basis for consent exemption: *Section 251 of the NHS Act 2006*, considered where obtaining consent is unfeasible, and the data use is in the public's interest. If neither informed patient consent nor S251 granted, transfer of secondary data must be anonymised. Prospective de novo data requires Research Ethics Committee review (now centralised through Integrated Research Application System), unless project is considered to be audit or service/therapy evaluation. However, distinction between "research" and "audit" can be problematic.	All NHS interactions captured with unique identifier: '*NHS Number'* Linkage undertaken by 'Trusted Third Party' for anonymisation and encryption [England: HSCIC; Wales: NWIS; Scotland: Accredited Safe Havens; N. Ireland: poor development of data linkage programs].	Access to potentially identifiable data determined by ethical review and must be in the public interest. Some organisations are explicit about industry access: HSCIC: data cannot be released 'solely for commercial purposes'; Farr Institute: industry access only permitted if in partnership with academic/NHS institution. Industry access to anonymised data is permissible subject to appropriate conditions of use. E.g. CPRD annual licences (primary care data); IMS health enquiries; 'accredited safe havens'.	Modern data service 'care.data' would centralise management of patient-level information across all health care settings; this could provide an extremely valuable resource. Program has stalled due to public concern around privacy. Initial 'pathfinders' underway; progress uncertain – unlikely to fulfil early ambitions. Number of organisations currently issuing consultations on how best to share data
The United States	The Health Insurance Portability and Accountability Act 1996 (HIPAA). Privacy rule: regulates use of protected health information (PHI) held by "covered entities" and business associates. PHI only used for treatment/payment/health care operations; otherwise patient consent must be obtained. For research organisation processing PHI has either: - Obtained patient consent ('individual authorisation' - Been granted a waiver of the consent requirements by the Institutional Review Board (IRB). IRB waiver will only be granted if research could not be conducted without the waiver, and there must be minimal risk to privacy.	The FDA 'Sentinel Initiative' encompasses data from each registered partner which is maintained between each health-plan firewall. This allows a single coordinating centre to submit FDA 'queries'.	Anonymisation through either 'safe harbour method' where unique identifying characteristics are removed, or 'expert determination method'. When data is rendered anonymous (risk of identification very small), it is no longer considered PHI and no longer subject to the constraints on access.	The Institute of Medicine (IOM) has issued guidance on a 'learning health system' in recognition of the increasing role of digital health data; IOM proposes data sharing models where level of data access is inversely related to information confidentiality. The production of RWD is expected to increase with the passage of the Patient Protection and Affordable Care Act.

	Data Protection – Health [Patient consent & Exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
	De novo patient data covered under the 'common rule' (not applicable to data collected for public health practice).			
France	Data Protection Act 1978 (updated last in 2009), Health Reform Act 2004, Public Health Code 2002. Authorisation for medical research using personal data is granted by CNIL (data protection authority). All data must be anonymised before transmission unless the study is for pharmacovigilance, the project requires it, *or* if the study is carried out by staff responsible for patient follow-up. All data subjects have the right to object to their data being used, and must be informed of uses and users. Data from medical files may only be used for statistics / evaluation in completely anonymised / aggregated form (unless the purpose is for reimbursement). Registry and observational studies must be submitted to the CCTIRS in the first instance prior to the submission to CNIL, via a consulting committee for consideration.	All citizens have a social security number (NIR), but this is deemed too sensitive to use for health records. Development is underway of unique health identifier (INS) – unclear how INS will match health insurance records (with anonymised NIR) Regular projects involving linking primary care data (SNIIRAM) to data on in-patient hospitalisations (PMSI) and survey data (ESPS).	There is a strong emphasis on protecting privacy of personal health data. Still, there is strong demand in France for RWD because of HAS value assessments and conditional reimbursement, but most commentators describe access as restrictive. Namely, the impracticable requirement that any access to data containing the unique identifier requires a decree from the Council of State. Access by companies is generally through contract research organisations, from whom companies will receive result reports. Seems access to PMSI for commercial organisations has been recently relaxed (e.g. PROSPERE and CONSTANCE).	There is some ambiguity in personal identification numbers for health care and how these will be used in the future; inconsistent uptake damages data linkage efforts. There is a call for greater transparency in data management, with recommendations that re-identification risk be assessed openly, with data containing no risk to be made publicly available. There is also a recommendation that data linkages be authorised through a more transparent system that assesses risk versus public benefit, rather than with a decree from the Conseil d'Etat (as currently); this would bring it more in line with UK and US.
Italy	Data Protection Code no.196 introduced in 2003. Permits processing of personal identifiable information if either: - Consent is attained (in writing) - The law authorises it. However, exemption is granted for 'special circumstances' through Section 41 of the Code, which allows that 'Garante' (Data Protection Authority) to authorise use without consent where obtaining it would require disproportionate effort.	Data linking is by 'TS' number, which is used for health and tax purposes. Despite strong data infrastructure, the fragmented nature of health service administration makes data sharing and linkage very difficult.	According to a 2013 'Authorisation' private entities may process health data without specific authorisation of the Garante, but only when consent has been obtained. Little information is available on the criteria for access to regional datasets, for which the applicant must apply to each region	The RWD environment in Italy would benefit from clear guidelines from public authorities on the criteria for approval of research projects and best practice for data linkage projects. AIFA's reimbursement requirements make this particularly important, including reliance on MEAs that use "Drug

	Data Protection – Health [Patient consent & Exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
	Health care professionals and 'public health care bodies' (including universities acting as such) may access health data without consent. For prospective observational studies ethical review must be attained by regional committees; strong heterogeneity has been observed.		individually (regions own and regulate datasets). For all observational research studies using there must be a written commitment to summarise the results and put them in the public domain.	Monitoring Registers" (Dec 2011: 78 therapeutic indications, for 66 active compounds).
Sweden	Collection and use of personal data is regulated through Personal Data Act 1998. Identifiable data can only be collected for explicit and legitimate reasons. Data can be processed only with consent, with exceptions: (1) protect subject's vital interest; (2) pubic interest task; (3) interest is of greater weight than risk. Patient health noted as special case. Sectoral laws: Patient Data Act (2008): written consent is not required but patients must be informed of their participation and have right to withdraw ('opt out'). National aggregated data can be used for statistics, quality analysis and research, upon permission from Ethical Review Board (governed by Ethical Review Act 2003). Same principles apply for research involving non-routinely collected data.	The "personnummer" (Personal Identity Number) is used for all official purposes (tax, social welfare, health care, education, income etc.). Sweden performs amongst the highest in terms of regular data linkage projects. Measures in place to ensure identifiable data from linked datasets is protected. Moreover, government data analysts and external researchers with approved projects are only provided de-identified datasets.	Pharmaceutical Register Act 2005: collection and processing of national register of prescribed medicines. Patients cannot require data to be excluded but access needs explicit consent. Exemption for consent: prescriber can access data to ensure accurate treatment. The National Board of Health and Welfare in Sweden (the Socialstyrelsen) has a process for requests to access patient-level data for research purposes, including when data linkage is required. Sweden does not rule out access to data by commercial companies.	Sweden boasts good medical records, rich datasets, and strong integration, but their use to inform decision-making is relatively low. Sweden is considering the introduction of new legislation which sets out the conditions for personal data access for research and analysis more clearly. Also, emerging "three party agreements" between pharmaceutical companies, HTA body (TLV) and county councils to collect efficacy and outcome data – pilot phase, currently with 8 drugs.
Germany	Federal Data Protection Act: Bundesdatenschutzgesetz (BDSG), introduced 2003 and revised 2009. The collection and processing of personal data is only allowed if the data subject has expressly offered their (written) consent. Provision in the law for use of data for scientific research purposes is through a special rule which describes, as in other countries, permitted use where: - obtaining consent is impossible or would require disproportionate effort, and	Public Health Insurance Act 2003: introduced the electronic health card (eGK) and a new universal health identifier: the Krankenversichertennummer. This provides capability for data linkages, but roll out has been problematic.	Personal medical data that is collected is only allowed to be used for the purpose for which is originally collected. Access to payer claims data is limited. Access to individual data by third parties is only allowed if the patient has given their consent, and strict confidentiality procedures must be adhered to.	Patient privacy concerns in German data protection can be seen through the long and protracted issuance of electronic health insurance cards for all residents, which has been ongoing for many years and halted at various times for confidentiality concerns.

	Data Protection – Health [Patient consent & Exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
	- where the scientific interest of the research significantly outweighs the risk to privacy Public and private bodies which process personal data must appoint in writing a 'data protection officer' who must comply with various duties. Health data is considered in the BDSG as a special category of personal data and is subject to stricter rules.	Data linkage projects are generally undertaken at the state rather than the national level, and only when authorised by law. For national projects, approval from each individual state required. Generally only de-identified data is shared/provided with researchers.	Private organisations can only access reimbursement data through an academic collaboration.	
The Netherlands	Dutch Personal Data Protection Act (2001). Prohibition of use of personal data concerning health does not apply under some circumstances – such as to ensure proper treatment or for insurance companies to assess performance of insurance agreement. Also, prohibition of processing personal data does not apply if consent granted or important public interest. For research or statistics purposes, prohibition also does not apply where disproportionate effort to obtain express consent. For de novo patient data, research protocol must be submitted to an accredited Ethics Committee for approval. Two major types of routinely collected patient data: 1) electronic health records (EHR), and 2) insurance claims data. The creation of central database ('Dutch National Healthcare Information Hub') was planned via an opt-out system – but halted in 2011 for privacy reasons; plans to re-launch on an opt-in basis.	There is no unique identifying number for patients for health care specifically. However, since 2009, all care providers and health insurers must refer to the 'citizen service number' when exchanging information about patients and in electronic patient records. This could in principle facilitate direct linkages across providers and between EHR and claims data – but latter has been rare. There are several individual data linkage programmes: 'Mondriaan' (subject to various 'Privacy Enhancing Techniques and Procedures' (PETs) that protect privacy both in the anonymisation of data and in preventing that data being re-identifiable) and PHARMO Record Linkage System	'Mondriaan': research must have a scientific aim, and all source dataset organisations must agree with the research proposal. Proposal might require a review by a Medical Ethics Review Committee PHARMO: conducts analyses of patient data to derive real-life insights into the performance of medicines. Not clear from the PHARMO website to what extent data is available or how data access or in-house services are facilitated, but mentions several collaborations (including with the pharmaceutical industry) for pharmacoepidemiological studies	The Netherlands produces relatively high-quality data as well as providing explicit guidance on how to use of RWE for cost modelling, outcomes research, and cost-effectiveness. The creation of the 'Dutch National Healthcare Information Hub' would be an important step toward a centralised platform for data linkage. Little work has been done to link insurance claims data to EHR to study costs and cost-effectiveness

	Data Protection – Health [Patient consent & Exemptions for use of data for secondary purposes]	**Data Linkage**	**Access**	**Governance ideals and changes in the environment**
Australia	Privacy Act 1988: governs the protection of personal information. Permits the handling of health information for health and medical research purposes, without individuals' consent in certain circumstances: providing public interest in the research activities substantially outweighs the public interest in the protection of privacy. The Privacy Amendment (Enhancing Privacy Protection) Act 2012: does not permit agencies to use or disclose identifiable records of personal information for research and statistical purposes, unless specifically authorised or required by another law, or the individual has consented to the use or disclosure. There are also some regulations at State and Territory level. The Personally Controlled Electronic Health Records (PCEHR): national system of shared electronic health records which can be viewed by patients and their authorised health care providers. Enrollment in the PCEHR began on July 2012. PCEHR Act also authorised the PCEHR system operator to prepare and provide de-identified data for research and other public health purposes.	Australian Rheumatology Association Database (ARAD): performs data linkage with various registries, as well as with international data sets Population Health Research Network (PHRN): national data linkage network. Data Linkage Unit (DLU) staff create unique Linkage IDs to allow data about the same person to be linked across different data collections.	TGI may disclose personal information to approved third parties (who need similar privacy and confidentiality laws as TGI staff) Skin and Cancer Foundation: data only disclosed for purpose it was collected. ARAD: Only de-identified patient data will be made available to third parties. No ARAD patient personal information will be released to third parties without explicit patient consent. PHRN: Only researchers who have approval from a Human Research Ethics Committee (HREC), who have signed confidentiality agreements with data custodians and who have sanctioned data security plans in place are allowed to access PHRN data.	The National Health and Hospitals Reform Commission (NHHRC) was established by the Australian government in 2008; in 2009 they prepared the report "A healthier future for all Australians" with over 100 short and long term recommendations to transform the Australian health system. This sets out a number of data governance recommendation, including the use of unique personal identifiers, standardised health record systems, legislation to balance confidentiality and privacy with access to data. There are two key sources of RWD: PHRN and PCEHR. The PCEHR system may provide de-identified data for research and other public health purposes but a framework for linkage, access and use has yet to be developed.

Source: OHE Consulting, from publicly available information

Data protection:
National data protection arrangements come into play where data is potentially identifiable to patients. All countries had similar wording with respect to how data should be handled appropriately, which are not covered in the comparative grid. However, these include: only collecting the data necessary for a specific purpose, keeping information secure, ensuring that the data are relevant and up to date, allowing subjects to view their own data on request, and only holding as much information as needed for as long as needed.

Positive examples of governance for RWD can be seen in countries where there is a clear and transparent recognition of the ethical concerns around patient anonymity, alongside an understanding of the benefits to the public of research. In general, data collection and access for the purposes of service evaluation and audit do not require patient consent. It is for the processing of data for research purposes that countries differ. Whilst in general the principal of 'consent or anonymise' is applicable to all countries, the more sophisticated systems have provisions, written into law, for exemption of the consent criteria on a case-by-case basis based on trading off the risk to privacy with the public interest of the research. For example in the UK this is facilitated through Section 251 exemption, in the U.S. by IRB review, and in Italy by consideration by the 'Garante'. In some other countries, data must be fully anonymised before sharing for research, or else new authorising legislation or a decree must be issued for every new use (such as is the case currently in France and Germany).

There are different models of acquiring patient consent. Whilst for some types of research informed individual consent must be obtained from each patient, some countries have in place 'opt-out' models in certain circumstances. For example, in Sweden patient data contribute to quality registers; patients must be informed of this use of their data and must have the right to withdraw. Consent is thereby implied rather than required to be collected explicitly from each patient.

As described at the beginning of this section, planned changes to the E.U. environment for protection of data may have a significant impact for RWD governance. By implementing an E.U. Data Protection Regulation (DPR) (a Regulation is addressed to all Member States and applied in full, without the need for national legislation) to replace the current 1995 Directive (addressed to all Member States requiring national authorities to draw up legislation in order to conform to the Directive within a specific timeframe), the heterogeneity in RWD governance across Europe will be reduced. The initial proposals for the DPR represented a positive and facilitative move forward in terms of recognising the benefits of data for research; however the proposed amendments to the DPR by the Committee on Civil Liberties, Justice and Home Affairs removes the exemptions for consent for use of identifiable research, and this would hinder research dramatically (Fears et al., 2013).

Data linkage:
Data linkage requires unique patient identifiers to be recorded for all health care service interactions. Good models for facilitating these data linkages appear to be those that have a (single) 'Trusted Third Party' to act as a trusted source for encryption, de-identification and linkage. Certain countries, such as the UK, are relatively advanced in this regard, which may be a helped by the fact it is a single payer system. However in

the U.S., where there are multiple providers, bringing together data across providers is facilitated through the 'Sentinel initiative', where common data models and data queries can take place whilst protecting proprietary of data and protecting patient confidentiality. The situation in Italy provides an example of how the fragmented nature of health service administration means that data sharing and linkage is difficult; data linkage in Germany is similarly inhibited.

Some countries, such as Sweden, use the social security number as the patient identifier, which paves the way for linkage projects incorporating information from other sectors. However, this capability has caused data protection concerns in some countries such as France, who are attempting to move away from the use of social security numbers for this reason.

Access:
Where data can be fully anonymised, they are no longer considered personal data and data protection rules do not apply. Access to data that is potentially identifiable depends heavily on the purpose for which the data were collected. This is why consideration of the ultimate use of the data is so important in the initial data collection phase. Most countries draw a line between access to data for research versus evaluation, with the latter being subject to lower levels of scrutiny and patient consent. Access to potentially identifiable data is generally restricted to purposes which are 'in the public interest' (addressing the risk versus benefit concern), and there are many examples of the requirement for industry to partner with research organisations to undertake research on their behalf.

We identified some commentary around the perceived difficulty in establishing whether the motivation for data access is in the public interest or for commercial interests. In Sweden, there are plans to draw out these criteria more clearly in legislation.

In a paper by Di Iorio and colleagues which considers the impact of EU-level changes to data protection laws, the authors questions the legitimacy of the notion that the use of data for management of health care services should be more important than scientific research, which is critical for understanding health benefits, without which fundamental rights to health may be compromised (Di Iorio et al., 2014).

12. Ideal framework for the governance of RWD in health care

We now consider an ideal framework for the governance of RWD in health care, with a view to develop policy recommendations to support this favourable model. By assessing the information collected through the country case studies, observing best practice, and consulting the literature as well as the legal expertise of our collaborators, we aim to assemble and describe the key elements of an ideal framework for data governance in health care composed of general principles that should be transferable across jurisdictions.

Before considering the key elements for an 'Ideal Governance Framework' for RWD in health care, it is worth considering: what are the core elements of governance in general, and what constitutes 'Good Governance'?

As described earlier, governance has been articulated as: *"…the processes, roles, standards and metrics that ensure the effective and efficient use of data and information in enabling an organisation to achieve its goals"* (Gartner, 2014). In setting out the principles of an ideal governance framework, it is therefore useful to articulate our goals, or objectives. The main objective for the use of RWD could be considered to be to demonstrate value and make better decisions about health care. Therefore, a good governance framework will be one that sets out the appropriate processes for developing, accessing, and using RWD to deliver reliable, actionable evidence that will improve health sector decision-making.

12.1. What is good governance?

According to the Independent Commission on Good Governance in Public Services in the UK, the core values underlying good governance can be defined by six core principles, which are summarised in Figure 6.

Figure 6 Core principles of Good Governance

Source: Independent Commission on Good Governance in Public Services, 2004. Established by the Office for Public Management (OPM) and the Chartered Institute of Public Finance and Accountancy (CIPFA), in partnership with the Joseph Rowntree Foundation (The Independent Commission on Good Governance in Public Services, 2004)

The themes described in Figure 6 are all transferable to the health care setting in all jurisdictions. The central aim for governance around the collection and use of health care

data must be to focus on the interest of users of health care services: both those patients from whom data is collected as well as the general public who could benefit as patients or potential patients from the research or service evaluation that good quality data can facilitate. Supporting this central aim must be a strong, well defined, effective and transparent governance structure that in all respects engages effectively with stakeholders and is accountable to the general public.

There are common themes in the general literature about what constitutes good governance, which are also well summarised in a 'Good Governance Guide' produced by a collaboration between experts in Australia, who propose that good governance: is accountable, is transparent, follows the rule of law, is equitable and inclusive, is effective and efficient, and is participatory (MAV et al., 2014). The guide also emphasises that good governance is *not* about making the 'correct' decisions, but about creating the *best processes for making those decisions*. This last point is particularly important for health care; we have demonstrated that governance in the management of health care data is about striking an appropriate balance between risks and benefits. An ideal governance framework should provide an optimal environment for striking that appropriate balance.

12.2. Key principles of a governance framework for RWD

The landscape for the collection and use of RWD is becoming ever more prominent for a wide range of stakeholders in health care due to the shifting paradigm in the development, licensing, and assessment of health technologies. On the regulation side, traditionally RWD has only been employed post-launch to monitor safety. However, the remit of regulators is shifting on two fronts. First, regulators are moving toward alternative licensing models, as treatments become more targeted and public pressure for earlier access to drugs for life threatening diseases becomes stronger. RWD collection will be pivotal in this progressive move toward earlier and more iterative assessments by regulators to reduce uncertainty. Second, regulators are increasingly being challenged to consider effectiveness in the real world rather than limiting assessments to the relatively clean efficacy results provided by clinical trials. This is closely linked with regulators' closer monitoring of benefits versus risks to patients over a medicine's lifecycle. For instance, the European Medicines Agency's (EMA) Roadmap vision and the most recent pharmacovigilance legislation give Europe's regulatory agency the ability to assess how a new medicine performs in clinical practice.

The same can be said on both fronts for reimbursement decision-makers, as those tasked with conducting HTA must work more closely with the regulators and similarly conduct earlier value assessments under greater uncertainty, with a view to re-visiting those assessments as further data is collected. Clearly, it is important that methodological advances in evaluating RWD and generating RWE are made, for which various international initiatives have been formed to develop assessment methods (Garrison et al., 2007; IMI, 2015). However, a pre-requisite for utilising RWE effectively is a strong foundation for collecting and accessing the raw data, and ensuring it is of high quality and as useful as possible. If manufacturers are expected to adapt to the shifting paradigm for early and more iterative adoption and assessment of drugs, then collection of and access to data to facilitate this must be built into drug development plans. A favourable governance framework is therefore critical.

Good data governance is essential for making the best use of personal health information, enabling a learning health system where knowledge flows effectively and

efficiently between research and clinical care, and in ensuring public trust. We have demonstrated considerable variation in local approaches to data governance, which reflects the fact that the law is often not completely clear-cut or prescriptive. Heterogeneity across jurisdictions in data protection practices hinders organisations whose focus is not bound to national markets, and impedes the development of international studies comparing health care service performance and quality (Oderkirk et al., 2013).

Protection of patient and citizen privacy is at the heart of the governance arrangements that have been established for the health care data environment. However, recognition of the need for a *proportionate* approach to governance means that systems must be both rigorous and transparent in their approach, but at the same time flexible. The *risk* to patient confidentiality (of which there are varying degrees) must be set against the *benefit* that could accrue from the research that those data could facilitate; not to consider these benefits is unethical in itself, and will impede progress in monitoring and improving health care treatments and services. It is for this reason that the literature on data governance in health care and advancing an appropriate model for such includes frequent use of the word 'balance' in terms of these risks and benefits.

The need for a proportionate system is highlighted by Sethi and Laurie (2013), who propose a flexible and accessible governance model that pays due regard to both privacy and public interests in research. The authors highlight the prevalent culture of caution among data custodians, finding that they often do not take account of the flexibilities within the law that can support data linking and sharing (Sethi and Laurie, 2013). The same authors were involved in the Scottish Informatics Programme (SHIP), a significant work stream whose output included a research paper on 'Information governance of use of health-related data in medical research in Scotland: towards a good governance framework' (Laurie & Sethi, 2012). The key components put forward and covered in the paper are: (a) guiding principles and best practice, (b) safe, effective, and proportionate governance, (c) roles and responsibilities of data controllers, and (d) researcher training. They argue that the tendency to polarise the options for data sharing into *consent or anonymise* is unhelpful, and does not represent a proportional approach.

In an article that looks forward to how the legal landscape may change in the future, Rosenbaum (2010) promotes the concept of data "stewardship": the existence of mechanisms for responsibly acquiring, storing, safeguarding, and using data (with data governance being the process by which responsibilities of stewardship are conceptualised and carried out). Rosenbaum discusses the evolution of information technology for the storage and use of health data and how this progress makes it possible to manage research in a safe and secure environment. Rather than ask *when*, she asks *how long* it will take for social and legal realignment to assure full use of this technology (Rosenbaum, 2010).

In 2007, the American Medical Informatics Association produced a white paper outlining recommendations towards a national framework for the use of health care data (Safran et al., 2007). Among these was a recommendation that the focus of ongoing discussions be on data access, use and control, "– *not on ownership"*. It was emphasised that a focus on ownership diverted attention from the needed development of sound policies and practice, which must apply across the continuum on data users. The need to increase public awareness of the benefits associated with secondary use of health data was also emphasised, as well as developing a consensus on privacy policies and security.

Data linkage activity which can capture health care utilisation and its impact on health is paramount in advancing health-related research (Oderkirk et al., 2013). This requires that person-level data be collected, managed, and shared in a way that is ethical, accountable, and transparent; policies and laws that enable these data sharing and linkage activities are required in order to strengthen information infrastructure. Below we outline our proposed key components of an ideal governance framework for RWD in health care.

12.3. Balancing public and privacy interest for health care data

As described, the balancing of public and privacy interests—of advancing our understanding of medical treatments through evaluation / research, on the one hand, and the protection of individuals' privacy, on the other—is the unifying thread which reconciles all of the national endeavours to formalise these aims into law and national frameworks for governance. Of course, these fundamental objectives are not mutually exclusive. There is clearly a 'public' value in protecting privacy and having confidence in the health care data management system, and likewise for individual patients whose data is being collected, it is important to know that the information will be put to good use for the benefit of themselves and/or society more generally. Nevertheless, it is useful to categorise them broadly in this way in order to explore the risks to person or to society if they are not managed appropriately.

Whilst RWD is collected alongside clinical practice, and therefore risk to patients is not of the same kind experienced in research of an interventional nature, risks to patients of a system that fails to protect the privacy interests of patients can be significant. These risks could be of a financial nature, for example, if the data is used by health insurance companies to discriminate on coverage level or prices, or if it is accessible to employers. It could also cause psychological harm, such as embarrassment, stigma and stress (OECD, 2013). Additionally, data that is not managed correctly could enable identity theft. On a more general level, there may be a loss of confidence in a health care system or even government if privacy interests of data subjects are not adequately protected. An extreme data protection policy would be one that did not allow any access outside of those directly involved in a patient's care, or one where all data are stripped any type of identifier thus rendering it completely anonymous; this would inhibit any linkage between datasets. The emphasis that is placed on the protection of privacy may differ according to the situation. For example, in early access schemes where critically ill patients with few other treatment options are provided a drug on the basis of accruing further evidence, patients may be willing to accept a greater level of risk. Whilst this issue of a 'willingness to share *spectrum'* (which may change according to a patient's need) is important, we leave this issue to one side for now.

Protection of privacy and the risks associated with disclosure of personal information must be set against the potential benefits arising from making use of that information. A benefit can also be framed as avoiding a harm, and potential harms arising from a restrictive policy around data collection and data access are foregone opportunities in evaluation and research, which could improve population health. The sharing of RWD can enable safety monitoring, service evaluation, and effectiveness / cost-effectiveness research. Where the infrastructure that can facilitate data linkage exists, this can impart rich information on a patient's status across the care continuum. The benefits arising from the outcome of this service evaluation and research activity may be diverse but

include tangible improvements to the services offered to patients. Also, companies can use this information to better understand the impact and uptake of their products, and ensure that this information feeds into research and development, thus having a positive impact on innovation. Restricting access to and linkage of datasets would impede these capabilities. Whilst the risks associated with this opportunity cost for research are less immediate and visible than privacy concerns, they deserve high recognition.

There is clearly a need for a governance framework that provides a facilitative research environment, but which pays due regard to privacy issues and maintains public trust in the system, which requires an appropriate balance between consent, anonymisation, and authorisation. RWD becomes RWE after a series of activities which facilitate the transformation of raw data into analysis and results. It is the *evidence* generated as the product of this process that is of value to stakeholders in health care. By setting out this value chain of RWD, we will draw out the specific elements that should be addressed by a governance framework, through which the balance between privacy and public interests—summarised in Figure 7—should be at the fore. We set out the elements of data governance for RWE in Figure 8.

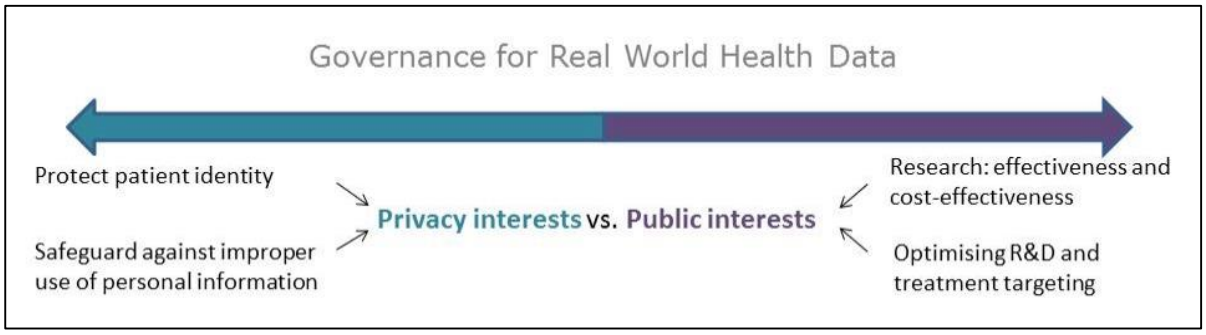


Figure 7 Balance of privacy and public interests

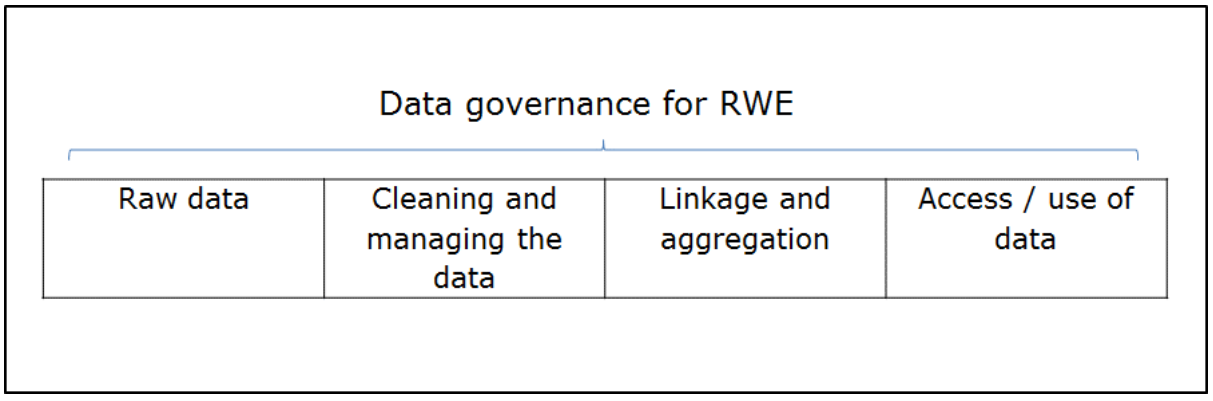


Figure 8 Framework: Key elements of data governance for RWE

12.4. Recommendations for an ideal governance framework for RWD

We have shown that national policies for the collection and use of health care data differ country to country, and that often the legal framework is not completely prescriptive. We propose an aspirational governance framework that could guide the management of data access and use, and the processes that would facilitate constructive interactions among the relevant stakeholders, whilst maintaining accountability and public trust. By setting out the relevant stakeholders and their key roles against the various steps of our framework, we illustrate the shared responsibilities between stakeholders and recognition of their shared values.

Table 7 Elements of a governance framework

Value chain → ***Actors / Stakeholders*** ↓	Routinely collected /De novo *Raw data*	*Cleaning and managing*	*Linkage and aggregation*	*Access / use of data*
Government as Regulator [public policy and legislation]	✓ Data protection legislation (health 'special case') ✓ Equitable patient selection and the protection of vulnerable subjects	✓ Data management: Recognised data stewardship entities	✓ Privacy rules ✓ Develop a clear set of nationally agreed and implemented standard rules to optimise interoperability of health record systems	✓ Managing re-identification risk ✓ Criteria for different uses (& different users)
Data subjects: Patients	✓ Patient consent ✓ Facilitative opt-in / opt-out consent models for research			
Data Collectors • Health care providers • Independent data collectors such as IMS Health or professional organisations	✓Unique patient identifiers (UPIs) ✓ Patient information ✓ Data quality assurance ✓ Data ownership: responsibility for data			
Data Controllers / Providers • Government departments • 'Trusted third parties' • Other agencies		✓ Process for de-identification ✓ Security arrangements: 'Privacy Enhancing Techniques and Procedures' (PETs) ✓Training of staff ✓ Specified arrangements for how long data are kept	✓ Unique patient identifiers ✓ Pseudonymisation ✓ Preparation for sharing	✓ Approval panels ✓ Confidentiality and data use agreements ✓ Balancing benefits of linkage for research with risk for re-identification
Data Users • Payers/ insurers (and their agencies, e.g. HTA bodies) • Health care providers • Private and public researchers, e.g. Pharmaceutical companies & Academic researchers	✓ Approval of data collection activities to be based on intended use ✓ Clear and transparent criteria for de novo data projects			✓ Audit / Service evaluation and quality monitoring ✓Degree of access, level of data, and mode of access ✓ Cost of access ✓ Appropriate experience/qualifications, and funding to conduct research

Source: OHE Consulting

For each step of the framework we outline the key elements of a governance framework, and provide our suggestion for the ideal scenario. In addition, we provide a heat map of how the individual countries perform against the key criteria that we set out in our governance framework. This is based on the OHE team's assessment of the information obtained as part of this project, and is reflective of the country comparison and the individual country assessments.

Raw Data

RWD can take various forms. Routinely collected data is that which is already collected for other purposes, such as electronic health records, and health care utilisation datasets (generally used for administrative or payment purposes). These datasets are likely to be structured in content and offer good population coverage, but arrangements for and rules around their use beyond the purpose for which it was collected must be considered carefully. We use 'de novo' data to describe the collection of further datasets (or further data fields in existing databases or registries) created for the purposes of a specific project. Both have the potential for data to be attached with patient identifying information.

Data protection legislation outlines the fair and lawful means by which personal data can be obtained and processed. Commonly, data protection requirements set by law include that the purposes of data collection be legitimate and specified explicitly, and that the data shall not subsequently be used in a manner that is incompatible with those purposes for which they were initially collected. However, health data represents a special case, and this is reflected in legislative clauses that refer to data purposes not only relating to the direct care of patient, but also to the processing of data for health care evaluation or assessment, prevention practices, and medical research. In some countries the use of patient data for scientific research purposes is considered compatible with the purpose for which the data were collected or processed initially i.e. for the improvement of patient health (Italian Data Protection Code section 99(1)). The criteria for processing personal data which are in the 'substantial public interest' or in the vital interest of the data subject can and should be set out, but there will always be room for interpretation. In addition, legislation can specify that data records be kept accurate, relevant, not excessive, complete, and up-to-date. In France, for example, the Health Insurance Reform Act 2004 specifies that if a patient has an electronic health record the health care professional must refer to and update or complete it. Legislation may also specify the clinical terminology that must be used to complete the record, to ensure data is of high quality.

Patient-level data is sensitive because of the personal-identifying information that is attached to the record. In the simplest form, data protection legislation will permit that identifiable data may be processed if (a) consent is obtained from the data subject, or (b) the law permits it. The notion described above whereby the processing of health care data is deemed to be a special case makes way for judgment to be passed on the systematic collection of data without patient consent, where it is impractical or impossible to do so, and when there is deemed to be substantial benefit in the processing of that data. In many countries this manifests as specific legislation that is passed for specific mandatory datasets, such as Sweden's Pharmaceutical Register Act 2005. Alternatively, there can be provisions within law to set aside the common law of duty of confidentiality for defined medical purposes, such as statutory exemption through Section 251 of the NHS Act 2006 in the UK and Section 41 of the Italian Data Protection Code, whereby patient identifiable data may be processed without the

requirement for consent. This statutory exemption should be granted through a rigorous examination of the necessity of using identifiable data, of its relation to improving patient care, of its being in the public's interest, and the committee should also be satisfied that the effort of obtaining consent from all subjects would be prohibitive or 'disproportionate'. The decision should be reached by a committee with representatives from a wide range of stakeholders including medical, legal, bioethical and public representatives. Terms of reference and criteria should be made public.

As well as appropriate governance for data collection being paramount for complying with legal and ethical requirements in each country, the way in which data is collected has an important influence on how the data can subsequently be used. We have framed the grounds for the collection and use of RWD as a balance between privacy risk and the public benefit to the data's application in research or evaluation. The potential uses of RWD are wide, and each use will involve different risks and benefits. In the same way, the preparation of data to facilitate its use for these various functions can be in varying degrees of identifiability (and therefore be associated with varying degrees of privacy risk), which must be set against the benefit of their use. Models of anonymisation must reflect these considerations.

Where patient consent is deemed appropriate, there can be various methods of obtaining that consent, which may be more or less appropriate depending on the scenario. For some data collection activities, particularly where data is collected for a specific research project, obtaining consent should be through an 'opt-in' system whereby patients are fully informed of the research study and willingly contribute their data. However, in scenarios where data is routinely collected from patients, for example to contribute to national datasets, an 'opt-out' system is often employed, such as the Swedish national quality registers, whereby patient data is entered by default and all patients are provided with sufficient information on how and for what purposes their data will be used, with the option to opt-out. This means that the data may be used at a later stage for service evaluation and research. However, it should be noted that clear communication with the public is very important in order to avoid a break-down in trust and harm to future data collection activities, the likes of which has been observed in the UK in the roll out of care.data and in the Netherlands in the implementation of their national data exchange point: the Dutch National Healthcare Information Hub. For patients who lack decision-making capacity such as the severely disabled, the elderly or young children, consent should be provided by a legal representative on their behalf. A white paper by Goldstein and Rein (2010) presents a useful outline of consent options for electronic health information, presenting the issues around and implications of the five models: No consent, Opt-out, Opt-out with exceptions, Opt-in, or Opt-in with exceptions.

Where de novo data collection is proposed, its consideration should involve careful assessment of the intended use of the data. There is often a line drawn between 'audit'/ 'service evaluation'/'quality assurance' activities on the one hand and 'research' on the other: the former of which generally does not invoke the need for ethical approval for data collection but the latter does. When ascertaining the differences between these two data collection purposes the following themes are generally referenced:

> (1) Intent. Primary research aims to achieve generalisable results, whereas audit / service evaluation measures standards of care. In other words, research is to find out what you should be doing, whereas audit is to investigate planned activity.

(2) Clinical support. In audit / service quality evaluation, treatments have a firm basis of support in the clinical community.

(3) Allocation of treatment. Audit / evaluation does not involve allocation of treatment by protocol. If randomisation is used, it is research.

Where ethical review is required, research Ethics Committees consider the societal benefit of the research and the risk to patients. Committees should be composed of multiple stakeholders, and the criteria used should be clear, transparent, and replicable. Where data collection is to be conducted on a national basis, there should ideally be one central review board. The danger of a system of local review boards is duplication of effort on the part of the applicant, inconsistent levels of or criteria for approval, and therefore poor coverage: this situation can be seen in Italy.

As indicated, where data collection is for the assessment of service quality, legal requirements for the collection of data are less rigorous due to the necessary nature of the evaluation activity in ensuring patients are treated optimally and according to specified standards. When it is deemed by a commissioner or HTA body that further RWE is required to reduce uncertainty around a treatment on its use in everyday practice, such activity could be regarded as service evaluation; this may have the implication of reducing the barriers or administrative requirements that may otherwise impede data collection, which may be critical given the time limited nature often attached to such arrangements. On the other hand, if the legal requirements are such that this would impede the way in which data can later be linked and evaluated, then care should be taken that the ethical requirements set at the outset of the data collection project are appropriate. There should be a recognition that the responsibility for managing the collection of data to reduce uncertainty around treatment effectiveness can fall to a range of stakeholders. In some countries, such as France in their CED decisions, this responsibility falls on manufacturers. Governance arrangements should accommodate this and recognise the valid and legitimate reasons that industry has in collecting this data.

Where identifiable information is required in order to facilitate research or the linkage of datasets, consent must be considered. Below we outline a consent model for consideration of the various uses of RWD.

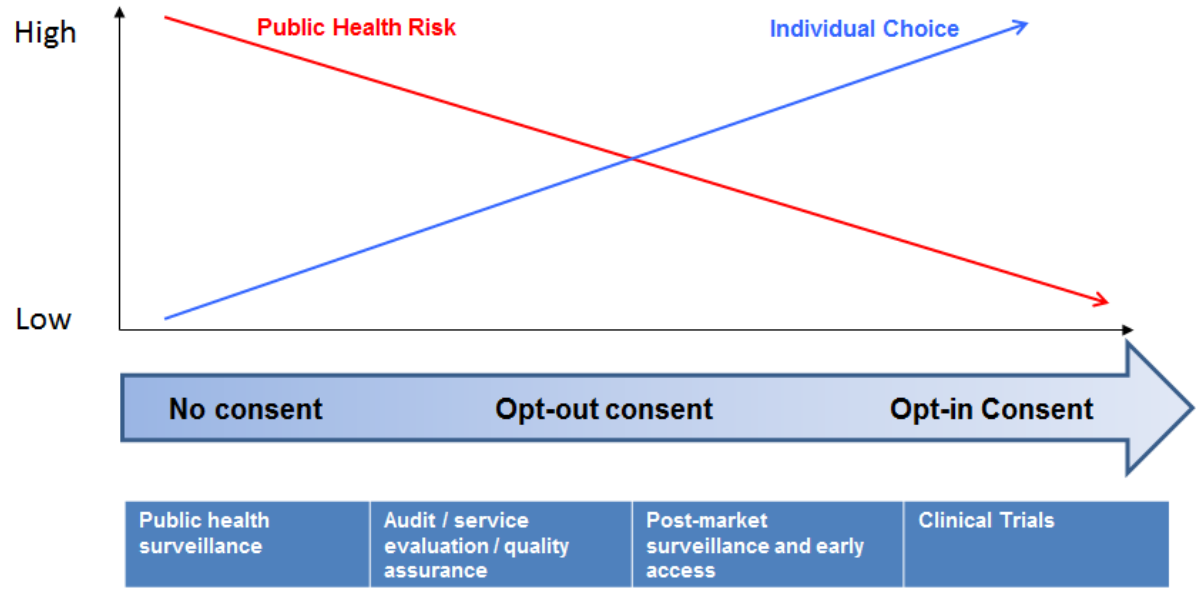


Figure 9 Consent model for individually identifiable health data

Precisely where the activities outlined in Figure 9 fit within this consent spectrum varies by country, but falls generally along the lines described. The objective of public health surveillance activities is to protect public health, the risk to which is deemed higher than the risk to privacy in the way that data is processed and managed. Legal authority is driven by legislation that mandates public authorities to access individually identifiable health data. Next, service evaluation activity (also labelled audit or quality assurance) is considered part of health care operations when used within the health care system, and as such is generally collected with no requirements for patient consent, or an 'opt-out' model. On the other end of the spectrum, clinical trials are entered into willingly by patients, for which an opt-in model of consent is completely appropriate. As discussed, activity around post-market surveillance and early access fit between these extremes.

With regard to 'ownership' of data, the most clear and positive systems of data collection and management place the patient at the heart of their data, with the ability to view and contribute to their own record. This is likely to increase public trust, though even in countries where such strategies are in place, the IT infrastructure to support it is generally behind. In legal terms, 'ownership' cannot be an absolute value / right associated with RWD, but it is important that there be a legal construct to establish patient's rights to include or exclude their data where this may be attributed to them personally, and also to support the rights and obligations of those processing the data. To consider data which has been collected and managed appropriately as a 'public' good, belonging to and serving the general population from whom it has been collected (rather than belonging to those who happen to be managing it) appears to offer a strong case for its use to advance health care through well-conducted evaluation and research. This stance is reflected by Safran and colleagues, who call for emphasis to be placed not on ownership of data but on discussions of data access and use (Safran et al., 2007).

'Heat map' for dimension "Routinely collected / De novo Raw Data"

	Routinely collected /De novo *Raw data*	UK	US	FR	IT	SW	GER	NL	AU
Government as Regulator	✓ Data protection legislation (health 'special case')								
	✓ Equitable patient selection and the protection of vulnerable subjects								
Data subjects: Patients	✓ Patient consent								
	✓ Facilitative opt-in / opt-out consent models for research								
Data Collectors	✓ Unique patient identifiers (UPIs)								
	✓ Patient information								
	✓ Data quality assurance								
	✓ Data ownership: responsibility for data?								
Data Users	✓ Approval of data collection activities to be based on intended use								
	✓ Clear and transparent criteria for de novo data projects								

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement.

Source: Based on OHE Consulting interpretation. See the country comparison section and the relevant country assessment for further details.

IDEAL FRAMEWORK for raw data:

- ***Data protection legislation***: Clear data protection requirements that recognise the legitimacy of health care data utilisation beyond the direct care of patients.
- ***Data quality assurance***. Requirements that records are accurate, and up-to-date. Patient identifiers which conform to national standards should be used and stored with the record.
- ***Patient consent***. Where patient consent is not feasible, the collection of data for purposes beyond direct care can be supported with relevant legislation. Requirements that new legislation be passed for each new dataset poses prohibitive restraints on legitimate and worthwhile data collection activities. Greater flexibility can be administered through a legislative framework that grants statutory exemption for the requirement of consent where this would be too burdensome and where the purpose of the exemption is in the interest of the public. This should be decided after careful assessment by an ethical review board. This kind of regulation can be government-sanctioned but privately administered by a government entity.
- Where data collection is to be collected on a routine basis across a large patient cohort, an ***opt-out, rather than opt-in, system of patient consent*** may serve to maximise coverage and allow patients to contribute data more easily.
- ***Patient information:*** There must be clear communication to data subjects of potential future uses of their data. Not explaining simply and clearly the rights of

patients to opt-out or 'object' to their data being collected and later used for purposes not aligned with their own care can damage public trust (HSCIC, 2015)[9].

- ***Approval of data collection activities to be based on intended use***. This relates to de-novo data collection. The requirements for new data collection activities should be cognizant of the future intended use of the data. For example data collection activities that often form part of MEAs or risk-sharing arrangements between payers and manufacturers should be recognised as essential to the appropriate and optimal treatment of patients. Clear and transparent roles for the various actors in the collecting and eventual sharing of data should be well set out, which will enable access to data without harm or impact on privacy and public interest positions.
- ***Clear and transparent criteria***. The criteria of Ethics Committees for data collection projects ('de novo' data) should be clear, transparent, and replicable. For national projects, there should ideally be a central ethical review board whose decision is accepted by the relevant national and local parties; this would reduce duplication of effort and promote consistent coverage.
- ***Data ownership***. Responsibility (to be distinguished from 'ownership') for the data after collection passes to the data controller, who must act in the interest of patients and the public as specified by law.

Cleaning and managing data

Data controllers are the organisations responsible for collecting, managing, and linking patient data. In order for the public to have trust in a system that collects and manages patient data, that system and those organisations that work within it must demonstrate strong and robust processes and meet quality criteria that give the public and data users confidence in the quality and security of the data held.

'Heat map' for dimension "Cleaning and managing the data"

	Cleaning and managing data	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓Data management: Recognised data stewardship entities								
Data Controllers	✓Process for de-identification								
	✓Security arrangements: 'Privacy Enhancing Techniques and Procedures' (PETs)								
	✓Training of staff								
	✓ Specified arrangements for how long data are kept								

[9] As an example, there have been recent media reactions to the patient 'objection' process confusion for health data in the UK, whereby a "flaw" in the wording provided by HSCIC around the type two objections (objections to data flowing from the HSCIC) unintendedly would prevent data flows for direct care purposes such as cancer screening and electronic prescriptions, and therefore have not been actioned.

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement.

Source: Based on OHE Consulting interpretation. See the country comparison section and the relevant country assessment for further details.

IDEAL FRAMEWORK for cleaning and managing data:

- ***Recognised data stewardship entities***. Data stewardship entities that manage the acquisition, storage, aggregation, and de-identification of data. The interests of those entities must be aligned with those individuals whose data is being collected. These come under various names, for example 'Trusted Third Parties'. These organisations must comply with the relevant legislation for the countries in which they operate.
- ***De-identification of data***. Where appropriate, data can be de-identified by removing any personally identifiable information and replacing the unique patient identifier (which in some countries is used across different sectors of the economy and therefore highly sensitive) with a pseudonym. Where data is not managed by one single entity, care should be taken that the algorithm for the pseudonymisation process is replicable for other datasets so that they may be linked, or else that the pseudonymisation process be reversible when desirable.
- ***Data quality***. In the same way that individuals and organisations collecting data from patients have a responsibility to ensure that the data are relevant, up-to-date, and accurate, so should those organisations processing patient data ensure that the quality and integrity of the data is maintained.
- ***Security arrangements***. Security arrangements for the protection of confidential patient data should be assured through sound security processes, ranging from physical and technical computing protections and to the legal, security, and confidentiality training of staff involved in processing the data. Such processes and techniques are often called 'Privacy Enhancing Techniques and Procedures' (PETs), which should be implemented for the anonymisation of data as well as in preventing loss of anonymity at a later date.
- ***How long data are kept***. In many countries, it is specified through data protection legislation that data should be kept 'no longer than necessary'. This is difficult to define, but the importance of rich longitudinal data that follows a patient over time through the care pathway and its benefits for research should be considered.

Appropriate data controls and processes build the foundation for appropriate and useful access to health care data.

Linkage and aggregation

The ability to link data across datasets is incredibly important for research. This can be facilitated by the use in most countries of a unique patient identifier, which may either have been created specifically for health care, or be an identifier used more broadly for other services such as social security numbers or national person numbers. When being prepared for sharing, data can be made completely anonymous by sharing only aggregated data. Even pseudonymous data has the potential for re-identification of

patient identity under certain situations, which must be considered at the later data sharing stage. The ability for central linkage of datasets may be impeded in countries where there are multiple data custodians each managing distinct datasets.

'Heat map' for dimension "Linkage and aggregation"

	Linkage and aggregation	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓ Privacy rules								
	✓ Develop a clear set of nationally agreed and implemented standard rules to optimise interoperability of health record systems								
Data Controllers	✓Unique patient identifiers								
	✓Pseudonymisation								
	✓ Preparation for sharing								

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement. Note: For Italy, we have used both amber and red for one entry due to the fragmented nature of the system. For Germany we have used both amber and red for one entry as process of pseudonymisatoin is irreversible. For the Netherlands we have used both amber and red for one entry as there is no unique patient identifier; however, there is now a 'citizen service number', but data linkage has been rare.

Source: Based on OHE Consulting interpretation. See the country comparison section and the relevant country assessment for further details.

<u>IDEAL FRAMEWORK for linkage and aggregation:</u>

- ***Develop a clear set of nationally agreed and implemented standard rules to optimize interoperability of health record systems***. This is key for datasets to be compatible with one another.
- ***Data linkage by trusted third party***. Common organisational and technical barriers to data linkage arise when there is no single group or organisation that has the responsibility or technical expertise required to manage the linking process. This could be minimised if linkage is undertaken by a single trusted third party. Where pseudonym IDs are created to facilitate the sharing of data with reduced risks whilst still allowing for linkage of datasets, the pseudonym IDs are common to the linked datasets and indicate that the records belong to the same person while protecting anonymity. This is more feasible in systems where management of this process is centralised. Whilst pseudonymisation helps to reduce the potential identifiability of data, there will always remain some residual risk of jig-saw re-identification. Therefore, it is still appropriate for requests for non-aggregated data to be examined

by information governance panels (often through ethics committees) which consider the balance between the risk to patient confidentiality and the public interest in the research. This process is considered below.

Access / use of data

Permissions and processes for access to health data vary substantially among countries. Provisions for data access will vary according to the manner in which that was collected; therefore, the legislative considerations discussed under the governance arrangements for raw data are paramount.

'Heat map' for dimension "access / use of data"

	Access / use of data	UK	US	FR	IT	SW	GER	NL	AU
Government as Regulator	✓ Managing re-identification risk								
	✓ Criteria for different uses (& different users)								
Data Controllers	✓Approval panels								
	✓ Confidentiality and data use agreements								
	✓ Balancing benefits of linkage for research with risk for re-identification								
Data users	✓Audit / Service evaluation and quality monitoring								
	✓Degree of access, level of data, and mode of access								
	✓Cost of access								
	✓ Appropriate experience/qualifications, and funding to conduct research								

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement. Note: For Italy, we have used both amber and red for some entries due to the fragmented nature of the system. For Sweden we have used both green and amber for one entry as it seems access to data can be problematic. For the Netherlands we have both amber and red for one entry due to legal barriers to use electronic health records to monitor national health care quality over the next years.

Source: Based on OHE Consulting interpretation. See the country comparison section and the relevant country assessment for further details.

IDEAL FRAMEWORK for access / use:

***Forms of data access*.** Different access arrangements may be employed to achieve the needed balance between protection of private information and informing real-world research:

- An often used model involves the potential data user applying for access and following privacy review and contracting, from the data provider. In this scheme, the data provider may offer information at varying levels of detail and scrutiny:
    - Data may be provided at the aggregate level in which there is no information about individual patients. Data at this level may be provided freely since the risks are low.
    - Where the data provider has capacity for such services, analyses may be conducted in-house by the data provider, the results of which are then shared with the applicant. Similarly, this would involve minimal risk to privacy.
    - Data may be provided at the level of individual patients but with most or all individually identifying elements removed (e.g. social security numbers). This level should require a routine data use agreement form in which the data user agrees to protect the privacy of individuals in the dataset and not attempt to discover their identities.
    - Data may be provided at the level of individual patients with most or all individually identifying elements intact. Clearly, this level of information carries greater risk to privacy. However, this may be justifiable in some cases when investigators specifically need the patient identifiers to link the dataset to other data sources for research. This level of data should require the highest level of scrutiny, including a data use agreement, justification that the benefits of research outweigh the risks, review by a privacy board, and perhaps ongoing scrutiny for the duration that the data user possesses the data.
    - Data at the individual patient level could alternatively be provided to researchers in a physical space, which allows for direct control and monitoring of data use in cases where those data are highly sensitive.
- Another model which is able to allow access to individual patient data, data linkage across data providers, while protecting individual privacy, is the distributed network model. This could help to overcome the difficulties that can arise when there are multiple data custodians.
    - In this model, a consortium of data providers mutually agrees to share data and work to develop a common data framework. Data is coded uniformly across the consortium (e.g. date of birth would be coded: "MM-DD-YYYY"). Each data provider stores their own data behind a firewall protected server. Data users may write standardized code which is sent to each data provider, analysed on site, within each data provider's server (protecting patient privacy) and the aggregate results are sent back to the data user.

- ***Approval panels / ethical review.*** Ethical review boards (also called institutional review boards) which grant access to health care data must be assured that the interest to society of the research project significantly outweighs the risk of violation of personal integrity of the individual that the processing may involve. A 'consent or anonymise' approach is too polarised and not a proportionate system. This risk of re-identification can be minimised with

requirements for security procedures, training of staff that will process the data, and carefully written confidentiality agreements which assure correct use and reporting of data and which carry with it sanctions for inappropriate use. Approval panels should be composed of representatives with a broad range of relevant expertise and standpoints. The criteria used by committees to grant access to data should be clear, consistent, and transparent.

- The onus should be on data custodians to communicate how information is being shared and with whom in order to ensure public trust and transparency.
- ***Data use agreements and confidentiality requirements****.* Permission for data access should be granted with contractual requirements around the protection of confidentiality. The agreement should clearly define the scope and define duration of use.
- ***Affiliation of the data user****.* The type of organisation requesting access to data may influence the potential risk associated with its distribution (both realised and perceived). However, whilst the organisation's remit may influence their motivation for requesting access, this should not be the only consideration by data providers. Where the appropriate safeguards are in place, authorisation should be based on careful consideration of the motivation for and outputs of the research facilitated, rather than on the basis of the organisation's status. This is particularly important where manufacturers are tasked by HTA agencies or regulators with assessing the evidence for their products in routine practice.
- ***Access costs****.* Arrangements for the cost of data access will vary according to the nature of the data controller. For many datasets collected and held on a national basis, data charges are based only to recover the costs of data extraction and cleaning. Cost of access should be fair and not excessive, but in recognition of the need for the sustainability of the system.

The following table combines all the individual heat maps above into one table.

'Heat map' for all dimensions

	Routinely collected /De novo Raw data	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓ Data protection legislation (health 'special case')								
	✓ Equitable patient selection and the protection of vulnerable subjects								
Data subjects: Patients	✓ Patient consent								
	✓ Facilitative opt-in / opt-out consent models for research								
Data Collectors	✓ Unique patient identifiers (UPIs)								
	✓ Patient information								
	✓ Data quality assurance								
	✓ Data ownership: responsibility for data?								
Data Users	✓ Approval of data collection activities to be based on intended use								
	✓ Clear and transparent criteria for de novo data projects								
	Cleaning and managing data	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓Data management: Recognised data stewardship entities								
Data Controllers	✓Process for de-identification								
	✓Security arrangements: 'Privacy Enhancing Techniques and Procedures' (PETs)								
	✓Training of staff								
	✓ Specified arrangements for how long data are kept								
	Linkage and aggregation	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓ Privacy rules								
	✓ Develop a clear set of nationally agreed and implemented standard rules to optimise interoperability of health record systems								
Data Controllers	✓Unique patient identifiers								
	✓Pseudonymisation								
	✓ Preparation for sharing								
	Access / use of data	**UK**	**US**	**FR**	**IT**	**SW**	**GER**	**NL**	**AU**
Government as Regulator	✓ Managing re-identification risk								
	✓ Criteria for different uses (& different users)								
Data Controllers	✓Approval panels								
	✓ Confidentiality and data use agreements								
	✓ Balancing benefits of linkage for research with risk for re-identification								
Data users	✓Audit / Service evaluation and quality monitoring								
	✓Degree of access, level of data, and mode of access								
	✓Cost of access								
	✓ Appropriate experience/qualifications, and funding to conduct research								

Colour Key: green = aligned with recommended; amber = ok but with room for improvement; red = very problematic/ barrier. Squares are blank where it was felt that there was insufficient information to make a judgement. Source: Based on OHE Consulting interpretation.

13. Conclusion

The evidence that is used to support decision-making in health care is becoming increasingly diverse, reflecting the increased complexity of the regulatory and reimbursement processes. Increasingly, the importance of understanding the impact of health care interventions in real-world settings is being recognised. In this report, we described the process by which RWD (the raw data) is transformed into RWE (the insight), and assessed the rules and roles for information governance along this process By examining and comparing the data governance models in place within the health care sectors of various countries, we set out and proposed a framework for good governance.

Appropriate and facilitative governance arrangements for RWE are imperative to facilitate evidence collection to meet the demands of regulators and HTA bodies, and to make the most of health care information and the role it can play in improving patient care. Problems arise due to the fact RWD is being used for purposes beyond those for which it was originally collected – to directly manage the care of the patient. As a result, legal frameworks are playing catch-up in order to accommodate these new secondary uses of data which clearly benefit patients and society but in a different way. With the general progressive move toward evidence-confirmatory pathways for the regulation and HTA of medical products, legislation that permits the utilisation of RWD for activities such as monitoring care quality and research to generate RWE, is becoming ever more important. This is evident through the increased reliance on and appetite for managed entry agreements, whose primary goals are usually one or more of: matching performance with payment, managing use, or to generate RWE. Research scientists and others, such as the companies tasked with providing the data as part of these arrangements, should be given every opportunity to support these goals. We have provided recommendations for an ideal governance framework that could lead to a more facilitative environment for the transformation of RWD into RWE.

APPENDIX

1. Pro-formas for data extraction

US & UK 'deep dives': pro-forma

1. **Brief overview** of the health system and collection / management of patient data

2. **Routinely collected patient data**
    a. **Core legislation governing the collection / use of routinely collected patient data.** Review and summarise key documentation outlining principles of governance and data protection.
    b. **Datasets.** Overview of what data is collected, and from what parts of the health service
    c. **Information providers.** Who 'holds' the data (likely to be a mix of public and private organisations), what data do they hold, how is it collected, and what are the core governing principles in handling the data?
    d. **Data linking.** To what extent can / are patient data linked across databases – how and by whom? What are the major organisations involved?
    e. **Data access.** To what extent is data shared, with whom is it shared, how does permitted access differ according to organisation (i.e. access by pharmaceutical companies versus access by public bodies / academic institutions), what are the processes involved in being granted permission to access data, what are the costs involved in data access (where available), and in what form is data access granted (e.g. raw data / in-house data analysis services only?)

3. **Collecting de novo patient data**
    a. **Governance arrangements for research to collect new data.** Review and summarise key documentation outlining research ethics and governance for the collection of new patient data (i.e. setting up registries, pragmatic clinical trials, etc.).
    b. **Research application process.** Process by which application for new data collection is considered, and governing principles of the committees that grant approval.

4. **Data use.** What are the rules governing the use of RWE?

5. **Suggested principles or guidance for data governance, and the adapting environment for such.** Summarise any key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

Remaining countries: pro-forma

1. **Brief overview** of the health system and collection / management of patient data

2. **Core legislation and governance arrangements for the collection and/or use of patient data**
   a. **Routinely collected patient data.**
   Core legislation governing the collection / use of routinely collected patient data. Review and summarise key documentation outlining principles of governance and data protection.
   b. **Collecting de novo patient data.**
   Governance arrangements for research to collect new data. Review and summarise key documentation outlining research ethics and governance for the collection of new patient data and governing principles of the committees that grant approval.

3. **Data linking.** To what extent can patient data be linked across datasets? What are the organisations involved, and what are the core governing principles under which they operate?

4. **Data access.** To what extent is data shared, with whom, and what are the principle governance issues in the preparation / sharing of this data?

5. **Data use.** What, if any, are the rules governing the use of RWD? [To cover contract arrangements between data suppliers and recipients, rules around use for HTA, etc.]

6. **Governance ideals and changes to the environment.** Summarise any key national documentation that contains advice or commentary on ideal governance frameworks, as well as information on any imminent changes to the governance environment.

2. CPRD Access License Template: Details of permitted and restricted use

Permitted Use

3.1 The licence granted in clause 2.1 is subject to:
(A) payment of the licence fee in accordance with clause 4;
(B) the use of Data or any other information by the Licensee or its Affiliated Companies (if any) in accordance with this Licence being restricted to Medical and Health Research Purposes on a non-profit making basis. For the avoidance of doubt this shall not prevent the Licensee from:
(1) recovering its reasonable direct operating costs associated with that research; or
(2) recovering a profit from any application of the results of the Licensee's research provided that such profit is solely attributable to the value added by the Licensee in its analysis or interpretation of the Data.
(C) the Nominated Users complying with the requirement to undergo training in accordance with clause 6;
(D) the Licensee complying with the restrictions set out in clause 9.
3.2 These restrictions on use shall survive the termination or expiry of the Licence.
3.3 In the event the Licensee has any doubts as to the scope of the licence granted in clause 2.1, it shall contact the Licensor for clarification.

Restrictions on Use of Data

9.1 The Licensee shall not use or attempt to use the Data, or any information obtained by the Licensee in accordance with the provision of the Services, whether on its own or in conjunction with any other data in any other form, for:

(A) identifying, contacting or targeting patients;
(B) identifying, profiling, contacting or targeting general medical practitioners or general medical practices; or
(C) studying the effectiveness of advertising campaigns or sales forces,

and the Licensee shall ensure that reports, papers or statistical tables that are published or released to third parties as a result of use of the Data cannot be used to identify or enable others to identify patients, contributing general medical practitioners or contributing general medical practices. If at any time the Licensee considers that there is information in the Database accessible via the Services which could be used to identify any individual, general medical practitioner or general medical practice, the Licensee will inform the Licensor immediately in writing by way of a notice delivered in accordance with clause 23.

9.2 The Licensee shall not sell, transfer (except as permitted under this Licence to Affiliated Companies), trade or otherwise dispose of any Data downloaded from the Database by the Licensee save that, with the written permission of the Licensor, Data may be supplied to regulatory authorities by the Licensee for audit purposes. For the avoidance of doubt this does not preclude inclusion of Data in papers published by the Licensee or its Affiliates in medical or scientific journals or in presentations of a medical or scientific nature provided that the Data so included are limited to no more than are strictly necessary to support the relevant paper or presentation.

9.3 Subject to clauses 7.5, 9.4 and 9.5 the Licensee shall not permit any third party to access, study, analyse, refer to or otherwise use the Data (with the exception of Affiliated

Companies), or permit any third party to reproduce any Data downloaded from the Database by the Licensee.

9.4 The Licensee shall not permit any contractor access, study, analyse, refer to or otherwise use the Data, save with the written permission of the Licensor. Any contractor granted such permission shall, before being given access to the Data sign and return to the Licensor a confidentiality agreement in a form to be supplied by the Licensor. The Licensee shall ensure that any Data transferred to a contractor under this clause is returned to the Licensee at the expiry of the contract between the Licensee and the contractor.

9.5 The Licensee will not use the Data for projects where the results may be communicated to third parties without first obtaining approval from the Licensor of a protocol describing the project, unless the Licensor has informed the Licensee in writing that the submission of such a protocol is unnecessary. The Licensor will, if appropriate, pass any protocol submitted to it by the Licensee to the Independent Scientific Advisory Committee for advice and the Licensor will then revert to the Licensee to confirm whether the protocol has been accepted. The procedure for submission of protocols will be available on the Website or, if unavailable on the Website, will be made available to the Licensee on request. The decision of the Licensor under this clause shall not be interpreted as the views of the UK Licensing Authority acting via the Licensor.
9.6 The Licensee shall be entitled to send up to 300 case histories (or such greater number as the Licensor may in writing agree) to external experts approved by the Licensor for review provided that:

(A) The case histories and number of such case histories shall be strictly limited to those required for the purposes of the relevant project;
(B) The experts shall be made aware of the confidential nature of the information provided;
(C) The maximum number of experts conducting such reviews per project shall be ten (10) unless otherwise agreed by the Licensor;
(D) The experts shall not be permitted to make any copies of the case histories, other than those strictly necessary for the purposes of their review; and
(E) Upon completion of their review the Licensor shall procure that the case histories sent to each expert and any copies are either returned to the Licensor or destroyed, and the Licensee shall notify the Licensor once this has been done.

3. Good Practice in Secondary Data Analysis (Germany)

Good Practice in Secondary Data Analysis (GPS) (AGNES et al., 2008)

Guideline 8: Data protection
The data protection provisions in force for protecting informational self-determination should be observed when planning and conducting secondary data analyses.

The data protection provisions in force, including the principle of data avoidance and data scarcity, which requires collecting and storing only those data that are absolutely necessary, (§ 3a of the German federal data protection law [BDSG] refers) and, if applicable, other regulations relevant to the data bodies used must be observed. All persons who deal with personal data in connection with a research project must be informed of the content, scope and capacity of the relevant legal provisions. In research with personal data, both the individual's right to informational self-determination as well as the right to freedom in science and research must be observed.
Recommendation 8.1 – Purpose of data provision
The purpose of data provision (in terms of data protection) is to answer the research questions (see Guideline 2) and must be set down in writing.

Recommendation 8.2 – Pseudo-anonymization and anonymization
Use should be made of the means of pseudo-anonymization and anonymization contained in the German federal law on data protection (§ 3a BDSG data avoidance and data scarcity). The involvement of a data custodian should be considered here.

Recommendation 8.3 – Depseudo-anonymization and re-identification
It is important to stipulate in writing in the general contractual conditions whether depseudo-anonymization is intended, and if so, in which cases. In the analysis, appropriate means (technical and contractual) should be employed to prevent unreliable re-identification

Recommendation 8.4 – Transfer of personal data to third parties
As a rule, any transfer of personal data is done by the data owner only.

Recommendation 8.5 – Personal data linkage with external data sources
All personal data linkages with external data sources that are not explicitly provided for require compliance with data protection provisions.

Recommendation 8.6 – Persons responsible for data protection
In every secondary analysis, national and international standards of data security and data protection should be observed. Within a research division, a person should be appointed as the person responsible for data processing, who monitors compliance with these standards. The person in question must have appropriate qualifications for these duties.

Recommendation 8.7 – Deletion deadlines
If, for reasons of data protection, the data provided for secondary data analysis has to be deleted or anonymized after the purpose of the study has been achieved, this must be done in accordance with the retention requirements for baseline and analysis data sets specified in Recommendations 6.2 and 6.7.
Similarly, when setting deletion deadlines, an opportunity to check the results obtained from secondary utilization as specified under Guideline 7 must also be provided.

Recommendation 8.8 – Co-operation with persons responsible for data protection
The need to make contact with the legitimate persons responsible for data protection should be borne in mind as early as planning the secondary data analysis.

REFERENCES

Abrahamsson, K, 2015a. National Patient Registry. Joint Programming Initiative - Data mapping project. Available at: http://www.jpi-dataproject.eu/Home/Database/361?topicId=1 [Accessed 27 January 2015a].

Abrahamsson, K, 2015b. Swedish Pharmaceuticals Registry. Joint Programming Initiative - Data mapping project. Available at: http://www.jpi-dataproject.eu/Home/Database/388?topicId=1 [Accessed 22 January 2015b].

ACSQHC, 2014. *Framework for Australian clinical quality registries*. Sydney: 978-1-921983-71-9.

ADELF, 2007. Recommandations de déontologie et bonnes pratiques en épidémiologie (version France 2007). Association des épidémiologists de langue française. Available at: http://adelf.isped.u-bordeaux2.fr/LinkClick.aspx?fileticket=BHPxc8buLII%3d&tabid=534 [Accessed 10 February 2015].

AGNES, DGSMP, DGEpi, and GMDS, 2008. *GPS - Good Practice in Secondary Data Analysis: Revision after Fundamental Reworking*. Working Group or the Survey and Utilization of Secondary Data (AGNES) and Working Group for Epidemiological Methods.

AHRQ, 2014. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality. Available at: http://www.ahrq.gov/research/data/hcup/index.html [Accessed 23 November 2014].

AIFA, 2007. Linee guida per la classificazione e conduzione degli studi osservazionali sui farmaci. Agenxia Italiana del Farmaco. Available at: http://www.agenziafarmaco.gov.it/it/content/linee-guida-studi-osservazionali [Accessed 11 February 2015].

AIFA, 2015a. Registers Drugs monitored. Agenxia Italiana del Farmaco. Available at: http://www.agenziafarmaco.gov.it/it/content/registri-farmaci-sottoposti-monitoraggio [Accessed 18 February 2015a].

AIFA, 2015b. The Italian Medicines Agency. Agenxia Italiana del Farmaco. Available at: http://www.agenziafarmaco.com/en/node/4111 [Accessed 18 February 2051b].

AIHW, 2015. Australia's Health System. Australian Institute of Health and Welfare. Available at: http://www.aihw.gov.au/australias-health/2014/health-system/ [Accessed 2 February 2015].

AIRTUM, 2015. Welcome to AIRTUM. Associazone Italiana Registri Tumori. Available at: http://www.registri-tumori.it/cms/en/english-home [Accessed 18 February 2015].

AMAI, 2014. AMAI. Informatics Professionals. Leading the way. Americal Medical Informatics Associaion. Available at: http://www.amia.org/ [Accessed 7 December 2014].

Andersen, M.R. and Storm, H.H., 2013. Cancer registration, public health and the reform of the European data protection framework: Abandoning or improving European public health research? *European Journal of Cancer*

ANSM, 2015. Médicaments et produits biologiques. Agence nationale de sécurité du médicament et des produits de santé. Available at: http://ansm.sante.fr/Activites/Medicaments-et-produits-biologiques/Reglementation-francaise/%28offset%29/3 [Accessed 10 February 2015].

ARAD, 2011. ARAD Governance Document. Australian Rheumatology Association Database. Available at: https://arad.org.au/Documents/ARADgovernanceMar30_2011.pdf [Accessed 29 January 2015].

ARAD, 2015a. ARAD Access Policy. Australian Rheumatology Association Database. Available at: https://arad.org.au/Documents/ARA_Access_policy_May_2012.pdf [Accessed 29 January 2015a].

ARAD, 2015b. ARAD external researcher confidentiality agreement. Australian Rheumatology Association Database. Available at: https://arad.org.au/Documents/ARAD_CA_for_external_researchers2010.pdf [Accessed 29 January 2015b].

Armstrong, D., Kline-Rogers, E., Jani, S.M., Goldman, E.B., Fang, J., Mukherjee, D., Nallamothu, B.K., and Eagle, K.A., 2005. Potential impact of the HIPAA privacy rule on data collection in a registry of patients with acute coronary syndrome. *Archives of Internal Medicine*, 165 (10) pp. 1125-1129

Artmann, J, Giest, S, and Dumortier, J, 2015. eHealth Strategies. Coutry Brief: France. European Commission, DG Information Society and Media, ICT for Health Unit. Available at: http://ehealth-strategies.eu/database/documents/france_countrybrief_ehstrategies.pdf [Accessed 10 February 2015].

Ashjari, R and Strom, H, 2-11-2005. Registration of pharmaceutical prescriptions. Bird & Bird. Available at: http://www.twobirds.com/en/news/articles/2005/registration-of-pharmaceutical-prescriptions [Accessed 23 January 2015].

ATIH, 2015. Accès aux applications de l'Agence technique de l'information sur l'hospitalisation. Agence Technique de l'Information sur l'Hospitalisation. Available at: https://pasrel.atih.sante.fr/cas/login?service=http%3A%2F%2Fstats.atih.sante.fr%2Fsihnat%2Fconnexion_sihnat%2Findex.php [Accessed 10 February 2015].

Australian Government, 2014. Privacy fact sheet 17: Australian Privacy Principles. Australian Government Office of the Australian Information Commissioner. Available at: http://www.oaic.gov.au/privacy/privacy-resources/privacy-fact-sheets/other/privacy-fact-sheet-17-australian-privacy-principles [Accessed 28 January 2015].

Australian Government, 2015. Health information and the Privacy Act. Office of the Australian Information Commissioner. Available at: http://www.oaic.gov.au/privacy/privacy-act/health-and-medical-research [Accessed 28 January 2015].

Baccetti, E, 2007. Proposed new regulation on observational studies of medicinal products. Bird & Bird. Available at: http://www.twobirds.com/en/news/articles/2007/italy-proposed-regulation-observational-studies-medicinal-products [Accessed 11 February 2015].

Barron, A.J., Klinger, C., Shah, S.M.B., and Wright, J.S., 2014. A regulatory governance perspective on health technology assessment (HTA) in France: The contextual mediation of common functional pressures. *Health Policy*

Bevan, G., Karanikolos, J., Nolte, E., Connolly, S., and Mays, N., 2014. *The four health systems of the United Kingdom: how do they compare?* THe Health Foundation and Nuffield Trust. 978-1-905030-78-1.

BMJV, 2009. Federal Data Protection Act in the version promulgated on 14 January 2003 (Federal Law Gazette I p. 66), as most recently amended by Article 1 of the Act of 14 August 2009 (Federal Law Gazette I p. 2814). Bundesministerium der Justiz und für Verbraucherschutz [Federal Ministry of Justice and Consumer Protection]. Available at: http://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.html [Accessed 26 February 2015].

Bras, P. and Loth, A., 2013. *Rapport sur la gouvernance et l'utilisation des données de santé*. Ministere des Affaires Sociales et de la Santé.

CCMO, 2015. Centrale Commissie Mensgebonden Onderzoek. Available at: http://www.ccmo.nl/ [Accessed 18 February 2015].

CCTIRS, 2015. Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé. Organismes Indépendants et Organismes sous Tutelle.Enseignment supérieur et recherche. Available at: http://www.enseignementsup-recherche.gouv.fr/cid20537/cctirs.html. [Accessed 10 February 2015].

Clark, S. and Weale, A., 2011. *Information governance in health. An analysis of the social values involved in data linkage studies*. London: Nuffield Trust.

CNIL, 1978. Loi Informatique et Libertes. Act N.78-17 of 6 January 1978 on Information Technology, Data files and Civil Liberties. Commission Nationale de l'Informatique et des Libertés. Available at: http://www.cnil.fr/fileadmin/documents/en/Act78-17VA.pdf [Accessed 6 February 2015].

CNIL, 2007. *CNIL conclusions du 20 Février 2007 sur l'utilisation du NIR comme identifiant de santé*.

CODEX, 2014. Handling personal information. CODEX rules & guidelines for research. Available at: http://www.codex.vr.se/en/manniska3.shtml [Accessed 22 January 2015].

CODEX, 2015. Research with registers. CODEX rules & guidelines for research. Available at: http://www.codex.vr.se/en/manniska6.shtml [Accessed 26 January 2015].

Corrao, G., 2014. Towards the rational use of Healthcare Utilization Databases for generating real-world evidence: new challenges and proposals. *A position paper from the Italian Society of Medical Statistics and Clinical Epidemiology Working group on Observational Studies*. *Epidemiology Biostatistics and Public Health*, 11 (3) pp. e10328-1-e10328-6

Coudert, F., 2008. *Study on Legal Framework of Interoperable eHealth in Europe: National profile France*. Brussels: European Commission Directorate General Information Society. SMART 2007/0059

CPRD, 2014a. CPRD Access Licence Template. The Secretary of State for Health, acting through the Clinical Practice Research Datalink Division within the MHRA. Available at: http://www.cprd.com/_docs/CPRD%20Access%20Licence%20Template.pdf [Accessed 24 November 2014a].

CPRD, 2014b. CPRD Governance. Clinical Practice Research Datalink. Available at: http://www.cprd.com/governance/ [Accessed 18 November 2014b].

CPRD, 2014c. Governance relating to the use of CPRD for research. Clinical Practice Research Datalink. Available at: http://www.cprd.com/ISAC/governance.asp [Accessed 20 November 2014c].

CPRD, 2014d. Observational Data. Clinical Practice Research Datalink. Available at: http://www.cprd.com/ObservationalData/CodedData.asp [Accessed 18 November 2014d].

De Lusignan, S. and Seroussi, B., 2013. A comparison of English and French approaches to providing patients access to Summary Care Records: scope, consent, cost. *Stud Health Technol Inform*, 186 pp. 61-62

Department fo Health, 2014. Personally Controlled Electronic Health Record System Operator: Annual Report 2012-2013. Australian Government Department of Health. Available at: http://www.health.gov.au/internet/main/publishing.nsf/Content/PCEHR-system-operator-annual-report2012-2013 [Accessed 29 January 2015].

Department of Health, 1997. *The Caldicott Committee. Report on the Review of Patient-Identifiable Information*.

Department of Health, 2005. *Research Governance Framework for Health and Social Care. Second edition*. Crown Copyright 2005.

Department of Health, 2010. Memorandum of Understanding with Medicines Australia. Australian Government Department of Health.Pharmaceutical Benefits Scheme. Available at: http://www.pbs.gov.au/info/industry/useful-resources/memorandum [Accessed 18 January 2013].

Department of Health, 2014. Step 7: Entering agreements to share risk. Australian Government Department of Health.The Pharmaceutical Benefits Scheme. Available at: http://www.pbs.gov.au/info/industry/listing/listing-steps/g-entering_agreements [Accessed 12 February 2015].

Destasis, 2015. Health. Destasis.Statistisches Bundesamt. Available at: https://www.destatis.de/EN/FactsFigures/SocietyState/Health/Health.html [Accessed 25 February 2015].

Destefano, F. and Vellozzi, C., 2012. *Lessons learned exercise: ECDC Vaccine Adverse Event Surveillance and Communication (VAESCO II) project 2009-2011*. Facilitator's Report.

DGEpi, 2004. *Good Epidemiologic Practice (GEP)*. German Society for Epidemiology in collaboration with GMDS, DGSMP and DR-IBS.

DH, NISCHR, CSO, and PHA, 2011. *Governance arrangements for research ethics committees: A harmonised edition*. Leeds: 15917.

DHHS, 2014. Benefit of EHRs. Department of Health and Human Services. Available at: http://www.healthit.gov/providers-professionals/electronic-medical-records-emr [Accessed 18 November 2014].

Di Iorio, C.T., Carinci, F., and Oderkirk, J., 2014. Health research and systems–governance are at risk: should the right to data protection override health? *Journal of medical ethics*, 40 (7) pp. 488-492

Doupi, P, Renko, E, Giest, S, and Dumortier, J, 2010a. eHealth Strategies. Country Brief: Sweden. European Commission, DG Information Society and Media, ICT for Health Unit.

Available at: http://ehealth-strategies.eu/database/documents/sweden_countrybrief_ehstrategies.pdf [Accessed 26 January 2015a].

Doupi, P., REnko, R., Giest, S., and Dumortier, J., 2010b. *eHealth Strategies. Country Brief: Sweden*. European Commission, DG Information Society and Media, ICT for Health Unit.

eDRIS, 2013. *Technical Consultation Paper on the Design of the Data Sharing and Linking Service*.

eDRIS, 2014. Use of the National Safe Haven. ISD Scotland. Available at: http://www.isdscotland.org/Products-and-Services/eDRIS/Use-of-the-National-Safe-Haven/ [Accessed 20 November 2014].

Eftimovska, E, 2014. Appendix 1.1: The Swedish e-Health Landscape Surrounding the SRQ Registry. Karolinska Institutet. Available at: http://srq.nu/srqny/wp-content/uploads/2014/10/Appendix-1.1-Swedish-e-health-Landscape-4-30-14-FINAL-1.pdf [Accessed 21 January 2015].

Elm, E.v., Altman, D.G., Egger, M., Pocock, S.J., G+©tzsche, P.C., and Vandenbroucke, J.P., 2007. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*, 335 (7624) pp. 806-808

EMIF, 2015. Welcome to the EMIF website. European Medical Infromation Framework. Available at: http://www.emif.eu/ [Accessed 1 July 2015].

Emilsson, L., Lindahl, B., Koster, M., Lambe, M., and Ludvigsson, J., 2015. Review of 103 Swedish Healthcare Quality Registries. *Journal of Internal Medicine*, 277 (1) pp. 94-136

EPN, 2015. Vetting the ethics of research involving humans. Etikproving av forskning som avser manniskor. Available at: http://www.epn.se/en/start/startpage/ [Accessed 26 January 2015].

epSOS, 2015. epSOS as a Pilot is over. European Patients Smart Open Services. Available at: http://www.epsos.eu/home.html [Accessed 12 June 2015].

EU-ADR, 2015. Welcome to the EU-ADR website. EU-ADR Alliance. Available at: http://euadr-project.org/ [Accessed 11 June 2015].

EUREC, 2015a. National Information: France. European Network of Research Ethics Committee. Available at: http://www.eurecnet.org/information/france.html [Accessed 10 February 2015a].

EUREC, 2015b. National Information: Sweden. European Network of Research Ethics Committees. Available at: http://www.eurecnet.org/information/sweden.html [Accessed 26 January 2015b].

EUREC, 2015c. National Information: The Netherlands. European Network of Research Ethics Committees. Available at: http://www.eurecnet.org/information/netherlands.html [Accessed 23 February 2015c].

EUREC, 2015d. National Informtion: Germany. European Network of Research Ethics Committee. Available at: http://www.eurecnet.org/information/germany.html [Accessed 19 February 2015d].

European Commission, 2010. *Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions*. Brussels: COM(2010) 609.

European Commission, 2012. *Proposal for regulation of the European Parliament and of the Council on protection of individuals with regard to the processing of personal data and on the free movements of such data*. Brussels: COM(2012) 11 final.

European Observatory on Health Systems and Policies, 2015. Health Systems in Transition (HiT) profile of France. The Health Systems and Policy Monitor. Available at: http://www.hspm.org/countries/france25062012/livinghit.aspx?Section=2.1%20Overview%20of%20the%20health%20system&Type=Section [Accessed 10 February 2015].

European Parliament, 2012. *Draft Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011–C7-0025/2012–2012/0011(COD))*. Committee on Civil Liberties, Justice and Home Affairs.PR\922387EN.doc.

Fagot, JP, 2012. Outcome and safety monitoring especially for new drugs. PIPERSKA workshop. Available at: http://www.piperska.org/sites/default/files/Day_3_Fagot.pdf [Accessed 5 February 2015].

Farr, 2014. Frequently Asked Questions. Farr Institute @ Scotland. Available at: http://www.farrinstitute.org/centre/Scotland/182_Frequently-Asked-Questions.html [Accessed 20 November 2014].

FDA, 2014. Mini-sentinel Distributed Database "At A Glance" as of July 2014. Food and Drug Administration. Available at: http://www.mini-sentinel.org/about_us/MSDD_At-a-Glance.aspx [Accessed 20 November 2014].

Fears, R., Brand, H., Frackowiak, R., Pastoret, P.P., Souhami, R., and Thompson, B., 2013. Data protection regulation and the promotion of health research: getting the balance right. *QJM* pp. hct236

Ferrario, A. and Kanavos, P., 2013. *Managed entry agreements for pharmaceuticals: The European experience*. EMINET.

Ford, D.V., Jones, K.H., Verplancke, J.P., Lyons, R.A., John, G., Brown, G., Brooks, C.J., Thompson, S., Bodger, O., and Couch, T., 2009. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research*, 9 (1) pp. 157

France, G., Taroni, F., and Donatini, A., 2005. The Italian health–care system. *Health economics*, 14 (S1) pp. S187-S202

G-BA, 2015. The Federal Joint Committee: Who we are and what we do. Gemeinsamer Bundesausschuss. Available at: http://www.english.g-ba.de/structure/ [Accessed 25 February 2015].

Garante, 2003. Personal Data Protection Code. Legislative Decree no.196 of 30 June 2003. Translation provided by Italian Trade Agency. Available at: http://www.italtrade.com/personalDataProtectionCode.pdf [Accessed 16 February 2015].

Garante, 2009. Guidelines on the Electronic Health Record and the Health File - 16 July 2009 [1672821]. Garante per la protezione dei dati personali. Available at:

http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1672821 [Accessed 18 February 2015].

Garante, 2012. General Authorisation to Process Personal Data for Scientific Research Purposes - 1 March 2012 [1884019]. Garante per la protezione dei dati personali. Available at: http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1884019 [Accessed 17 February 2015].

Garante, 2013. Authorisation No. 2/2013 Concerning - Processing of Data Suitable for Disclosing Health or Sex Life [2941268]. Garante per la protezione dei dati personali. Available at: http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/2941268 [Accessed 16 February 2015].

Garante, 2015. Italian Legislation. Garante per la protezione dei dati personali. Available at: http://www.garanteprivacy.it/home_en/italian-legislation [Accessed 16 February 2015].

Garrison, L.P., Neumann, P.J., Erickson, P., Marshall, D., and Mullins, C.D., 2007. Using Real–World Data for Coverage and Payment Decisions: The ISPOR Real–World Data Task Force Report. *Value in Health*, 10 (5) pp. 326-335

Garrison, L.P., Towse, A., Briggs, A., de Pouvourville, G., Grueger, J., Mohr, P.E., Severens, J.H., Siviero, P., and Sleeper, M., 2013. Performance-based risk-sharing arrangements - good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. *Value in Health*, 16 (5) pp. 703-719

Gartner, 2014. IT Glossary. Information Governance. Gartner, Inc. Available at: http://www.gartner.com/it-glossary/information-governance [Accessed 18 December 2014].

GBE des Bundes, 2015a. Behavioural and Risk Aspects of Health. Gesundheitsberichterstattung de Bundes. Available at: https://www.gbe-bund.de/gbe10/abrechnung.prc_abr_test_logon?p_uid=gast&p_aid=4711&p_sprache=E&p_knoten=TR5800 [Accessed 26 February 2015a].

GBE des Bundes, 2015b. Diseases / Health Problems: Endocrine Diseases - Diabetes. Gesundheitsberichterstattung des Bundes. Available at: https://www.gbe-bund.de/gbe10/trecherche.prc_them_rech?tk=8500&tk2=12000&p_uid=gast&p_aid=16330398&p_sprache=E&cnt_ut=5&ut=12100 [Accessed 26 February 2015b].

Gematik, 2012. For a health-care system with a future. The electronic health insurance card. Gematik.de. Available at: http://www.gematik.de/cms/media/infomaterialpresse/_Broschuere_Englisch_2.pdf [Accessed 25 February 2015].

Goldstein, M. and Rein, A., 2010. *Consumer options for electronic health information excahnge: policy considerations and analysis*. Washington: A White Paper prepared for the Office of the National Coordinator for Health IT.

Government, 1998. Data Protection Act. The National Archives. Available at: http://www.legislation.gov.uk/ukpga/1998/29/contents [Accessed 7 November 2014].

Government, 2006. National Health Service Act. The National Archives. Available at: http://www.legislation.gov.uk/ukpga/2006/41/contents [Accessed 13 November 2014].

Government, 2012. Modernising health and care public bodies - *The Health and Social Care Act 2012*. Factsheet. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/138265/B6.-Factsheet-Streamlined-arms-length-bodies-250412.pdf [Accessed 13 November 2014].

Government of the Netherlands, 2015a. Health insurance. Government.nl. Available at: http://www.government.nl/issues/health-insurance/standard-health-insurance [Accessed 18 February 2015a].

Government of the Netherlands, 2015b. The Citizen Service Number (BSN). Identification documents. Available at: http://www.government.nl/issues/identification-documents/the-citizen-service-number [Accessed 22 February 2015b].

GRACE Initiative, 2015. Good ReseArch for Comparative Effectiveness. Promoting High Quality Observational Research. Available at: http://www.graceprinciples.org/ [Accessed 10 February 2015].

Gray, B., Bowden, T., Johansen, I., and Koch, S., 2011. Electronic Health Records: An International Perspective on "Meaningful Use". *Commonwealth Fund pub.1565.Vol.28*

Green, D, Irvine, B, Clarke, E, and Bidgood, E, 2013. Healthcare Systems: France. CIVITAS. Available at: http://www.civitas.org.uk/nhs/download/france.pdf [Accessed 28 January 2015].

Grossman, C. and McGinnis, J.M., 2011. *Digital Infrastructure for the Learning Health System:: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary.* National Academies Press.

Hager, E. W. and Mirsch, A, 2014. Data protection in Sweden: overview. Practical Law.Thomson Reuters. Available at: http://uk.practicallaw.com/8-502-0348?service=ipandit# [Accessed 22 January 2015].

Hallgren Elfgren, I.M., Grodzinsky, E., and T+Ârnvall, E., 2013. Swedish diabetes register, a tool for quality development in primary health care. *Primary health care research & development*, 14 (03) pp. 250-257

Ham, C., Heenan, D., Longley, M., and Steel, D., 2013. *Integrated care in Northern Ireland, Scotland, and Wales. Lessons from England*. The King's Fund.978 1 909029 13 2.

HAS, 2011. *Les études post-inscription sur les technologies de santé - Principes et mèthods*. Haure Autorité de Santé. Unite methdologie et etudes post-inscription.

Haut Conseil de la Santé Publique, 2012. *Pour une meilleure utilisation des bases de données nationales pour la santé publique et la recherche*.

HCSP, 2012. *Pour une meilleure utilisation des bases de données nationales pour la santé publique et la recherche*.

HDN, 2013. Germany. Health Data Navigator. Available at: http://www.healthdatanavigator.eu/national/germany [Accessed 19 February 2015].

HDN, 2015. France Data Source. Health Data Navigator.EuroREACH. Available at: http://www.healthdatanavigator.eu/component/content/article/80-data-source/127-france-data-source [Accessed 5 February 2015].

HealthCore, 2014. Uncommon Depth of Knowledge, Empowered at Its Core. Subsidiary of WellPoint Inc. Available at: http://healthcore.com/home/research_enviro.php?page=Research%20Environment [Accessed 20 November 2014].

Henriksson, C., Larsson, M., Herlitz, J., Karlsson, J.E., Wernroth, L., and Lindahl, B., 2014. Influence of health-related quality of life on time from symptom onset to hospital arrival and the risk of readmission in patients with myocardial infarction. *Open heart*, 1 (1) pp. e000051

HIQA, 2010. *International Review of Unique Health Identifiers for Individuals*. Health Information and Quality Authority.

HMO, 2011. IRB Review of Multi-Site Research. HMO Research Network. Available at: http://www.hmoresearchnetwork.org/en/Tools%20&%20Materials/IRB/HMORN_IRB_SOP.pdf [Accessed 21 November 2014].

Hoeksma, J, 2010. Germany suspends e-health card project. ehi Primary Care. Available at: http://www.ehi.co.uk/news/primary-care/5551 [Accessed 19 February 20015].

Holtorf, A-K, Matuszewski, K, Nuijten, M, and Vauth, C, 2009. Germany - Pharmaceutical. ISPOR Global Health Care Systems Road Map. Available at: http://www.ispor.org/htaroadmaps/germany.asp [Accessed 25 February 2015].

HQIP, 2014a. Clinical databases and registries. Healthcare Quality Improvement Partnership. Available at: http://www.hqip.org.uk/support-for-clinical-databases-and-registers/ [Accessed 18 November 2014a].

HQIP, 2014b. NCAPOP: Data access request process. Healthcare Quality Improvement Partnership. Available at: http://www.hqip.org.uk/national-clinical-audit-and-patient-outcomes-programme-data-access-request-process/ [Accessed 20 November 2014b].

HQIP, 2014c. What is the difference between clinical audit and research? Healthcare Quality Improvement Partnership. Available at: http://hqip.org.uk/what-is-the-difference-between-clinical-audit-and-research-3/ [Accessed 24 November 2014c].

HRA, 2014a. Is it research? Health Research Authority. Available at: http://www.hra.nhs.uk/resources/before-you-apply/is-it-research/ [Accessed 24 November 2014a].

HRA, 2014b. Research Governance Frameworks. Health Research Authority. Available at: http://www.hra.nhs.uk/resources/research-legislation-and-governance/research-governance-frameworks/ [Accessed 24 November 2014b].

HRA and MRC, 2014. Is my study research? HRA Decision Tool. Available at: http://www.hra-decisiontools.org.uk/research/ [Accessed 24 November 2014].

Hripcsak, G., Bloomrosen, M., FlatelyBrennan, P., Chute, C.G., Cimino, J., Detmer, D.E., Edmunds, M., Embi, P.J., Goldstein, M.M., and Hammond, W.E., 2013. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. *Journal of the American Medical Informatics Association* pp. amiajnl-2013

HSC BSO, 2014. Honest Broker Service. Business Services Organisation. Available at: http://www.hscbusiness.hscni.net/services/2454.htm [Accessed 18 November 2014].

HSCIC, 2013a. A guide to confidentiality in health and social care. Version 1.1. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/media/12822/Guide-to-confidentiality-in-health-and-social-care/pdf/HSCIC-guide-to-confidentiality.pdf [Accessed 13 November 2014a].

HSCIC, 2013b. Hospital Prescribing: England 2012 - Data Quality Statement. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/catalogue/PUB12651/hosp-pres-eng-2012-qual.pdf [Accessed 18 November 2014b].

HSCIC, 2014a. Data Access Request Service (DARS). Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/dars [Accessed 20 November 2014a].

HSCIC, 2014b. Data Access Request Service. Service Charges 2014/15. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/media/14839/DARS---Service-charges/pdf/DARS-Service_Charges.pdf [Accessed 24 November 2014b].

HSCIC, 31-7-2014c. Data Access Request Service: Products. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/media/14837/DARS---Products/pdf/DARS_-_Products.pdf [Accessed 17 November 2014c].

HSCIC, 2014d. Data flows transition - Commissioner access to SUS. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/dataflowstransitionmanual [Accessed 13 November 2014d].

HSCIC, 2014e. Diabetes Data Set. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/article/2116/Diabetes-Data-Set [Accessed 18 November 2014e].

HSCIC, 2014f. FAQs on legal access to personal confidential data. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/article/3638/Patient-data-access-FAQs [Accessed 13 November 2014f].

HSCIC, 2014g. Information Governance (IG). HSCIC.gov. Available at: http://systems.hscic.gov.uk/infogov [Accessed 7 November 2014g].

HSCIC, 2014h. Information Governance Alliance - work programme. Health and Social Care Information Centre. Available at: http://systems.hscic.gov.uk/infogov/iga/ourwork [Accessed 2 November 2014h].

HSCIC, 2014i. Looking after your health and care information. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/article/3388/Looking-after-your-health-and-care-information [Accessed 20 November 2014i].

HSCIC, 2014j. Registers of approved data releases. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/dataregister [Accessed 24 November 2014j].

HSCIC, 2014k. Rules for sharing information. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/article/3399/Rules-for-sharing-information [Accessed 14 November 2014k].

HSCIC, 2014l. Secondary Uses Service (SUS). Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk/sus [Accessed 13 November 2014l].

HSCIC, 23-1-2015. Data sharing: HSCIC statement on patient objections. Heath and Social Care Information Centre. Available at: http://www.hscic.gov.uk/5342 [Accessed 14 February 2015].

Hughes, B and Kessler, M, 2013. RWE Market Impact on Medicines: A lens for Pharma. IMS Health White Paper.RWEWHI0513/MIC3030. Available at: http://www.imsconsultinggroup.com/deployedfiles/consulting/Global/Content/Our%20Latest%20Thinking/Static%20Files/rwe_market_impact_on_medicines.pdf [Accessed 19 August 2014].

Humana, 2014. Comprehensive Health Insights: Data Assets. A Humana Company. Available at: http://www.competitive-health-analytics.com/research/data_assets.asp [Accessed 26 November 2014].

Hunton & Williams, 2-2-2015. Debate Over the Progress of the EU General Data Protection Regulation. Hunton & Williams Privacy & Information Security Law Blog. Available at: https://www.huntonprivacyblog.com/2015/02/02/debate-progress-eu-general-data-protection-regulation/ [Accessed 2 June 2015].

ICO, 2012. Anonymisation: managing data protection risk. Code of practice. Information Commissioner's Office. Available at: http://ico.org.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Practical_application/anonymisation-codev2.pdf [Accessed 25 November 2014].

ICO, 2014. Data protection - looking after the data you hold about patients. Information Commissioner's Office. Available at: http://ico.org.uk/for_organisations/sector_guides/health [Accessed 7 November 2014].

IHIE, 2012. *IHIE - Building Effective Data Governance Models, Policies and Agreements in a HITECH World*.

IIGOP, 2013. *Information: To share or not to share? The Information Governance Review*. Department of Health.

IMI, 2015. Workpackage 1: Developing a framework for the assessment of development strategies that provide evidence of relative effectiveness. IMI Get Real. Available at: http://www.imi-getreal.eu/About-GetReal/Workpackage-1 [Accessed 12 January 2015].

IOM, 2013. *Key capabilities of an Electronic Health Record System: Letter Report*. Washington D.C.: Institute of Medicine: Committee on Data Standards for Patient Safety.

IOM, 2014. *Discussion framework for clinical trial data sharing : guiding principles, elements, and activities*. Institute of Medicine (U.S.).

IRAS, 2014a. IRAS Guidance Project Filter. Integrated Research Application Sysmte (IRAS). Available at: https://www.myresearchproject.org.uk/help/hlpcollatedqsg-sieve.aspx [Accessed 24 November 2014a].

IRAS, 2014b. Welcome to the Integrated Research Application System (IRAS). IRAS. Available at: https://www.myresearchproject.org.uk/SignIn.aspx [Accessed 24 November 2014b].

ISB, 2013. Standard ISB 1523. Anonymisation Standard for Publishing Health and Social Care Data. Infomation Standard Board for Health and Social Care. Available at: http://www.isb.nhs.uk/library/standard/128 [Accessed 18 November 2014].

ISD, 2013. Data Sharing and Linking Service. Analysis of responses to the Technical Consultation on the design of Data Sharing and Linking Service. eDRIS. Available at: http://www.isdscotland.org/Products-and-Services/EDRIS/DSLS-consultation/DSLS-Consultation-Analysis-Report.pdf [Accessed 20 November 2014].

ISD, 2014a. Health and Social Care Data Integration. NHS NSS. Available at: http://www.isdscotland.org/Products-and-Services/Health-and-Social-Care-Integration/ [Accessed 20 November 2014a].

ISD, 2014b. Topics. ISD's Major Work Programmes. Information Services Division's. Available at: http://www.isdscotland.org/Health-Topics/index.asp [Accessed 18 November 2014b].

ISPOR, 2007. The Netherlands - Reimbursement Process. ISPOR Global Health Care System Road Map. Available at: http://www.ispor.org/htaroadmaps/netherlands.asp [Accessed 25 February 2015].

Jansen, T and Hinzpeter, B, 2014. Data protection in Germany: overview. Practical Law.Thomson Reuters. Available at: http://uk.practicallaw.com/3-502-4080?q=germany+data+protection [Accessed 19 February 2015].

Janusinfo, 2015. Decisions imposing level. Janusinfo.Stockholms lans landsting. Available at: http://www.janusinfo.se/Nationellt-inforande-av-nya-lakemedel/Nationellt-inforande-av-nya-lakemedel/Beslut-om-inforandeniva/ [Accessed 27 January 2015].

Lantieri, A. and Pelsy, F., 2014. *Overview of the national laws on electronic health records in the EU Member States. National Report for France*. Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth service. Contract 2013 63 02. Milieu Ltd.

Laurie, G. and Sethi, N., 2011. *Information Governance Of Use Of Health-Related Data In Medical Research In Scotland: Current Practice and Future Scenarios*. SHIP Core Programme 2 Working Paper No.1. SHIP Core Programme 2 Working Paper No.1, University of Edinburgh.

Laurie, G. and Sethi, N., 2012. *Information Governance Of Use Of Health-Related Data In Medical Research In Scotland: Towards a Good Governance Framework*. SHIP Working Paper No.2, University of Edinburgh. No 2012/13

Luce, B.R., Drummond, M., Jönsson, B., Neumann, P.J., Schwartz, J.S., Siebert, U.W.E., and Sullivan, S.D., 2010. EBM, HTA, and CER: clearing the confusion. *Milbank Quarterly*, 88 (2) pp. 256-276

Mattsson, T., 2014. Patient safety at odds with patient privacy? The case of national and regional quality registries for incapacitated elderly in Sweden. *Lex Medicinae* (IV EAHL Conference) pp. 69-82

MAV, VLGA, LGV, and LGPro, 2014. What is Good Governance? Municipal Association of Victoria (MAV), Victorian Local Governance Association (VLGA), Local Government Victoria (LGV) and Local Government Professionals (LGPro). Available at: http://www.goodgovernance.org.au/about-good-governance/what-is-good-governance/ [Accessed 28 November 2014].

Miani, C., Robin, E., Horvath, V., Manville, C., Cave, J., and Chatawat, J., 2015. *Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe*. Cambridge: RAND Europe.

Mondriaan, 2011a. Mondriaan Privacy. Foundation Mondriaan Health Research Data. Available at: http://mondriaanfoundation.org/privacy.html [Accessed 20 February 2015a].

Mondriaan, 2011b. Review of research proposals and data requets. Foundation Mondriaan Health Research Data. Available at: http://mondriaanfoundation.org/toetsing.html [Accessed 21 February 2015b].

Moroni, C, 2014. Italy's AIFA: Agenda for Regulatory Leadership. Interview with Dr Sergio Pecorelli, Chairman of AIFA. Pharmaceutical Executive. Available at: http://www.pharmexec.com/italys-aifa-agenda-regulatory-leadership [Accessed 18 February 2015].

MRC, 2014. Glossary. Medical Research Council (MRC) Data and Tissues Tool Kit. Available at: http://www.dt-toolkit.ac.uk/glossary.cfm [Accessed 14 November 2014].

Navarria, A., Drago, V., Gozzo, L., Longo, L., Mansueto, S., Pignataro, G., and Deago, F., 2015. Do the current performance-based schemes in Italy really work? "Success fee": a novel measure for cost-containment of drug expenditure. *Value Health*, 18 (1) pp. 131-6

NCHVS, 2009. *Health Data Stewardship: What, Why, Who, How (An NCHVS Primer)* . Hyattsville, MD: National Committee on Vital and Health Statistics.

NCI, 2013. PHARMO Record Linkage System. National Cancer Institute, at the National Institute of Health. Available at: http://epi.grants.cancer.gov/pharm/pharmacoepi_db/pharmo.html [Accessed 22 February 2015].

NCI, 2014. Surveillance, Epidemiology, and End Results Program. National Cancer Institute. Available at: http://seer.cancer.gov/ [Accessed 22 November 2014].

NCIN, 2014a. Cancer Outcomes and Services Dataset (COSD). National Cancer Research Institute. Available at: http://www.ncin.org.uk/collecting_and_using_data/data_collection/cosd [Accessed 18 November 2014a].

NCIN, 2014b. SACT Chemotherapy Dataset. National Cancer Intelligence Network. Available at: http://www.chemodataset.nhs.uk/home [Accessed 18 November 2014b].

NHHRC, 2009. *A healthier future for all Australians - Final Report of the National Health and Hospitals Reform Commission*. Commonwealth of Australia. 1-74186-940-4.

NHMRC, 2014. National Statement on Ethical Conduct in Human Research (2007) - Updated March 2014. Australian Government National Health and Medical Research Council. Available at: https://www.nhmrc.gov.au/guidelines-publications/e72 [Accessed 29 January 2015].

NHS, 2008. *NHS Number Programme Implementation Guidance*.

NHS, 2014. What is Section 251? NHS Health Research Authority. Available at: http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/what-is-section-251/ [Accessed 13 November 2014].

NHS Choices, 2013. Your records. NHS. Available at: http://www.nhs.uk/nhsengland/thenhs/records/healthrecords/pages/overview.aspx [Accessed 7 November 2014].

NHS England, 2014a. CCGs to help develop care.data programme. NHS England News. Available at: http://www.england.nhs.uk/2014/10/07/ccgs-care-data-programme/ [Accessed 25 November 2014a].

NHS England, 2014b. Information Governance. NHS Engalnd Technology, systems and data. Available at: http://www.england.nhs.uk/ourwork/tsd/ig/ [Accessed 14 November 2014b].

NHS England, 2014c. Information governance bulletin. NHS England technology, systems and data. Available at: http://www.england.nhs.uk/ourwork/tsd/ig/ig-bull/ [Accessed 14 November 2014c].

NHS England & HSCIC, 2013. *NHS Hospital Data and Datasets: A consultation*. Leeds: NHS England and the Health and Social Care Information Centre.

NHS European Office, 2014. Data Protection. NHS Confederation. Available at: http://www.nhsconfed.org/regions-and-eu/nhs-european-office/influencing-eu-policy/data-protection [Accessed 7 November 2014].

NICE, 2011. *Medical Technologies Evaluation Programme: Methods guide*.

NICE, 2013. *Guide to the methods of technology appraisal 2013*.

NICE, 2014. NICE calls for a new aproach to managing the entry of drugs into the NHS. National Institute of Health and Care Excellence. Available at: https://www.nice.org.uk/news/press-and-media/nice-calls-for-a-new-approach-to-managing-the-entry-of-drugs-into-the-nhs [Accessed 13 February 2015].

NIGB, 2011. The Care Record Guarantee. National Information Governance Board for Health and Social Care (NIGB). Available at: http://webarchive.nationalarchives.gov.uk/20130513181011/http://nigb.nhs.uk/pubs/nhscrg.pdf [Accessed 1 July 2015].

Nordqvist, E, 28-7-2014. About the Swedish eHealth Agency. eHalsomyndigheten. Available at: http://www.ehalsomyndigheten.se/Om-oss-/Uppdrag-och-verksamhet/Other-languages1/Swedish-eHealth-Agency/ [Accessed 23 January 2015].

Oderkirk, J., Ronchi, E., and Klazinga, N., 2013. International comparisons of health system performance among OECD countries: opportunities and data privacy protection challenges. *Health Policy*, 112 (1) pp. 9-18

OECD, 2013. *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*. OECD Publishing.

OECD, 2014a. How does Italy compare? OECD Health Statistics 2014. Available at: http://www.oecd.org/els/health-systems/Briefing-Note-ITALY-2014.pdf [Accessed 10 February 2015a].

OECD, 2014b. How does Sweden compare? OECD Health Statistics 2014. Available at: http://www.oecd.org/els/health-systems/Briefing-Note-SWEDEN-2014.pdf [Accessed 23 January 2015b].

Optum, 2014. *Optum Core Datasets*. Eden Prairie.

Payne.C, 2013. *Expenditure on healthcare in the UK: 2011*. Office for National Statistis:

PCORI, 2014a. Common Data Model (CDM) Specification, Version 1.0. Data Standards,Security and Network Infrastructure (DSSNI) Task Force. Available at: http://pcornet.org/wp-content/uploads/2014/07/2014-05-30g-PCORnet-Common-Data-Model-v1-0-RELEASE.pdf [Accessed 22 November 2014a].

PCORI, 2014b. PCORnet: The National Patient-Centered Clinical Research Network. Patient-Centered Outcomes Research Institute. Available at: http://www.pcori.org/content/pcornet-national-patient-centered-clinical-research-network [Accessed 22 November 2014b].

Persson, U., Willis, M., and +ûdegaard, K., 2010. A case study of ex ante, value-based price and reimbursement decision-making: TLV and rimonabant in Sweden. *The European Journal of Health Economics*, 11 (2) pp. 195-203

Pharmafakt, 2015. Datenschutzrichtlinien. Pharmafakt GmbH. Available at: http://www.pharmafakt.de/index-Dateien/Page915.htm [Accessed 26 February 2015].

PHARMO, 2015. Who we are. The PHARMO Institute. Available at: http://www.pharmo.nl/about-us/who-we-are [Accessed 18 February 2015].

PHRN, 2015. What type of data is used? Population Health Research Network. Available at: http://www.phrn.org.au/about-us/data-linkage/what-type-of-data-is-used/ [Accessed 1 February 2015].

Ramesh, R, 2014. NHS patient data to be made available for sale to drug and insurance firms. The Guardian. Available at: http://www.theguardian.com/society/2014/jan/19/nhs-patient-data-available-companies-buy [Accessed 25 November 2014].

ResDAC, 2013. Introduction to CMS Data. Research Data Assistance Center. Available at: http://www.resdac.org/sites/resdac.org/files/Introduction%20to%20CMS%20Data%20%28Slides%29.pdf [Accessed 23 November 2014].

ResDAC, 2014. Find a CMS Data File. Research Data Assistance Center. Available at: http://www.resdac.org/cms-data [Accessed 23 November 2014].

Rice, T., Rosenau, P., Unruh, L., Barnes, A., Saltman, R., and van Ginneken, E., 2013. United States of America: health system review. *Health systems in transition*, 15 (3) pp. 1-431

Richesson, R.L., Hammond, W.E., Nahm, M., Wixted, D., Simon, G.E., Robinson, J.G., Bauck, A.E., Cifelli, D., Smerek, M.M., and Dickerson, J., 2013. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *Journal of the American Medical Informatics Association*, 20 (e2) pp. e226-e231

RKI, 2004. The Robert Koch Institute: Tasks and Aims of the Robert Koch Institute. Robery Koch Institut. Available at: http://www.rki.de/EN/Content/Institute/institute_node.html [Accessed 26 February 2015].

RKI, 2015a. German Centre for Cancer Registry Data (ZfKD). Zentrum fur Krebsregisterdaten. Available at: http://www.krebsdaten.de/Krebs/EN/Home/homepage_node.html [Accessed 20 February 2015a].

RKI, 2015b. Information about applying for data use. Zentrum fur Krebsregisterdaten. Available at: http://www.krebsdaten.de/Krebs/EN/Content/ScientificUseFile/Information_about_application/information_about_application_node.html [Accessed 20 February 2015b].

Roggemans, M, 2012. White Paper: Clinical Research in France - an Introduction. CROM SOURCE. Available at: http://www.cromsource.com/wp-content/uploads/2012/12/Clinical-Trials-in-France.pdf [Accessed 10 February 2015].

Rolandsson, O., Norberg, M., Nystr+Âm, L., S+Âderberg, S., Svensson, M., Lindahl, B., and Weinehall, L., 2012. How to diagnose and classify diabetes in primary health care: lessons learned from the Diabetes Register in Northern Sweden (DiabNorth). *Scandinavian journal of primary health care*, 30 (2) pp. 81-87

Rosenbaum, S., 2010. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health services research*, 45 (5p2) pp. 1442-1455

Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel-Fox, S., Tang, P.C., and Detmer, D.E., 2007. 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) pp. 1-9

SAIL, 2013. Data Management Policy. Secure Anonymised Information Linkage (SAIL). Available at: http://www.saildatabank.com/media/20689/SAIL%20data%20management%20policy%20v1%200.pdf [Accessed 18 November 2014].

SAIL, 2014a. Data Acquisition Process. Secure Anonymised Information Linkage Databank. Available at: http://www.saildatabank.com/data-acquisition-process [Accessed 20 November 2014a].

SAIL, 2014b. Governance. Secure Anonymised Information Linkage Databank. Available at: http://www.saildatabank.com/governance.aspx [Accessed 18 November 2014b].

SAIL, 2014c. SAIL Datasets. Secure Anonymised Information Linkage Databank. Available at: http://www.saildatabank.com/data-dictionary/sail-datasets [Accessed 18 November 2014c].

Santarlasci, B., Messori, A., Pelagotti, F., Trippoli, S., and Vaiani, M., 2005. Heterogeneity in the evaluation of observational studies by Italian ethics committees. *Pharmacy World and Science*, 27 (1) pp. 2-3

Scottish Government, 2012. *Joined up data for better decisions: A strategy for improving data access and analysis. Answering the important questions for Scotland through legal, secure, ethical and efficient data linkage*.

Scottish Government, 2013. Proposal for Joint Data Linkage and Informatics Centre. Paper 3. Available at: http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.scotland.gov.uk%2FResource%2F0044%2F00445801.docx&ei=rNdtVL_TDcKbNumBgQg&usg=AFQjCNGi6QrKMZFKVF1X9oO6DJUEIqKZxA&sig2=-PviZan-0UKolKcSsP-z8Q&bvm=bv.80120444,d.eXY [Accessed 20 November 2014].

Scottish Government, 2014. Information Governance. eHealth. Available at: http://www.ehealth.scot.nhs.uk/information-governance/ [Accessed 18 November 2014].

Sethi, N. and Laurie, G.T., 2013. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Medical law international*, 13 (2-3) pp. 168-204

Smart Card Alliance, 2006. German Health Card. Smart Card AllianceHealthcare Council. Available at: http://www.smartcardalliance.org/resources/pdf/German_Health_Card.pdf [Accessed 19 February 2015].

Smeets, H.M., de Wit, N.J., and Hoes, A.W., 2011. Routine health insurance data for scientific research: potential and limitations of the Agis Health Database. *Journal of clinical epidemiology*, 64 (4) pp. 424-430

Socialstyrelsen, 2011. National Information Structure for health and social care in Sweden
social care in Sweden. Socialstyrelsen.The National Board of Health and Welfare. Available at: http://www.socialstyrelsen.se/e-health/Documents/The_National_Information_Structure_for_health_and_social_care_in_Sweden_summary.pdf [Accessed 22 January 2015].

Socialstyrelsen, 2013. *Quality and efficiency in swedish health care - regional comparisons 2012*. Swedish National Board of Health and Welfare and the Swedish Association of Local Authorities and Regions.978-91-7164-949-2.

Socialstyrelsen, 2015a. Ehealth is one of the National Board of Health and Welfare's responsibilities. Socialstyrelsen eHealth. Available at: http://www.socialstyrelsen.se/e-health [Accessed 22 January 2015a].

Socialstyrelsen, 2015b. If you want to order data or statistics. National Board for Health and Welfare. Available at: http://www.socialstyrelsen.se/register/bestalladatastatistik [Accessed 27 January 2015b].

Socialstyrelsen, 2015c. National Quality. National Board for Health and Welfare. Available at:
http://www.socialstyrelsen.se/register/registerservice/nationellakvalitetsregister [Accessed 27 January 2015c].

Socialstyrelsen, 2015d. Order individual information for research purposes. National Board for Health and Welfare. Available at:
http://www.socialstyrelsen.se/register/bestalladatastatistik/bestallaindividuppgifterforforskningsandamal [Accessed 27 January 2015d].

Socialstyrelsen, 2015e. Step-by-step in research on registers. National Board of Health and Welfare. Available at:
http://www.socialstyrelsen.se/register/bestalladatastatistik/bestallaindividuppgifterforforskningsandamal/steg-for-stegvidforskningparegister [Accessed 27 January 2015e].

Socialstyrelsen, 2015f. Swedish cancer registry. Socialstyrelsen.The National Board of Health and Welfare. Available at:
http://www.socialstyrelsen.se/register/halsodataregister/cancerregistret/inenglish [Accessed 22 January 2015f].

Socialstyrelsen, 2015g. The National Board of Health and Welfare. Socialstyrelsen. Available at: http://www.socialstyrelsen.se/english [Accessed 22 January 2015g].

Socialstyrelsen, 2015h. The National Patient Register. Socialstyrelsen.The National Board of Health and Welfare. Available at:

http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish [Accessed 22 January 2015h].

Socialstyrelsen, 2015i. The Swedish medical birth register. Socialstyrelsen.The National Board of Health and Welfare. Available at: http://www.socialstyrelsen.se/register/halsodataregister/medicinskafodelseregistret/inenglish [Accessed 22 January 2015i].

Statistics Netherlands, 2013. Growth in spending on health care in the Netherlands and the OECD levelling off. CBS. Available at: http://www.cbs.nl/en-GB/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2013/2013-3998-wm.htm [Accessed 20 February 2015].

Swedish Government, 1998. *Personal Data Act (1998:204)*.

Swedish Institute, 2015. Health care in Sweden. sweden.se. Available at: https://sweden.se/society/health-care-in-sweden/ [Accessed 22 January 2015].

Tang, P., 2010. *Health IT Policy Committee Letter to David Blumenthal, National Coordinator for Health Information Technology. [recommendations on the nationwide information network governance issue by HealthIT Policy Committee Governance Workgroup.*

Taylor, K, 2014. Buidling trust in the collection and use of real world health data. Deloitte. Available at: http://blogs.deloitte.co.uk/health/2014/06/building-trust-in-the-collection-and-use-of-real-world-health-data.html [Accessed 24 November 2014].

The Independent Commission on Good Governance in Public Services, 2004. *The Good Governance Standard for Public Services*. Office for Public Management Ltd (OPM) & The Chartered Institute of Public Finance and Accountancy (CIPFA). London: Hackney Press Ltd.1 898531 86 2.

TIPharma, 2015. The Mondriaan Project. TIPharma, Netherlands. Available at: http://www.tipharma.com/pharmaceutical-research-projects/completed-projects/mondriaan-project.html [Accessed 18 February 2015].

Touraine, M., 2014. *Commission open data en santé*. Ministre des Affaires sociales et de la Santé.

Truven Health Analytics, 2014. Better Understand Health Economics and Treatment Outcomes. Databases and Tools. Available at: http://truvenhealth.com/your-healthcare-focus/life-sciences/data-databases-and-online-tools [Accessed 16 November 2014].

Twitter, 2015. Transparance Santé. Twitter.com. Available at: https://twitter.com/OpenDataSante/media [Accessed 10 February 2015].

UMC Utrecht, 2015. AGIS Health Database. UMC Utrecht Julius Center. Available at: http://portal.juliuscentrum.nl/research/en-US/cohortsandprojects/cohortsprojects/agishealthdatabase.aspx [Accessed 19 February 2015].

US Census Bureau, 2014. *Current Population Survey, 2014 Annual Social and Economics Supplement.* Washington D.C.: US Census Bureau.

Wammes, J., Jeurissen, P., and Westert, G., 2015. *'The Dutch Health care System, 2014' in: International Profiles of Health Care Systems*. The Commonwealth Fund.

Willis, M., Persson, U., Zoellner, Y., and Gradl, B., 2010. Reducing uncertainty in value-based pricing using evidence development agreements. *Applied health economics and health policy*, 8 (6) pp. 377-386

Wilson, J. and Bock, A., 2012. The benefit of using both claims data and electronic medical record data in health care analysis. *Optum Insight*

Wolf, M.S. and Bennett, C.L., 2006. Local perspective of the impact of the HIPAA privacy rule on research. *Cancer*, 106 (2) pp. 474-479

Wonder, M., Backhouse, M.E., and Sullivan, S.D., 2012. Australian managed entry scheme: a new manageable process for the reimbursement of new medicines? *Value in Health*, 15 (3) pp. 586-590

Working Party, 2012. Working Document 01/2012 on epSOS. 00145/12/EN. WP189. Article 29 Data Protection Working Party. Available at: http://194.242.234.211/documents/10160/2133805/WP189 [Accessed 18 February 2015].