



## EUROPE

CHILDREN AND FAMILIES  
EDUCATION AND THE ARTS  
ENERGY AND ENVIRONMENT  
HEALTH AND HEALTH CARE  
INFRASTRUCTURE AND  
TRANSPORTATION  
INTERNATIONAL AFFAIRS  
LAW AND BUSINESS  
NATIONAL SECURITY  
POPULATION AND AGING  
PUBLIC SAFETY  
SCIENCE AND TECHNOLOGY  
TERRORISM AND  
HOMELAND SECURITY

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

Skip all front matter: [Jump to Page 1](#) ▼

### Support RAND

[Browse Reports & Bookstore](#)

[Make a charitable contribution](#)

### For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore [RAND Europe](#)

View [document details](#)

### Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND Web site is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).

This report is part of the RAND Corporation research report series. RAND reports present research findings and objective analysis that address the challenges facing the public and private sectors. All RAND reports undergo rigorous peer review to ensure high standards for research quality and objectivity.



# Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe

Céline Miani, Enora Robin, Veronika Horvath, Catriona Manville,  
Jonathan Cave and Joanna Chataway, with contributions from IBM



EUROPE

# Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe

Céline Miani, Enora Robin, Veronika Horvath, Catriona Manville,  
Jonathan Cave, Joanna Chataway, with contributions from IBM

The research described in this report was prepared for  
Pfizer Pharmaceuticals Ltd.

RAND Europe is an independent, not-for-profit policy research organisation that aims to improve policy and decisionmaking for the public interest through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**RAND**® is a registered trademark

© Copyright 2014 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page ([www.rand.org/publications/permissions.html](http://www.rand.org/publications/permissions.html)).

RAND OFFICES

SANTA MONICA, CA • WASHINGTON, DC  
PITTSBURGH, PA • NEW ORLEANS, LA • JACKSON, MS • BOSTON, MA  
DOHA, QA • CAMBRIDGE, UK • BRUSSELS, BE  
[www.rand.org](http://www.rand.org) • [www.rand.org/randeurope](http://www.rand.org/randeurope)

# Preface

---

In October 2013 RAND Europe and IBM were invited to support Pfizer to assess the real-world data policy landscape in Europe and this report documents the results of a short study conducted over six weeks.

Real-world data (RWD) is an umbrella term for different types of data that are not collected in conventional randomised controlled trials. RWD comes from various sources and includes patient data, data from clinicians, hospital data, data from payers and social data. Through its use alongside traditional data sources such as clinical trials, RWD has the potential to provide new insights into medicines and their effects in the context of different patient groups.

There are already examples of ways in which research has contributed to the provision, construction and capture of RWD to improve health outcomes. However, to maximise the potential of these new pools of data in the healthcare sector, stakeholders need to identify pathways and processes which will allow them to efficiently access and use RWD in order to achieve better research outcomes and improved healthcare delivery.

Current efforts to improve access to RWD and facilitate its use take place in a context of resource scarcity. This provides incentives for the healthcare industry, care providers, policymakers and other stakeholders to align and coordinate their respective activities and carve out viable and productive regulation and practice.

RAND Europe and IBM have worked together through a multi-method approach to assess different RWD pathways that the health and healthcare sector has explored and the options going forwards. Based on a literature review, case studies and a small set of interviews of experts from public and private organisations, the study outlines possible strategies to illustrate how RWD standards development could facilitate RWD-based research.

For more information about RAND Europe or this document, please contact:

Prof Joanna Chataway

Tel. +44 1223 353329

[chataway@rand.org](mailto:chataway@rand.org)

RAND Europe

Westbrook Centre, Milton Road

Cambridge, CB4 1YG.



# Table of Contents

---

Preface.....	iii
Table of Contents.....	v
Tables.....	vi
Executive Summary.....	7
Acknowledgements.....	13
Abbreviations .....	15
<b>1. Introduction: The rise of RWD in health and healthcare research .....</b>	<b>17</b>
1.1. Origins and aims of the report .....	17
1.2. What is real-world data? .....	20
<b>2. Moving from data to evidence: the state of RWD in Europe .....</b>	<b>27</b>
2.1. RWD is useful for building the evidence base for drug development and post-market studies in the pharmaceutical and medical sectors .....	27
2.2. RWD is increasingly used to improve healthcare service delivery .....	29
2.3. Factors that influence access to and use of RWD .....	35
<b>3. Overcoming existing barriers through the development of standards to improve access to and use of RWD .....</b>	<b>41</b>
3.1. Overcoming content and quality issues .....	41
3.2. Overcoming methodological issues .....	49
3.3. Improving governance .....	57
3.4. Developing privacy best practices.....	60
3.5. Summary of findings: main barriers to access to and use of RWD and strategies to overcome them .....	63
<b>4. Conclusion .....</b>	<b>65</b>
4.1. Summary of findings .....	65
4.2. Understanding the role of RWD in the current political, societal and economic context.....	66
References .....	75



## Tables

---

Table 1 Examples of RWD.....	21
Table 2 Examples of responsibility for data management.....	23
Table 3 Areas for RWD use in health and healthcare.....	24
Table 4 The EuResist initiative.....	30
Table 5 National Cancer Institute of Milan: analytics and big data to improve cancer treatment.....	31
Table 6 The Shared Care platform: using RWD to improve the efficiency of homecare delivery in Southern Denmark.....	32
Table 7 Partnership to reduce readmissions for COPD.....	33
Table 8 Achmea Health Database.....	34
Table 9 Summary of the Danish EHR system: ‘my health summary’.....	45
Table 10 Regional initiatives in Spain and Italy.....	46
Table 11 The EPIRARE Project <sup>1</sup> : feasibility of patient-level data collection at the European scale.....	53
Table 12 Hypergenes, European Network for Genetic-Epidemiological Studies.....	55
Table 13 The epSOS project: a first attempt to combine e-health records across countries.....	56
Table 14 The CPRD (Clinical Practice Research Datalink).....	58
Table 15 Improving access to and use of RWD: from barriers to enablers.....	63
Table 16 PESTL analysis of the RWD landscape in Europe.....	70

## Executive Summary

---

Real-world data (RWD) is an umbrella term for different types of data that are not collected in conventional randomised controlled trials. In the healthcare sector, RWD can be obtained from many sources and includes patient data, data from clinicians, hospital data, data from payers and social data. In order to account for these varied datasets in this study, and to analyse their collective promise, we offer the following working definition:

RWD is any data not collected in conventional randomised controlled trials. It includes data from existing secondary sources (eg databases of national health services) and the collection of new data, both retrospectively and prospectively.

In the view of many analysts and researchers, RWD has significant potential to improve the ways drugs are discovered and developed. Moreover, the assessment of the value of medicines and treatments in real-world settings may be made less resource intensive with RWD-based methodologies. Whilst the health and healthcare sector is an early adopter of such methodologies, the rate, direction and use of data generation is influenced by a series of factors: technological advances, for example, or data protection policy. The continuing impact of such factors will occur in ways that are difficult to predict and sometimes contradictory. The pace of technological change and the pace of change in governance arrangements, capabilities and in the building of relationships necessary to allow for successful generation and use of RWD do not necessarily move in tandem and this can create significant disappointment and frustration for all those involved. The computing technology underpinning the collection and use of data is advancing fast. Whilst necessary, these developments alone are insufficient for the successful use of data. To be of use to health researchers and innovators, and to be acceptable as evidence, data needs to be processed, analysed and presented in a coherent form. Its use and form also needs to be acceptable to the broad range of stakeholders involved in health innovation. The complex interaction of computing technology, the practice of health research and innovation, and governance and standards can make it difficult to understand what is happening in the field and what constitute enabling and constraining forces.

In that context, this report provides an overview of the use of RWD in health research and innovation in Europe and attempts to develop learning and understanding about its future potential in European research. This report was commissioned by Pfizer in October 2013 and is based on research carried out

by RAND Europe and IBM over six weeks in the autumn/winter of 2013. The objectives of the work were the following:

- construct an evidence base on the establishment and evolution of standards governing the collection and use of RWD and identify the different ways in which standards have been applied
- understand the factors that have enabled or limited access to and use of RWD
- learn lessons from the use of RWD in different contexts
- identify opportunities for increasing access to RWD and contributing constructively to standard setting in Europe.

The study used a combination of approaches to survey and understand the current use of RWD and the potential opportunities for using this type of data going forwards. These approaches included a review of the academic and grey literature and also a small number of in-depth interviews with stakeholders from a range of public and private organisations in Europe. In addition, case study examples of the use of RWD in the healthcare sector were compiled. Finally, a workshop, engaging both the study team and individuals from Pfizer, brought together insights from these three streams of research to explore the options for the future and provide contextual information as to the feasibility of these opportunities.

By investigating the current forms and uses of RWD in Europe, this study has highlighted their significant potential for assessing the (short- or long-term) impact of different drugs or medical treatments and for informing and improving healthcare service delivery. Although the potential of RWD use seems quite clear, this research reveals barriers that restrict further development towards its full exploitation:

- the absence of common standards for defining the content and quality of RWD (absence of common terminology, incomplete datasets, lack of data quality assurance systems)
- methodological barriers (absence of standards for RWD analysis and for data linkage) that may limit the potential benefits of RWD analysis
- governance issues underlying the absence of standards for collaboration between stakeholders active in the field of RWD, and limitations of incentives for data sharing
- privacy concerns expressed predominantly by clinicians and patients and binding data protection legislation which can be seen to restrict access and use of data.

These issues are being addressed – although in a somewhat uneven fashion – by current initiatives from both public and private stakeholders at the regional, national and European scale. For example, the issues of data quality are being tackled through European and international initiatives aiming to improve the standardisation of terminology. Elsewhere, the development of international research coalitions is facilitating knowledge and best practice sharing and accelerating the development of common frameworks that guide RWD collection and use. This contributes both to the improvement of data quality and to researchers' analytical capabilities. In addition, a strong push towards the development of electronic health records – and eHealth infrastructures more broadly – has been observed in some European countries (Nordic countries, France, Belgium, the UK) and has been actively supported by various EU funding programmes. Such initiatives offer great potential for the automated and routine collection of patient data. Finally, access barriers – which are related to both governance and data protection issues and mostly faced by private companies – are being overcome through the implementation of strategic partnerships

among stakeholders and the development of online consent management architecture. Engaging directly with academics on specific research projects, with physicians in exchange of technological and analytical services, but also with data vendors are examples of access strategies that have been extensively explored by private stakeholders in the healthcare sector.

The insights gained from the evidence can be summarised along the lines of a PESTL analysis, which explores the Policy, Economic, Social, Technological and Legal domains in relation to RWD. In each category we outline the main drivers, enablers, barriers and alternative approaches to use of RWD that have emerged from the analysis (Table I).

**Table I. PESTL analysis of the RWD landscape in Europe**

Area	Drivers	Enablers	Barriers	Alternatives
Policy	<p>European Commission (EC)’s push for the development of eHealth infrastructures and use of EHR</p> <p>EC’s drive for the creation of Pan-European datasets and improved interoperability</p> <p>National healthcare reforms aiming to greater efficiency in service management and provision.</p>	<p>EU funding instruments</p> <p>Regional and National data infrastructure.</p>	<p>EC’s data protection regulation</p> <p>Fragmentation of national approaches to health reform</p> <p>Disparities between national eHealth systems</p> <p>Governance issues regarding the design and implementation of RWD standards.</p>	<p>Reliance on data collected in countries with easiest rules for access</p> <p>Involvement in EU-funded research projects in partnership with relevant public and private stakeholders.</p>
Economic	<p>Resources constraints and need to develop efficient pathways to analysis</p> <p>Incentives for collaboration to pool resources</p> <p>Development of a market for data.</p>	<p>New synergies within the data value chain (eg with insurance companies)</p> <p>National authorities encouraging data input.</p>	<p>Fragmented markets presenting different characteristics</p> <p>Issues surrounding cost sharing for data access and use</p> <p>Conflicts of interest.</p>	<p>Routine collection of publicly available data</p> <p>Funding to academia for research in databases</p> <p>Participation in research-minded consortia to spread the cost of data</p>

Area	Drivers	Enablers	Barriers	Alternatives
				access and analysis Engagement in disease specific research projects with direct access to self-reported patient data.
Social	<p>Increased familiarity with sharing data</p> <p>Increased attention to the burden of a chronically ill and ageing society</p> <p>Enthusiasm for new cures for illnesses</p> <p>Willingness to access personalised health services.</p>	<p>Positive media coverage</p> <p>Interaction with stakeholders (eg rare disease groups)</p> <p>Practitioners care about improving outcomes for patients.</p>	<p>Increased suspicions about data use and potential breaches</p> <p>Privacy risks due to linking different datasets</p> <p>Regulation surrounding consent management</p> <p>Image problem of pharmaceutical companies or insurers.</p>	<p>Development of personalised and stratified health services offer</p> <p>Communication around the positive effects of RWD-based research.</p>
Technological	<p>Increased technological capabilities for data storage and analysis</p> <p>Increasing capacity to link distinct datasets</p> <p>Push towards standardisation of terminologies.</p>	<p>Machine learning, including natural language processing</p> <p>National/patient identifier systems</p> <p>Social media and apps for self-reported data collection.</p>	<p>Limits of analytical capabilities for the treatment of data</p> <p>Inconsistency of existing databases and limited development of data quality insurance standards.</p>	<p>Leveraging methods and tools developed in other sectors</p> <p>Exploration of the potential of apps/partnerships with device manufacturers.</p>
Legal	<p>EU level and national level debate on data protection, use and access.</p>	<p>Potential of using RWD to improve health services efficiency might influence existing regulation to</p>	<p>Privacy and data protection likely to be strengthened</p> <p>Ethical standards for research</p> <p>Fragmented standards</p>	<p>Efforts on transparency and ethical commitments</p> <p>Publication of RWD-based</p>

Area	Drivers	Enablers	Barriers	Alternatives
		facilitate data access	for access to databases.	research results.
		Technological advances reduce the burden of work for consent documentation collection.		

In this landscape, strategies that seek to optimise RWD access and use have to align interests of the three parties typically involved in healthcare: the organisations (payer, provider), the professionals (clinicians) and the patients. Indeed, a disproportionate advantage, or disadvantage between the three generally leads to slow adoption or refusal to change custom and practice. We therefore think that strategic partnerships between those stakeholder groups are key to defining better routes to access and improved use of data. A variety of collaborations can be developed to overcome existing barriers and facilitate RWD access and use, depending on the kind of data that is needed and the scope of their use. Those partnerships would rest on both non-monetary and monetary agreements and leverage a broad range of incentives at different levels of the health systems, from patient-level initiatives to collaborations with national health organisations.



## Acknowledgements

---

We gratefully acknowledge the very helpful contributions of Adam Heathfield and colleagues at Pfizer who supported our work throughout, contributing to the discussion through regular meetings and conference calls and identifying participants and relevant stakeholders for interview. We would like to thank Emma Harte and Jennie Corbett for the research support they provided throughout the project. We thank John Kevin Dean, Jonathan Grant, Rebecca Schindler, and Steven Wooding, for the very insightful comments and suggestions they provided on earlier drafts of this report in their role as quality assurance assessors. This work would not have been possible without the participation of seventeen interviewees who generously donated their time to inform the study.

This is an independent report commissioned and funded by Pfizer. The views expressed in this report are those of the authors alone and do not necessarily represent those of Pfizer. The authors are fully responsible for any errors.





## Abbreviations

---

COPD	Chronic Obstructive Pulmonary Disease
CPRD	Clinical Practice Research Datalink
EHRs	Electronic Health Records
EC	European Commission
EMA	European Medicines Agency
EMRs	Electronic Medical Records
EU	European Union
GP	General Practitioner
HES	Hospital Episode Statistics
ICD	International Classification of Diseases
NHS	National Health System
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
RCT	Randomised control trial
RWD	Real-World Data



# 1. Introduction: The rise of RWD in health and healthcare research

---

## 1.1. Origins and aims of the report

The promise of real-world data (RWD) has been discussed for many years and a number of analysts and academics have signalled its potential to contribute to improved health products and outcomes (eg Holtorf et al., 2008). Advances in computing allow us to collect, share, analyse and use large quantities of data routinely at a relatively low cost – as never before. The increased use of new technologies in the healthcare sector has changed the ways in which patient level information are collected, stored and used. There are currently approximately 500,000 different types of medical devices on the market (European Commission, 2012), and the development of intelligent medical products (active implantable medical devices, networked medical devices and so on) represents a tremendous opportunity to collect a high volume of patient-level information in real time (West, 2012). In this context, RWD can be used in conjunction with randomised controlled trials (RCTs) and other medical data to provide insights into real-world clinical outcomes. A range of stakeholders in health research, innovation and care delivery hope that the combination of laboratory data and RWD can be used to help develop more targeted drugs and to encourage better use of those drugs by clinicians and patients. Data relating to patient experience in using drugs and to the contexts and settings in which drugs are used could potentially play a role in the way that trials are designed and conducted, the processes of drug registration and post-marketing benefit risk assessment, as well as create novel incentives for open health research (Eichler et al., 2012; Eichler et al., 2013). In the context of the severe cost and productivity challenges that health researchers and innovators have experienced in recent years, the prospect of data and mechanisms that could improve efficiency at multiple levels of the health research ecosystem without the cost of clinical trials is welcomed by many.

Although the bulk of evidence concerning new medicines and interventions will continue to be developed using clinical trials, there is evidence that clinicians and patients are excited by the potential of RWD to make treatments more effective, and the prospects opened up by this development. However, because patient and clinician data is sensitive, hard to collect and difficult to interpret, relationships and data sharing mechanisms between suppliers and users of RWD need to be in place for the potential benefits of RWD to be delivered. Moreover, the organisational and institutional pathways to data collection are evolving and often unclear, and privacy norms are not yet fully established. A recent study by the Organisation for Economic Co-operation and Development (OECD) notes that although RWD standards are emerging at the European level, the development of health information infrastructures has

been limited so far by the technical and legislative challenges posed by data protection (OECD, 2013). Our study confirms the difficulties in agreeing norms and pathways that would allow not only for access to data but for data to be used efficiently and effectively, while respecting privacy concerns. For those reasons, the way in which the use of RWD will evolve is uncertain and the rate and direction of its use in health innovation and healthcare is intimately related to a number of policy issues.

With the move towards personalised, stratified and genetic medicine, and the digitisation of most administrative and clinical data, researchers are confronted with an increasingly vast amount of data produced every day by a plethora of data sources which do not necessarily have the potential to be connected across systems (Chataway et al., 2012). RWD datasets can be specifically designed (eg adherence data collected specially from patients homes via dedicated devices), but the vast range of data collection devices that are shared, or work on shared infrastructure is growing. The sources of data are therefore less 'dedicated' and require a greater degree of trust and cooperation between those who own, administer and analyse the data. Technological advances alone are likely to prove insufficient, and the need to commonly define clear standards for RWD access and use is becoming a priority of the research and innovation policy agenda at the European and national levels (European Commission, 2007; Stroetmann, 2011).

With this challenging and rapidly evolving context in mind, the study has set out to explore the RWD landscape in Europe and more specifically in the European Union (EU), in order to achieve the following objectives:

- construct an evidence base on the establishment and evolution of standards governing the collection and use of RWD and identify the different ways in which standards have been applied
- understand the factors that have enabled or limited access to and use of RWD
- learn lessons from the use of RWD in different contexts
- identify opportunities for increasing access to RWD and contributing constructively to standard setting in Europe.

The study therefore aims to identify projects in which RWD is already being used and also find ways to establish standards for the collection and appropriate use of RWD that will allow the list of practical RWD applications to grow. This report is structured as follows: Chapter 1 presents the technological and analytical advances that are used by organisations in the health and healthcare domain, enabling the generation, analysis and use of a growing pool of data to support medical research objectives and better patient outcomes. Chapter 2 seeks to understand what types of RWD are being collected in Europe and their uses, and it describes the main categories of barriers to collecting and using RWD. Chapter 3 focuses on strategies and innovations that have enabled researchers and other relevant stakeholders to overcome those barriers. Finally Chapter 4 discusses the study findings in relation to the strategies that have been implemented to overcome current challenges. It then analyses the Policy, Economic, Social, Technological and Legal domains in relation to RWD to identify outstanding challenges in those fields and reassess the role of the various actors of the European healthcare research landscape in the future development of RWD.

## Research approach

The approach takes into account the multiple viewpoints and sectors that were subject to the study and the need to review the wide evidence base and define potential future directions of development. The approach is broken down into five tasks (Figure 1) and comprises the following methodologies:

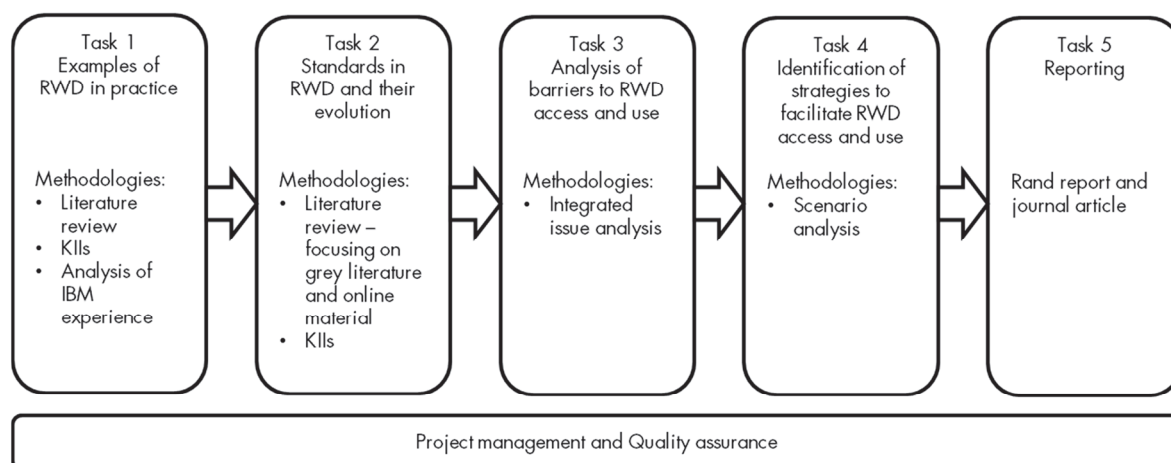
- (i) Firstly, we reviewed the available academic and grey literature through a structured approach covering academic and grey web-based databases. To ensure a wide coverage of the available academic literature, five different online academic databases were searched (Pubmed, EBSCO, Scopus, Opengrey and Google Scholar). After piloting a range of search term strings, the team opted to use very broad search terms to limit inappropriate selection and bias (the main search was a combination of 'real world data' or 'real world evidence' and European country names; while a complementary search combined 'healthcare' and 'database' to retrieve information on existing European databases). Search results were limited by language (English) and year of publication (after 2008). The study team screened a total of 935 sources for relevance to the questions investigated in the study and reviewed in depth a total of 43 articles. Other sources of evidence included documents retrieved through specific web sites and additional grey literature searches. The additional web sites included those of industry associations, websites of consultancies engaging with real-world data studies and the web sites of international organisations (such as the European Commission). Furthermore, the review also included references identified through 'snowball search', a process by which sources referred to in one or more relevant documents have been identified and retrieved by the researchers.
- (ii) Insights from the literature were complemented by information collected through 10 in-depth semi-structured interviews reflecting the views of additional stakeholders and policymakers. Interviewees were selected to provide insight about instances in which the collection and use of RWD has contributed to demonstrating or increasing the value of treatments and care, as well as those in which standards (or lack thereof) have influenced research opportunities. The selection of interviewees has been supported by existing networks of Pfizer and RAND.<sup>1</sup>
- (iii) Thirdly, 22 case studies, supported by seven interviews, were developed as a background to the current situation regarding the use of RWD in healthcare and other sectors, drawing on IBM's expertise and track record with RWD applications. Case study selection aimed to provide good coverage of the pharmaceutical value chain and the integral healthcare ecosystem, with a focus on European initiatives. In addition, a number of retail, automotive, banking and insurance cases were provided where additional insights could be extracted for the pharmaceuticals sector.
- (iv) Bringing together the insights from the first two sets of activities, the study team then performed an integrated issues analysis to interpret these outcomes, and identified the key barriers to RWD collection, use and access. This analysis also fed into the thematic analysis of the Policy, Economic, Social, Technological and Legal (PESTL) domains in relation to RWD.

---

<sup>1</sup> It was not within the scope and budget of this study to interview representatives of all stakeholder groups involved in RWD-based research. For instance, we did not seek to interview representatives of user/patient groups. It was also not within the scope and budget of this study to interview representatives from every European country.

- (v) Finally, we further developed a set of possible approaches to improve access to and use of RWD, through a scenario workshop where researchers worked closely with staff from Pfizer to explore the different potential routes to increase the collection of RWD and facilitate access and use. This exercise fed into the discussion and the PESTL analysis.

**Figure 1 Summary of research approach**



Details regarding the different methodological approaches can be found in Annex A, the literature review search strategy is outlined in Annex B, the list of interviewees’ affiliations can be found in Annex C, and the set of scenarios is described in Annex D. Findings have been analysed across workstreams and synthesised within Chapters 1–4 of this report. Contributions of interviewees have been made pseudonymously, with identifiers (Int1, Int2, etc) used throughout the report to reference evidence from interviews.

## 1.2. What is real-world data?

In 2007, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) created a Real-World Data Task Force, with the aim of developing a framework to assist healthcare decision-makers in dealing with RWD especially in relation to coverage and payment decisions (Garrison et al., 2007). According to the task force, real-world data can be defined as:

Data used for clinical, coverage, and payment decision-making that are not collected in conventional randomised controlled trials (RCTs).

This type of data includes data from existing secondary sources, such as the databases of national health services. It also takes in the collection of new data, whether carried out retrospectively or prospectively (eg disease registries, medical records). In order to make this wide-ranging classification explicit, we use the following working definition:

RWD is any data not collected in conventional randomised controlled trials. It includes data from existing secondary sources (eg databases of national health services) and the collection of new data, both retrospectively and prospectively.

While the above definition means that some types of data (such as all types of data recorded in controlled trials) are explicitly excluded from the category, there is a wide range of sources and indicators that are included under the umbrella term. These inform different research and application areas, ranging from social media to physicians' notes and from genomic data to health insurers' data. Table 1 offers examples of categories of data and indicators included within these categories.

**Table 1 Examples of RWD**

Area	Content
Resource use	Cost, contacts with the health system, treatment
Health outcomes	Observational indicators, eg medical records, hospital statistics, insurance data
Patient behaviour	Compliance with treatment, outcomes, preferences (from reported data or social media)
Population health	Clinical data – both structured and unstructured content (eg dictation, medical history, labs, images) Events that can be linked to public health: weather, disease outbreaks, local events Physiological data from at-home monitors and bedside monitors and sensors
Sales, marketing and distribution	Data from insurance claims and pharmacies, digital marketing data, sales and product data
Longitudinal Patient Record	Medical history (diagnosis codes, physicians' notes, images), pathology, disease, drug information, environmental factors, product information, social media activities over time.
Consumer engagement and analytics	Blogs, chatrooms, patient communities
Health monitoring and intervention	Streaming data from monitoring devices, personal devices, apps



## RWD and big data

'Big data' has been defined as a 'term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning'.<sup>2</sup> As with RWD, then, big data inheres in unstructured datasets. While life sciences and healthcare research generally approach research questions through known structured data, unstructured datasets could be increasingly used in the near future. These developments could bring RWD and big data even closer. However, whilst the two concepts may appear interchangeable, it is useful to introduce a distinction: the focus of the definition of RWD is on the methods through which it is created (ie outside the framework of a RCT), but that of big data is a more operational one, focusing on the characteristics of the dataset and the analysis that it makes possible.

An example of a big data application relevant for the pharmaceutical industry is that of the Strategic Intellectual Property Insight platform (SIIP) database.<sup>3</sup> This cloud-based database aggregates worldwide patent data and scientific literature, with more than 30 million documents and over 200 million annotated chemical compounds, enabling insight into complex chemical and biological patents. When the concepts are compared there is some overlap. Data from several of the RWD sources – for example from personal or bedside monitors or social media activity – may also be seen as constituting big data, as they present large, relatively unstructured datasets that need support from specific software and rely on machine learning to discern patterns. However, other data sources included in RWD analytics include data collected over relatively small and structured samples and sets that do not necessitate big data-powered algorithms for efficient analysis. In this study, the focus will be on the smaller and structured datasets, but relevant insights from big data projects will not be excluded.

## eHealth infrastructures

The development of RWD generation, collection and analytics heavily relies on electronic systems and infrastructures. This has led to the term 'eHealth' – an umbrella term describing the use of such structures. As Marconi (2002) puts it, the term 'eHealth' describes the 'application of Internet and other related technologies in the healthcare industry to improve the access, efficiency, effectiveness, and quality of clinical and business processes utilized by healthcare organizations, practitioners, patients, and consumers in an effort to improve the health status of patients.' eHealth comprises institutional structures, data architecture systems, competence centres and legal frameworks. Competencies of eHealth authorities cover a range of eHealth instruments, including ePrescriptions, telehealth and patients' electronic health records (EHRs) systems (Stroetmann et al., 2011).

---

<sup>2</sup> Jonathan Stuart Ward, Adam Barker Undefined By Data: A Survey of Big Data Definitions  
<http://arxiv.org/abs/1309.5821>

<sup>3</sup> IBM BAO Strategic IP insight platform (SIIP), <http://www-935.ibm.com/services/us/gbs/bao/siip/>. Last accessed 19/01/2014.

## Electronic health records

In this report, we adopt the OECD definition of electronic health records (EHRs): an EHR is ‘the longitudinal electronic record of an individual patient that contains or virtually links records together from multiple Electronic Medical Records (EMRs) which can then be shared across health care settings (interoperable). It aims to contain a history of contact with the health care system for individual patients from multiple organisations that deliver care’ (OECD, 2013).

## Ownership structures and data pooling – two key characteristics of RWD

RWD can be found in a variety of forms and the data are embedded in multiple sources. As a consequence their generation, analysis and sustainability rely on a broad range of stakeholders who own and curate the data or otherwise interact, compete or collaborate with each other (see Table 2 for examples of stakeholders involved in data management).

**Table 2 Examples of responsibility for data management**

Source	Type of data	Content	Party responsible for managing the data
General practice records	Primary	Medical history, prescriptions, outcomes, symptoms, diagnoses	GPs/National health systems
Electronic health records	Primary	Pathology, treatment, outcomes, symptoms, diagnoses, lab results, imaging	Hospitals; National health systems
Point of care records	Primary	Resource use, pathology, treatment.	Hospitals/GPs/National health systems
Disease registry	Primary	Patient characteristics, diagnoses, medical history, treatment, adherence	Disease registry/ National health systems
Pharmacy records	Primary	Prescription, adherence	Pharmacies/National health systems
Social media, blogs, chat rooms, patient communities	Primary	Pathology, treatment, adherence, outcomes	Individual patients/platforms
NHS databases	Secondary	Patient characteristics, cost, resource use	National Health systems
Health insurer databases	Secondary	Patient characteristics, cost, resource use, outcomes, treatment, claims	Private or public health insurers

Source: IBM, 2013.

Ownership structures have the potential to define the setup of possible collaboration structures regarding data sharing and use. For example, a post-marketing study conducted by Franchi et al. (2013) demonstrated (albeit on a small scale) that it is feasible to pool administrative and clinical data on epilepsy from different sources to ensure long-term follow-up of patients with the condition. Their study analysed data from hospitals (including hospital discharge diagnoses), prescription records from pharmacies, prescription records for diagnostic tests and other data collected for the national health system. In this instance, sharing and combining many previously siloed data pools is likely to improve partner collaboration and partnership outcomes between different actors within the healthcare ecosystem (IBM, 2013). Further examples of data, data use and opportunities for data pooling are presented in Table 3.

**Table 3 Areas for RWD use in health and healthcare**

Use case	Description of use case	RWD leveraged
<i>Life sciences</i>		
Drug development	In an area traditionally dominated by randomised control trials, RWD analysis can be used to assess the efficacy of different medical treatment and inform drug development strategies.	Medical history, demographics, pathology, regulatory filings, product information.
Post-market studies	RWD analysis improves understanding of safety and effectiveness of drugs and devices once they are on the market. It uses large sets of post-market observational health data to gain insights into diseases, products and patient populations, in areas such as health outcomes research, drug effectiveness, and drug safety.	Outcome data (ie hospital (re-)admissions, mortality rates), adherence, monitoring device data, pharmacy data.
<i>Healthcare</i>		
Healthcare service delivery	Decisions at the individual and system levels increasingly incorporate evidence from RWD analytics. RWD is used to support personalised decisions on treatment options and healthcare delivery strategies, building on healthcare coverage, quality and costs analysis.	Service utilisation data, treatment uptake rates, treatment outcomes, insurer data, diagnoses, treatment, cost data.
Longitudinal patient record	The ability to enable a healthcare provider to pull information about a patient from multiple sources.	Medical records, imaging, lab results, drugs, treatment, service utilisation data, diagnoses.
<i>Customer insights</i>		
Claims and premium analysis	Insurance companies collect and leverage RWD in a variety of ways. Some of the most relevant uses of data analytics for the insurance sector as a whole include risk management and the fight against fraud; and these companies build on RWD to support payment	Insurer data, medical history, outcomes data, cost data.

Use case	Description of use case	RWD leveraged
	decisions.	
Customer insights	Increasing understanding about the customers requires the ability to analyse the customers' (healthcare provider, patient) needs with regard to health, treatment, education, finances and decision-making.	Sales data, marketing data, cost data.

In this report we focus on life sciences and healthcare uses of RWD as they are relevant to a broader range of stakeholders, including public and private research institutions, payers, care providers and services users (ie patients).<sup>4</sup>

---

<sup>4</sup> Cases of RWD use for business development and customer insights purposes can be found in Annex C, with examples from different sectors.



## 2. Moving from data to evidence: the state of RWD in Europe

---

The literature review and the analysis of the industry's case studies identified the areas in which RWD use has been pioneered in the health and healthcare sector. These areas include drug development (including post-market phase), and improved healthcare delivery via assessment of the evidence. In this chapter we summarise these insights, which illustrate some of the potential that RWD use could have in innovation for the health and healthcare sector. We also present the main boundaries and limitations of RWD-based approaches, which emerged from the interviews and the integrated issue analysis.

### 2.1. RWD is useful for building the evidence base for drug development and post-market studies in the pharmaceutical and medical sectors

This review has found solid evidence on the practice of using RWD to assess the (short- or long-term) impact of different drugs or medical treatments after their introduction in the market, and somewhat less evidence for its use in product development.

#### Using RWD for drug development

RCTs have traditionally been the preferred setting for product development in the healthcare industry. However, RWD can also be used to assess the efficacy of different medical treatments and inform drug development strategies. For instance, a research team including researchers directly affiliated with GSK and Novartis studied the relative efficacy of drugs used to treat Chronic Obstructive Pulmonary Disease (COPD) and the relationship between these data and the results of clinical trials, using data from the Optimum Patient Care Research Database (Price et al., 2013). Another study, sponsored by Novartis and called the European Cubicin Outcomes Registry and Experience project (EU-CORESM), is gaining access to a registry that gathers data from 118 institutions. The study considers the characteristics of the patient population and the relative efficacy of treatment for skin and soft tissue infections (Gonzales Ruiz et al., 2011). The findings will also be used by the company for in-house research, going beyond the scope of the EU-CORESM study.

#### Using RWD for treatment evaluation

##### *Analysing the long-term outcomes of an intervention*

An example of medium-scale project using a disease-specific database can be found in a study using data from 870 patients to assess long-term outcomes of transcatheter aortic valve interventions (TAVIs) based on the UK TAVI registry, which has been set up to capture the outcomes of all such procedures executed

in the UK (Moat et al., 2011). The study, one of the first of its kind to concentrate on a mid-to long-term time horizon, tracked survival and mortality rates for the interventions at 30 days, 1 year and 2 years after the event. It found that while a substantive proportion of these high-risk patients were deceased within the first year, overall the survival rates were encouraging.

#### *Analysing the long term effects of drugs*

Several of the cases examined for the study are concerned with post-market drug risk assessment. Such cases have used RWD to gain an in-depth understanding of specific issues, including the long-term effects of different treatment options on a determined patient group, such as those registered in a disease-specific registry.

We also identified occasions in which studies have drawn on large national datasets to assess the impact of drugs or medical treatments. These included a Danish study that evaluated the net clinical benefit of new oral anticoagulants versus no treatment in a 'real world' atrial fibrillation (AF) population (Banerjee et al., 2012). The study used a long-term database covering all Danish patients discharged with AF over ten years (between 1997 and 2008), looking at patients' clinical histories, including pharmacotherapy, and premorbid risk stratification scores for stroke/thromboembolism. The analysis was further facilitated by linking the existing dataset to the unique personal identifier and Danish biobanks in order to assess the effects of three drugs compared to usual treatment and inform healthcare decision making. Studies at the national level can, then, draw on databases linked across multiple identifiers and databases (depending on the maturity of the e-infrastructure of the individual countries).

There are also examples of cross-border initiatives offering added dimensions by including a cross-national set of patients. The studies building on the EU-ADR database for example used eight databases<sup>5</sup> in four European Countries (Denmark, Italy, Netherlands and the UK) where both clinical information and drug prescriptions are recorded for large-scale drug safety monitoring. The database contains information about 30 million patients. The studies looked at drug safety across a range of diseases including acute myocardial infarction; acute renal failure; anaphylactic shock; bullous eruption; and rhabdomyolysis (Coloma et al., 2011; Coloma et al., 2012).

A further example is supplied by the VAESCO<sup>6</sup> project which has supported studies in the areas of vaccine safety surveillance. This study involved seven databases from European countries (Italy, Spain, Finland, UK, Sweden, Norway, Denmark and the Netherlands), covering at least 26.67 million patients.<sup>7</sup> Its aim was the development of vaccine safety and best practices, evaluation of strategies and new methods

---

<sup>5</sup> Health-Search (HSD, Italy); Integrated Primary Care Information (IPCI, Netherlands); Peditnet (Italy) and QResearch (United Kingdom) are general practice (GP) databases; Aarhus University Hospital Database (Denmark), PHARMO (Netherlands), and the regional Italian databases of Lombardy and Tuscany are record-linkage systems where drug dispensing data is linked to their medical registries as well as registries of hospital discharge diagnoses.

<sup>6</sup> Vaccine Adverse Event Surveillance and Communication (VAESCO), <http://vaesco.net/vaesco.html> Last accessed 10/04/2014.

<sup>7</sup> Health Search Database (HSD); General Practice Research Database (GPRD); Norwegian Patient Register, Danish Civil Registration System (DCRS); Integrated Primary Care Information Project (IPCI); National Health Registers in Sweden; the Finnish Care Register for Health Care (HILMO); BIFAP; Sicily Regional Database (Italy).

and to facilitate data collection through common aims and standards, and to provide information on vaccination safety.<sup>8</sup>

## 2.2. RWD is increasingly used to improve healthcare service delivery

Healthcare delivery decisions at the individual and system levels increasingly incorporate evidence from data analytics. In some cases RWD is used to support personalised decisions on treatment options. These options are then tailored to the patient's specific genotypic characteristics and outcome probabilities. In other examples, data are incorporated into studies investigating questions related to health services coverage, quality or costs with a view to informing national or regional healthcare delivery strategies.

### RWD for assisting doctors and patients in choosing between treatment options

In some cases RWD and big data analytics are synthesised for initiatives involving personalised medicine and which require treatment decisions to be based on the individual characteristics of the patient. These cases use large datasets on treatment outcomes. In the EuResist project, algorithms processing genotypic information across a multinational database are used with other genetic and response indicators in order to determine the best course of treatment for individuals with HIV infection (Table 4). The initiative aims to develop a system capable of predicting how patients are likely to respond to a specific method of treatment and consequently recommend a certain treatment out of a portfolio of options. In pursuit of this aim, the project builds on databases of genotypic information, which are combined with data on drug resistance.

---

<sup>8</sup> VAESCO, "About Us," <http://vaesco.net/vaesco/about-us.html>, accessed 04/12/2013 Population figures for Finland, Spain and Sicily are not obtainable from the project documentation. While some of the results and data from the project are available through the website, data from the individual countries continues to be subject to authorisation by the relevant national organisations.



**Table 4 The EuResist initiative**

Case description

The EuResist project builds on the collaboration between pharmaceutical companies, governmental institutions, private companies and other partners, including a European Economic Interest Grouping (Karolinska Institutet Sweden; Max Planck Institute for Informatics, Germany; University of Siena, Italy; Informa s.r.l., Italy; Cologne University, Germany). The project's new technologies and mathematical models aimed to provide a more efficient way to choose the best drugs and drug combinations for any given HIV genetic variant. It created an online system that helps doctors to choose a HIV treatment with the highest probability of halting virus replication and impairing the evolution of drug resistance, building on the following objectives:

- Integrate biomedical information from three large genotype-response correlation databases, thus collecting the required critical mass of data on the clinical implications of HIV drug resistance.
- Develop and validate a number of different engines for effective prediction of the response to treatment based on the integrated biomedical information.
- Combine the different engines into a predictive system and make it publicly available on the internet, with a sponsor-based exploitation plan.

EuResist is the first freely available data-driven computational method that predicts the success of a treatment regimen against any given HIV genotype, based not only on viral genotype information, but using analytic technologies to take into account treatment response information from clinical practice. It is also the only system providing the global medical community with an estimate of activity for combination therapy, rather than for individual drugs.

Type of real-world data used

- New mathematical prediction models have been developed to use both the patient's own history and the wealth of information that EuResist researchers have amassed. The recent expansion of the EuResist Integrated Data Base (EIDB) to include information from more than 60,000 patients, 150,000 therapies and 500,000 viral load measurements, makes it the world's biggest database centred on HIV resistance and clinical response information.
- Prediction models analyse information from large genotype-response databases: the ARCA database (one of the biggest in the world, based in Italy), AREVIR database (Germany) and data from the Karolinska Infectious Diseases and Clinical Virology department (Sweden).

Benefits of real-world data use

- The system's predictions are nearly 76 per cent accurate, outperforming other commonly used HIV resistance prediction tools, as well as human experts in the field.
- To simulate real practice in HIV-specialised care, the EVE (Engine versus Experts) study compared EuResist with 10 international experts who were confronted with 25 case histories, where all the clinical and virological information was available. EuResist's predictions outperformed nine out of ten human experts.

In a similar initiative targeting rare types of cancer (Table 5), genotypic information is analysed together with treatment outcomes data to support decisions on the course of treatment, which can then be tailored to the individual characteristics of the patient.

**Table 5 National Cancer Institute of Milan: analytics and big data to improve cancer treatment**

Case description

A collaborative project between IBM and the National Cancer Institute of Milan has been using genomics and analytics technology to improve the treatment of rare tumours, sarcomas and cancers of the head and neck.

New clinical genomics analytics technology provides medical personnel and healthcare management with a broader overview of the treatment processes performed and their appropriateness, as well as insight into the effects of the care – both in terms of clinical efficacy and economic viability.

The system analyses the individual characteristics of the patient and the specific profile of the disease, associating this information with knowledge derived from the analysis of previously treated clinical cases and specific guidelines as defined by the Rete Oncologica Lombarda (ROL) oncological expert panel.

Types of real-world data used:

- clinical data
- genomics
- guidelines
- patient data.

Benefits of real-world data use:

- The use of RWD has facilitated the development of personalised care services.
- Better patient outcomes have been achieved.
- The institute is looking to expand the project to collect data, analyse practices, and share deep insight with other interested centres in the Lombardy Oncology Network — a network of 14 medical centres in the Lombardy region and the National Rare Cancer Network – a cross-Italian national effort to improve treatment for patients with rare cancer.

### Analysing RWD to optimise the efficiency of healthcare services delivery

RWD analytics are particularly useful in supporting innovative ways to improve and optimise healthcare delivery. One potential area for innovation is that of expanding the potential range of healthcare services by aggregating data for decision-support and supporting telemedicine, as illustrated by the strategy for home care implemented in Southern Denmark (Table 6). In this case, the new system was set up with the aim of improving outcomes for chronically ill patients. The strategy includes linking data across healthcare databases to create a holistic view of each patient; but also creates a platform that can integrate data from home monitoring and telemedicine applications and offer access to different healthcare

professionals that can use the data to support their decisions. Furthermore, the automation of processes supports trends toward process optimisation and an efficient use of time, while the business intelligence and analysis potential of the linked database may offer commercial value to the region.

**Table 6 The Shared Care platform: using RWD to improve the efficiency of homecare delivery in Southern Denmark**

#### Case description

Today, 80 per cent of expenses in the Danish healthcare system relate to chronic illness, and estimates indicate that Denmark has a productivity loss of between 3 and 5 billion Danish kroner per year due to work absences related to chronic illness. These patients are often seen by many different professionals within the healthcare, social services and community ecosystem.

The Region of Southern Denmark (RSD) intends to improve the quality and comprehensiveness of care for patients with chronic illness and enable real-time communication among patients, physicians, pharmacists, mental health professionals and specialists so that all parties have better insight into patient care plans.

RSD's programme includes two types of patients:

- Group 1: disadvantaged patients with a limited ability for self-care. Currently patients have to keep all the involved care providers informed with details about their illness. With the new platform, the care providers are automatically kept abreast of relevant information.
- Group 2: patients with the ability for self-care. They are active and interested in taking responsibility for their illness, but until recently they did not have the right tools. Now they have the possibility to gain insight into their treatment and, together with their care team, gain a more complete overview of their situation.

The programme is supported by the development of the Shared Care platform. It includes a range of tools that support cross-sectoral communication and integrate telemedicine and data from home monitoring in the clinical process. It reaches stakeholders from across the healthcare sector, ranging from general practitioners, to municipalities, hospitals, specialists and the patient.

#### Real-world data used

- Integration of data generated by all actors in the healthcare system: general practitioners, municipalities, hospitals, specialists and patients.
- Integration of telemedicine and data from home monitoring.

#### Benefits of real-world data use

- The consultancy firm McKinsey & Co concluded in 2010 that the region had spent DKK 2.1 billion on eHealth in 2009 and that the savings potential through coordination of ICT efforts and through consolidation could amount to DKK 330–490 million annually.
- An analysis from 2012 by the consultancy firm Boston Consulting Group concluded that digitalising communication with patients could release DKK 174 million annually at

regional level.

- Analyses by the Regional eHealth Organisation indicate that a cross-section of the regional ICT projects could release resources for patient treatment amounting to DKK 150 million annually.
- Fully deployed telemedicine ulcer assessment is expected to release up to DKK 262 million annually in municipalities and DKK 45 million in regions.

In other cases RWD analytics has been leveraged to optimise current processes in healthcare delivery and limit associated costs, for instance by reducing the number of UK patients that have to be readmitted to hospital with Chronic Obstructive Pulmonary Disease (COPD) following their discharge (Table 7). In this case, the computing assets of the NHS enabled the analysis of multiple types of standardised patient and treatment data. The analysis supported the optimisation of the treatment process for patient outcomes and cost implications for hospitals.

**Table 7 Partnership to reduce readmissions for COPD**

Case description

Chronic Obstructive Pulmonary Disease (COPD) exacerbations requiring inpatient admission have a significant impact on patients' quality of life, mortality risk and resource utilisation.

The UK National Health Service (NHS) is limiting reimbursement for readmissions for COPD occurring within 30 days of discharge. Reviewing the frequency and impact of such admissions is an important activity for decision makers in the NHS and pharmaceutical companies. As a result, the NHS, in partnership with a pharmaceutical company, hospitals and IBM, has sought to develop a system aiming to facilitate RWD use to support the optimisation of the treatment process for patient outcomes and limit the readmission rate for the same reasons (and associated costs).

Type of real-world data used

An anonymised version of admissions data from English hospitals submitted to the NHS (Hospital Episode Statistics (HES) data). Data include:

- diagnoses (ICD10)
- procedures (OPCS4)
- Healthcare Resource Grouping (HRG)
- admission and discharge dates
- any critical care information
- details of patients, hospitals, consultants and general practitioners (GP)
- visits to specialist (ambulatory, outpatient).

Benefits of real-world data use

- improve patient outcomes

- reduce readmissions
- reduce cost of healthcare.

Studies assessing healthcare delivery also include research investigating the evidence on the uptake of existing services and their delivery, for instance the effect of uptake of cardiac rehabilitation (CR) treatments on survival. Working with the database hold by private insurance Achmea (Table 8), Van Engen et al. (2013) looked at this relationship in the Netherlands, and were able to demonstrate that despite the efficacy of the CR treatment, most Dutch patients did not receive this type of care. Furthermore, the data analytics allowed the researchers to make recommendations about populations that should be specifically targeted by CR treatment initiatives, such as women, patients with long travelling distances to the nearest CR provider and patients with comorbidities.

**Table 8 Achmea Health Database**

Geography	Netherlands
Type	Records on payments on the provision of all medical care to its patients
Sample size	1.2 million
Content	1) Data on patients (lack clinical information eg on tests outcomes; adherence; contain only demographic patient data and only most recently diagnosis data) 2) Data on practitioners 3) Data on health services provided – outpatient diagnostic and therapeutic provision: - prescription drugs - dosage - costs
Use	Drug utilisation, disease course, diagnostic and prognostic studies, effects of (drug) treatments; studies with long follow-up time (if patient does not switch); etiologic studies (relationship between risk factor and outcome).
Advantages	High reliability (economic motivation, very well audited) but limited by specific content; long follow-up time; potential to link clinical and patient data Linked to other databases, eg regional and national registries on cancer/death/psychiatric conditions (virtual Mondrian project for linking is in progress).
Access	Database developed for in-house use and shared with research organisations

RWD has also supported studies focusing on evaluating the quality of care. Franzke et al. (2009) collected data on patients with acne. This research allowed the capture of raw data and aspects of subjective patient experience and socioeconomic factors reported by patients with acne vulgaris. The data were then used to analyse the patients' trade-offs in choosing between doctor prescribed medication and the acquisition of medical products through self-medication.

## Assessing the cost-effectiveness of medical treatments using RWD

RWD is also being used to inform decisions related to the burden of medical treatment costs. For example, the databases maintained in the Swedish national and Italian regional healthcare systems have been used to assess the burden of costs related to cardiac diseases or cancer (eg Lothgren et al., 2013; Roggeri et al., 2013). Lothgren et al. (2013) have simulated cost implications per patient and examined the budget implications of different drugs used by patients with bone tumours. While the researchers could not directly access the relevant data, they triangulated available sources to estimate cost burden per patient and at the system level in Austria, Sweden and Switzerland. They thereby determined the drug with the lowest administration and collateral costs. Roggeri et al. (2013) used a set of Italian regional databases (administrative databases of seven local healthcare units located in four different regions: Veneto, Toscana, Abruzzo and Puglia), linked with socio-economic datasets to assess the direct healthcare costs and resource needs associated with acute coronary events. Included in the study was information on demographic characteristics, prescriptions of drugs reimbursed by the national health system, hospital discharge records, outpatient visits and diagnostic-therapeutic procedures.

## 2.3. Factors that influence access to and use of RWD

As demonstrated by the examples in the previous sections, there is strong interest for RWD in the health and healthcare domain and researchers are using RWD to address a variety of issues. However, research has also shed light on a series of barriers that may have an impact on the future use of technology and analytic capabilities (ABPI, 2013) and prevent researchers from exploiting the full potential of RWD. The lack of shared standards with regard to content and quality of the data, methodological challenges, and the lack of shared standards with regard to governance structures and privacy practices constitute significant concerns shared among most industries and public sector bodies engaging in RWD analytics. Those categories of barriers are described in turn.

### *2.3.1. Standards defining the content and quality of RWD are yet to be adopted*

#### Common terminology standards are still lacking

The development of coding and language standards to refer to specific medical conditions shapes the ability of various stakeholders to use and compare data. However, such development remains at an infant stage in the EU. For example, one interviewee (Int10) reported that in the field of rare diseases less than five per cent of diseases have a unique code that can be used in health records. In order to ensure all rare diseases have a unique code, an update of the international classification of diseases (ICD) is required. This classification and the associated coding system would then need to be incorporated at the national level for the full potential of the data to be realised.

#### Existing data remains incomplete

Data are still often fragmented and siloed within and across organisations, reducing the visibility and research potential of datasets. Such issues have been raised even with regard to some of the most developed databases, including the Swedish national registries. According to Friberg et al. (2012), some of

the registries lack granularity: the National Swedish Drug Registry, for example, does not differentiate between different types of anticoagulants. In another Swedish study, Lothgren et al. (2013) note that certain types of events are not reported in registries, and that event rates have to be extracted from RCT data.

Similarly, Augustin et al. (2010) investigated the co-morbidity and age-related prevalence of patients with psoriasis, in a study based on the database of a German nationwide statutory health insurance scheme. The authors emphasised the limitations resulting from the particular setup of the database, which in this case did not allow for correction for confounding factors, such as smoking. However, they also underlined the fact that these limitations may limit the interpretation of the data, but not its accuracy.

### Data quality assurance systems remain underdeveloped

The trustworthiness and usability of data in supporting research and decision making are affected by inconsistent data quality. Harmonised strategies to facilitate the development of data quality standards are crucial to address current content limitations of available datasets (eg comorbidity, genetics, details on treatment, resource requirements). These limitations include (Int10; Avillach et al., 2013):

- consistency of data entry
- coding errors
- discontinuities in data collection (missing data points)
- inaccuracy (eg misdiagnosis).

### 2.3.2. Methodological barriers are limiting efficient use of RWD available

Several methodological obstacles still make it difficult to collect and use RWD efficiently.

#### Analytic capabilities are limited

Interviewees emphasised the strong disparities existing in Europe regarding analytic capabilities. The technology to collect and analyse these huge pools of data exists but the absence of a common methodology still makes it hard to exploit its full potential.<sup>9</sup>

#### Dataset linkage poses methodological challenges

Data linkage is a challenging and time-consuming activity (Coloma et al., 2011). However, integrating different data pools, including medical datasets (phenotypic data and genomic data), socioeconomic datasets or information on patient behaviour and preferences, for example, is crucial for research (eg to analyse the effects of socioeconomic conditions on medical treatments and mortality).

#### A fragmented Europe

Today's challenges of accessing, compiling and integrating data from often siloed sources are partly technological in nature, but they are mainly due to other factors. These include the legacy of national

---

<sup>9</sup> The lack of analytical expertise and standards can also constitute a barrier to drawing correct conclusions from RWD (see Morgan Jones et al., *forthcoming*); however this argument was not explicitly voiced in the interviews,

health systems, the presence or absence of coherent strategies regarding patient-level data collection and use (including codified and transparent standards) and the existence of coalitions/partnerships pushing in that direction.

Most interviewees highlighted the issues of fragmentation in the European Union and the difference between more digitalised countries and those lagging behind when asked about the main factors impeding access to and use of RWD in Europe. The 28 national member states all have health systems, data collection and data storage practices that operate differently.

### *2.3.3. Governance structures are influencing access to data*

#### **There is a lack of clear pathway to access**

The fragmentation of data sources and the plethora of stakeholders involved in the generation, control and analysis of RWD constitute a challenge for RWD access (Int3). From public administrations collecting patient- and provider-level data, to private companies – such as medical devices companies – collecting data for internal use, data controllers<sup>10</sup> play specific roles in the health and healthcare ecosystem, and therefore may have different incentives to collect, process and release their data. Besides, given that there is only limited standardisation of dataset access rules, there is no such thing as a typical access pathway. Access to RWD and the type of data available is therefore highly dependent on the kind of interactions that are developed between stakeholders (data controllers and data processors) and the governance mechanisms that shape these interactions. This highlights the need to incentivise individuals or organisations in charge of uploading the data, eg the care providers.

#### **Data access is often only granted to academic researchers**

Most databases are made accessible to academics upon request, but limit access to other groups, and the industry in particular. Similarly, access to some RWD sources that are currently being developed might remain dependent on the involvement of academic researchers, as this is believed to be the case for data curated by the newly created Farr Institute<sup>11</sup> in the UK (Int8).

#### **Lack of awareness among professionals can be detrimental**

In some instances, RWD initiatives can be hampered by lack of awareness or engagement amongst healthcare professionals (data controllers). The relatively low uptake of the French personal health record (the Dossier Médical Personnel) highlights this issue: it took six years of research and piloting for the French EHR to be ready for the large-scale implementation phase which started in 2011. By 2013, two

---

<sup>10</sup> A data controller can be defined as follows: “A data controller is the individual or the legal person who controls and is responsible for the keeping and use of personal information on computer or in structured manual files.” (Data Protection Commissioner, “Are you a data controller?”, <https://www.dataprotection.ie/docs/Are-you-a-Data-Controller-/43.htm> Last accessed 13/03/2014.

<sup>11</sup> The Farr Institute is described in Section 3.2.



years after the official national launch, only 412,137 records have been created<sup>12</sup> which falls below the target of the first year (500,000). Reported reasons for the slow uptake include a lack of political visibility and continuity (Cour des Comptes, 2012), and insufficient awareness campaigns targeting patients and professionals (Santi, 2012).

#### *2.3.4. Privacy concerns are limiting the amount of data available and the scope of their use*

##### Ethical concerns amongst health professionals

Ethical issues and the fear of seeing their patients' private information exposed to misuse have sometimes prevented GPs from sharing data with both public and private entities (Int3). For instance, one of the interviewees mentioned the failure of a publicly funded initiative that aimed to build a national electronic record system in the Netherlands to facilitate patient level information exchange between different care providers (GPs, hospitals, pharmaceutical and insurance companies) (Int7). The government had to abandon the initiative as GPs and patients groups were firmly opposed to it for data privacy reasons. More recently, the NHS had to postpone its 'care.data' scheme – a major infrastructure development aiming to collect and share (for scientific and commercial use) information from GPs and surgeries as well as information from hospitals and genomic information. The postponement followed concerns from patients' groups and GPs (represented by the Royal College of GPs) as well as the British Medical Association (BMA), regarding privacy and consent management issues. These organisations requested to delay the launch of the platform to give patients more time to evaluate and consider opt-out options.<sup>13</sup> Medical information is automatically collected as patients are enrolled into the system by default in the UK. This means that their personal medical data are automatically collected by the physicians unless they explicitly refuse the collection of this information. Such information is stored by the GPs who then decide whether to share these data or not. If the practitioner decides to share it, patients also have to be informed of that choice (Int7).

##### Medical data protection: a major concern for the public and European regulators

The public can be particularly sensitive to the risk of data breaches, especially when it comes to medical data<sup>14</sup> which have been obtained in the framework of an intimate and confident relationship (Westin, 1976; Gosting and Hodge, 2002). This sensitivity was mentioned by several of our interviewees as a major obstacle in accessing RWD and was thought to have led to a medical data regulation that is more constraining than the regulation of other types of personal data (OECD, 2013; Int3; Int8). At the European scale, the General Data Protection Regulation (the 'Proposed Regulation') was adopted by the

---

<sup>12</sup> Dossier médical personnel, <http://www.dmp.gouv.fr/> Last accessed 09/12/2013.

<sup>13</sup> <http://www.theguardian.com/society/2014/feb/21/nhs-plan-share-medical-data-save-lives> Last accessed on 24/02/2014.

<sup>14</sup> European Barometer, Attitudes on data protection and identity in the European Union. [http://ec.europa.eu/public\\_opinion/archives/eb\\_special\\_359\\_340\\_en.htm](http://ec.europa.eu/public_opinion/archives/eb_special_359_340_en.htm) Last accessed on 09/12/2013.

European Parliament on 21 October 2013. This new legislation is a revision of the 1995 Data Protection Directive,<sup>15</sup> and aims to facilitate the harmonisation of national data privacy regulations. If approved in its current form by the European Council and the European Parliament in April 2014, it will probably be introduced in national legislation by 2015. The General Data Protection Regulation could lead to more stringent rules in countries that have adopted 'liberal policies' towards personal data use without explicit prior consent (principally Nordic countries and the UK but also Belgium and Italy (OECD, 2013)). The new regulation is likely to limit medical data collection and processing (Ploem et al., 2013; Int8). The amendments<sup>16</sup> to articles 81 and 83 of the draft regulation are key, as they remove the clause that allows for exception from consent in the use of identifiable medical data for medical research (Frears et al., 2013). In its current form, the new regulation states that the 'processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes shall be permitted only with the consent of the data subject' (art. 81.2).<sup>17</sup> Moreover, article 83 imposes tougher standards for personal data processing (Box 1). Several interviewees mentioned the changing privacy legislation as a potential obstacle to the use of, and access to, RWD for scientific research. Similarly, several research groups, institutes and foundations (eg the Economic and Social Research Council, the Wellcome Trust and the Medical Research Council in the UK; the Inserm in France) have expressed concern regarding the negative impact of these new restrictions on the progress of medical research and urged the European Parliament to re-introduce the exception of explicit consent for research purposes (Rabesandratana, 2014).

---

<sup>15</sup> Directive 95/46/EC.

<sup>16</sup> The text voted by the European Parliament in October 2013 is a revision of the Commission's Draft Regulation Proposal of January 2012. Amendments in the current text were introduced by the Parliamentary Committee for Civil Liberties, Justice and Home Affairs (LIBE).

<sup>17</sup> Articles 81.2 of the LIBE amended proposal

## Box 1 New European Data Protection Regulation

European regulations are the most stringent European legislative instruments. They are directly binding upon the member states and directly applicable. Their effect is thus stronger than the directives which have to be implemented by the member states once voted through. The new regulation's main features are the following:

- **Creation of the European Data Protection Board** in charge of monitoring compliance with EU Regulation (arts. 64 and 65)
- **More stringent rules on obtaining valid consent of data subjects** (arts. 7 and 8)
- **Removal of the exceptions from consent** for the use of identifiable sensitive data for research (art. 81)
- **Tougher standards for personal data processing** as 'data enabling the attribution of information to an identified or identifiable data subject must be kept separately from the other information under the highest technical standards, and all necessary measures are taken to prevent unwarranted reidentification of the data subjects' (art. 83.1)
- **Appointment of a data protection officer** by any company doing business in the EU and processing data of more than 5,000 individuals (arts. 32 to 34)
- **Access to information by individuals** is facilitated by the new regulation. The text establishes a 'right to erasure' (art. 17).

While monitoring data breaches presents methodological hurdles (in many European countries mandatory notification regimes are not implemented; breached organisations are often unaware of having suffered an incident), some web sites are already monitoring and exposing major data breaches to the public (Robinson et al., 2013). For instance, Information is Beautiful shows the inventory of publicly exposed major data breaches worldwide (when losses are greater than 30,000 records) by organisation, sector (including the healthcare sector), degree of data sensitivity and method of leak (accidentally published, hacked, inside job, stolen computer, stolen media or poor security).<sup>18</sup> The site also gives access to the full data breach report and provides information on the time lag between the occurrence of the data breach and its reporting to competent authorities.

### Consent management constrains access to data

Data access can be restricted by data sharing forms and consent management issues. The specific nature of personal medical data has led to the implementation of restrictive regulations regarding the use of this information in several countries (France, Portugal, Spain) where informed consent is mandatory for the use of medical information by researchers (OECD, 2013).

---

<sup>18</sup> Information is beautiful, [www.Informationisbeautiful.net](http://www.Informationisbeautiful.net) (last accessed 20 January 2014). All these information are freely available on the web site from 2004 onwards.

### 3. Overcoming existing barriers through the development of standards to improve access to and use of RWD

---

In Chapter 2 we described a number of barriers that may restrict the realisation of the potential of RWD-based analysis. In this section, we discuss regional, national and European examples of strategies that were implemented to overcome these barriers. Attempts to build on existing international standards for disease classification are emerging at the European scale (especially in the field of rare diseases) and are contributing to the development of strong methodological standards to facilitate data use and international comparisons. Attention has also been given to the spread of eHealth infrastructures to facilitate RWD collection, linkage and storage across Europe. The European Commission has also been particularly active in incentivising the development of Pan-European databases that can be used for medical research. Finally, different tools enabling effective consent management have also been developed to enable stakeholders to solve privacy issues and facilitate the collection of patient-level data in a more routine way.

#### 3.1. Overcoming content and quality issues

##### *3.1.1. Terminology standards: standardisation on its way*

Progress needs to be made on multiple levels to change the way that health information is classified, captured and stored by different stakeholders. This view was corroborated by all of the interviewees who commented on this specific point. In the field of rare diseases, such progress has recently been made (Aymé and Rodwell, 2013). Some countries, such as France and Germany, are already using the International Classification of Disease (ICD) system, and the World Health Organization will soon publish guidelines to enhance the harmonisation of classification standards and facilitate international comparisons in the field of rare diseases (Int10).

Box 2 presents recent updates with regard to terminology standards. Such standards (namely ICD10, OPCS4, and HRG) were considered as the main enabler in facilitating data interpretation and linkage in the UK NHS COPD project presented in Table 7 (Section 2.2.).

## Box 2 Terminology standards

The use of international codification standards (and their development), is required to facilitate the harmonisation process and enhance interoperability between national databases (Avillach et al., 2010). In a recent publication, the OECD reports that there is a strong national drive in that respect, with 19 countries using ICD-10 codes and five countries reporting SNOMED codes (OECD, 2013). Thirteen countries are using DICOM standards for the electronic storage of medical images. There is also some consistency in the use of international standards for laboratory tests and medications, with 13 countries using LOINC codes for laboratory results and 12 using WHO ATC codes for medications.

The fact that vocabulary and coding standards continue to evolve and the inherent challenges notwithstanding, existing international standards have already provided EHRs with a head start. Standards that are particularly applicable are:

- ICD 9/10 (International Statistical Classification of Diseases and Related Health Problems): a code set containing diagnostic and operative procedural data that is used to classify and record diseases and other health problems.
- CPT 4/5 (Current Procedural Terminology): a set of approximately 7,800 numerical codes that provide uniform descriptions of medical, surgical, radiology, laboratory, anaesthesiology and valuation/management services.
- LOINC (Logical Observation Identifiers Names and Codes): a database containing over 48,000 universal names and identification codes for specifying laboratory and clinical test results.
- SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms): a healthcare terminology featuring over 344,000 terms covering most areas of clinical information, including diseases, findings, procedures, microorganisms, and pharmaceuticals.
- DICOM (Digital Imaging and Communication in Medicine): terminology to transmit information in medical imaging.
- ATC CS (Anatomic Therapeutic Chemical Classification System): international classification system used for the classification of drugs. It has been developed by the WHO.

It is worth noting though that standardisation is not necessarily considered a positive process in the innovation field (Farrell and Saloner, 1985). Indeed, in other sectors and contexts, trade-offs between standardisation and innovation can be at stake, and interoperability or matching techniques might be preferred alternatives to coding and classification standards. However, this point was not made explicitly in the literature review or the interviews, where the evidence indicated a preference towards standardisation across systems and countries. Nevertheless, developing interoperability solutions remains at the core of some international initiatives such as Integrating the Health Enterprise (IHE) (Box 3).

### Box 3 Integrating the Healthcare Enterprise (IHE)<sup>19</sup>

IHE is an initiative by healthcare professionals and industry built on existing standards (HL7, SNOMED CT, DICOM). It was created to help overcome the deficiencies in those standards, in their application in EHRs, and for the legitimate sharing of permitted records across domains (between professions; between organisations; between countries).

- IHE can be used for direct system-system interoperability if all systems are IHE-compliant; however it also allows for IHE-compliant intermediary systems to link non-compliant or part-compliant systems. IHE allows for the secure publication of records, where the record remains in the original Data Controller's system of record, but is visible in other systems where there are meaningful use or access permissions.
- IHE underpins the strategy of Turkey in building its health record interoperability approach and is implemented in Austria, parts of the USA, and is gaining traction in the UK NHS (SEPT, Lancashire commencing implementation of IHE-centric architectures).
- IHE profiles for quality, research and public health are being developed.

#### National patient identifiers support and offer alternatives to standardisation of datasets

Parallel to the trends in the standardisation of datasets, further developments in technology and practice support identification across platforms in other ways. As long as fragmentation remains an obstacle to working across systems, the issue of non-aligned technology and standards will need to be addressed. To this end several solutions are emerging: the advent of semi-structured, unstructured and Natural Language Processing (NLP) capabilities is one such example. Another is a relatively well established, and growing, adoption of terminology standards in EHRs and hospital files. Innovations are also occurring at the level of practices in healthcare management, through the adoption of national Patient/Provider IDs, enabling greater data collection. Several European countries<sup>20</sup> already report having implemented both a unique national identifying number (instead of using identifiable information such as name, address) that will be used to ensure the anonymity of patients, to build their electronic health record; and also a unique identifying number for healthcare professionals entering data into an EHR system (OECD, 2013). These tools facilitate the unambiguous identification and subsequent integration of health information.

---

<sup>19</sup> Integrating the Healthcare Enterprise, <http://www.ihe.net/> Last accessed 27/02/2014.

See also: IHE quality, research and public health profiles, [http://wiki.ihe.net/index.php?title=Profiles#IHE Quality.2C Research.2C and Public Health Profiles](http://wiki.ihe.net/index.php?title=Profiles#IHE_Quality.2C_Research.2C_and_Public_Health_Profiles) Last accessed 27/02/2014.

<sup>20</sup> Namely Belgium, Denmark, Estonia, Finland, France, Norway, Poland, Slovakia, Slovenia, Sweden, Switzerland and the United Kingdom (England).

### *3.1.2. Systematically collecting patient data through electronic health records (EHR) has the potential to improve the content of datasets ('completeness')*

EHRs contain patient-level information collected during clinical care (socio-economic information, medical history, details about drug use, diagnostic test results, physician notes, etc). Such information is particularly useful for research (eg clinical trial recruitment, combination with phenotypic and genomic information), drug development and post marketing. The OECD (2013) and the European Commission (Stroetmann, 2011) report that a majority of European countries are now designing or implementing strategies for the development of national EHR systems in order to facilitate patient data collection and the sharing of data amongst primary and secondary care providers. Some initiatives build on previous developments introduced at the regional scale (eg in the UK, Spain, Belgium or Switzerland) while others are implementing more centralised strategies (eg Sweden, Austria, France). To illustrate contemporary examples of successful implementations, Table 9 presents the main features of the national Danish EHR system and Table 10 describes regional initiatives taking place in Spain and Italy.

**Table 9 Summary of the Danish EHR system: 'my health summary'**

Geography	Denmark
Type	Building on hospital data and infrastructure of the National Patient Register, <sup>2</sup> and on patient data available from other national and regional sources, 'my health summary' is an EHR containing all existing relevant patient information and is supplemented with relevant data continuously (eg updated through every contact with the health system).
Sample size	Universal coverage
Content	<ul style="list-style-type: none"> <li>- Summary of hospital admissions (back to 1995)</li> <li>- Recent notes from hospital charts</li> <li>- Summary of medication prescribed over the last two years</li> <li>- Overview of personal wishes in relation to organ donation and receiving life-prolonging treatment (living wills)</li> <li>- Status of laboratory tests ordered by physicians</li> <li>- Contact information for the personal GP.</li> </ul>
Use	It is used by most care providers and, since 2010, patients can access their data through the national electronic portal Sundhed.dk. Appropriate availability of patient data is expected to allow for greater patient safety, improved treatment quality and a more efficient workflow. In addition, patients are guaranteed the possibility of a more active involvement in their own treatment.
Advantages	Country wide IT infrastructure with the possibility to be linked to regional systems.
Access	<p>Relevant providers and patients can directly access the data. No information available with regard to access for research purposes.</p> <p>More generally in Denmark, to conduct research projects based on register data, permission from the Data Protection Agency is required. Multicentre trials can only be approved by a single research ethics committee. Clinical trials concerning drugs must be approved by the Danish Medicines Agency.<sup>3</sup></p>

Sources:

1 Doupi P, Renko E, Giest S, Dumortier J (2010) Country Brief: Denmark, EHealth Strategies, European Commission.

2 For further information, see: Elsebeth Lyng, Jakob Lyng Sandegaard and Matejka Rebolj, "The Danish National Patient Register", Scandinavian Journal of Public Health, 2011, 39:30.

3 Olejaz M, Nielsen AJ, Rudkjobing A, Birk HO, Krasnik A, Hernandez-Quevedo C. (2012) Denmark Health System Review, Health systems in Transition vol14 No2, European Observatory on Health Systems and Policies.



**Table 10 Regional initiatives in Spain and Italy**

	<b>DIRAYA System<sup>1</sup></b>	<b>Social and Health Information System<sup>3</sup></b>
Geography	Andalucía (Spain)	Lombardy (Italy)
Type	Data are collected by all health care providers from primary to secondary care and the network also links pharmacies.	General Practitioners, hospitals and pharmacists linked to the regional health service are required to collect information on patients by a regional law (law no. 18, of 31 July 2007, article 1). <sup>4</sup>  In 2009, the same law was extended to private healthcare practitioners.
Sample size	97% of the Andalusian population of 8,302,923. <sup>2</sup>	Approximately 9,500,000 people are included in the database (almost full province coverage as the population of Lombardy is 9,900,000 people). <sup>5</sup>
Content	Patient background, allergies, medication, diagnosis, tests. Electronic prescriptions are in the database as well as appointment schedules. <sup>2</sup>	All aspects of patient history from primary to secondary care and pharmaceutical needs. <sup>6</sup>
Use	Healthcare providers can obtain patient information to enable more effective and efficient treatment; patients can book (primary healthcare) appointments online and link e-Prescriptions to their accounts.	This database was established in order to improve patient care and the quality of the healthcare service reaching citizens all the while increasing the efficiency of the healthcare system. <sup>5</sup>
Advantages	This database offers a large, regional sample size. It is a good example of an integrated healthcare system with easy to access patient information which includes over 7.5 million records (based on figures from 2012). <sup>2</sup>	Twenty-nine public hospitals and five medical institutions offering hospitalisation. <sup>7</sup> A large network of information is available from a number of public and private healthcare sources (approximately 7,200 GPs and paediatricians and 2,600 pharmacies <sup>5</sup> ) which can give a complete overview of the patients' medical and healthcare history. Originally, data collection started in the latter years of the 1990s and was extended to include all healthcare providers in 2009.
Access	Healthcare professionals directly dealing with the patient can access relevant	Access to data is not straightforward. Even within the SISS system, only GPs have access

	DIRAYA System <sup>1</sup>	Social and Health Information System <sup>3</sup>
	information. No information available with regard to access for research purposes.	to all the data while other healthcare professionals need the permission of the patient to view certain records. Furthermore, patients show or hide as much information as they choose. <sup>8</sup> No information available with regard to access for research purposes.

Sources:

1 Karl A. Stroetmann, Jörg Artmann, Veli N. Stroetmann with Denis Protti, Jos Dumortier, Sarah Giest, Uta Walossek and Diane Whitehouse, European countries on their journey towards national eHealth infrastructures, 2011, p. 22, [http://www.ehealth-strategies.eu/report/eHealth\\_Strategies\\_Final\\_Report\\_Web.pdf](http://www.ehealth-strategies.eu/report/eHealth_Strategies_Final_Report_Web.pdf) Last accessed 02/12/2013.

2 Junta de Andalucía, "Healthy Andalusia," 29 October 2012, [http://www.juntadeandalucia.es/salud/channels/temas/temas\\_es/\\_\\_\\_QUIENES\\_SOMOS/C\\_7\\_Healthy\\_Andalusia/healthy\\_andalusia\\_navegable?idioma=es&perfil=ciud&tema=/temas\\_es/\\_\\_\\_QUIENES\\_SOMOS/C\\_7\\_Healthy\\_Andalusia/&contenido=/channels/temas/temas\\_es/\\_\\_\\_QUIENES\\_SOMOS/C\\_7\\_Healthy\\_Andalusia/healthy\\_andalusia\\_navegable&desplegar=/temas\\_es/\\_\\_\\_QUIENES\\_SOMOS/#punto\\_8](http://www.juntadeandalucia.es/salud/channels/temas/temas_es/___QUIENES_SOMOS/C_7_Healthy_Andalusia/healthy_andalusia_navegable?idioma=es&perfil=ciud&tema=/temas_es/___QUIENES_SOMOS/C_7_Healthy_Andalusia/&contenido=/channels/temas/temas_es/___QUIENES_SOMOS/C_7_Healthy_Andalusia/healthy_andalusia_navegable&desplegar=/temas_es/___QUIENES_SOMOS/#punto_8) Last accessed 04/12/2013.

3 Nasi, G, Cucciniello M, Ongaro E, Galli D and CGuerrazzi, "Coordination Practice: Regional Electronic Patient Record in Lombardy", 2013, [http://www.cocops.eu/wp-content/uploads/2013/06/Italy\\_Health\\_Electronic-Patient-Record\\_Lombardy.pdf](http://www.cocops.eu/wp-content/uploads/2013/06/Italy_Health_Electronic-Patient-Record_Lombardy.pdf) Last accessed 29/11/2013.

4 SISS, "Medical Facilities", SISS, "Medical Facilities," <http://www.siss.regione.lombardia.it/ds/Satellite?c=Page&childpagename=SISS%2FSISSLayout&cid=1213350945296&p=1213350945296&pagenam=SISSWrapper> Last accessed 03/12/2013.

5 Lombardia Informatica, "Health Care Information System," <http://www.lispa.it/cs/Satellite?childpagename=Lispa%2FLLLayout&c=Page&p=1213366334319&pagenam=LlWrapper&rendermode=previewnoinsite&cid=1213366334319> Last accessed 04/12/2013.

6 SISS, "SISS Overview", <http://www.siss.regione.lombardia.it/ds/Satellite?c=Page&childpagename=SISS%2FSISSLayout&cid=1213347263759&p=1213347263759&packedargs=locale%3D1194453881584&pagenam=SISSWrapper> Last accessed 03/12/2013.

7 SISS, "Public hospitals and public IRCCS," <http://www.siss.regione.lombardia.it/ds/Satellite?c=Page&childpagename=SISS%2FSISSLayout&cid=1213357542383&p=1213357542383&pagenam=SISSWrapper> Last accessed 05/12/2013.

8 Gemalto, "Lombardy's service card: Making the delivery of services significantly more efficient," [http://www.gemalto.com/govt/customer\\_cases/italy-lombardy-service-card.html](http://www.gemalto.com/govt/customer_cases/italy-lombardy-service-card.html) Last accessed 05/12/2013.

National and regional EHRs are a first step towards the collection of large and complete medical datasets. However, the development of standards to improve the quality of data collected through EHRs and facilitate their interoperability is also needed in order to optimise the value of EHRs now and in the future (Box 4).

#### Box 4 What's next for EHRs?<sup>21</sup>

- Technological developments, especially in the field of medical devices, will likely enable a rising adoption rate for information collection through EHR and thus a shift from manual input processes to automated ones, although this trend can be limited by contextual factors, such as attitudes, security and privacy concerns. The less difficult it is for consumers and practitioners to adopt and maintain personal health records as it places less burden on the individual.
- The sharing of EHR information between care providers and between these and the patient might progressively shift from printed and portable media to consumer-authorized and secure automated transactions. The easier it is for consumers to share data with providers and providers to access that information on a just-in-time basis, the more valuable the EHR will become for healthcare provision. However, strict security mechanisms will have to be developed to ensure personal data protection.
- The functionality of EHRs may expand beyond the role of an aggregator of personal information, to include educational, collaborative, compliance, analytic, and service access functions. The greater the healthcare value that EHRs can deliver to system stakeholders, the more compelling their use is likely to become.
- Some of the obstacles for the adoption and implementation of EHRs include:
  - security and privacy practices
  - ensuring full integration between datasets (currently discrete or only loosely integrated)
  - ensuring benefits at the level of patients and healthcare professionals
  - legal and policy obstacles to the internationalisation of databases
- EHR use will likely expand from today's principal patient base of specific groups (mainly the elderly and chronically ill) to patients in every demographic. It is predicted that as the uses of EHRs scale up and expand to a larger patient base then others in the sector, such as vendors, payers, and providers are likely to follow.

Source: IBM, 2008

#### *3.1.3. More robust data quality assurance processes are being developed at the European level*

As highlighted earlier in this report, concerns over the quality of RWD – especially data extracted from patient registries – are one of the main barriers to its use for medical research and need to be addressed. In that respect, valuable insights can be driven from initiatives developing data quality insurance processes in the field of rare diseases. It is worth mentioning that data quality verifications might be easier to process

---

<sup>21</sup> Although the data in Box 4 are based on a report published in 2008, most of the obstacles remain outstanding, and none of the foreseen developments have been observed yet.

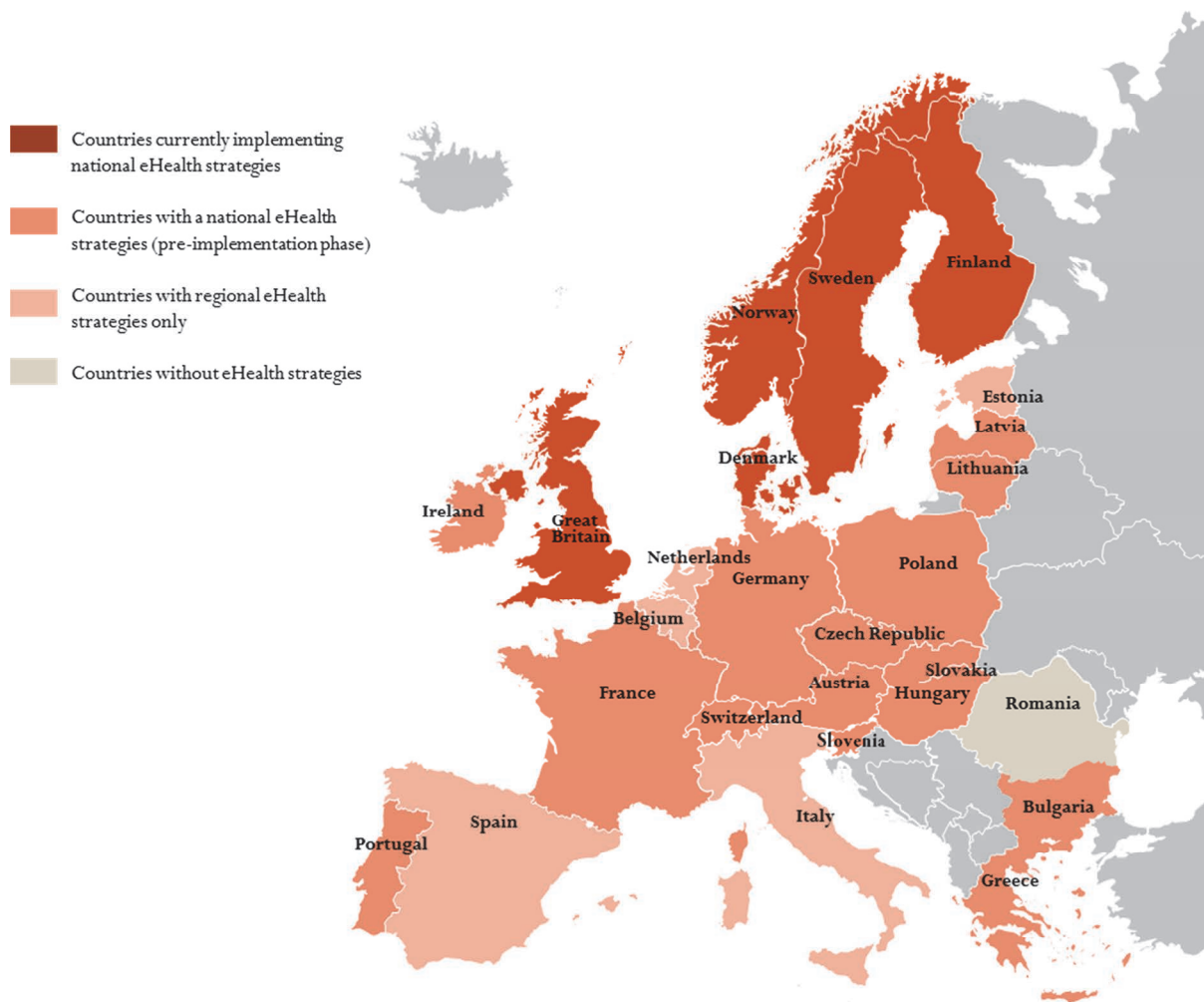
with a smaller data sample, which is often the case with rare disease registries. Nevertheless, future initiatives targeting the development/improvement of RWD quality procedures might draw on the lessons learnt from smaller-scale projects. For example, an international working group has been formed to develop improved data quality procedures for the European Cystic Fibrosis Society Patient Registry (ECFSPR), which stores demographic and clinical data from national and individual Cystic Fibrosis registries (23 countries and 26.000 patients involved). The group is in charge of drafting a standardised procedure document for data quality checks that are requested to allow national organisations to upload their data to the ECFSPR. The working group is composed of representatives from each participating country (mostly the national registries' data manager). The drafting of the new quality insurance framework is based on the review of existing national data quality insurance procedures and discussions around their relevance. The final list of data quality controls will then be used by each registry data manager for data control before uploading the data onto the European platform. Based on this shared list, the next step will be to develop software that can proceed to automatically process these controls (Zolin and Gulmans, 2013). Although at an infant stage, such strategies rest on the collaboration of multiple national stakeholders willing to build common solutions at the European scale.

## 3.2. Overcoming methodological issues

### *3.2.1. National strategies are supporting the development of analytical capabilities*

National governments are now aware of the importance of developing standards for data collection and datasets linkage to improve healthcare provision. As mentioned above, efforts have been made to invest in national platforms for the collection and use of RWD and of EHRs in particular. However, the European landscape is quite fragmented with some countries clearly leading in that field (namely the Scandinavian countries and the UK – see Box 5), and most initiatives are still in pilot phase (see Figure 2).

Figure 2 eHealth strategies across Europe



Source: Adapted from Stroetmann et al. (2011).

There are, then, not yet many concrete results from the use of eHealth infrastructures and the ways in which information that can be derived from it might be used for research purpose. However, the development of eHealth infrastructure facilitates medical data collection, which is a first step towards use and application of the data. Some initiatives have been implemented in several European countries (Box 4) but they remain highly dependent on a country's technological capabilities and regulatory environment, as the development of infrastructure for data linkage is slowed down by data privacy concerns (eg recombination of health and socio-economic datasets might threaten patient anonymity).

## Box 5 Examples of initiatives in the development of methodological standards for data collection and data use

**National strategy for digitalisation of the Danish healthcare sector 2013–2017** ‘Making e-Health work’ provides a legal and institutional framework for the development of RWD standards. The National Board of eHealth is responsible for research-based health surveillance, rational use of IT in the Danish healthcare system and prevention and control of infectious diseases, biological threats and congenital disorders. The strategy aims to develop a national standardised framework for data collection that will allow researchers to link different databases for research purpose.

In July 2013, the UK Minister for Science and Universities has announced a £20 million investment for the creation of the **Farr Institute**.<sup>1</sup> This new platform will be funded by the Medical Research Council and will involve 19 UK Universities and four eHealth informatics research centres (eHIRCs): in London, Manchester, Dundee and Swansea. This initiative aims to combine various kinds of expertise (IT, healthcare) to link and analyse different health datasets. The MRC hopes to facilitate the development of fruitful collaborations between research centres, IT and pharmaceutical companies through the creation of this new platform.

In terms of data collection, codification and sharing, important progress has been made in the field of rare diseases with 588 national and regional rare diseases registries (among 33 countries). Most of them have been developed by public authorities or research centres in France (130), Germany (105) and the UK (70). The **Orphanet network**<sup>2</sup> provides access to these registries to public and private entities; it also develops classification standards to facilitate international comparisons and enhance database interoperability (Int10).

Note:

1 MRC Communication, £20 million for new health informatics research institute, MRC website, July 2013

<http://www.mrc.ac.uk/Newspublications/News/MRC009207> Last accessed 09/12/2013.

2 Orphanet <http://www.orpha.net/consor/cgi-bin/index.php> Last accessed 09/12/2013.

### *3.2.2. Shared analytical standards are being developed through the creation of platforms and fora that rely on stakeholders’ expertise to collect and use RWD*

International research projects are a means to enhance sharing of best practice regarding data collection, data sharing and data use. While most of them have been created to address big data issues, their structures and aim could potentially be transferred to the RWD domain. The recent creation of the Global alliance to enable responsible sharing of genomic and clinical data (Global alliance to enable responsible sharing of genomic and clinical data, 2013) is an example of platform that facilitates the collaborative development of standards for data exchange, data use and data analysis. This alliance, genomics oriented and originally US-centric, sees the number of its European members increase rapidly with more and more research centres (eg the Wellcome Trust Sanger Institute (UK), the Lund University

(Sweden), International Rare Diseases Research Consortium (International Consortium<sup>22</sup>), national health bodies (eg National Institute for Health and Welfare (Finland)) and regulatory centres (eg Centre for Genomic Regulation (Spain)) joining the initiative. With a mission to increase data sharing, improve interoperability and develop ‘harmonised approaches’ (p. 9) the alliance sets up an ambitious standard-setting programme that includes the development of methodological standards. A number of commercial hubs also offer analytical services and capabilities to a range of organisations: for instance the BT for Life Sciences R&D Connect cloud provides pharmaceutical companies with services and structures that seek to optimise data analysis.<sup>23</sup>

### 3.2.3. *Overcoming lack of inter-operability in RWD systems across Europe through EU-funded joint research projects*

Since 2004, the EU has been very active in the internationalisation and linkage of national and/or regional databases, by creating incentives for the development of methodological standards for data collection and processing through the funding of a number of Pan-European research projects. This has led to the harmonisation of RWD collection standards, although most of these projects remain focused on particular diseases or health conditions. The Second Programme of Community Action in the Field of Health 2008–2013 makes explicit reference to the importance of developing IT infrastructures for the collection of health data to improve knowledge generation for the diagnosis and treatment of patients.<sup>24</sup> The generation of more comprehensive patient level data will be enhanced by the development of platforms to collect and link medical and historical patient-specific data as well as demographic data.<sup>25</sup> The total budget for this programme is €321m and it is managed by the European Commission and the Executive Agency for Health and Consumers (EAHC). The European Parliament and the European Council also encouraged relevant actors to undertake joint actions with other community programmes such as the FP7 Health Programme<sup>26</sup>. The main objective of the FP7 Health Program is to improve the health of European citizens and boost the competitiveness of health-related industries and businesses, while also addressing global health issues. Various research projects aiming at favouring Pan-European collaboration for the sharing and use of medical data have been funded under this scheme. The EPIRARE

---

<sup>22</sup> Members include institutions from Australia, Canada, China, Finland, France, Germany, Georgia, Italy, Korea, Spain, the UK, the US and the European Commission.

<sup>23</sup> BT for Life Sciences R&D Connect, [http://www.globalservices.bt.com/uk/en/industries/pharmaceutical/r\\_and\\_d\\_connect](http://www.globalservices.bt.com/uk/en/industries/pharmaceutical/r_and_d_connect) Last accessed 27/02/2014.

<sup>24</sup> Decision No 1350/2007/EC of the European Parliament and of the Council of 23 October 2007 establishing a second programme of Community action in the field of health (2008–13).

<sup>25</sup> European Commission, Horizon 2020 Work Programme 2014–2015, Annex 6 to the Decision [http://ec.europa.eu/research/horizon2020/pdf/work-programmes/information\\_and\\_communication\\_technologies\\_draft\\_work\\_programme.pdf#view=fit&pagemode=none](http://ec.europa.eu/research/horizon2020/pdf/work-programmes/information_and_communication_technologies_draft_work_programme.pdf#view=fit&pagemode=none) Last accessed 12/12/2013.

<sup>26</sup> Decision No 1350/2007/EC of the European Parliament and of the Council of 23 October 2007 establishing a second programme of Community action in the field of health (2008–13).

project presented in Table 11 illustrates current debates on RWD and describes efforts to promote RWD standards.<sup>27</sup>

**Table 11 The EPIRARE Project: feasibility of patient-level data collection at the European scale**

Geography	Pan-European
Type	Research exploring different types of resources, needs and expectations of existing registries. It is a feasibility study into a number of issues associated with registering rare diseases and the creation of such an EU platform which aims to collect information on rare disease patients. Information is contributed by both existing and expired rare disease registers and the information is drawn from these.
Sample size	n/a
Content	n/a The EPIRARE project is working towards promoting the establishment of rare disease registries through workshops and collaboration with stakeholders. The current EPIRARE study is looking into the feasibility and practical and ethical issues surrounding collecting data on rare disease patients and the establishment of an EU-wide platform on rare diseases where relevant information could be collated and shared with the appropriate stakeholders. The study began in 2011 and is currently ongoing. It seeks disease registries already established in Europe to participate in a survey. The objective of the project is to examine the legal basis and other criteria of established registries in order to determine best practice for a potential rare diseases registry. Specific information on participating registries is confidential.
Use	The aim is to boost support for rare disease registers for epidemiological purposes and this is meant to inspire rare disease registration.
Advantages	Insight from a number of different countries and healthcare systems. Could provide a good measure of how practices change from one country to the other.

Source: EPIRARE, <http://www.epirare.eu/> Last accessed 29/11/2013.

One limitation of most of these EU funded research projects is that they focus on the collection of new data or the harmonisation of existing datasets on specific diseases or medical treatments. The projects do not necessarily insist on the integration of different data sources (socio-economic, medical, environmental, etc), which present yet more opportunities for future medical research. Furthermore, they often involve countries that are already ahead in the development of standards and tools for data collection and use at the national scale (Int10). In that sense, such projects do not necessarily create incentives for less advanced

<sup>27</sup> See also the articles describing the EU-ADR project on adverse drug events, in Avillach et al., 2010, Avillach et al., 2013, Coloma et al., 2011 and Coloma et al., 2012.



countries to develop technological and methodological standards for RWD collection and use. However, they do constitute a first step towards the development of common methodologies for both data collection and use and for the creation of Pan-European disease specific datasets, which can further be combined with other Pan-European databases. They also represent a first attempt to overcome issues related to the fragmentation of databases. Such an attempt has been demonstrated by the EuResist project (see Table 4, Section 2.2.) through which non-profit, multidisciplinary partnerships have been developed to expand the integrated data pool and advance science, and where collaboration has been facilitated by focusing on one disease area. The Hypergenes project<sup>28</sup> (Table 12), which mostly uses big data (genomics), is another example of successful European initiative.

---

<sup>28</sup> Hypergenes network, <http://www.hypergenes.eu/> Last accessed 09/12/2013.

**Table 12 Hypergenes, European Network for Genetic-Epidemiological Studies**

Case description

Hypergenes is co-funded by the FP7 Health Programme (HEALTH-2007-2.1.1-2 ‘Molecular epidemiological studies in existing well characterised European (and/or other) population cohorts.’).

The Hypergenes initiative aims to develop an exhaustive model to disentangle the genetic bases of a complex disease using population genetic epidemiology as a methodological tool.

Essential hypertension (EH) was chosen as the disease model, both because of long-term experience in investigating the genetics of EH and because the cardiovascular complications remain the major cause of death in the EU. Its impact in term of cost and disability are a devastating burden for patients, for their relatives and for the human potential of the EU.

Designing a comprehensive genetic epidemiological model of complex traits also aims to translate genetic findings into improved diagnostic accuracy and new strategies for early detection, prevention and eventually personalised treatment of a complex trait.

Type of real-world data used

Clinical, environmental and genomic data.

Benefits of using RWD

- Find genes responsible for EH
- Develop an integrated disease model, taking the environment into account
- Test the predictive ability of the model to identify individuals at risk.

Lessons learned

The Hypergenes approach foresaw the creation of a Biomedical Information Infrastructure (BII), providing the project itself an infrastructure. Such a data warehouse had to store existing and newly created harmonised and standardised information (data and knowledge), at the same time as providing efficient access to it.

The BII enabled the collection, integration, harmonisation and correlation of data described in diverse formats and vocabularies, scattered in disparate geographies. Furthermore, the BII design targets both research and clinical environments by having a single standard-based warehouse as a source of multiple marts that serve specific needs in research and healthcare.

BII allows integration, harmonisation and standardisation of clinical, environmental and genomic data. The warehouse service stores data in its richest format using a set of constrained internationally-recognised standards such as HL7 Clinical Document.

Architecture (CDA), the Pedigree and Genetic Variation standards.

The essence of the BII lays in its methodology, which can be replicated in future projects similar to Hypergenes, and it is open to other potential uses: for example, for biobanks warehousing, integration of electronic health records and pharmaco-surveillance.

### 3.2.4. Developing Pan-European eHealth infrastructure to reduce fragmentation across countries

In addition to the growing number of academic network-led European research projects, the first European eHealth Action Plan was adopted in 2004 (European Commission, 2004), leading to the development of the eHealth Initiative in 2007 (European Commission, 2011). The overall objective of this programme was to foster cooperation between member states to facilitate the access to healthcare services and improve the quality of care at the European scale. In that perspective, efforts have been made to reinforce interoperability between national health records. This has led to the implementation of the European Patient Smart Open Services (epSOS) Pilot Project (Thorp, 2010) in 2008 to develop an eHealth infrastructure at the European scale. (Table 13) This represents the first attempt to develop European wide health information infrastructures, enrolling 22 EU member states and 3 non-EU member states.

**Table 13 The epSOS project: a first attempt to combine e-health records across countries**

Geography	Twenty-two European Union states and three non-member states: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Luxembourg, Malta, Norway, Portugal, Poland, Slovakia, Slovenia, Spain, Sweden, Switzerland, The Netherlands, Turkey, UK.
Type	A pan-European project aimed at developing a cross-border infrastructure of electronic health records. The project is open to pharmacies, hospitals, healthcare facilities, physicians and individuals from the participating countries.
Sample size	Forty-seven consortium members in 25 different countries. Precise figures for a sample size are unclear and as the project is in the pilot phase, more beneficiaries can be added. <sup>2</sup>
Content	Patients' medical histories and other data and cross-border e-Prescription capabilities. A list of all prescribed medicines taken by the patient is kept.
Use	Designed to facilitate cross border health care in Europe, making it easier for people to access medication whilst abroad, health care professionals can access patient information in an easier more accurate fashion.
Advantages	Large breadth of information spanning 25 countries – a large pool of information.
Access	The project will finish at the end of June 2014. The results of the project will be made available by the project partners and should outline best practice recommendations, technical information, software and organisational models. However, it is unclear whether this includes patient data.

Sources: epSOS, <http://www.epsos.eu/> Last accessed 02/12/2013.

### 3.3. Improving governance

#### 3.3.1. *Buying existing data remains the most common strategy to access RWD*

As mentioned by two interviewees (Int2, Int8), the simplest way to obtain data access is to ‘buy access’ through the means of an intermediary, and upon request, this intermediary delivers access to existing datasets or builds new ones, especially designed for the pre-defined research purpose. Private companies have used this system, buying the services of consultancies specialised in RWD, such as Evidera<sup>29</sup>, or data vendors such as Cegedim<sup>30</sup> – whose biggest clients are pharmaceutical companies (Int1) – to collect patient-level information, generally extracted from EMRs. These intermediaries directly collect data that are anonymised and pooled and then sell it to public or private body for research purpose. Pricing might also differ for public and private entities (Int1). The amount of data available is strictly limited by the research question. In exchanging access to data, the intermediaries also participate in the standardisation of collection methodologies by providing common software to practitioners especially GPs whose data are often lacking consistency and clarity. One of our interviewees stated that they generally offer financial incentives to data controllers (small amount of money to GPs) in exchange of the data. For instance, Cegedim provide guidelines on how to use and collect the data and allow benchmarking. It has also been stressed by our interviewees that other kinds of incentives (non-monetary) are at stake in motivating practitioners to share their data (such as the desire to identify missing information in their database or to contribute to research and the development of better treatments for their patients). A recent example includes a contract between Pfizer and Optum<sup>31</sup> to collect lung cancer related data from French hospital and national databases. This contributed to the evidence base on the economic burden of lung cancer (Int2).

#### 3.3.2. *Engaging with academic partners is another common option for private stakeholders*

From the private sector point of view, another common access strategy in Europe is to involve an academic partner in the research project, as most databases are made accessible to academics upon request.<sup>32</sup> A recent project on adherence to treatment after acute myocardial infarction (Boggon et al., 2011) in England, involved the linkage of four independent datasets. It was initiated by AstraZeneca, in partnership with the Clinical Practice Research Datalink (CPRD, formerly GPRD, see Table 14)<sup>33</sup>, and

---

<sup>29</sup> Evidera, <http://www.evidera.com/> Last accessed 05/12/2013.

<sup>30</sup> Cegedim, <http://www.cegedim.com/Pages/default.aspx> Last accessed 05/12/2013.

<sup>31</sup> Optum, <http://www.optum.com/> Last accessed 05/11/2013.

<sup>32</sup> See for example Juliusson et al. (2012) on population based registries in Sweden, Holmberg (2012) on cancer registries, and Smeets et al. (2011) on insurance data in the Netherlands. Those three articles are described in Annex B.

<sup>33</sup> Clinical Practice Research Datalink <http://www.cprd.com/intro.asp> Last accessed 09/12/2013. See also Williams et al 2012.

The Myocardial Ischaemia National Audit Project (MINAP) managed by University College London<sup>34</sup>. Data from CPRD and MINAP were linked to the Hospital Episodes Statistics (HES) data and the Office of National Statistics (ONS) Mortality data. Access to CPRD and MINAP data was guaranteed by general funding of the study and specific grants to researchers involved in the analysis.

### 3.3.3. Creating new rules of access for publicly-funded databases

Some publicly funded databases are creating pathways to access for all stakeholders, including the private sector. In the UK, the CPRD is leading the way (Table 14). It offers market pricing structure and tailored pricing upon request. Data are extracted by the CPRD team against a query specification. A trusted third party is in charge of linking different datasets before the CPRD can share the data with its clients. The CPRD also has the ability to identify potential sites which could be used for clinical trials based on the available patient population and the disease characteristics required. According to the CEO, currently 18 out of the top 22 pharmaceutical companies are using this service to source and recruit patients to clinical trials.

**Table 14 The CPRD (Clinical Practice Research Datalink)**

Geography	UK
Type	Originally general practice data (GPRD); now links into a range of datasets including secondary care hospital statistics (HES), ONS death certification and socioeconomic classification, National Cancer Intelligence Network and the Myocardial Infarction National Audit Program, pollution-level data and cohort studies such as Avon Longitudinal Study.
Objectives	Maximise the use of anonymised NHS clinical data can be used and linked to improve observational research and lead to better research outputs.
Sample size	Records from over 12 million patients.
Content	Diagnosis, prescription, costs, use of services, medical history, etc.
Collection	CPRD team provides services to GPs who share their data.
Use	Epidemiology, drug safety, harm–benefit assessment, pharmacogenetic studies, pragmatic randomised controlled trials, health services research, public health, etc. Over 890 original research papers have used CPRD data so far.
Advantages	Substantive linkage programme enabling access to secondary care data, disease-specific registry data and socio-economic data. Longitudinal data (some over 30 years).
Access	Market pricing structure. Tailored pricing upon request. Data are extracted by the

<sup>34</sup> Myocardial Ischaemia National Audit Project <http://www.ucl.ac.uk/nicor/audits/minap> Last accessed 09/12/2013.

	CPRD team against a query specification. A trusted third party is in charge of linking different datasets before the CPRD can share the data with its clients.
--	--

### 3.3.4. Engaging clinicians: Incentivising information sharing by targeting the data controller

#### Providing business solutions for care providers

The routine collection of RWD has the potential to make RWD use more viable and sustainable. However, care providers (eg GPs) might be reluctant to share personal medical information to public or private bodies and to adopt the infrastructure that would facilitate data sharing as highlighted earlier in this report. Providing physicians with technological tools to strengthen in-house analytics capabilities and improve the efficiency of internal organisational and administrative process to incentivise data sharing. One interviewee (Int1) commented indeed that an incentive for general practice managers to use a specific analytical suite and to share their patients' data is the provision of additional services in exchange of data sharing. A data vendor can for instance provide a benchmark on the practice's data compared to other practices and can deliver support to populate missing information in the practice's database. One case study revealed that the Belgian government has been providing financial incentives (about 875 euros/per GP) for GPs to subscribe to a system that would enable the systematic use of EHRs. In turn the EHR could be augmented with additional decision aids seeking to identify certain categories of patients and improve treatment. Such a model has also been used by private companies for the construction of specific registries: some companies are now specialising in selling electronic medical record, clinic management and reporting system all together.<sup>35</sup> By selling this system to specialist health centres, these companies are building a powerful vehicle for establishing and ensuring access to a sustainable flow of data, while health centres buying the system benefit from clinical management capabilities that are adapted to their specialties.

#### Giving clinicians more control over the data collection process

Supporting a sense of ownership among clinicians might contribute to gaining access to data. One interviewee identified the fact of being in control of the data as an incentive for practitioners to sign up to data collection, with the option to opt out instantly would they wish to do so (Int6). As a result, it might be worth giving physicians direct control of the kind of data that is uploaded and used through the development of dedicated software.

Moreover, physicians might be more supportive of a data collection process that they contributed to build. This is, for example, illustrated by the experience of the Rizzoli Orthopaedic Institute in

---

<sup>35</sup> Information was taken from an anonymised case study.

Bologna, Italy, where clinicians supported the design of an analytics solution that increased more effective treatment and more accurate determination of hereditary bone disease risks.<sup>36</sup>

### 3.4. Developing privacy best practices

#### 3.4.1. *Addressing the concerns of health professionals with regard to privacy issues by involving a third party to collect and anonymise the data*

The sharing and related access to medical data depends on the ability of the data user to process and use it in an anonymised way, to respect patients' privacy, and in a securitised way to avoid data breaches. As a result, engaging with third parties that are in charge of collecting consent and making sure that data are anonymised is a common strategy to gain access to observational data. Our findings suggest that reputational effects have a strong impact on physicians' willingness to share their patients' data. In Belgium, for instance, the recent standardisation of EHR formats, driven by the Belgian government in collaboration with EHR vendors, has allowed broader data access. Although only a minority of physicians were strongly reluctant to share personal data, data transfer and hosting of the central analytics environment was entirely tasked to a Trusted Third Party for data depersonalisation.<sup>37</sup> The reputation of this established company inspired the trust required for the implementation of the scheme.<sup>38</sup>

Furthermore, recent initiatives in the field of rare diseases have aimed to sustain data collection through engaging more directly with the patients. For example, the development of an International Dysferlinopathy Registry is based on the establishment of a tripartite relationship between the register curator, the patients and their medical doctors. The registry is opened to all patients worldwide; they are in charge of registering online and filling in relevant documentation relative to informed consent. They also provide their GP's contact information and some personal medical information (by responding to a standardised questionnaire). Once consent is collected, the registry curator contacts the patient's doctor who then uploads genetic and biomedical information on the server. Patients are incentivised to enrol in exchange for receiving relevant information about their condition as well as invitations to participate to clinical trials or research projects, and the curator is responsible for making sure that data are stored in a secured and anonymised way. Access to these data is open to all and subject to approval (Blandin et al., 2013).

---

<sup>36</sup> "Rizzoli Orthopaedic Institute looks for genetic insights in the family tree" [https://www-01.ibm.com/software/success/cssdb.nsf/CS/JSTS-7X9W5S?OpenDocument&Site=default&cty=en\\_us](https://www-01.ibm.com/software/success/cssdb.nsf/CS/JSTS-7X9W5S?OpenDocument&Site=default&cty=en_us). Last accessed 26/02/2014.

<sup>37</sup> The term depersonalisation refers to "the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour, be attributed to an identified or identifiable individual" (Fischer-Hubner, 2001:112)

<sup>38</sup> See also other elements of this case study in Section 3.3.4.

### *3.4.2. Responding to the public concerns by improving communication around the social and individual benefits of RWD-based research*

Several large-scale initiatives have recently shown the willingness of patients to give access to their personal data in exchange for better knowledge of the disease they are affected by and better access to, and management of, relevant treatment. Examples include the Parkinson's Genetics initiative and the electronic platform PatientsLikeMe<sup>39</sup> which counts among its partners a number of pharmaceutical companies (including Merck, Novartis). The first example was an ambitious US-based project involving the Michael J Fox foundation, the Parkinson's Institute and the genomics company 23andme<sup>40</sup> which pooled resources to mobilise a cohort of patients whose DNA was then analysed. It took eight months to produce some of the research outputs, a timeline that was considered much shorter than the average research project length in the same field. As for PatientsLikeMe, this for-profit social networking platform enables patients to submit online their personal and clinical data that will later be sold to a range of organisations in exchange for the opportunity to engage with patients similar to them, and access to information that could help them manage their condition. Both projects demonstrate the eagerness of some patients to give access to data provided they see their interest in doing so. Alongside those established organisations, other recent examples of citizen-centred data initiatives and start-ups contribute to accelerate the expansion of a data market where patients, via apps, actively choose to share their data.<sup>41</sup>

### *3.4.3. Designing alternative approaches to consent management within the limits of regulatory frameworks*

#### 'Liberal' national data frameworks

A few European countries have made anonymised personal data available to researchers without prior informed consent. In Nordic countries and Belgium for instance, identifiable personal data can be shared and processed without prior consent (OECD, 2013). In the UK, the NHS data sharing platform (care.data)<sup>42</sup> intends to make 'pseudonymised' patients data (including information consultations, notes on prescriptions, but also information regarding mental health condition, drinking and smoking habit as well as diseases) from GPs and hospitals available to private companies after approval of the Health and Social Care Information Centre. The scope of use of these data is quite broad and includes health intelligence, health improvement, audit, health service research and service planning (Ramesh, 2014). However, patients' consent remains mandatory for the processing of sensitive data in most countries (Cheuk et al., 2013). In that case, the development of efficient consent management tools can facilitate

---

<sup>39</sup> Patientslikeme, <http://www.patientslikeme.com/>. Last accessed 05/12/2013.

<sup>40</sup> 23andme, <https://www.23andme.com/> Last accessed 05/12/2013.

<sup>41</sup> See for instance the healthpump developed by Pumpco, <http://pumpco.co.uk/healthpump.html> Last accessed 27/02/2013.

<sup>42</sup> Care.data – modern data service for the NHS, <http://www.england.nhs.uk/ourwork/tsd/care-data/> (last accessed 20/01/2014).



data collection while complying with privacy regulations and issues of consent (especially when informed consent is mandatory).

### New consent management tools

Engaging directly with the patients and incentivising data sharing in exchange for personalised healthcare services is an option that is likely to be further explored in the near future, and will partly be facilitated by the use of social media. Bearing that in mind, innovative solutions have been developed to secure access to RWD while making sure that privacy protection is guaranteed. The development of electronic consent management suites or “eConsent” systems (O’Keefe et al., 2006) for instance is key in enabling access to RWD, especially when informed consent is mandatory. If several academic papers are referring to this issue and are investigating methodologies to build efficient consent management systems (Kluge, 2004; Heinze et al., 2013; Khan, 2013), their effective implementation remains at an infant stage (Box 6). Yet, they may represent a more efficient way to guarantee access to RWD while complying with existing privacy protection regulation.

### **Box 6 Example of consent management tool in Germany**

The implementation of a Regional Health Information Network in Germany (Rhine-Neckar Region) has led to the development of methodological tools to equip healthcare institutions with opt-in consent management solutions. This consent management tool proposes two different services. The Consent Management Service stores information about existing patients and answers any queries they have regarding their consent status. Second, the Consent Creator Service is an electronic tool that enables new patients to create consent. This new tool would reinforce interoperability between existing consent management platforms through the use of international standards, the overall objective being to expand its use to other regions in order to facilitate the integration of consent management platforms at the national scale.

Source: Heinze et al., 2011

### 3.5. Summary of findings: main barriers to access to and use of RWD and strategies to overcome them

The following table provides a brief overview of the study's findings. It summarises the main barriers to access to and use of RWD and presents the types of strategies that are implemented to improve RWD access and use.

**Table 15 Improving access to and use of RWD: from barriers to enablers**

Barriers	Strategies to improve access and use
<i>Content and quality issues</i>	
Terminology issues	Standardisation of codes
Incomplete data	Longitudinal data collection
Data quality issues	Processes for data quality assurance
<i>Methodological barriers</i>	
Limited analytics capabilities	National eHealth strategies
Lack of analytical standards	Transnational and multisectoral coalition of experts
Linkage challenges	European projects and best practices
Fragmentation	Interoperable systems
<i>Governance structures</i>	
Lack of clear pathway	Buying the data
Access granted to academics only	Hiring/partnering with academics
Lack of data controller engagement	Incentives for clinicians
<i>Privacy practices</i>	
Ethical concerns among professionals	Trusted third party, depersonalisation tools
Ethical concerns among the public	Communication campaigns
Consent management	Liberal national strategies and innovative consent management tools



## 4. Conclusion

---

### 4.1. Summary of findings

By investigating current forms and uses of RWD in Europe (Chapter 2), this study has highlighted their significant potential for assessing the (short- or long-term) impact of different drugs or medical treatments and for informing and improving healthcare service delivery. Although the potential of RWD use seems quite clear, this research reveals barriers that limit further development towards the full exploitation of this type of data. We have identified factors limiting the potential benefits driven from RWD analysis. These include the absence of common standards for defining the content and quality of RWD (absence of common terminology, incomplete datasets, lack of data quality assurance systems) and several methodological barriers (absence of standards for RWD analysis and for data linkage). Access to this type of data is also restricted by the lack of governance standards including the absence of standards for collaboration between stakeholders active in the field of RWD, and by the limitations of incentives for data sharing. Finally, privacy concerns (eg ethical concern among health practitioners) and privacy practices (eg personal data protection regulation) influence the amount of data available and the scope of its use.

These issues are being addressed – although in a somewhat uneven fashion – by current initiatives from both public and private stakeholders at the regional, national and European scale (Chapter 3). For example, the issues of data quality are being tackled through European and international initiatives aiming to improve the standardisation of terminology (eg ORPHANET in the field of rare disease). The development of international research coalitions is also facilitating knowledge and best practice sharing and accelerating the development of common frameworks that guide RWD collection and use. These initiatives thus contribute both to the improvement of data quality and to researchers' analytical capabilities. In addition, a strong push towards the development of EHRs – and eHealth infrastructures more broadly – has been observed in some European countries (Nordic countries, France, Belgium, the UK) and has been actively supported by the EU over the last six years through various schemes (eg FP7 Health Programme). Such initiatives offer great potential for the automated and routine collection of patient data. Finally, access – which is related to both governance and data protection issues and problematic mostly for private companies – can be granted through the implementation of strategic partnerships among stakeholders (eg engaging directly with academics on specific research projects, with physicians in exchange of technological and analytical services, but also with data vendors) and the development of online consent management architecture.

## 4.2. Understanding the role of RWD in the current political, societal and economic context

The growing role of RWD in developing an understanding of certain diseases and improving the quality of care calls for a more systematic analysis of the Policy, Economic, Social, Technological and Legal domains in relation to RWD. Although we are unable to fully develop this in the context of this report, in this section we synthesise the insights gained from the evidence along the lines of a PESTL analysis and use this framework to summarise some of the main issues in each category. In each category we outline the main drivers, enablers, barriers and alternative approaches to use of RWD that have emerged from the analysis. Drawing on evidence and analysis in this report we also make some assessments about trends and key issues. Table 16 summarises the five areas of analysis. We hope that use of the PESTL categories here may assist in identifying issues that could be addressed in further studies and initiatives.

**Policy environment:** The European Commission has substantially engaged with RWD promotion and development, while individual member states also demonstrate awareness of the need for RWD standards by developing strategies to improve information infrastructure and enable data use.

The main driver behind the development of an EU-level approach to RWD appears to be the European Commission's push for the development of eHealth infrastructures and use of EHRs. The European Commission has been particularly active in the development of methodological standards to facilitate the collection and use of patient data. The European Health Strategy for the 2008–2013 programming period has put emphasis on the development of eHealth infrastructures and on the funding of research projects aiming to promote the adoption of international terminology and coding standards to enable data sharing and international comparisons (European Commission, 2007).

Another important driver is the increased support for the development of EU-wide datasets and enhanced interoperability. Enhancing interoperability between European datasets constitutes part of the Strategy objectives (Objective 3 in European Commission, 2007). Cross-country research projects are actively supported by different EU programmes to link existing registries, develop new ones and pool resources, paving the way to more standardised strategies for the collection and use of RWD. These recent developments are creating new opportunities for research through improved data collection and enhanced interoperability; however they remain quite fragmented, potentially hindering the pace of scientific advances. Reflexions on data harmonisation, data linkage and interoperability are therefore taking place at the European level, aiming to give guidance on data harmonisation.

At the same time, another strong driver, related to the first one, can be identified in the national-level reforms that are taking place in the individual EU member states. National healthcare reforms are aiming for greater efficiency in service management and provision, with national and regional strategies to develop eHealth making provisions from the technological, methodological and governance perspectives.

All of the above-mentioned initiatives can count on the availability of funding through dedicated budget lines and strategic programmes, besides national public and private sector resources. However, progress on the objectives of interoperability, infrastructure creation and data access could be influenced by data protection legislation, at the European level in particular. Furthermore, the persistence of national

approaches to health reform, an area in which there is a limited EC mandate, could jeopardise harmonisation (although not interoperability) objectives. Finally, the existing disparities in national eHealth systems, coupled with governance issues regarding the design and implementation of RWD standards, constitute two likely persistent obstacles to Pan-European coverage by any initiative taken in the direction of a facilitated data collection, analysis and access.

*Economic context: Leveraging cross-sector and cross-country partnerships could enable stakeholders to take advantage of the development of data markets.*

The healthcare sector has experienced severe cost and productivity challenges in recent years, requiring developing new solutions to optimising limited financial resources. Optimising pathways and tools for interpreting the available evidence in order to increase the system's efficiency is therefore crucial. Limited resources, together with the existence of significant economies of scope and scale inherent in larger datasets, have created a context which incentivises collaboration to pool resources across and within sectors. Finally, the development of a market for data in some European countries has enabled the marketisation not only of data but also of analytical tools and services.

These developments appear to be further enabled by new synergies within the data value chain, for instance between general practices and databases developers and entities active in biomedical research fields. While national and European level initiatives aim at the creation and input of data.

However, despite the market-driven growth and increasing sophistication of RWD tools and analysis, controversies surrounding the definition of cost sharing mechanisms across the value network and fragmented markets presenting highly different characteristics could slow down its development.

The routine collection of publicly available data to create databases that can form the basis of multiple analysis could contribute to mitigate financial risks, as could a more efficient leveraging of existing partnership structures: for instance increasing collaboration between industry and academia by funding studies working with data from sources available to academic researchers; or intensifying participation in research-oriented consortia to leverage economies of scope and scale in access and analysis, such as the EU-ADR project. One further promising area appears to be in the area of disease-specific research projects, which often collect data directly from the patients.

*Social factors: There is growing support for tools that help address the needs of patients, but security concerns are likely to determine scope of RWD use.*

There are several factors that suggest that in the future the social support for data-powered health solutions will increase. Policymakers and the media are conscious about the challenges raised by the rise of chronic conditions and an ageing society in terms of social and healthcare and costs. With increasing concerns over dependency ratios (less available young workforce available to take care of older members of society), data-driven technological solutions to social and health problems such as telemedicine, personal monitors and home automation, are increasingly accepted (May et al., 2011). Enthusiasm for advances in medicine that offer new cures for illnesses and a trend towards leveraging genetics and big data to find personalised solutions for individual patients also support the public acceptance of data use. Social acceptance is further enabled by flagship initiatives regarding some enthusiastic patient groups that

advocate for appropriate leverage of data-powered tools to improve their members' circumstances, in particular patients with rare diseases. Finally, practitioners who are enthusiastic about tools that improve outcomes and processes for patients are likely to be supportive of RWD use. Favourable media coverage of technological progress, inventions and data analytics reinforce these trends.

At the same time, there is some evidence suggesting that social support could be lessened by increased media coverage of high-profile information security and data breach cases that could potentially undermine patients' trust in electronic records of personally sensitive information (Rynning, 2007). In particular, incidents related to re-identification of patients, which is made possible by the linking of datasets even if data managers adhere to anonymisation guidelines; and an increasing awareness of benefits accrued to private sector companies through the use of personal data (facilitated by compulsory informed consent procedures), could jeopardise the uptake of RWD tools and services.

While information security and privacy best practices are likely to become even more important components within private companies strategies, social acceptance can also be supported by efficient communication regarding the benefits of RWD research and its applications.

*Technological advances: Technology has been and will likely remain the key enabler in extracting value from RWD.*

Recent advances in computing power, storage, data integration and analytics have facilitated the collection and effective use of high volumes of data (Cate et al., 2013). The generation and dissemination of new types of RWD will likely continue, driven by the emergence of ever more instrumented, interconnected and intelligent devices, instruments such as natural language processing and the introduction of advance analytics to gain insight through improved dataset linkage (IBM, 2013). This trend is further supported by the progressive uptake of potential RWD tools in a number of European countries, such as national/patient ID systems; and health-related social media and mobile applications, which can be a fertile ground for self-reported patient data collection.

Technology-related progress in RWD analytics at present seems to be limited by two main factors: on the one hand by the maturity of the technology, for example by the limits of the analytical and processing power of software and hardware. On the other hand, the inconsistency of existing databases and poor development of data quality assurance standards also constitute limitations to the potential of technological tools (ABPI, 2013; McKinsey, 2011).

Approaches to overcome these barriers could lie in the increased use of big data or through learning from other sectors that routinely manipulate unstructured data. Alternatively, there may be potential in intensifying the collaborations between the companies conducting or utilising RWD and those manufacturing the analytical tools, in an effort to steer advances in technology development in a direction that best supports the specific interests of the health and healthcare sector.

*Legal framework: Data protection regulation and associated issues such as privacy and ownership can restrict access.*

The rising interest in patient data from a variety of stakeholders poses inevitable issues around regulation of access to patient-level data, in relation to the concepts of privacy, consent and data security. A recent

warning letter by the US Food and Drug Administration (FDA) to the personal genome service firm 23andme (FDA 22/11/2013)<sup>43</sup> questioning the use of and claims made for the genome tests provided and highlighting the risks of directly sharing results with the patients. The FDA judgment raises a host of questions about the scope of regulatory authority and there has been extensive comment in US and European media about the decision. It should be noted that the decision seems likely to have been based on concerns about the claims and inferences being made on the basis of data rather than on objection to collection of data in and of itself.<sup>44</sup> It is also the case that the scope of the FDA and European Medicines Agency (EMA) differ and therefore they may not act in similar ways. Furthermore, regulatory controls in terms of accountability and transparency from data controllers and users will likely be tightened in the coming years, as illustrated by the proposed EU Data Protection Regulation.

At the same time, the benefits of using RWD to improve health services efficiency might influence existing regulation to facilitate data access in order to operate more efficient trade-offs between privacy and healthcare delivery. Some of the additional burdens in terms of consent management could be balanced out by technological advances and consent management tools reducing the burden placed on physicians.

Overall, access to RWD is highly dependent on but can be facilitated by stakeholders' ability to implement rigorous ethics framework that demonstrate awareness of the regulatory environment and the development of suitable consent management tools. This could be further supported by a strategy of systematically publishing RWD-based research findings, which would constitute an efficient and low-cost tool to enhance transparency and accountability while illustrating the public value of data collection.

---

<sup>43</sup> Food and Drug Administration Warning letter to 23andme <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm> Last accessed 09/12/2013.

<sup>44</sup> The FDA and me: Medical testing firms find it is in their interest to cooperate with regulators. *Nature*, 5 Dec 2013, Vol. 504, p. 7



**Table 16 PESTL analysis of the RWD landscape in Europe**

Area	Drivers	Enablers	Barriers	Alternatives
Policy	<p>EC’s push for the development of eHealth infrastructures and use of EHR</p> <p>EC’s drive for the creation of Pan-European datasets and improved interoperability</p> <p>National healthcare reforms aiming to greater efficiency in service management and provision.</p>	<p>EU funding instruments</p> <p>Regional and National data infrastructure.</p>	<p>EC’s data protection regulation</p> <p>Fragmentation of national approaches to health reform</p> <p>Disparities between national eHealth systems</p> <p>Governance issues regarding the design and implementation of RWD standards.</p>	<p>Reliance on data collected in countries with easiest rules for access</p> <p>Involvement in EU-funded research projects in partnership with relevant public and private stakeholders.</p>
Economic	<p>Resources constraints and need to develop efficient pathways to analysis</p> <p>Incentives for collaboration to pool resources</p> <p>Development of a market for data.</p>	<p>New synergies within the data value chain (eg with insurers)</p> <p>National authorities encouraging data input.</p>	<p>Fragmented markets presenting different characteristics</p> <p>Issues surrounding cost sharing for data access and use</p> <p>Conflicts of interest.</p>	<p>Routine collection of publicly available data</p> <p>Funding to academia for research in databases</p> <p>Participation in research-minded consortia to spread the cost of data access and analysis</p> <p>Engagement in disease specific research projects with direct access to self-reported patient data.</p>
Social	<p>Increased familiarity with sharing data</p> <p>Increased attention to</p>	<p>Positive media coverage</p> <p>Interaction with</p>	<p>Increased suspicions about data use and potential breaches</p>	<p>Development of personalised and stratified health</p>

Area	Drivers	Enablers	Barriers	Alternatives
	the burden of a chronically ill and ageing society Enthusiasm for new cures for illnesses Willingness to access personalised health services.	stakeholders (eg rare disease groups) Practitioners care about improving outcomes for patients.	Privacy risks due to linking different datasets Regulation surrounding consent management Image problem of pharmaceutical companies or insurers.	services offer Communication around the positive effects of RWD-based research.
Technological	Increased technological capabilities for data storage and analysis Increasing capacity to link distinct datasets Push towards standardisation of terminologies.	Machine learning, including natural language processing National/patient identifier systems Social media and apps for self-reported data collection.	Limits of analytical capabilities for the treatment of data Inconsistency of existing databases and limited development of data quality insurance standards	Leveraging methods and tools developed in other sectors Exploration of the potential of apps/PPPs with device manufacturers.
Legal	EU level and national level debate on data protection, use and access.	Potential of using RWD to improve health services efficiency might influence existing regulation to facilitate data access Technological advances reduce the burden of work for consent documentation collection.	Privacy and data protection likely to be strengthened Ethical standards for research Fragmented standards for access to databases.	Efforts on transparency and ethical commitments Publication of RWD-based research results.

Strategic partnerships: shaping the future of the RWD landscape.

In this landscape, strategies that seek to optimise RWD access and use have to align interests of the three parties typically involved in healthcare: the organisations (payers, providers), the professionals (clinicians) and the patients. Indeed, disproportionate advantage, or disadvantage between the three generally leads to slow adoption or refusal to change custom and practice. We therefore think that strategic partnerships between those stakeholder groups are key to defining better routes to access and improved use of data.

These partnerships could fall into three main categories:

- **Enhanced collaboration with European or national public organisations** (eg Public Health Agencies, National Health Data Repositories, research institutes) **and/or payers would guarantee access to a large amount of data and a high coverage of the population.** Defining pathways to accessing such data, which would include EHRs, would allow a variety of research applications. Besides, collaborating with public bodies might also allow relevant stakeholders to get access to non-medical data that could be linked to medical datasets, such as environmental and socio-economic data (eg Eurostat data).
- **Strategic partnerships with some care providers, charities and patient groups would secure access to disease specific data.** Access to these data is dependent on the development of effective and harmonised solutions for data collection that would minimise the data input burden and incentivise data controllers, namely care providers and patient organisations. Indirect incentives provided by both public and private stakeholders through support to the development of eHealth infrastructures and standards for data collection, enabling the delivery of better outcomes for both the patients (eg better care, patient safety) and the health and healthcare organisations (eg efficiency of services, reduced waste) could be developed. They would include investment in data processing tools and systems that are deemed to enhance the efficiency of processes of healthcare systems (eg electronic platforms, user-friendly software). This strategy would be particularly appropriate for research on diseases such as diabetes or breast cancer, with strong patient networks and vocal patient associations interested in advancing research.
- **Key partnerships with the data owners (patients) would also involve the development of relationships with new players in the RWD landscape (eg IT companies, software and app developers, social medias)** to facilitate data collection and the development of suitable consent management tools to ensure that anonymity and data protection is guaranteed. Such a strategy would allow access to a rather disparate and messy pool of data, although the large quantity of data points collected could compensate the relatively low quality of the data. As a result, efforts should also be dedicated to the development of data quality assurance standards; the progressive standardisation and systematisation of data collection mechanisms is a first step towards the creation of new datasets that could then be used for research. In that framework, incentives for data sharing could be either monetary (eg small amount of money in exchange for medical data) or non-monetary (eg data is shared in exchange for personalised healthcare services) and data could be used, for example, to research healthy lifestyle behaviours, or diet.

A variety of collaborations can be developed to overcome existing barriers and facilitate RWD access and use, depending on the kind of data that is needed and the scope of their use. Those partnerships would rest on both non-monetary and monetary agreements and leverage a broad range of incentives at different

levels of the health systems, from patient-level initiatives to collaborations with national health organisations.



## References

---

- Association of the British Pharmaceutical Industry (ABPI), *Big Data Road Map*, London: 2013.
- Augustin, M., K. Reich, G. Glaeske, I. Schaefer and M. Radtke, “Co-morbidity and age-related prevalence of psoriasis: analysis of health insurance data in Germany”, *Acta Dermato-Venereologica*, Vol. 90, No. 2, 2010, pp. 147–151.
- Aymé, S. and C. Rodwell, “2013 Report on the State of the Art of Rare Disease Activities in Europe”, Scientific Secretariat of the European Union Committee of Experts on Rare Diseases (EUCERD), 2013.
- Bakerly, N. D., A. Woodcock, J. P. New, M. Gibson, W. Wu, D. Leather and J. Vestbo, “The Salford Lung Study Protocol: A Prospective, Randomized Real-World Effectiveness Trial In Chronic Obstructive Pulmonary Disease (COPD)”, Dollars and sense: identifying the costs, utilisation and burden of respiratory disease management, *American Thoracic Society International Conference*, 2013.
- Banerjee, A., D. Lane, C. Torp-Pedersen and G.Y Lip. "Net clinical benefit of new oral anticoagulants (dabigatran, rivaroxaban, apixaban) versus no treatment in a 'real world' atrial fibrillation population: A modelling analysis based on a nationwide cohort study." *Thrombosis and Haemostasis*, Vol. 107 No. 3, 2012, pp. 584–589.
- Blandin, G., L. Rufibach, B. Von Rekowski, C. Beroud and M. Krahn, “A new international web-based registry for Dysferlinopathy involving participation of patients and their doctors”, Background paper, 2<sup>nd</sup> EPIRARE International Workshop, 2013.
- Boggon, R., T. P. van Staa, A. Timmis, H. Hemingway, K. K. Ray, A. Begg, C. Emmas and K. A. Fox, “Clopidogrel discontinuation after acute coronary syndromes: frequency, predictors and associations with death and myocardial infarction—a hospital registry-primary care linked cohort (MINAP-GPRD)”, *European Heart Journal*, Vol. 32, No. 19, 2011, pp. 2376–86.
- Caldicott, F., *Information: to share or not to share? The Information Governance Review*, London: Department of Health, 2013.
- Cate, F., H., P. Cullen and V. Mayer-Schonberger, *Data protection principles for the 21<sup>st</sup> Century: Revising the 1980 OECD Guidelines*, Organisation for Economic Cooperation and Development (OECD), 2013.
- Chao, C., W. Jen, Y. Chi and B. Lin, “Improving patient safety with RFID and mobile technology”, *International Journal of Electronic Healthcare*, Vol. 3, No. 2, 2007, pp. 175–192.

Chataway J., C. Fry, S. Marjanovic and O. Yaqub, "Public–private collaborations and partnerships in stratified medicine: making sense of new interactions", *New Biotechnology*, Vol. 29, No. 6, 2012, pp. 732–740.

Cheuk, A., J. Halpert, K. Lucente, P. McCormack and S. Thiel, *Data Protection: Laws of the World*, DLA Piper, 2012.

Coloma, P. M., G. Trifirò, M. J. Schuemie, R. Gini, R. Herings, J. Hippisley-Cox, G. Mazzaglia, G. Picelli, G. Corrao, L. Pedersen, J. van der Lei and M. Sturkenboom, "Electronic healthcare databases for active drug safety surveillance: is there enough leverage?", *Pharmacoepidemiology and Drug Safety*, Vol. 21, No. 6, 2012, pp. 611–621.

Cour des Comptes, *Le Coût du dossier médical personnel depuis sa mise en place*, Paris: Cour des Comptes, 2012.

Department of Health, *The NHS Constitution: the NHS belongs to us all*, London: Department of Health, 2013.

Eichler, H. G., E. Abadie, A. Breckenridge, H. Leufkens and G. Rasi, "Open Clinical Trial Data for All? A View from Regulators", *PLoS Med*, Vol. 9, No. 4, 2012.

Eichler, H. G., F. Pétavy, F. Pignatti and G. Rasi, "Access to patient-level trial data – a boon to drug developers", *New England Journal of Medicine*, Vol. 369, No. 17, 2012, pp. 1577–1579.

van Engen-Verheul, M., H. Kemps, R. Kraaijenhagen, N. de Keizer and N. Peek, "Cardiac rehabilitation uptake and its determinants in the Netherlands", *European Journal of Preventive Cardiology* Vol. 20, No. 2, 2013, pp. 349–356.

European Commission, *e-Health – Making Healthcare Better for European Citizens: An action plan for a European e-Health Area*, 2004.

European Commission, *Together for Health: A strategic approach for the EU 2008–2013*, White Paper, 2007.

European Commission, *Attitudes on Data Protection and Electronic Identity in the European Union*, Special Eurobarometer 359, Brussels: European Commission, 2010.

Farrell, Joseph, and Garth Saloner. "Standardization, compatibility, and innovation." *The RAND Journal of Economics*, 1985, pp. 70–83.

Fears, R., H. Brand, R. Frackowiak, P. Pastoret, R. Souhami and B. Thompson, "Data protection regulation and the promotion of health research: getting the balance right", *QJM: Monthly Journal of the Association of Physicians*, Vol. 107, No 1, 2013.

Fischer-Hubner, S., and Fisher-Hubner, G., "It-Security and Privacy: Design and Use of Privacy-Enhancing mechanisms", Issue 1958, *Berlin Springer*, 2001.

Franchi, C., G. Giussani, P. Messina, M. Montesano, S. Romi, A. Nobili, I. Fortino, A. Bortolotti, L. Merlino and E. Beghi, "Validation of healthcare administrative data for the diagnosis of epilepsy", *Journal of Epidemiology and Community Health*, Vol. 67, No. 12, 2013, pp. 1019–1024.

Froomkin, A. M., "The Death of Privacy?", *Stanford Law Review*, Vol. 52, No. 5, 2000, pp. 1461–1543.

Garrison Jr, L. P., P. J. Neumann, P. Erickson, D. Marshall and C. D. Mullins, "Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report", *Value in Health*, Vol. 10, No. 5, 2007, pp. 326–335.

Global alliance to Enable Responsible Sharing of Genomic and Clinical Data, *Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data*, White Paper, 2013.

Gonzalez-Ruiz, A., A. Beiras-Fernandez, H. Lehmkuhl, R. A. Seaton, J. Loeffler and R. L. Chaves, "Clinical experience with daptomycin in Europe: the first 2.5 years", *Journal of Antimicrobial Chemotherapy*, Vol. 66, No. 4, 2011, pp. 912–919.

Gostin, L. O. and J. G. Hodge, "Personal privacy and common goods: A framework for balancing under the national health information Privacy Rule", *Minnesota Law Review*, Vol. 86, 2002, pp. 1439–1479.

Hawn, C., "Take Two Aspirin and Tweet Me in the Morning: How Twitter, Facebook, And Other Social media are Reshaping Healthcare", *Health Affairs*, Vol. 28, No. 2, 2009, pp. 361–368.

Heinze, O., M. Birkle, L. Koster and B. Bergh, "Architecture of a consent management suite and integration into IHE-based regional health information networks", *BMC Medical Informatics and Decision Making*, Vol. 11, No. 1, 2011.

Henderson S, S. Lin, H. Fraser, P. Lindell, T. Yu, "Collaborative innovation: Partnering for success in Life Sciences", IBM Institute for Business Value, Somers (NY): IBM, 2010.

HIMSS Privacy and Security Committee, *Social media in healthcare: Privacy and Security considerations*, HIMSS White Paper, 2012. As of January 21, 2014: <http://www.himss.org/ResourceLibrary/GenResourceReg.aspx?ItemNumber=17244>

Holtorf, A. P., J. B. Watkins, C. D. Mullins and D. Brixner, "Incorporating observational data into the formulary decision-making process—summary of a roundtable discussion", *J Manag Care Pharm*, Vol. 14, No. 3, 2008, pp. 302–08.

IBM, *Healthcare 2015 and Personal Health Records: A Standards Framework*, Somers (NY): IBM Global Services, 2008.

IBM Institute for Business Value, *Analytics: The real-world use of big data in healthcare and life sciences. How innovative healthcare and life sciences organisations extract value from uncertain data*, Somers (NY): IBM Global Services, 2013.

Janzek-Hawlat, S., E. Ammenwerth, W. Dorda, G. Duftschmid, W. Hackl, A. Hörbst and W. Gall, "The Austrian e-Medikation Pilot Evaluation: Lessons Learned from a National Medication List", In: *MEDINFO 2013: Proceedings of the 14th World Congress on Medical and Health Informatics*, IOS Press, Vol. 192, 2013.



Juliusson, G., V. Lazarevic, A. S. Hörstedt, O. Hagberg and M. Höglund, “Acute myeloid leukemia in the real world: why population-based registries are needed”, *Blood*, Vol. 119, No. 17, 2012, pp. 3890–3899.

Lothgren, M., E. Ribnicsek, L. Schmidt, W. Habacher, J. Lundkvist, A. M. Pfeil and A. Bracco, “Cost per patient and potential budget implications of denosumab compared with zoledronic acid in adults with bone metastases from solid tumours who are at risk of skeletal-related events: an analysis for Austria, Sweden and Switzerland”, *European Journal of Hospital Pharmacy: Science and Practice*, 2013.

Marconi, J., “E-Health: Navigating the Internet for Health Information Healthcare”, Advocacy White Paper. Healthcare Information and Management Systems Society, May, 2002.

May, C.R., et al. "Integrating telecare for chronic disease management in the community: what needs to be done?" *BMC Health Services Research* 11.1, 2011, 131.

McKinsey, *How Can Pharmacos Take Advantage of the Real-World Data Opportunity in Healthcare?*, McKinsey IT Insights, 2011.

Moat, N. E., P. Ludman, M. A. de Belder, B. Bridgewater, A. D. Cunningham, C. P. Young and M. J. Mullen, “Long-term outcomes after Transcatheter aortic valve implantation in high-risk patients with severe aortic stenosis”, *Journal of the American College of Cardiology*, Vol. 58, No. 20, 2011, pp. 2130–2138.

Morgan Jones, M., A. Hall, D. Brooker, S. Castle-Clarke, E. Winpenny, D. Jahagirdar, J. Exley & J. Chataway, “The future of public health: a horizon scan”, Santa Monica: RAND Corporation, *Forthcoming*.

National Institute for Health Research, *NIHR Information Strategy: Turning Data into Information to enable insight*, Southampton: NIHR, 2013.

OECD, *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, Paris: OECD Publishing, 2013.

Ploem, M. C., M. L. Essink-Bot and K. Stonks, "Proposed EU data protection regulation is a threat to medical research", *BMJ*, Vol. 346, 2013.

Price, D., Y. Asukai, J. Ananthapavan, B. Malcolm, A. Radwan and I. Keyzor, "A UK-based cost-utility analysis of indacaterol, a once-daily maintenance bronchodilator for patients with COPD using real world evidence on resource use", *Applied Health Economics and Health Policy*, Vol. 11, No. 3, 2013, pp. 259–274.

Rabesandratana, T., “EU Privacy Protection Bill would hamper research, scientists warn”, *Science* Vol. 343, No. 6174, 2014, pp. 959–960.

Ramesh, D. “NHS patient data to be made available for sale to drug and insurance firms”. The Guardian, 19 January 2014. As of 20 January 2014:

<http://www.theguardian.com/society/2014/jan/19/nhs-patient-data-available-companies-buy>

Raviprakash A. V., B. Prabu, N. Alagumurthi and V. Soundararajan, RFID: Rx to Healthcare Industry, *ICFAI Journal of Supply Chain Management*, Vol. 6, No. 2, 2009, pp. 7–25.

Rynning E. Public Trust and Privacy in Shared Electronic Health Records. *European Journal Of Health Law* [serial online]. Vol. 14, No. 2, July 2007, pp. 105–112.

Robinson, N., Horvath, V., Roosendaal, A., Cave, J., “Data and Security Breaches and Cyber-Security Strategies in the EU and its International Counterparts”, Brussels: European Parliament, Directorate-General for Internal Policies, 2013.

Roggeri, D. P., A. Roggeri, E. Rossi, E. Cinconze, M. De Rosa and A. P. Maggioni, “Direct healthcare costs and resource consumption after acute coronary syndrome: a real-life analysis of an Italian subpopulation”, *European Journal of Preventive Cardiology*, 2013, pp. 1–7.

Santi P., "Le dossier médical personnel, très mal ficelé", *Le Monde*, 27 August 2012.

Smeets, H. M., et al. "Routine health insurance data for scientific research: potential and limitations of the Agis Health Database." *Journal of Clinical Epidemiology* Vol. 64 No. 4, 2011, pp. 424–430.

Stroetmann, K. A., J. Artmann and V. N. Stroetmann, *European countries on their journey towards national eHealth infrastructures: Final European Progress Report*, Brussels: DG INFSO and eHealth Strategies, 2011.

Thayer, S., W. Wei, E. Buysman, L. Brekke, W. Crown, M. Grabner and R. Cuddihy, “The INITIATOR Study: Pilot Data on Real-World Clinical and Economic Outcomes in US Patients with Type 2 Diabetes Initiating Injectible Therapy”, *Advances in Therapy*, Vol. 30, No. 12, 2013, pp. 1128–1140.

The Danish Government, *Making eHealth Work: National strategy for digitalisation of the Danish healthcare sector 2013–2017*, Copenhagen: The National ehealth Authority, 2013.

Thorp, J., “Europe’s E-Health Initiatives: An Overview of European Interoperability Initiatives”, *Journal of AHIMA*, Vol. 81, 2010, pp. 56–58.

Wan-Chu, W., “Ethical Risks Inhabited in Social Networking Sites: A case study on patientslikeme.com”, IACSIT Press, Singapore, Vol. 45, 2012.

West, D., M., “Improving Health Care through Mobile Medical Devices and Sensors”, Centre for Technology and Innovation at Brookings, 2012.

Westin F. A., *Computers, health records, and citizen rights*, University of Michigan, 1976. As of 21 January 2014:

[http://eric.ed.gov/ericwebportal/custom/portlets/recorddetails/detailmini.jsp?\\_nfpb=true&\\_ericextsearch\\_searchvalue\\_0=ED143358&ericextsearch\\_searchtype\\_0=no&accno=ED143358](http://eric.ed.gov/ericwebportal/custom/portlets/recorddetails/detailmini.jsp?_nfpb=true&_ericextsearch_searchvalue_0=ED143358&ericextsearch_searchtype_0=no&accno=ED143358)

Williams, T., T. Van Staa, S. Puri and S. Eaton, “Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource”, *Therapeutic Advances in Drug Safety*, Vol. 3, No. 2, 2012, pp. 89–99.

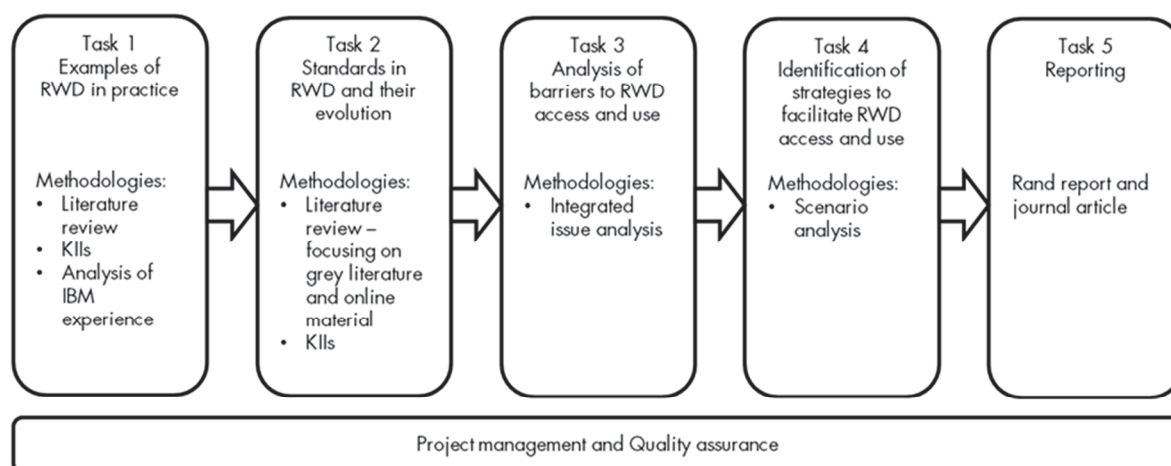
Zolin, A. and W. Gulmans, “Improving quality of European Cystic Fibrosis Society Patient Registry Data”, Working paper presented at the 2<sup>nd</sup> EPIRARE International Workshop, 2013.

## Annex A Research approach

---

To review the evidence base and define potential future directions of development for RWD access and use, the research team has used an approach based on multiple methods. In this section we briefly outline the approach (summarised in Figure A.1) and the ways in which the methodologies supported each of the tasks within the project.

**Figure A.1 Summary of research approach**



In the first phase of the study, two parallel workstreams aimed to explore the available evidence on current examples of use of real-world data in practice on the one hand and standards for RWD and their evolution on the other. These two tasks were supported by a review of the available academic and grey literature through a structured approach covering academic and grey web-based databases (the search strategy and outcomes are outlined more in detail in Annex B of this report). The study team screened a total of 935 sources and reviewed in depth a total of 43 articles.

Insights from the literature were complemented by information collected through a total of 10 in-depth semi-structured interviews with stakeholders and policymakers (the list of interviewees' affiliations can be found in Annex D of the report). The interviews and the literature reviews mutually provided insights for the other workstream by indicating additional references and providing a background to the interview protocols.

In addition to these two activities which supported both workstreams, a series of case studies were developed, drawing on IBM's expertise and track record with RWD applications to illustrate the current situation regarding the use of RWD in the healthcare and life sciences sector and other industries (see case studies overview in Annex C).

Building on the insights from the first two sets of activities, in the second phase of the research the study team performed an integrated issues analysis to interpret outcomes from the literature review, the interviews and the case studies. This approach consisted of mapping issues and underlying drivers in a logical sequence and ensured that all elements of the problem were kept present and that strategic options defined by the process were relevant. This exercise also mapped options and pathways to developing and implementing standards around RWD and fed into the PESTL analysis presented in the discussion section.

In the fourth part of the project, the study team developed a set of scenarios in order to test different data access strategies and to explore ways to engage with the key stakeholders in each of these possible strategies. Developing scenarios served as a consistency check on the analysis conducted in the first half of the project while directing attention to the main drivers of a possible future. These were fine-tuned through a workshop in which researchers worked closely with the client and concentrating on the drivers of possible future developments. In this study, the definition of the scenarios was structured along the strategic partnerships that would characterise each of these potential futures, and on the type of access to data (whether it is dominated by centralised or decentralised forms of data control). The analysis pinpointed the most important stakeholders and outlined the salient risk factors as well as potential limitations to data access and use. The three scenarios illustrate polarised and naive visions of the future. Thinking about developing coherent narratives around each of these extreme pathways forced the team to establish a hierarchy of factors that are likely to shape future strategies.

In the closing stage of the project, insights and supporting evidence were summarised in the present research report.

## Annex B Literature review: search strategy and extraction tables

The review of the academic and grey literature was performed using a rapid evidence assessment approach. This section of the report provides an overview of the search and review activities.

### Search strategy

To ensure a wide coverage of the available academic literature, five different online academic databases have been searched (Pubmed, EBSCO, Scopus, Opengrey and Google Scholar). After piloting a range of search term strings, the team opted to use very broad search terms to limit inappropriate selection and bias. Search results were also limited by language (English) and year of publication (after 2008).

Four out of the five databases screened yielded limited (less than one hundred per query) results. These search outputs have been reviewed in their entirety. The fifth database, Google Scholar, was the only one producing a high number of results (>10,000). Out of this sample, only the first 200 references for this database (sorted by relevance) have been downloaded to the reference manager EndNote. Table B.1 provides a summary of the search terms and screening activities.

**Table B.1 Search strategy and outputs**

Date	Search terms	Grey literature		Academic/peer reviewed		
		Opengrey	Google Scholar	Pubmed	EBSCO Host	Scopus in Elsevier
01/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world data" )			87 (87)		
04/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world data" )			4 (4)		
01/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world data" )				71 (71)	
01/11/13	("challenge*" OR "obstacle*" OR "barrier*" OR "enabler*" OR "strategy") AND ("Real world data" )				50 (50)	

Date	Search terms	Grey literature		Academic/peer reviewed		
		Opengrey	Google Scholar	Pubmed	EBSCO Host	Scopus in Elsevier
04/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world evidence" )				8 (8)	
01/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world data" )					80 (80)
04/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world evidence" )					8 (8)
01/11/13	"real world data"	0				
04/11/13	"real world evidence"	1 (0)				
01/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world data" )		10, 900 (198)			
04/11/13	("Europe" OR "EU" OR "Sweden" OR "UK" OR "Britain" OR "England" OR "Italy" OR "France" OR "Germany" OR "Spain" OR "Poland" OR "Dutch" OR "Netherlands" OR "Romania" OR "Denmark" ) AND ("Real world evidence" )		1,030 (198)			
04/11/13	"databases" AND "healthcare"		74,600 (198)			

Notes:

RECORD in each cell: number or hits/records; (number of hits included in Endnote)

EBSCO host searches included the following databases: CINHAL, EconLit, ERIC, GreeFILE, PsycARTICLES, PsychINFO, Health Source -Consumer, Health Source -Nursing/Academic Edition, Military and Government Collection.

Other sources of evidence included documents retrieved through specific web sites and additional grey literature searches. Relevant web sites included those of industry associations, such as ABPI; <http://realworlddata.org/>; the King's Fund library, web sites of consultancies engaging with real-world data studies and the web sites of international organisations (such as the EU or the OECD). Furthermore, the review also included references identified through 'snowball search', whereby sources referred to in one or more relevant documents have been identified and retrieved by the researchers.

## Screening

All the results have been screened and groups created in EndNote, distinguishing between potentially relevant and not relevant sources.

Articles were excluded from the study upon meeting one or more of the following exclusion criteria:

- databases that were very small (covering data from less than 200 individuals)
- databases that were overly context specific
- studies analysing US data
- studies using data from countries other than Europe and the US.

The screening involved mapping the abstracts according to a matrix of specific categories- relating to main questions as formulated in the Terms of Reference and the Kick-off meeting of the project. These included the country covered by the individual study, its research design, and its coverage of specific databases.

In total, the research team screened 935 sources.

Table B.2 provides a summary of the mapping exercise on the complete set of articles.

**Table B.2 Summary of mapping exercise**

Which main question does it relate to?								Does this study mention a specific market								Research type/design						
Access to rare disease registries	RWD used to validate a theoretical framework	Data used to build theoretical models	Standards for RWD in HTA	Linkage across countries	Other areas relevant to the use of RWD	Use of RWD to assess the effectiveness of treatment	Example from other sector	Other (please describe)	EU	UK	France	Sweden	Poland	Spain	Germany	Italy	Other (specify)	Survey	Qualitative	Quantitative	Review	Empirical/ user perspective
31	25	8	9	9	29	75	27	n/a	15	37	13	19	0	13	26	19	51	9	13	109	27	47

The study team has defined the relevance of the articles screened according to the extent to which they presented applicable insights on challenges and enablers with regard to access, use, linkage, regulation and development of RWD, or discussed databases that had the potential to be used in a variety of research projects. The most relevant articles (n=43) were selected for full text review, and fully extracted.



## Extraction

We extracted the articles that had been determined to be the most relevant using a standardised review template. Categories for data extraction were the following: full reference, source of RWD, rule of access to data, sample size, disease/treatment under study, source of funding for the study, summary of the study, and relevance for our project (see Table B.3).

**Table B.3 Example of data extraction**

Author and Title	Avillach, P., Coloma, P. M., Gini, R., Schuemie, M., Mougin, F., Dufour, J.-C., Mazzaglia, G., Giaquinto, C., Fornari, C. & Herings, R. (2013). "Harmonisation process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project". <i>Journal of the American Medical Informatics Association</i> , 20, 184–192.
Data source	EHR databases
Access to data	n/a
Sample size	19,647,445
Disease	Acute myocardial infarction (AMI); acute renal failure (ARF); anaphylactic shock (AS); bullous eruption (BE); and rhabdomyolysis (RHABD).
Funding	FP7- EU-ADR Project
Summary (max 3 paragraphs)	<p>This paper describes the work of the ADR project, launched in 2008, in defining the procedure used for harmonising the extraction from eight European EHR databases of five events of interest deemed to be important in pharmacovigilance. This study shows how event extractions may differ across databases and how different choices impact on the estimated incidence of a given event, even when a common language (UMLS) can be used across the databases.</p> <p>The databases included in the project are: Health Search/CSD Patient (HSD, Italy), Integrated Primary Care Information (IPCI, The Netherlands), Pedianet (Italy), and QRESEARCH (UK). These are general practice (GP) databases where both clinical information and drug prescriptions are recorded. The Aarhus University Hospital Database (Aarhus, Denmark), PHARMO Network (The Netherlands), and the regional Italian databases of Lombardy and Tuscany are all comprehensive record linkage systems in which drug dispensing data of a well-defined population are linked to a registry of hospital discharge diagnoses and various other registries. The databases are heterogeneous in both structure and content. In order to address privacy and data protection worries, the research used a decentralised approach.</p>
Relevance for Pfizer	<p>The study illustrates some limitations in linking data across countries. The study draws attention that to the fact that currently available data sources even when comparable across countries, often do not capture sources of bias and residual differences, including the effects of immigration and ethnic variation. Furthermore, even in the case of comparable databases it can be challenging to create meaningful queries due to the limitations in semantic interoperability. Furthermore, it is important to note that for the linking of databases across countries, the project had to use a decentralised approach as the data controllers in each country were reluctant to give up control of the databases.</p> <p>Finally, the article also mentions other EU projects on combining databases such</p>

	as Pharmacoepidemiological Research on Outcomes of Therapeutics (PROTECT), which has been recently funded to link healthcare databases throughout Europe under the umbrella organisation of the Innovative Medicines Initiative (IMI), the VAESCO or ARITMO projects.
--	---



## Annex C Case studies

---

To inform and illustrate the analysis, we were provided with 22 illustrative cases studies by IBM. The case selection aimed to provide good coverage of the pharmaceutical value chain and the healthcare ecosystem, with a focus on European initiatives. In addition, a number of retail, automotive, banking and insurance cases were provided when it was felt that additional insights could be extracted for the health and healthcare sector. The development of the cases was supported by IBM documents including large-scale surveys on big data use in healthcare (IBM 2013), and by seven key informant interviews with subject matter experts in IBM research laboratories (Table C.1).

**Table C.1 Key informants' roles for task 2**

<b>Organisation</b>	<b>Geographical scope</b>
IBM Institute for Business Value (IBV)	Global
IBM Sales & Distribution	European Union
IBM Sales & Distribution	Nordic countries
IBM Sales & Distribution	United Kingdom
IBM Strategy & Transformation	Global
IBM Global Business Services	Benelux
IBM Research Haifa Labs	Israel

### Overview of RWD cases

The corpus of cases collated contained 17 detailed healthcare and life sciences initiatives covering the pharmaceutical value chain and the healthcare ecosystem (eg cases about decision aids, EHRs, data management, service delivery, etc). Relevant cases have been incorporated in the report.<sup>45</sup>

IBM also investigated five cases in other sectors in an attempt to gain insight from other industries that would inform health and healthcare RWD analytics. Those were marginally relevant, as they tended to be based on big data analytics rather than RWD analytics and mainly focused on customer insights issues. Cases that demonstrated successful use of RWD, or linkage of big data and RWD, are presented in Tables C.2-4.

---

<sup>45</sup> Case studies exclusively based on big data analytics were not included in the report.

**Table C.2. Major Italian Bank leverages client data to improve customer retention**

Case outline	<p>Changing client behaviours are detected by analysing branch teller notes, call centre notes and client emails. Retention schemes were applied based on the insights gleaned.</p> <p>The bank also monitored social media sentiment to measure the impact of targeted campaigns.</p>
Geography	Italy
Key benefits sought	<p>Reduced attrition from 6% to 3%.</p> <p>Optimised offers and cross sell to increase average products per customer from 1.4 to 2.2.</p>
Data Sources leveraged	<p>Branch teller notes, call centre notes, client emails</p> <p>Social media analysis</p>
Lessons learned	Internal and external data can be leveraged to predict and prevent customer churn

**Table C.3. Santam: Predictive analytics improve fraud detection and speed up claims processing**

Case outline	<p>At Santam, South Africa's largest short-term insurance provider, fraud losses accounted for 6 to 10 per cent of annual premium costs. Furthermore, fraud led to poor operational efficiency. Because agents had to handle and investigate both high- and low-risk claims, all claims took a minimum of three days to settle, and Santam began to feel its reputation for customer service suffer.</p> <p>Santam gained the ability to catch fraud early with an advanced analytics solution that captured data from incoming claims, assessed each claim against identified risk factors and segmented claims to five risk categories, separating likely fraudulent claims and higher-risk from low-risk cases.</p>
Geography	South Africa
Key benefits sought	<p>Savings of up to \$2.5m in pay-outs to fraudulent customers.</p> <p>Nearly \$5m in total repudiations.</p> <p>Reduced claims processing time on low-risk claims by nearly 90%.</p> <p>Cut operating costs by reducing the number of mobile claims investigations.</p>
Data Sources leveraged	Claims data
Lessons learned	RWD, predictive analytics and risk segmentation helped the company identify patterns that led to focused fraud detection whilst improving operational efficiency (claim turnover).

**Table C.4. Major Eyewear retailer: Customer intelligence advanced analytics to track, segment and score customers at individual level**

Case outline	<p>This large eyewear retailer generated large amounts of data, the majority of which was housed and managed by outside data and marketing vendors. Lacking a holistic understanding and view of the customers, marketers struggled to nurture customer relationships, seize business opportunities, personalise campaigns and acquire new customers.</p> <p>After a successful proof of concept (POC), the company deployed advanced analytics technology built on a high-performance platform that integrated online and physical customer data from multiple internal and external sources.</p> <p>The resulting 360-degree customer view not only helped the retailer identify its most profitable sales channels, but also segment, track and score customers down to the individual level based on thousands of behavioural attributes, and refine and personalise marketing campaigns.</p>
Geography	USA
Key benefits sought	<p>Developing more personalised campaigns.</p> <p>Conducting faster and richer customer segmentations based on customer attributes</p> <p>10% anticipated improvement in marketing effectiveness.</p>
Data Sources leveraged	Multiple internal and external sources
Lessons learned	Integration of various data sources to establish rich customer profiles combined with analytics for granular segmentation provide the foundation for customer centricity

It has proven difficult to find European RWD cases in healthcare and life sciences with sufficient detail to provide the insights needed for this study. Despite a perceived enthusiasm for RWD initiatives, only a few RWD initiatives outside the IBM's engagement sphere were found. Possible explanations for lack of well documented RWD industry cases outside of the IBM's engagement sphere include:

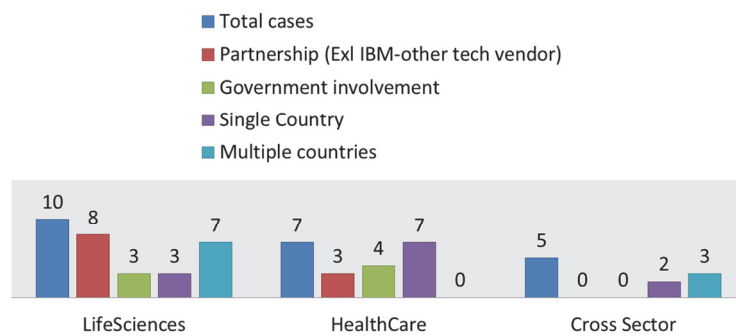
- The relatively recent emergence of the contemporary big data/analytics field as a focus of healthcare and life sciences investments.
- Lag of the Europe region in adoption by healthcare and life sciences organisations, in part due to a more fragmented EU landscape in terms of initiatives, data and legislation. This was voiced in a number of industry key informant interviews.
- The close association to competitive advantage of these initiatives is probably keeping them from become publicised, let alone in a sufficiently detailed matter.

In the life sciences area of market access and product efficacy demonstration, it is particularly hard to find anything beyond anecdotic evidence. It is also an area where IBM has less presence and involvement.

Initiatives in this domain are often conducted in house or in confidential partnerships. This is reflected in scarce references from professional service providers.

Although the sample size is small, some sector differences are observable in regarding partnerships, government involvement in cross-border activity.

**Figure C.1 Stakeholder involvement in RWD initiatives**



Partnering for RWD initiatives is strongly present in life sciences. Both healthcare and life sciences partnerships seek government involvement in the forms of grants or participating institutions.

The difference in geographic span of healthcare and life sciences initiatives is remarkable; life sciences seem to be more prone to international collaboration. This was supported by some interviewees who mentioned the regional or national focus of healthcare systems. Factors such as language barriers, uncoordinated standards adoption and legislative differences may also contribute to national focus of healthcare initiatives.

## Annex D Interviews

---

The primary goal of key informant interviews (KIIs) was to obtain qualitative description of experts' knowledge, perceptions or experiences of a range of factors which are not easily identifiable or documented in the published literature. They sought to further our understanding of the more salient issues related to RWD access and use.

### Selection of interviewees

Interviewees have been selected to provide insight about instances where the collection and use of RWD has contributed to demonstrating or increasing the value of treatments and care, as well as where standards (or lack thereof) have influenced research opportunities. The selection of interviewees has been supported by existing networks of Pfizer and RAND.

It was not within the scope and budget of this study to interview representatives of all stakeholder groups involved in RWD-based research. For instance, we did not seek to interview representatives of users/patients groups. It was not within the scope and budget of this study either to interview representatives from every European country.

We focused instead on stakeholders who were believed to have both the technical and overview knowledge of RWD matters. This potentially included representatives from academia, research consortia, healthcare providers, health insurance sector, data processing industry and government.

Tables D.1 indicates the affiliations of the key informants.



**Table D.1 Key informants' organisations**

Organisation	Sector	Geographical scope
Evidera	Data consultancy	Global
Cegedim	Data consultancy	Global
Independent consultancy	Academia/consultancy	Netherlands
CPRD	Data vendor (Public)/Research	United Kingdom
Department for Business, Innovation & Skills	Government	United Kingdom
Orphanet	Consortium (Public and Private)	France (coordination)/Europe
Pfizer Belgium	Pharmaceutical company	Belgium
Pfizer Germany	Pharmaceutical company	Germany
Pfizer Netherlands	Pharmaceutical company	Netherlands
Department of Health, UK	Government	United Kingdom

### Conducting and analysing the interviews

After selecting candidates for the interviews, the interviewees were approached by email and telephone, and where necessary with an introductory letter from Pfizer to attest the importance of the research. Reminders were sent when necessary.

Interviewees were not offered any financial compensation for their participation.

Most interviews were conducted by telephone by two interviewers and lasted on average 45 minutes. KIIs were conducted as semi-structured interviews following an interview protocol comprising questions about example of current RWD access and use and RWD standards, but also leaving space for individual follow-up questions based on the particular expertise of the interviewee. The interview protocol drew on emerging findings from the rapid evidence assessment and was designed to target specific themes relevant to the workstream it fed into (Box 7).

The 10 interviews were recorded, but not transcribed verbatim. Analyses were informed by key themes guiding the interviews as described above while also seeking to identify additional emerging themes. Quotes inserted in the report have been anonymised.

## Box 7 Interview protocol

### **Background to the project:**

RAND Europe and IBM have been commissioned by Pfizer to conduct a study which aims to further assess the real-world data (RWD) policy landscape in the EU.

This study seeks to:

- investigate and analyse ways in which organisations have contributed to use of RWD in product or service development and the ways in which standards have impacted on the nature of that engagement. This will include examples both where the relevant community perceive that this has and has not been successful.
- understand more systematically how standards are shaping the way in which RWD is being used in product development and to improve health outcomes and the opportunities that exist for pharmaceutical companies to contribute to the development of standards.

We have conducted a review of the literature and would like to supplement the information gathered with telephone interviews with experts across different stakeholder groups, such as yourself. Taking part in the interview is entirely voluntary. All information collected will be kept strictly confidential. We sent you some information in advance of this call, may I ask you if you have read and understood the consent form? With your permission, we will record this interview for the purpose of writing up the notes and analysis. Quotes will be anonymised.

### **Interview guide:**

- 1) Can you please tell us a little bit about your background and your organisation, role and engagement in RWD use/regulation/production?
- 2) Are you aware of the existence of a formal RWD agenda and/or task force in your organisation? In your department?
- 3) Which kind of RWD data sources can you access today (or you know are accessible)? Could you give some examples?
- 4) How can those databases been used? Could you give some examples?
- 5) What are the RWD data sources/pools for which access is desired?
- 6) How could those databases be used?
- 7) What are the main enablers experienced in developing/using/accessing RWD sources?
- 8) What are the main challenges experienced in developing/using/accessing RWD sources?
- 9) What is done to facilitate/enable direct RWD sources access?
- 10) Are you aware of any partnership that facilitate the access to/use of RWD?
- 11) What approaches does your organisation develop to advance access to RWD? Is your organisation developing a strategy to improve access/use of RWD?
- 12) Are you aware of any standards or framework which shape the way in which RWD is being used? At which level (organisational? regional? national? international?)
- 13) Has your organisation actively contributed to standards elaboration/evolution?
- 14) Do you perceive differences in the progress between Europe and the Rest of the world?
- 15) Is there anything we haven't covered which you think we should consider in our study?



## Annex E Scenarios

---

In order to deepen the understanding of the possible future challenges of the RWD landscape in Europe, the study team developed a set of three scenarios which served as a basis for analysis during the workshop with the client. Developing scenarios served as a consistency check on the analysis conducted in the first half of the project while directing attention to the main drivers of a possible future.

The aim of scenarios is to build several pictures of a specific domain of interest (RWD and data access and management in this case) in which hypotheses can be set to obtain a range of future views. Scenario thinking aims to identify new developments, risks or impacts which might otherwise have been missed. It is a means of stimulating more informed and deeper conversations about the future direction of a certain policy area. Building scenarios is therefore an exercise in both discipline and creativity. The discipline is needed to structure the set of scenarios so that they reflect the issues requiring exploration. Creativity is needed in filling out the scenarios so that they become meaningful, consistent and plausible.

In this study, the definition of the scenarios was structured along the strategic partnerships that would characterise each of these potential futures, and on the type of access to data (whether it is dominated by centralised or decentralised forms of data control). The analysis elaborated the role of pharmaceutical companies in the changing contexts, pinpointed the most important stakeholders and outlined the salient risk factors as well as potential limitations to the data. The three scenarios illustrate polarised and naive visions of the future. Thinking about developing coherent narratives around each of these extreme pathways forced us to establish a hierarchy of factors that are likely to shape strategies. Table E.1 summarises the three scenarios.

**Table E.1 Three possible scenarios for RWD and the pharmaceutical industry**

Scenario	Scenario 1: Strategic partnership with payers (public or private)	Scenario 2: Strategic partnership with patients and/or patient organisations or care providers	Scenario 3: Strategic partnership with data vendors
Type of access to data	Centralised	Decentralised	Decentralised
Vision	Strategic alliance with public bodies (Health and Healthcare agencies, HTAs) and/or private bodies (insurance companies) to access RWD.	Access to RWD facilitated by the development of direct relationships with data owners. Patients and patients' organisations are giving/selling data to pharmaceutical companies.	New players have emerged in the RWD landscape (data vendors like social medias, IT companies selling both data and software to collect the data). Pharmaceutical companies are strengthening their relationships with data vendors to get access to this huge amount of data.
Role of the pharmaceutical companies	<ul style="list-style-type: none"> <li>Working closer with public health agencies, pharmaceutical companies might expand their activities, integrating missions regarding outcomes assessment (health bond)</li> <li>Working closer with insurance companies, pharmaceutical companies potentially develop links between drug use and insurance reimbursement policy.</li> </ul>	<ul style="list-style-type: none"> <li>Pharmaceutical companies to provide services to people to 'manage' and monitor their health.</li> <li>Investment in Apps and e-health</li> <li>Building relationships with patients groups and associations. Need to invest in advocacy to facilitate relationship and data access.</li> </ul>	<ul style="list-style-type: none"> <li>Wave of M&amp;A between IT companies and pharmaceutical companies to access data.</li> <li>Pharmaceutical companies to provide in-house designed individual health monitoring services based on IT tools, apps</li> </ul>
Relevant stakeholders	<ul style="list-style-type: none"> <li>EU Funded projects building pan-European databases,</li> <li>National Health Data repository (ex: CPRD)</li> <li>Insurance companies</li> <li>National Health Systems</li> </ul>	<ul style="list-style-type: none"> <li>Individuals</li> <li>Patients associations (Patients like me, Quantified Self, etc)</li> <li>Apps developers</li> <li>Charities</li> </ul>	<ul style="list-style-type: none"> <li>IT companies owning amounts of potentially relevant data (Facebook, Google, Cegedim, Apple, etc).</li> <li>Medical device providers (Philips) are collecting data on users</li> <li>Key partners in the retail industry (supermarket for nutrition data)</li> </ul>
Risks	<ul style="list-style-type: none"> <li>Competition between data providers</li> <li>High cost of partnership</li> </ul>	<ul style="list-style-type: none"> <li>Need to negotiate consent terms and ownership of data with wider range of stakeholders</li> <li>Reliability of the data which is self-reported</li> </ul>	<ul style="list-style-type: none"> <li>Strengthening relations with private companies might increase suspicions regarding the use of RWD. Need for safeguards to respect individuals' privacy</li> </ul>

Scenario	Scenario 1: Strategic partnership with payers (public or private)	Scenario 2: Strategic partnership with patients and/or patient organisations or care providers	Scenario 3: Strategic partnership with data vendors
Sample	<ul style="list-style-type: none"> <li>Public payers: access to data about sick patients mostly – unless pharmaceutical companies are pursuing missions of public interest (evaluating health policy outcomes – in that case, they might get access to other types of administrative databases – merging socio economic and health datasets.</li> <li>Private payers: access to claims data, but the uninsured are marginalised.</li> </ul>	<ul style="list-style-type: none"> <li>Stratification of the population by disease, and therefore access to certain groups of patients only;</li> <li>Not capturing the less “vocal” and less connected patients</li> </ul>	<ul style="list-style-type: none"> <li>Need to develop standards to facilitate interoperability between different databases (if large number of data vendors)</li> <li>Quality of the data is at stake (disparate datasets, incomplete health indicators selection)</li> </ul>