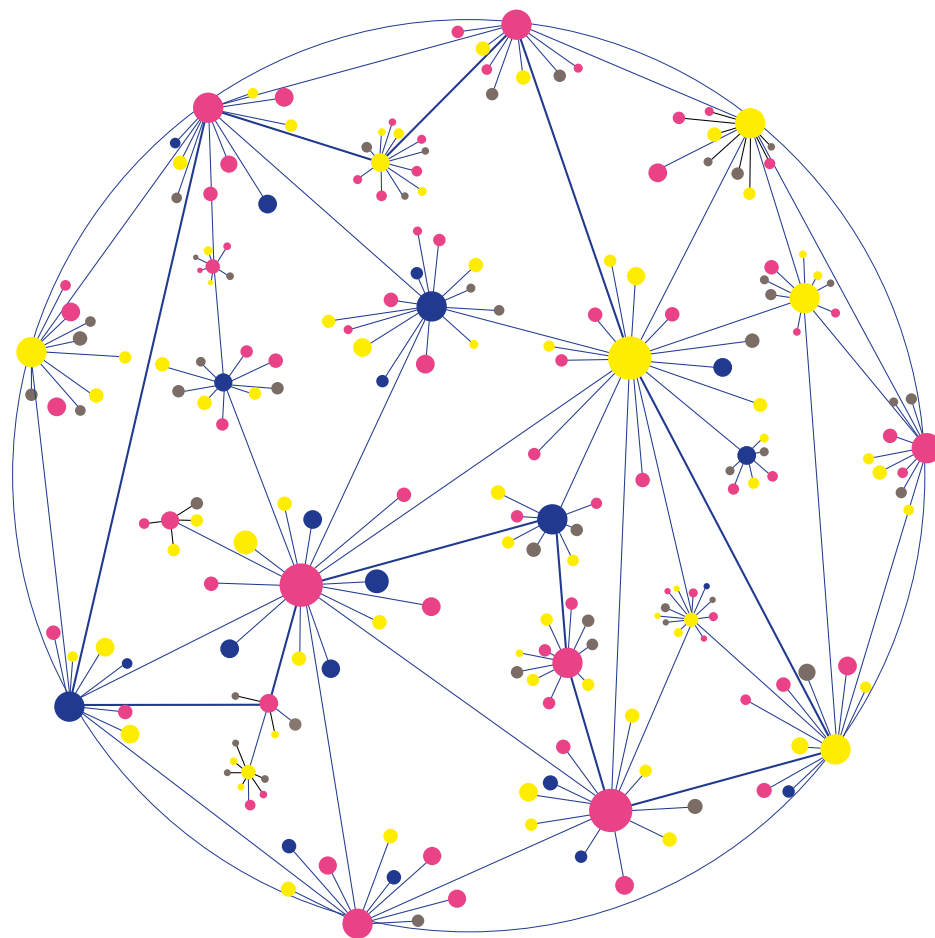


Proposals for an Enabling Data Environment for Health and Related Research in Ireland

Rosalyn Moran



Proposals for an Enabling Data Environment for Health and Related Research in Ireland

A discussion document

Rosalyn Moran,
Health Research Board



Published by:

Health Research Board
Grattan House
67-72 Lower Mount Street
Dublin 2
DO2 H638
Ireland

t 353 1 234 5000
f 353 1 661 2335
e hrb@hrb.ie

w www.hrb.ie

First published 2016

© Health Research Board 2016

ISBN: 978-1-903669-23-5

Please cite this publication as:
Moran, R (2016)

*Proposals for an Enabling Data Environment for Health
and Related Research in Ireland*
Health Research Board, Dublin.

Contents

Foreword	v	3 Attitudes, Practices and Incentives – Data Access, Sharing and Linkage	31
Acknowledgements	vii	3.1 Introduction	31
Executive Summary	1	3.2 Attitudes/Culture	31
Recommendations towards the implementation of the DASSL model	4	3.3 Data Management Practices	32
1 The HRB Data Project – Introduction and Methods	9	3.4 Data Linkage and Needed Services	34
1.1 Introduction to the HRB Data Project	9	3.5 Incentives to Facilitate Data Sharing and Encourage Compliance	35
1.2 Project Aims, Methods and Procedures	10	3.6 Conclusions	35
1.3 The Wider Data Ecosystem in Ireland	11	4 The DASSL Model and the Proposed Research Data Trust	37
1.4 Raising Awareness of Benefits of Open Access to Research Data	13	4.1 Introduction	37
2 Health and Related Data – Safeguarding Provisions for Safe Access and Use	15	4.2 The DASSL Model and the Research Data Trust	38
2.1 Introduction	15	4.3 A Walk-Through of the Proposed New Infrastructure	49
2.2 Health and Health-Related Data Collections in Ireland and their Availability for Research	16	4.4 Conclusions and Recommendations	50
2.3 Safeguarding Provisions Relating to Consent and Other Guiding Principles	17	5 Recommendations and Conclusions	53
2.4 Safeguarding Provisions Relating to Types of Data	20	5.1 Recommendations and Conclusions	53
2.5 Safeguarding Provisions Related to Building a Robust Environment for Health and Related Research	22	5.2 Conclusions	55
2.6 Special Legislation	25	Appendices (see overleaf)	57
2.7 Striking a Balance – a Role for Guidance and Codes of Practice	26		
2.8 Conclusions and Recommendations	28		

List of Figures

Figure 1:	Players and initiatives in Irish data ecosystem – shared needs	12
Figure 2:	The DASSL model.	41
Figure 3:	Secure anonymised data access through SAIL (Forde et al. 2009)	43
Figure 4:	Trusted third party data linkage service, Scotland (Pavis 2015)	45
Figure 5:	The eDRIS service (Pavis 2015)	47

List of Boxes

BOX 1:	Caldicott principles	19
BOX 2:	The Irish Social Science Data Archive (ISSDA)	20
BOX 3:	The Central Statistics Office	21
BOX 4:	The SAIL technique for providing safe access to data	44

List of Appendices

Appendix 1:	Examples of Research Which Used Data Access, Storage, Sharing and Linkage	59
Appendix 2:	Summary of Anticipated Benefits Associated with the British Columbia Centre for Data Innovation by Stakeholder Group	62
Appendix 3:	The Research Data Agenda – Communication to Participants in the HRB Data Project	63
Appendix 4:	HRB Data Project – List of Informants	65
Appendix 5:	Listing of Health and Social Care Data Collections from HIQA 2014	67
Appendix 6:	Listing of Datasets in ISSDA	70
Appendix 7:	Demand for ISSDA Data, 2002–14	74
Appendix 8:	Guidelines for Data Sharing in the Public Sector – Office of the Data Protection Commissioner	75
Appendix 9:	The Sail Data User Journey	76
Appendix 10:	Abbreviations and Acronyms	77

Foreword



The HRB Data Project is a response to requests from the health research community to address challenges they experience in relation to accessing, sharing and linking data, which leads to valuable health research being inordinately delayed or in some cases abandoned. In certain cases data has to be collected anew, resulting in duplication of effort and costs as well as delays in securing research results for the benefit of people's health and the effective delivery of patient care.

Ireland has considerable data resources which could be harvested to advance medical treatments, enhance health service delivery and inform policy and planning across government and civic society. The HRB Data Project reveals that a range of complex cultural, social, technical and governance issues are the source of many concerns among data users.

The fundamental question is – how can researchers and policy makers avail of one of our most valuable national assets i.e. existing data, and use it in a safe, secure manner, protecting the privacy and confidentiality of the data subjects, and in accordance with existing legislation? The report presents a model along with proposals for the types of infrastructure and services required to enable safe access, usage and linkage of data.

Research Evidence Action, the HRB strategy for 2016 – 2020, commits to the exploitation of data in health research that contribute to improvements in health.

Accordingly, the HRB presents this discussion document as a key contribution to the conversation that needs to happen about how we maximise the use of our national data assets. And how we can enable the health research community, as well as the broader science, technology and innovation ecosystem, to use it to deliver benefits for the Irish people and enhance delivery of our health services. We hope that this report will stimulate much needed debate and action towards establishing the data infrastructure and services required to support world class research and innovation in Ireland.

Graham Love, PhD
Chief Executive

Acknowledgements

I would like to thank all the informants interviewed in the course of the HRB Data Project. They gave generously of their time, sharing their experiences and ideas, and they greatly influenced the shape of this report.

In particular, I would like to thank John Dunne of the Central Statistics Office, Pat O'Hara, National Statistics Board, and Mary Doyle, Department of Education and Skills, who provided strategic focus. I would also like to thank Peter Lennon, Department of Health, whose in-depth knowledge of legislative aspects of health information was invaluable.

The HRB Data Project has benefited from the knowledge and experience of colleagues involved in developing and implementing infrastructural solutions to data access, sharing and linkage internationally – especially in Northern Ireland, France, Scotland, Canada and Australia. Special thanks are due to the staff of the Northern Ireland Statistics and Research Agency, the Northern Ireland Longitudinal Study – Research Support Unit, Northern Ireland Health and Social Care Business Services Organisation – in particular Maire Bolly and Daryll Madine, who gave so generously of their time and expertise. Thanks also to staff of the Administrative Data Research Centre Northern Ireland whose experience in setting up the Centre has been very valuable. The learning from international colleagues has been invaluable in the development of the Data Access, Storage, Sharing and Linkage (DASSL) model that is proposed in this report.

Colleagues in the HRB have supported the project in many ways: I would like to extend my gratitude to the HRB Project Group (Teresa Maguire, Donna Tedstone, Patricia Clarke, Sarah Craig, Maura Hiney and Mairéad O'Driscoll); to the HRB Executive Team for their inputs; to colleagues who provided administrative and library support – in particular, Sharon Kelly and Mary Dunne, respectively.

I would like to acknowledge the helpful feedback provided by the peer reviewers of the report – Professors Tom Fahey, Royal College of Surgeons in Ireland and Professor Dermot O'Reilly, Queen's University Belfast.

I would like to say a special thank you to colleagues in the Central Statistics Office, in particular John Dunne, Garry Dunphy and Richard McMahon with whom various aspects of the report were discussed.

We hope that this report will stimulate needed debate and action towards establishing the data infrastructure and services required to support world-class research and evidence generation in Ireland.

Executive Summary

A great deal of data is collected within the Irish health system and the broader public administrative system, at considerable cost to the public purse. These datasets are among the most important assets held by the State. Existing datasets, including data collected by researchers, could be better exploited to unlock their tremendous potential to inform policy and practice, but also to 'support innovation and economic growth across multiple sectors'.¹ There are, however, a number of barriers hindering the exploitation of such data. In order to realise the potential of existing data, it is important to ensure that data can be stored, accessed and shared safely within a robust governance framework that protects privacy and confidentiality.

The present work, hereinafter referred to as the HRB Data Project, set out to explore these issues and also to explore possible means that could unlock the benefits of existing datasets for health researchers and policymakers. The issue became increasingly acute when the adoption of the *HRB Strategy 2010–2014* resulted in increased funding for research in population health and health services which, traditionally, had been under-resourced in Ireland. Researchers seeking to study in these domains often needed access to health and administrative datasets in order to undertake research that could help to inform policy and practice in relation to health and well-being. Frequently, they encountered access and related difficulties, which resulted in valuable research ideas being abandoned or delayed. The HRB Data Project focused on data issues arising in health and related research – in particular in the population health and health services domains.

Data issues relating to clinical trials were considered out of scope, however, since they operate under different laws, regulations and administrative provisions.²

Data in the health area are collected by a range of different players with different roles and mandates. For example, health service data such as hospital admissions are routinely collected in the course of health service delivery, while population-based health data such as disease registers, e.g., the National Cancer Registry Ireland, are often collected by researchers in collaboration with healthcare providers. Again, data generated by surveys and cohort studies are typically collected by researchers, whereas a lot of administrative and census-type data are collected by government departments and agencies. Obtaining access to these different types of datasets is often difficult for a variety of reasons – many of which are related to concerns regarding the safeguarding and safe use of sensitive information.

The HRB Data Project undertook interviews with individual players and agencies involved in the collection, use, storage, sharing and linkage of data in the health area, and with actual and potential key users of such information in health and other policy areas. Data-related needs, practices and attitudes were explored, as were respondents' views regarding both existing and required infrastructure for the safe sharing and linkage of data. The questions put to interviewees were designed to explore the culture around data usage and sharing, and to identify the measures needed to create a robust environment for safe use of data in health and related research areas.

1 BC Centre for Data Innovation 2014. BC Centre for Data Innovation Final Report of Working Group at http://www.gov.bc.ca/citz/down/BC_Centre_for_Data_Innovation-FINAL.pdf

2 http://ec.europa.eu/health/human-use/clinical-trials/information/index_en.htm#cti

This report is informed by an analysis of the interviews conducted, a review of the international literature, ongoing discussions with major stakeholders, in particular the CSO³ as well as visits to the Northern Ireland Statistics and Research Agency (NISRA) and also to the National Institute for Statistics and Economic Studies (INSEE)⁴ in France, in particular.

Chapter 2 provides an overview of health and related data resources in Ireland. There was general agreement among those interviewed that there should be greater access to these resources, particularly data collection funded from the public purse. Many reasons were given, chief among which were: expanding the opportunities to explore a broad range of research questions through the use of such data (see examples of studies in Appendix 1); using data to improve healthcare quality and health system performance; verifying research findings; preventing duplication of data collection and research effort. In addition, health researchers noted that some types of research, including research into the efficacy of treatments, often require large numbers of participants and the pooling and linkage of data from different studies as was required, for example, in order to establish the link between smoking and lung cancer.

The chapter identifies a number of safeguarding provisions which could support safe access to, and use of, data in the Irish context, in particular, sensitive data that are typically used in health research. These safeguarding provisions include the protection of data, as well as the development of data services and infrastructure to support researchers in the safe use of data, e.g. health research data hub, trusted third party service, safe haven etc. A discussion is needed regarding the necessity and utility of the introduction of special legislation to underpin the infrastructure and services identified here in an Irish context. Lessons learned from international literature suggest that such legislation will not necessarily result in good or better governance. Rather, what is needed is a robust authorising mechanism which can deliberate on the conformity of research projects to existing legislation, and also on their conformity

to guiding principles and best practice, to assist with navigating the ambiguous areas thrown up by legislative regimes.

Chapter 3 presents the findings from the interviews conducted, and the associated recommendations, under three headings: attitudes/culture; data management; data linkage and needed services. Difficulties experienced by researchers in linking data are summarised. The views of interviewees on how to incentivise data sharing and encourage compliance are reported in a final section. This chapter presents more than 20 recommendations including those which reflect interviewees' views regarding measures required to create a robust data environment.

In the course of the interviews conducted and the literature reviewed, it became clear that the issues of data access, sharing, storage and linkage – which were identified as key issues for the health research community – were also of immediate or emerging interest to a large number of institutions and government departments.

A confluence of factors has been instrumental in the development of this interest in data – recognition of data as a core societal asset, increasing demands for data to inform public policy and planning; international movements towards open data and the sharing of research and administrative data; growth of big data and data analytics allied with developments in supercomputing, which facilitates the analysis and interpretation of large datasets; a willingness by many governments, including the Irish Government and public administration, to exploit these developments for economic benefits (e.g., Action Plan for Jobs), scientific benefits (e.g., the strategic plans of Science Foundation Ireland (SFI), the Health Research Board (HRB) and the Economic and Social Research Institute (ESRI)), and social and cultural (e.g., open government) benefits.

Government departments and interest groups increasingly recognise that realisation of these potential benefits requires some essential pieces of national infrastructure as well as data services that can support safe use of data.

The research undertaken by the HRB Data Project has enabled us to identify the most important pieces of infrastructure and data services needed in order to serve the data access, sharing, storage and linkage needs of the health research and broader data community in Ireland.

3 The work was informed in particular by discussions with John Dunne from the CSO.

4 Centre d'Accès Sécurisé Distant aux Données was the division of INSEE visited.

Chapter 4 configures the requirements identified by the HRB Data Project into the DASSL model (see diagram on p.5), which allows for the safe and efficient access, storage, sharing and linkage of data. The model is informed by experience gained in the implementation of similar models in Northern Ireland, Scotland and France in particular. It is intended to stimulate discussion and debate on this area of national significance.

It is proposed that an entity – a research data trust (RDT) – would be established to provide the institutional and technical environment to respond in a concerted manner to the growing data-related needs for data access, storage, sharing and linkage within the Irish research environment and the broader data ecosystem. The operationalisation of the DASSL model would take place within this context. This would be a valuable addition to the Irish statistical system.

The DASSL model comprises seven elements – five related to infrastructure and services, and two related to the broad legislative and socio-cultural context needed to facilitate implementation, i.e., governance and public engagement. In Chapter 4, the rationale for the inclusion of each element, the benefits they offer, possible means of implementation, possible players, possible institutional arrangements and related governance issues are discussed. The elements of the model are briefly as follows:

1. **Governance** – The Health Information and Quality Authority (HIQA) notes that good information governance allows organisations and individuals to ensure that personal information...is handled legally, securely, efficiently and effectively.⁵ Learning from international developments and implementations in Scotland in particular, it is suggested that optimal governance can be achieved for research projects involving the sharing and linkage of data through adopting a principled, proportionate, risk-based approach to governance. This approach involves adjudication on research proposals involving data sharing and linkage by an agreed and suitable authorising entity,⁶ in the first instance, with reference to regulatory requirements. No

legal framework, however, can address each and every circumstance thrown up by complex research studies in a changing social and technical environment, and it is foolhardy to hold such expectations. Thus, importantly, the principled, proportionate, governance (PPG) approach deals with the 'spaces in-between legal provisions' (Sethi and Laurie 2013)⁷ where judgement calls are required to address ambiguities and grey areas which research has shown has resulted in a 'culture of caution' and lassitude in sharing. In navigating the grey areas, the Scottish (and other OECD countries) approach is guided by key principles e.g. privacy, public interest etc. and by robust and transparent policies, processes and procedures for holistic risk assessment which are informed by stakeholder engagement.

This principled, proportionate approach to achieving balance between the protection of privacy and individual interests and public benefit in the conduct of research is receiving growing acceptance internationally. It allows for an adaptable and efficient governance mechanism which can deal responsibly and proportionately with the complexities which modern health research poses, and inform the interpretations required by the extant legislative environment.

Proposals are put forward for discussion by all stakeholders regarding the kinds of structures which could deliver this model of governance for research projects (involving sharing and/or linkage of data) e.g. information governance review panel/research ethics committee and project approval boards. The forthcoming Health Information and Patient Safety Bill will be important to such discussions. Effective operation of such structures should ensure the highest standards in the safe use of data, and the protection of privacy and confidentiality. Sound governance is seen as essential, in order to maintain public confidence and trust, which are fundamental to a thriving research community.

2. **A health research data hub (HRDH)**, is proposed which would, with the agreement of data custodians, provide or facilitate safe access to protected (e.g. via third party service

⁵ <http://www.hiqa.ie/healthcare/health-information/information-governance>

⁶ The relationship of such an entity with the Office of the Data Protection Commissioner (perhaps a consultative group) and research ethics committees (perhaps a subcommittee) in the light of the Health Information and Patient Safety Bill will need to be discussed.

⁷ Sethi, N. and Laurie, G. (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

and data protection techniques) routinely collected health data, registry data and other research data files, in accordance with agreed governance arrangements. The hub would thus provide for safe access to data already collected at considerable cost, and allow for research that could provide invaluable inputs to policy and planning.

3. **A third-party data linkage service** would involve a unit or team within the RDT, which would facilitate research by integrating and linking datasets that hold personal information. The RDT would use a (possibly separate)⁸ third-party data service which employs advanced anonymisation and encryption to prevent re-identification of individuals. Thus, protected linked data would be made available to researchers. This report elaborates on how linking different datasets provides researchers with inexpensive access to larger datasets, and widens the range of variables available for inclusion in particular projects.
4. **A safe setting/haven** is a further piece of infrastructure identified as needed by interviewees. In essence, this is a 'locked-down' environment designed to allow researchers process sensitive data safely. It is proposed that the RDT would host such a facility. The anticipated expansion of demand from researchers as a consequence of developing the proposed health research data hub, coupled with growth in secondary data analysis, inter alia, would require investment in this critical piece of the national statistical infrastructure. It is expected that the facility would be used by a range of actors in the wider data ecosystem.
5. **A research support unit (RSU)** for data sharing would provide assistance to researchers in relation to data access and linkage aspects of the research process, thus maximising efficiencies and minimising delays. Already in operation in Northern Ireland, Scotland and France, it is considered to be a vital part of the proposed infrastructure, as it helps to ensure maximum and safe use of research and administrative data. Fifteen functions of such a unit are outlined in this report.

6. **Output checking and disclosure control** – the different pieces of infrastructure described above would be involved in the output of data. Most of these data have to be thoroughly checked by highly trained statisticians with expertise in disclosure control, so as to ensure that individuals or entities cannot be identified. This is a specialist and time-consuming activity, but is essential for good governance and data quality.
7. **Public engagement** would involve ongoing education, consultation and engagement with the public, in relation to the development and operation of a research data trust, inter alia. The general public would be represented on structures and committees, as appropriate. People who provide their personal information for research purposes need to be confident that their data will be used in the public interest; that the data will be held securely, and that their privacy and confidentiality will be respected. This is essential, as the trust of the general public is critical to the successful exploitation of Ireland's national data assets.

This report provides a walk-through of the steps a researcher might take in steering a project through the proposed infrastructure. A number of examples of research that would be facilitated by the DASSL infrastructure and services are provided in Appendix 1.

The DASSL model provides an important means by which personal and sensitive data can be safely harvested by researchers to complete research that can break new ground, help inform public policy and support economic growth and innovation across multiple sectors.

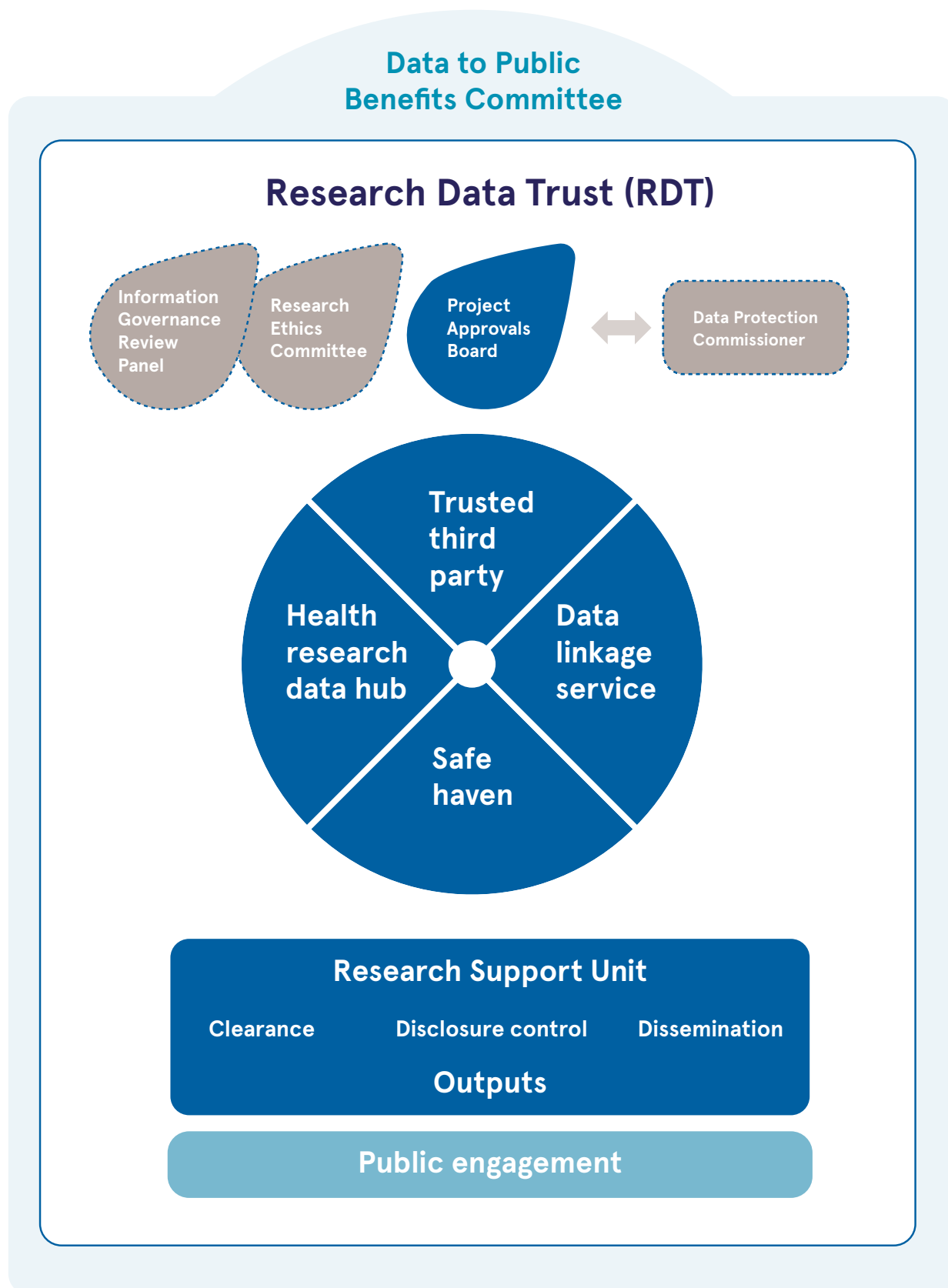
Chapter 5 brings together 33 recommendations based on the research carried out which for the most part, have been taken from the individual chapters; it ends with a section on conclusions.

Recommendations towards the implementation of the DASSL model

The infrastructure and services required to create an enabling research and data environment, as outlined above, will not happen automatically and will require strong leadership and concerted

⁸ A number of techniques exist to safely link data; some of these are described in Chapter 4.

The DASSL Model



The key elements of the DASSL model are outlined here. Their configuration and operationalisation will require further discussion and agreement.

and committed engagement by a large number of stakeholders. Consequently the following high level recommendations regarding the way forward towards implementation of the DASSL model are presented.

1. A high-level, possibly cross-governmental, 'Data to Benefits Committee' needs to be established to drive strategic discussions and action towards the implementation of the DASSL model components.
2. The seven elements of the DASSL model, along with the proposals and practical examples presented here in relation to establishing the (data access, storage, sharing and linkage) model need to be debated within the research community and other stakeholder groups. These seven elements are designed to ensure safe projects, safe researchers, safe data, safe settings and safe outputs. These debates should aim to develop shared understandings towards consensus on the implementation of the model.
3. It would be desirable that the National Statistics Board (NSB), HSE and other relevant agencies review the recommendations set out in the report and their implications for the use, sharing and linkage of data in Ireland.
4. In accordance with the research carried out it is recommended that all seven elements of the DASSL model would be best implemented within a research data trust (RDT) environment in order to ensure maximum efficiency and quality. Consideration needs to be given, however, to the immediate need to put in place trusted third party/data linkage and safe haven facilities within a robust, principled, proportionate governance framework.
5. Public engagement, transparency and the development of trust should be prioritised in the implementation of the DASSL model to ensure its successful adoption.

If we want to have a safe and trusted modern infrastructure that will enable researchers to unlock the significant value of currently under-exploited data for the public good, then the DASSL model or a similar model needs to be implemented, in order to facilitate not only health research but other research that serves national economic and social agendas.

The research undertaken by the HRB Data Project has enabled us to identify the most important pieces of infrastructure and data services needed in order to serve the data access, sharing, storage and linkage needs of the health research and broader data community in Ireland.



1

The HRB Data Project – Introduction and Methods

1.1

Introduction to the HRB Data Project

Health research in Ireland is inhibited by a lack of formal infrastructure to support access to, and linkage of, datasets. A great deal of data is collected in the health area, but researchers find it difficult to access and/or link these data; consequently, many opportunities to conduct studies that could inform policy and practice are missed. For example, without access to large datasets as a result of researchers sharing their data, the link between smoking and lung cancer would not have been established.⁹

A variety of data custodians with different mandates and roles collect and maintain health data. Population-based health data e.g., the cancer registry and health census data are typically collected by researchers. On the other hand, health services datasets e.g., data on hospital care and community care, mental health, public health and the primary care reimbursement scheme are collected as part of public health surveillance and/or in the course of operating health and social systems. Regardless of the original purpose of collection, secondary or further analysis of such data can have tremendous value for researchers by enabling evidence-based input to policy and practice. Typically, however, the data custodians of these valuable datasets do not have an explicit

mandate to support the research community, thus resulting in barriers to access and re-use of such data.

In addition, datasets generated by surveys and cohort studies can be re-used to address new research questions, if and when they are made available to researchers. For example, data from the Growing Up in Ireland (GUI) study have been used in over 100 studies exploring diverse topics such as the timing of solid food introduction and obesity risk,¹⁰ young people's use of information technologies, and the social and demographic characteristics of migrant children living in Ireland. Researchers have also linked Growing Up in Ireland data to census and administrative datasets. When data from one or more datasets can be linked, significant quantities of data can be inexpensively generated, thus allowing for powerful hypothesis testing. However, researchers who are endeavouring to link data experience a range of difficulties e.g., issues regarding data integration and data protection.

Many research questions with potential to inform policy and practice are not attempted, or are abandoned or inordinately delayed, because data cannot be accessed or linked in a timely manner. The purpose of the HRB Data Project is to explore these issues, identify the key players and their concerns, and formulate a way to address these infrastructural deficits within the Republic of Ireland. It should be noted that the access and linkage issues related to data from clinical trials are outside the scope of the present study, as they raise a number

⁹ A number of examples of research that would be facilitated by the DASSL infrastructure and services are provided in Appendix 1.

¹⁰ <http://www.ucd.ie/issda/data/growingupinirelandgui/guiregisterofuse/>

of different issues, and operate under different laws, regulations and administrative provisions¹¹ with associated norms.¹²

Internationally, research agencies and bodies have been working to open up access to data generated by research for re-use, and several countries have put infrastructure in place to facilitate access, sharing and linkage not only of health datasets but also of administrative datasets e.g., the UK, Australia, Canada.¹³ Economic as well as health-related agendas are driving these developments; for example, in the UK, the Medical Research Council led a mapping exercise¹⁴ on behalf of major funders and industry to determine what was needed in order for UK researchers to be in the best position to make use of electronic health records for research. A substantial investment in infrastructure and research followed. For example, four major e-Health Informatics Research Centres were established in 2012, and the Farr Institutes, UK Health Informatics Research Network and the Administrative Data Research Network followed. The aim is to position the UK at the forefront of research using linked electronic health and administrative records. In 2013, in the Canadian province of British Columbia, the Minister for Technology, Innovation and Citizens' Services together with the Minister for Health established a Working Group to examine the potential for more efficient use of public sector information for the purposes of research and innovation.¹⁵ The Group examined options to support acquisition, linkage and access to data from multiple sectors in a secure manner. Having examined initiatives in other jurisdictions, they found that 'there was a clear trend towards combining growing public sector data assets with emerging technologies, research techniques and analytical tools to unlock significant value for the benefit of citizens' (see Appendix 2 for summary of benefits).

The Organisation for Economic Co-operation and Development (OECD), recognising large datasets as a core asset for the economy and well-being, has launched a number of initiatives under the 'data-driven innovation for growth and well-being' banner.¹⁶ The OECD sees exploitation of data as capable of 'fostering new industries, processes and products and creating significant competitive advantage'.¹⁷

By contrast, with some exceptions, such as, for example, initiatives funded under The Strategy for Science, Technology and Innovation 2006-13 and the Department of Public Expenditure and Reform's (DPER) Open Government and Open Government Partnership activities, Ireland has lagged behind in recognising the potential value of data, particularly health-related data. Developments such as the Irish Social Science Data Archive, which provides researchers with access to social science and related datasets have been poorly resourced while the data linkage service offered by the Central Statistics Office (CSO) has only been provided on a limited basis. A structured and coherent approach is critical in order to exploit the benefits of data, much of which have been collected and financed from the public purse and at considerable expense.

The output from the HRB Data Project is intended to feed into the development and implementation of the new corporate strategy for the Health Research Board.¹⁸ The project involved collaboration with a number of HRB colleagues, including colleagues in the Research Funding Directorate and the Information Systems Directorate who work on various aspects of archiving, access, sharing and linkage of publicly funded research data and existing routinely collected data.

1.2 Project Aims, Methods and Procedures

The aim of the HRB Data Project was to explore the current situation vis-à-vis practice and culture in relation to data archiving, access, sharing and

¹¹ http://ec.europa.eu/health/human-use/clinical-trials/information/index_en.htm#cti

¹² It should be noted, however, that the culture regarding data sharing in clinical trials is changing rapidly towards greater openness, and it is conceivable that, in time, the infrastructure proposed here could serve as a model for data sharing in this domain, or be used by researchers re-using clinical trial data.

¹³ OECD (2013). *Strengthening Health Information Infrastructure for Health Care Quality Governance. Good practices, new opportunities and data privacy protection challenges*. OECD Health Policy Studies 2013, OECD Publishing.

¹⁴ <http://www.rcfprogress.mrc.ac.uk/>

¹⁵ BC Centre for Data Innovation 2014. 'BC Centre for Data Innovation Final Report of Working Group' at http://www.gov.bc.ca/citz/down/BC_Centre_for_Data_Innovation-FINAL.pdf

¹⁶ <http://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf> (Oct 2014 - printed synthesis).

¹⁷ <http://www.oecd.org/sti/ieconomy/data-driven-innovation.htm>

¹⁸ Proposed Project: Data archiving, access, sharing and linkage coordination project. HRB Internal Document. 2013. The 2016-2020 HRB Corporate Strategy was launched in January 2016 and reflected inputs from this work.

linkages among the health research community and data custodians in Ireland. The specific objectives were to:

- identify the main actors and agencies in the public health and health services research (PHHSR) ‘data’ communities that have an involvement in data collection, access, sharing and linkage in Ireland
- capture their practices, and
- identify barriers and facilitators to data re-use and sharing through exploring their knowledge, attitudes and beliefs.

Identification of main datasets and actors

In collaboration with colleagues in the HRB, actors and agencies involved in the collection, use, sharing and linkage of data were identified. These included, for example, researchers in population health and health services research with large datasets; custodians of national patient registers; HSE personnel and other health research organisations and researchers; repository holders, e.g., the Irish Social Science Data Archive (ISSDA) and the Irish Qualitative Data Archive; and government departments and agencies such as the Department of Health, CSO and Forfás.

Development of questionnaires

The international literature relating to the themes under investigation was scanned and the resulting information was used to develop questionnaires. The HRB’s Executive Team and the Data Project Group also contributed to the development of the questionnaires.¹⁹ Finally, input was sought and received from the International Population Data Linkage Network, and from participants in a networking session that took place at the Scottish Health Informatics Programme (SHIP) conference in Scotland in 2013.

Two questionnaires were developed – one for repository holders and one for researchers/holders of large datasets. These questionnaires were modified, as required, for interviews with various government/agency personnel and a set of questions focusing on possible solutions, titled ‘The way forward’, was explored in greater detail with these informants. In the case of interviewees from agencies within the broader data ecosystem,

for example Enterprise Ireland, the questions were used as a springboard for a wide-ranging discussion on needs and options.

Strategic, solutions-focused interviewing

A formative approach to the selection of interviewees and research instruments was adopted. Completed interviews shaped the direction of future work: later interviews explored possible infrastructural solutions with key strategic actors who would be influential in supporting or implementing possible infrastructural solutions.

After a number of interviews with researchers and repository holders had been carried out, it was clear that there was broad agreement that an infrastructure which could support data access, linkage etc. was needed in Ireland; that its absence was a major inhibitor to research, and that the situation needed to be addressed as a matter of urgency.

The Department of Health draft document *Research data and innovation plan for Healthy Ireland: Pathways and strategies to maximise impact* (July 2014) also highlights the need for access, sharing and linkage of data to inform the implementation and evaluation of the **Healthy Ireland** framework,²⁰ a cross-government framework for action to improve the health and well-being of Irish residents over the coming generation. Researchers need greater access to existing datasets and also need to be able to bring different datasets together to provide the evidence to underpin Healthy Ireland initiatives.

Having identified these gaps in provision, the focus of the project changed to exploration of the type of supportive data infrastructure that would best suit the Irish data ecosystem; the legal frameworks involved; the identification of possible key players, champions, collaborators and funders.

1.3

The Wider Data Ecosystem in Ireland

A scan of the national data ecosystem i.e., agencies and bodies involved in the production or use

¹⁹ The design of the project was shaped by inputs from the HRB’s Executive Team and a Data Project Group comprising the following HRB personnel: Teresa Maguire, Donna Tedstone, Maura Hiney, Patricia Clarke, Sarah Craig, Mairéad O’Driscoll and Ros Moran.

²⁰ *Healthy Ireland: A Framework for Improved Health and Wellbeing*, 2013-2025. Department of Health, 2013.

of data and data-related strategy (see Figure 1) revealed a significant range of data-related activities with regard to data access, sharing and linkages that were relevant to the HRB Data Project, including possible infrastructural solutions. For example:

- research and data-related strategy (Science Foundation Ireland (SFI), Forfás)
- government departments and agencies (Department of Health (DoH); Department of Public Expenditure and Reform (DPER); Department of Jobs, Enterprise and Innovation (DJEI); Department of Agriculture Food and the Marine (DAFM), Teagasc)
- library/information systems (Digital Enterprise Research Institute (DERI), Digital Repository of Ireland (DRI), Health Service Executive (HSE))

- jobs and enterprise (Enterprise Ireland (EI), Irish Centre for High End Computing (ICHEC) – addressing big data, data analytics, supercomputing, connected health);
- HSE (Information System Framework, Knowledge and Information Strategy), and
- EU (Directorate General Connect, Directorate General Research).

Given the range of relevant activities in other areas of the national data ecosystem, it was decided to (i) interview key actors in this wider ecosystem in the expectation that they might have some stake in initiating, providing or using a supportive infrastructure; (ii) visit other jurisdictions where tried and tested solutions were in place; (iii) interview strategists on the feasibility of the

Figure 1: Players and initiatives in Irish data ecosystem – shared needs



proposed infrastructural model developed as part of the project, and seek their input on how it might be implemented.

Findings from these activities have been integrated into the discussion in later chapters. The HRB Data Project work revealed that the issues of data access, sharing, storage and linkage identified as key issues for the health research community were of immediate or emerging concern to a large number of institutions, entrepreneurial initiatives and government departments. Interest in these issues manifests in increasing demand for data to inform policy and planning across government and agencies; moves towards open data and the sharing of research and administrative data; the growth of big data and data analytics allied with developments in supercomputing. A willingness by government and public administration to exploit these developments for economic and scientific benefits (e.g., Action Plan for Jobs, the HRB, SFI and ESRI strategic plans) and social and cultural benefits (e.g., Open Government Partnership) was observed.

1.4

Raising Awareness of Benefits of Open Access to Research Data

It was decided to use the interviews with health researchers and data custodians to raise awareness of the work of Science Europe and the National Steering Group on Open Access (chaired by the HRB)²¹ and the possible benefits of open access to research data for health research in the future, and to gather feedback on same. The interviewer guided interviewees through prepared briefing materials,²² and gathered information regarding their awareness and practices as well as their attitudes to data management. Many researchers found working through the questionnaire and the explanations provided by the interviewer to be a hugely valuable learning experience. A copy of the HRB Open Access Policy²³ was also provided to interviewees for their information.

A list of the study informants is set out in Appendix 4. In total, 59 informants contributed to the study. For the most part, face-to-face interviews were conducted using either the questionnaires or lists of questions.

Ten telephone interviews were conducted. Three key informants were interviewed a number of times, i.e., the CSO, National Statistics Board (NSB) and DoH. Typically, interviews were conducted at the interviewee's place of work or at the HRB, and lasted for 1.5 hours. Almost all interviews were taped with the interviewees' permission. It was agreed with interviewees that tapes would be used by the researcher to inform the write-up, but no attributions to them would be made in this text.

The following chapters present the main findings and themes emerging, with a focus on their implications and the recommendations arising. Chapter 2 discusses the different types of data in the health area and related governance issues. Safeguarding provisions and the necessary infrastructure to create an enabling environment for health research are identified for safe access to, and (re)use of, data. Interviewees' practices and views informed these deliberations.

Chapter 3 focuses on interviewees' attitudes to a range of data-related issues and the culture around data sharing and linkage. Interviewees' views on required developments and workable incentives for data sharing reinforced the findings discussed in Chapter 2 relating to the required infrastructure.

Chapter 4 configures the infrastructure needed to support health research, and proposes a data access storage sharing and linkage (DASSL) model for discussion and comment; this is based on models used mainly in Northern Ireland, Scotland and France. The proposed DASSL model has been shaped by discussions with the CSO and other key stakeholders. In addition, informants from the NSB and the DoH contributed to thinking regarding possible legislative and implementation frameworks. It is proposed that the DASSL model would be implemented via infrastructure and services housed in a newly established entity – the Research Data Trust.

Chapter 5 presents recommendations based on the preceding chapters; it ends with a section on conclusions.

²¹ Patricia Clarke of the HRB chairs this group.

²² Patricia Clarke, (Policy, Evaluation and External Relations (PEER)), HRB, provided the briefing materials. See Appendix 3.

²³ <http://www.hrb.ie/research-strategy-funding/policies-and-guidelines/policies/open-access/>

2

Health and Related Data – Safeguarding Provisions for Safe Access and Use

2.1

Introduction

Interviewees were asked a range of questions about:

- the types of data they collected, where the data were stored, if the data were shared and, if so, with whom and under what circumstances;
- practices in relation to obtaining consent from participants in research studies, data linkage and their knowledge and use of repositories and their data management practices; and
- the major gaps in the Irish context in relation to research infrastructure to facilitate access, (re)use, sharing, storage and linkage of data for health research and what agencies/individuals should be involved in the provision of such infrastructure.

Analysis of the interviewees' responses informed the discussion below as well as the discussion in Chapter 3.

There was general agreement among those interviewed that there should be greater access to research and health-related data, particularly data funded from the public purse. Many reasons were given. Chief among them were expanding the opportunities to explore a broad range of research questions through the use of such data, verifying research findings and preventing duplication of data collection and research findings. The issue

of greater and open access to research and related data is increasingly being debated in the research community as it is in the broader data ecosystem.^{24,25} In addition, health researchers have noted that some types of research, including research into the efficacy of treatments, often require large numbers of participants and the pooling and linkage of data from different studies. This increases the sample size and the validity of findings.²⁶ However, researchers typically have encountered a number of problems in their efforts

²⁴ See, for example, Minister Howlin launch of consultation on open data policy paper: <http://www.per.gov.ie/en/minister-howlin-launches-open-data-licences-paper-for-public-consultation-at-public-service-seminar/>

²⁵ For example, discussions about re-use of public sector information <http://data.fingal.ie/About/>, open data portals, e.g., <http://www.statcentral.ie/>, open government, the potential of 'big data' and pro bono mining of data held by voluntary organisations for social purposes, e.g., Datakind, <http://www.datakind.org/>. An interesting data project that has received funding (£500,000) from the Big Lottery Fund in Northern Ireland involves 'data journalists' working with voluntary groups to exploit administrative data in education, health and other public services to serve community needs. In one case, figures from every GP practice in Northern Ireland were mapped, and this helped to identify geographical areas where dementia was potentially under-diagnosed.

²⁶ For example, by combining data from a total of three million procedures carried out under general anaesthesia in every public hospital in Ireland and the United Kingdom, researchers established that accidental waking occurs in one in 19,000 operations. The research identified risk factors and consequences for patients, and made recommendations for changes in practice, e.g., use of a simple checklist at the start of each operation, to minimise the incidence of patients waking up. <http://www.nationalauditprojects.org.uk/NAP5report> and Silins *et al.* (2014) were able to increase statistical power by combining participant-level data from three large, long-running longitudinal studies in Australia and New Zealand. Consequently, they were able to demonstrate dose-response relations between the frequency of adolescent cannabis use and adverse young adult outcomes such as school completion, degree attainment, later cannabis dependence and suicide attempt. The researchers concluded that prevention or delay of cannabis use in adolescence is likely to deliver broad health and social benefits. [http://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366\(14\)70307-4/abstract](http://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(14)70307-4/abstract)

to link data – e.g., gaining legitimate access to identifiable data from another study and related data protection issues, the absence of unique personal identifiers, uncertainty regarding the legal framework which applies when data are linked, (leading to a culture of caution in sharing data), sourcing of the required skills to carry out the linkage, etc. These difficulties and obstacles deter the exploration of research questions.

This chapter presents an outline of health-related data collected in Ireland. This is followed by consideration of some important features of such data and possible infrastructural, legislative and good practice frameworks to facilitate their use. The HRB Data Project research findings show that legal access to, and processing of, data are critically dependent on the type of data collected, the circumstances under which the data were collected, and the intended use of the data. In order to carry out research in an ethical and legal manner, researchers and data custodians need in the first instance to understand the rights²⁷ citizens have in relation to the protection of their data, and to be cognisant and follow requirements relating to its use i.e. to follow the eight data protection rules for data custodians outlined on the Data Protection Commissioner's website.²⁸ Of course, ambiguities will arise in the application of these rules, and clarifications may need to be sought. Researchers need to take cognisance of legal requirements, lest they waste a lot of time pursuing research ideas which may prove impossible or unduly onerous to undertake.

2.2

Health and Health-Related Data Collections in Ireland and their Availability for Research

Many extremely valuable datasets collected in the health services, including some data collected on a routine basis – for example, patient admissions, discharges, diagnoses, and patient registers such as the Cystic Fibrosis Register – are not routinely available for research purposes. As a consequence, valuable insights, e.g., insights into effective patient care and efficient service delivery, are missed.

The wealth of potential health data available in the Irish context is captured in the catalogue of routinely collected health and social care data collections developed by HIQA,²⁹ which covers sources with national, regional and sub-regional coverage. The catalogue deals with routinely collected health and social care data, administrative collections, censuses, surveys, and national registers. It provides detailed information on over one hundred datasets gathered in Ireland, including the content of each dataset, coverage, the managing organisation, how to access the data, contact information etc. It also contains information on the nine major national health-related surveys including cohort studies, e.g., GUI and TILDA. See Appendix 5 for a complete listing of the datasets described.

The HSE, or its associated health facilities, is the data controller for much of the data (e.g., the routinely collected data, case registers etc.). The HSE (for a variety of historical and other reasons, many of which relate to concerns regarding capacity to provide safe access) has, until recently, not given due consideration to making these valuable national resources available for research purposes. An exception in this context is the work of Health Atlas Ireland³⁰ (part of the Health Intelligence Unit, HSE). Health Atlas Ireland provides an intermediary service which facilitates the use of certain HSE datasets. The HSE Primary Care Reimbursement Service (HSE-PCRS) pharmacy claims database is perhaps one of the most frequently requested routinely collected datasets in Ireland, and has been used to very good effect to inform Irish treatment regimes, pricing policy

²⁷ <https://www.dataprotection.ie/docs/A-guide-to-your-rights-Plain-English-Version/858.htm>

²⁸ <https://www.dataprotection.ie/docs/a-guide-for-data-contollers/696.htm>
 'A data controller who holds information about a person must:
 – get and use the information fairly;
 – keep it for only one or more clearly stated and lawful purposes;
 – use and make known this information only in ways that are in keeping with these purposes;
 – keep the information safe;
 – make sure that the information is factually correct, complete and up to date;
 – make sure that there is enough information – but not too much – and that it is relevant;
 – keep the information for no longer than is needed for the reason stated; and
 – give the person a copy of their personal information when they ask for it'.
 Data custodians need to be aware that onus on compliance with the core aspects of the Data Protection Acts at all times rests with them.

²⁹ <http://www.hiqa.ie/system/files/Data-Catalogue-2014.pdf>

³⁰ <https://www.healthatlasireland.ie/>

etc. However, gaining access to this database has proved problematic for some researchers. It is important that the HSE and other data custodians are facilitated (through provision of supportive infrastructure and services) to make their data available in a safe manner, so that the economic and health-related benefits of these data can be fully realised.

Anonymised data from a small portion of the HIQA catalogue have been repositied among the approximately 40 datasets in the Irish Social Science Data Archive (ISSDA) (see Box 2 below and Appendix 6 for a listing of data accessible through the ISSDA), and therefore are easily accessible to researchers. Datasets are made available in accordance with clearly set out protocols. The ISSDA also makes available selected data collected by the CSO, such as censuses, the Quarterly National Household Survey etc.

There is growing use of the data available through the ISSDA (see Appendix 7 for usage figures 2002–14) but, as noted, only a handful of the other HIQA health listings have been exploited for research purposes. As stated above, the difficulties of getting safe access to data have hindered researchers, as have the difficulties in safely linking data. In contrast, the Nordic countries have a culture where unique personal identification numbers are recorded on health and administrative databases (in effect turning them into registers), many of which can be accessed and linked through national statistical agencies, such as bodies similar to the CSO in Ireland. For example, Statistics Denmark, under protection from the Act on Processing of Personal Data and the set of conditions in place for use of data in scientific projects,^{31, 32} and through the provision of supportive infrastructure, protocols etc., enables a large volume of valuable health-related research to be carried out. The data made available through these legislative and institutional measures has brought the Nordic countries to the fore in register research, and has made possible such research as the internationally recognised study which has provided strong evidence against

the hypothesis that the MMR vaccination causes autism.³³ This research involved the linkage of six separate databases, and therefore was able to provide strong evidence above and beyond that provided by previous research using data from case series, ecological or cross-sectional studies. It resulted in increased uptake of the MMR vaccine internationally, with consequent public health and economic advantages.

In addition, it is well documented that the more datasets are used, the better their quality, since regular use identifies inaccuracies and deficiencies as they arise – deficiencies which can then be tackled. It is undoubtedly the case that the quality of all the Irish listings above would benefit from greater use – see, for example, the debate regarding Hospital Inpatient Enquiry (HIPE) data, Fogarty 2013.³⁴ The usage of unique identifiers, and the availability and linkage of data, has provided a considerable economic (pharma and related industries), social (evidence-based policy and practice) and research (increased international competitiveness) dividend for Denmark and other Nordic countries, which several countries are now emulating.

2.3 Safeguarding Provisions Relating to Consent and Other Guiding Principles

In the Irish context, data such as those from the HIQA catalogue could be made available and used if safeguarding provisions were put in place.

Such provisions could include, for example, legal, infrastructural, institutional or technical measures that can unlock the value of data and facilitate their safe usage. Such provisions help achieve good governance. McGrail *et al.*³⁵ describe data governance as being ‘about the processes and controls in place to cover the original collection of data, their protection from physical and technical systems, their disclosure and use, and ultimately their archiving or destruction’. HIQA notes that ‘information governance provides a consistent

31 Thygesen *et al.* 2011, pages 12–16 in http://sjp.sagepub.com/content/39/7_suppl/12.full.pdf+html

32 The Act specifies that use of the CRS or Civil Registration System for research purposes does not require informed consent of study participants. However, researchers are required to seek project approval by the Data Protection Agency, which ensures that safeguarding measures in the use of data are adhered to. See Schmidt *et al.* 2014 at <http://link.springer.com/article/10.1007/s10654-014-9930-3#page-1>

33 <http://www.nejm.org/doi/full/10.1056/NEJMoao21134#t=articleTop> data

34 http://www.medicalindependent.ie/29007/should_we_believe_the_hiipe

35 McGrail *et al.*, forthcoming, p.5 <http://www.springer.com/us/book/9783319236322#aboutAuthors>

way for people working in health and social care to deal with the many different legal provisions, guidance and professional codes of conduct that apply to handling personal health information'. HIQA also notes that good information governance allows organisations and individuals to ensure that personal information, such as that contained in a healthcare record, is handled legally, securely, efficiently and effectively.³⁶

In the following sections, a variety of provisions which facilitate safe access to, and use of, data in the conduct of research are discussed. The safeguarding and protection of privacy, confidentiality and security in the sharing of data was deemed vital by the researchers interviewed.

Of foremost consideration in collecting health data is the need to obtain consent from individuals participating in a research study.³⁷ In normal circumstances, under the Data Protection Acts 1988 and 2003, a researcher is required to obtain consent from a person providing information for the purposes of a research study. These data would typically include directly identifiable information such as the person's name and address.

In addition, many research studies involve the analysis of data that were collected for one purpose but are now to be used for another, i.e., to answer a different research question, thereby ensuring efficient use of research resources. For example, a researcher might wish to look at diabetes indicators in patients who were studied in the first instance in relation to cardiovascular health. This is referred to as 'secondary analysis of data',³⁸ and involves use of the data for purposes other than those for which they were originally collected. It is considered optimal to obtain consent from the research subject for the use of their data for secondary purposes.

However, in conducting research, it is not always possible to obtain consent for a variety of reasons; for example, the patients from whom data are to be collected may be physically unable to give consent. In spite of that, the planned research may be of considerable public health importance. HIQA (2012)³⁹ notes: 'There is a need to strike a balance between the service user's right to personal privacy and the desirability of making information available to improve the quality and effectiveness of care through audit and research'. More recently, the 2013 UK Government response⁴⁰ to the Caldicott Review into responsible sharing of patient data, accepted a seventh principle to the Caldicott Guardians⁴¹ guiding principles for the handling of patient-identifiable information, namely 'the duty to share information can be as important as the duty to protect patient confidentiality'.

A great deal of work has been carried out internationally on the development of guiding principles – see, for example, the OECD's Privacy Guidelines OECD 2013⁴² and OECD 2013⁴³, 2015.⁴⁴ These principles flow from laws related to data protection, human rights, freedom of information, common law duty of confidentiality etc., but importantly also, they flow from reflection and deliberation on experience gained, as Sethi and Laurie (2013)⁴⁵ (both academic lawyers) would describe it 'from navigating the regulatory spaces in between the legal architecture'.

³⁶ <http://www.hiqa.ie/healthcare/health-information/information-governance>

³⁷ Consent can be explicit, i.e., clearly and unmistakably stated, for example in writing, or implied; leaflets or notices may be handed out or displayed in healthcare settings to inform patients that their information may be used for local clinical audit. Implied consent is considered valid for the sharing of information within the circle of care, including for billing purposes. See HIQA 2012 - <http://hiqa.ie/system/files/Review-Secondary-Use-Health-Info.pdf>

³⁸ Secondary use of data is use of information for purposes other than those originally specified, e.g., data collected in the course of one research study being used for another, or data collected in the course of delivery of healthcare being used for purposes other than direct patient care, e.g., audit, performance monitoring, service planning, epidemiology etc.

³⁹ <http://hiqa.ie/publications/international-review-secondary-use-personal-health-information>

⁴⁰ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/251750/9731-2901141-TSO-Caldicott-Government_Response_ACCESSIBLE.PDF

⁴¹ A Caldicott Guardian is a senior person responsible for protecting the confidentiality of patient and service user information and enabling appropriate information sharing. Each NHS organisation is required to have a Caldicott Guardian; this was mandated for the NHS by Health Service Circular: HSC 1999/012. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf

⁴² OECD 2013. Privacy Guidelines. http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf. The OECD privacy guidelines served as a foundation to the development of country-specific privacy legislation, including the EU Data Protection Directive (Directive 95/46/EC) – see McGrail *et al.* forthcoming p.5 <http://www.springer.com/us/book/9783319236322#aboutAuthors>

⁴³ OECD (2013). *Strengthening Health Information Infrastructure for Health Care Quality Governance. Good practices, new opportunities and data privacy protection challenges*. OECD Health Policy Studies 2013, OECD Publishing

⁴⁴ OECD 2015. *Health Data Governance*. <http://www.oecd.org/publications/health-data-governance-9789264244566-en.htm>

⁴⁵ Sethi and Laurie (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

BOX 1

Caldicott principles which apply to the handling of patient-identifiable information are:

1. Justify the purpose(s) of every proposed use or transfer of data.
2. Don't use it unless it is absolutely necessary, and
3. Use the minimum necessary.
4. Access to it should be on a strict need-to-know basis.
5. Everyone with access to it should be aware of their responsibilities.
6. Understand and comply with the law and
7. The duty to share information can be as important as the duty to protect patient confidentiality.

An evolving interdisciplinary literature informed by lawyers, philosophers, regulators, data custodians, health service personnel and a variety of other stakeholders including the public is contributing to the development and implementation of guiding principles (relating for example to privacy, consent, anonymisation, public interest, security, access, sanctions), and best practice (i.e. examples of principles in action, guiding frameworks etc.) for the sharing and linkage of data – see for example the Scottish Government approach⁴⁶ and Laurie *et al.* 2015.⁴⁷ These aim to provide a common framework for decision-making by identifying the kinds of issues arising and their implications, and so feeding into the deliberative processes in relation to the authorisation of projects or not.

Guiding principles are not rules; rather, they are applied in the exercise of informed judgement and make for responsible decision-making in relation to the vetting of research studies. Extant laws have to be adhered to, and guiding principles, *inter alia*, help in navigating the grey areas between regulatory frameworks and also help to inform the interpretations required by the extant legislative environment.

In the Irish context, HIQA's *National Guiding Principles for National Health and Social Care Data Collections 2013*⁴⁸ provide important guidance

for data collections which cover, *inter alia*, data access, data quality, primary and secondary use of data, governance arrangements etc. The Irish Data Protection Commissioner (DPC) has provided guiding principles for the sharing of data, including the sharing of personal data between state agencies⁴⁹ (see Appendix 8 for explanations of these principles). The guiding principles are as follows:

1. demonstrable justification
2. explicit legal basis
3. authorisation
4. transparency
5. data minimisation
6. data access and security
7. data retention.

What would enable research in the Irish context would be the operation of an agreed designated authorising entity which could make determinations in accordance with good governance principles and procedures in relation to the conduct of research involving secondary use, sharing and linkage of data etc. Operating in this manner, such an entity would allow for rounded risk assessment which would address the issue of consent along with other guiding principles in making a determination about safe data use, thus providing for good governance solutions to fit particular circumstances and research scenarios.

⁴⁶ <http://www.gov.scot/Topics/Statistics/datalinkageframework>

⁴⁷ Laurie *et al.* (2015). *On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research?* <http://www.sciencedirect.com/science/article/pii/S1386505615300356>

⁴⁸ http://www.hiqa.ie/system/files/Guiding-Principles-Data-Collections_o.pdf

⁴⁹ https://www.dataprotection.ie/viewdoc.asp?m=m&fn=/documents/guidance/Data_Sharing_in_the_Public_Sector.htm

BOX 2

The Irish Social Science Data Archive (ISSDA)

The Irish Social Science Data Archive (ISSDA) is a trusted repository, preservation and dissemination service for social science data. It holds a range of key Irish and international datasets (more than 40, see Appendix 6), and makes them available for secondary analysis by students, academics, and researchers in the public and commercial sectors.

All data disseminated by the ISSDA have been rigorously checked to ensure that they have been fully anonymised, thus protecting individual data confidentiality. The ISSDA has well-developed data governance arrangements, with documentation on policy, procedures, protocol and user guides. The ISSDA provides access to a number of well-used survey datasets of interest to the health area, some of which have been funded by the HRB, e.g., GUI, TILDA, SLÁN, Survey of Income and Living Conditions (SILC).

The ISSDA was initially funded under Cycle 4 of the Programme for Research in Third-Level Institutions.⁵¹ Although it provides a national service, it is not in receipt of ongoing funding and is currently being supported by the UCD Library. The repository, which aims to respond to increasing demands and engineer itself for the future, requires an injection of resources. There was strong agreement among interviewees that the ISSDA is a very valuable part of the Irish research infrastructure and that it needs to be adequately resourced. It was felt by some that the HRB, Irish Research Council (IRC) and SFI should jointly invest in such an infrastructure. There would be synergies between the ISSDA and the structures proposed in Chapter 4.

We will return to the supportive role that such governance can play in enabling research where public good is balanced with privacy, but first let us consider two types of data files that provide for safe use of research data.

2.4 Safeguarding Provisions Relating to Types of Data

Two safeguarding provisions relating to types of data that can facilitate the conduct of research are described in this section – anonymisation or de-identification and the creation of research micro files (RMFs).

Anonymised microdata files (AMF)

Where personal data are wholly anonymised or de-identified, data cannot be linked to the subject, and reversibility is impossible (notwithstanding a lively debate in the literature around the possibility/impossibility of achieving complete anonymity). Wholly anonymised data are outside the requirements of the Data Protection Acts and the issue of consent for

their re-use does not arise. Thus, a researcher can access another researcher's or agency's anonymised dataset provided that the data controller is willing to share the data. Typically, agreements around sharing are made between the parties, and protocols are put in place to protect patient information (for example, access may be restricted to bona fide researchers) and to ensure best practice. The CSO makes a large body of anonymised data available through the ISSDA (see Box 2 below), which as we outlined above, can be accessed by researchers in accordance with a minimally demanding set of protocols. The CSO typically refers to such data as anonymised microdata files or AMF data. Similarly, researchers may lodge survey data for re-use in the ISSDA, and these AMFs are often used for health research. For example, the Growing Up in Ireland study has been made available as an anonymised dataset which has been used extensively by the health and social research communities.⁵⁰

⁵⁰ <http://www.ucd.ie/issda/data/growingupinirelandgui/guiregisterofuse/>

⁵¹ The PRTLTI was part of Ireland's EU Structural Funds Programme 2007–2013, which was co-funded by the Irish Government and the European Union – 'Investing in Your Future'.

BOX 3

The Central Statistics Office

The **Central Statistics Office (CSO)** was established in 1949 as Ireland's national statistics office. Its status was formalised in legislation with the enactment of the Statistics Act, 1993. The mandate of the CSO, as set out in that Act, is: 'The collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State'. The CSO is also responsible for coordinating the official statistics of other public authorities and for developing the statistical potential of administrative records. Particular attention is paid to the specialist needs of business and the research/academic community for more detailed and focused data. See <http://www.cso.ie/>

The **National Statistics Board** is responsible to Government for developing a national statistical strategy (through the Department of the Taoiseach). This responsibility includes the setting of priorities for the compilation and development of official statistics in Ireland and guiding the strategic direction of the Central Statistics Office. See <http://www.nsb.ie/>

Official Statistics are defined in the Statistics Act, 1993 as statistics compiled by the CSO or any other public authority whether under the Statistics Act, 1993 or otherwise.

Research micro files (RMFs)

While anonymisation should be strived for where possible, certain research in the public interest may require non-anonymised data. When data are stripped of identifiers and of information that carries a risk of disclosure – a person's home address, for example – there may not be sufficient detail to answer the research question of interest. Thus, anonymised data may be of limited benefit to researchers. Under the Statistics Act, 1993, the CSO can provide an enabling and safeguarding service to the research community by creating research micro files (RMFs). This involves the CSO taking the research data,⁵² processing it under the Statistics Act, 1993, and putting protections on the data towards anonymising/de-identifying it. Thus, the researcher is given access to a somewhat reduced but useful dataset⁵³ and a rigorous procedure for access and use is put in place.

Access to RMFs is provided under the legal framework of the Statistics Act, 1993 and use is for statistical purposes only. There is a statutory requirement for researchers working on such data files to be appointed as an Officer of Statistics by the CSO. The appointment is time limited,

signed off by the Director General of the CSO and relates to a specific body of work. The Officer is subject to the full rigour of the Statistics Act, 1993. Matching or linkage of RMF data to other CSO data is prohibited unless written permission from the Director General of the CSO is forthcoming.⁵⁴

The use of RMFs where sensitive data are at issue, in addition to requiring sign-off by the Director General of the CSO, would also customarily require that the DPC 'had noted the project with no concerns'. The appointment by the new DPC of additional resources to the health area is welcomed and will undoubtedly facilitate health researchers in their use of RMFs, inter alia. In addition, in the interests of good governance, the CSO would normally require approval by the data controller (for use of the data controller's data), in addition to requiring the approval of a research

52 For the most part, such data would be collected under the Statistics Act, 1993.

53 However, since a risk of indirect disclosure remains, the data would be treated as if they were personal data.

54 <http://www.isscop.ie/officialstatistics/> 'Official Statistics' are agreed between the Director General of the Central Statistics Office (CSO) and the head of a public authority which compiles statistics when the statistics 'are considered to be of sufficient public interest and satisfy the following criteria:

- produced by or on behalf of a public authority
 - continuous i.e. there should be a reasonable expectation that the published statistic will be updated with new data to provide comparability over time
 - where a statistic is produced as a "one-off" the Director General of the CSO, in consultation with the responsible public authority, may deem the statistic an Official Statistic if it is considered to be of public interest
 - in the public domain.
- The list of Official Statistics will be developed and updated over time'.

ethics committee. In a case where medical records are involved, there are further requirements, since medical records are treated as a special category under the Statistics Act, 1993.

At present, for RMFs based on medical records, the CSO would seek:

- a view from the data owner/provider/controller as to the bona fides of the research and the suitability of the data
- permission from the Minister of Health (which would typically encompass an opinion from a research ethics committee)
- the project to be noted by the DPC without any immediate objection (the DPC may raise an objection at a future date).

These requirements would have to be met before the Director General of the CSO will consider signing off on such a project.

While these requirements are somewhat cumbersome, they provide an avenue by which certain valuable studies in the health area can be carried out using RMFs, with the assistance of the CSO. There are, however, limitations to the use of the RMFs. This valuable service is primarily available for data which are used or have the potential to be used in the compilation of 'official statistics' (broadly speaking all the data in StatCentral). The research community needs to make a case for inclusion of datasets of particular value to research in the body of official statistics. There is scope to further develop official statistics in the context of the health sector, as highlighted in the recent National Statistics Board Strategy for Statistics 2015 to 2020.⁵⁵ For example, datasets of particular value to researchers, such as the PCRS data, would be a valuable addition to the set of official statistics. The research community needs to involve itself as stakeholder and advocate for the further development of official statistics in the health sector leveraging on the open data agenda and joined-up government initiatives.

2.5

Safeguarding Provisions Related to Building a Robust Environment for Health and Related Research

There is scope to expand access to data for research through greater use of the safeguarding provisions described in sections 2.3 and 2.4 above, but also through integrating them into a robust infrastructure involving mechanisms such as a 'trusted third party' or 'honest broker' services, and facilities/infrastructure such as a 'safe haven' and a 'health research data hub' – all operating within robust governance structures.

A health research data hub – providing safe access to data and use of a trusted third party service

As outlined above, valuable health and social care datasets in addition to administrative data are collected in the Irish context, and it would be desirable that safe means, which enjoy the confidence of the health community and related stakeholders, be created for the storage of, and access to, such data.

From the literature, we see that *federated/distributed* (e.g. Datashield,⁵⁶ Scotland; UKSeRP,⁵⁷ Wales) or *centralised* (e.g. SAIL – Secure Anonymised Information Linkage Databank, Wales; New Zealand) options have been developed to facilitate safe access, sharing and linkage of data.⁵⁸ In a federated solution, data typically rest with the custodian and data are dealt with on a project-by-project basis, whereas centralisation typically involves warehousing or banking of data. Both solutions can avail of the services of a trusted third party service which de-identifies data, so that a researcher never has access to the identifying information of any other data custodian. In this way, the trusted third party supports the operation of a data hub, through which a researcher can gain access to safe or protected data.

Access to, and use of, such protected data is only possible in line with adherence to sophisticated governance mechanisms (e.g. accredited and approved research, safe researcher training; this

⁵⁶ <http://www.datashield.ac.uk/publications/>

⁵⁷ UK Secure eResearch Platform – http://www.farrinstitute.org/centre/CIPHER/37_Innovative-Governance.html

⁵⁸ See Chapter 4 for further discussion.

will be discussed further in Chapter 4) which facilitate adherence to confidentiality guidelines and data protection legislation. Furthermore, analysis of such data is frequently done within a safe haven environment (see below). Thus, in this manner, safe access to data is made available to researchers, thereby enabling analysis of important research questions.

In the Irish context, at least initially, a hybrid model suggests itself: some data might be centralised (e.g. centralisation of a number of critical HSE datasets for monitoring, audit and research purposes), whereas a federated solution might better fit other datasets or data custodians. Using the services of a data hub, data could be directed in a secure manner from, for example, the HSE and other relevant data holders (e.g. HIPE, PCRS and register data) and processed for safe research use by the trusted third party service using privacy protection techniques.⁵⁹ The trusted third party would release the protected data to the data hub for safe use by researchers.⁶⁰

Thus, a health research data hub could facilitate safe access for researchers to HSE, health register and other health and health-related administrative datasets, subject, of course, to approval of the data custodians and other authorities, as appropriate. Having access to such a wide variety of health data would exponentially increase opportunities for health research, opening up valuable avenues for exploration, improved patient treatments and care etc.

HIQA's 2014 report *Recommendations for a more integrated approach to National Health and Social Care Data Collections in Ireland*⁶¹ recommends that oversight for all national social and care data collections should be assigned

to a specific organisation at a national level.⁶² It further recommends that the organisation responsible for national oversight 'should develop a quality framework to drive improvements in the data quality of all national data collections.' Improving access, as well as sharing and linkage of data, would be an important aspect of such an organisation's work.

As highlighted by HIQA above, there are approximately 109 health and social care data collections in Ireland. Having a health research data hub that would provide safe access to such valuable data would transform the health research environment in Ireland and would leverage significant economic advantage. In the United Kingdom, a number of similar entities have been established and have been in operation for several years: one example of a centralised approach is the Secure Anonymised Information Linkage (SAIL) Databank in Wales, which houses more than 500,000 datasets while the FARR Institute in Scotland enables a wide range of research using a federated approach (see details in Chapter 4 and see Appendix 1 for examples of research enabled by the service).

Using a third party and linkage service to safely link data

In addition to increasing access to data via the third party mechanism described above, a similar safeguarding provision could be put in place to facilitate the conduct of research studies that need to link two or more datasets. Linkage of data is a powerful tool which, inter alia, enables researchers to address certain research questions in an inexpensive and efficient manner, e.g., by combining datasets to create one large dataset with sufficient statistical power to test hypotheses.

A trusted third party service is required when data need to be linked, since each researcher can access only the identifiable data they have collected and obtained consent for. For example, a researcher might want to know how many people receiving methadone availed of psychiatric services in a given year. In such a case, using the most simple technique, the third party would accept the two datasets in question – the National Psychiatric Inpatient Reporting System (NPIRS) data managed

59 Models complying with data protection and related legislative and ethical requirements in different OECD countries have been developed (see OECD 2015 *Health Data Governance*. <http://www.oecd.org/publications/health-data-governance-9789264244566-en.htm> in Chapter 5 and also see discussion of SHIP and SAIL in Chapter 4 of this report). A discussion of the model most appropriate to the Irish legislative, ethical and cultural context needs to take place.

60 When (as is usual) such files are processed by the third party to a wholly anonymised (versus pseudonymised) status, then they could be accessed through the CSO, ISSDA or Health Atlas Ireland as well as through the RDT/hub.

61 The principal recommendation of this report is that a strategic framework, including a detailed roadmap, needs to be developed to inform policy development and lead to greater integration of national health and social care data collections. p.7. <http://www.hiqa.ie/publications/recommendations-more-integrated-approach-national-health-and-social-care-data-collection>

62 Having reviewed a number of international examples, HIQA noted that in the jurisdictions reviewed governance arrangement for national data collections were provided for through national legislation, national or bilateral agreements or agreements with district health authorities. HIQA 2014, p.18

by the HRB, and the methadone treatment data from the Central Treatment List (CTL) managed by the HSE National Drug Treatment Centre – and use the subject identifiers in both to create a subset of patients who appear in both datasets. The trusted third party (or linkage service) would make the linked data (minus identifiers) available in a protected form to the researchers, in accordance with robust governance mechanisms, thus ensuring safe linkage of data.⁶³ More sophisticated models involving the separation of the indexing from the linkage service are described in Chapter 4.⁶⁴

Example: Research proposed by Redmond *et al.*⁶⁵ is seeking the services of the CSO to link medication dispensing (PCRS⁶⁶) with hospital discharge data (e.g., HIPE), which will allow investigation of disruptions in older patients' medication, particularly long-term medicines used to treat chronic conditions following hospitalisation. Changes in medication can occur either inappropriately or inadvertently on discharge from hospital, and can have serious implications for patient health and well-being. The study being carried out by Redmond *et al.* will allow investigation of changed medicine regimes, relapses, re-hospitalisations etc. It will help inform plans around the sharing of information about medication between primary and secondary care settings for vulnerable patients.

The CSO provides a limited 'trusted third party' and linkage service for researchers, primarily in respect of official statistics and in accordance with strict protocols. However, its capacity to respond to such requests from the research community is limited, as the CSO mandate is focused on 'official statistics' which would form a subset of the potential datasets of interest to health researchers. Some augmentation of resources for the existing third party and linkage service provided by the CSO to researchers would be desirable. In addition there is a need nationally to develop

services, which would address the needs of actors in the Irish statistical system for linkage of non-official statistical data.

It can be expected that an efficient linkage service would result in far greater numbers of researchers exploiting the potential of linked data to the benefit of policy and practice in Ireland.

Example: Use of safeguarding provisions (trusted third party services) while linking a large number of datasets e.g., Community Health Index, dispensed prescribing data, hospital admissions, General Register Office mortality data, enabled McGowan *et al.* 2009⁶⁷ to establish elements of care in the provision of methadone treatment (e.g., history of co-prescription of benzodiazepines) which were likely to influence or be a marker for a person's risk of death, and thus are of great practical value to GP practice. Without the use of linked data and safeguarding provisions, it would not have been possible to carry out this research.

Safe havens – providing a safe environment for data analysis

A safe haven is an environment in which data are held securely and in which researchers are facilitated to manipulate data safely under controlled conditions. Access to such environments is heavily controlled and there is constant supervision of users. Users are provided with sealed work folders with no Internet access, and data cannot be downloaded or transferred. The governance arrangements around accessing data in a safe haven are very stringent, and are guided by detailed protocols and documentation. Typically, the CSO is involved in the provision of access to RMFs in tightly controlled environments, which could be considered safe havens.

A safe haven environment for the general research community needs to be created in Ireland (at least five operate in Scotland and

63 The Health Intelligence Unit, Health and Wellbeing Directorate, HSE are working on a similar concept. Personal communication from Howard Johnson, 29/04/15 'Health Research Hub'

64 See Chapter 4 for more details regarding techniques and methodologies used for safe linkage e.g., ADRN video link https://www.youtube.com/watch?v=nnxz3_XGMAE, see also ADRN, Appendix 2 <http://www.esrc.ac.uk/files/news-events-and-publications/publications/themed-publications/improving-access-for-research-and-policy/> and Figure 4 in Chapter 4 of present report for Scottish SHIP example of the separation of indexing and linkage services.

65 Title of study: Medication Reconciliation – unintentional discontinuation of long-term medication post-hospitalisation. Principal Investigator: Dr Patrick Redmond, RCSI in HRB Centre for Primary Care Research – personal communication

66 Primary Care Reimbursement Service – Drugs Payment Scheme

67 McGowan C, Kidd B and Fahey T (2009). *Factors associated with mortality in Scottish patients receiving methadone in primary care: retrospective cohort study.* BMJ338:b2225

more than 15 in the UK). The availability of such a secure environment is an important resource, particularly for health researchers, given the sensitivity of some of the datasets on which they work. In addition, safe havens play an important role in creating and preserving the confidence and trust of both the wider health community and the general public. Confidence and trust are essential to the functioning of a robust health research environment.

In this section we have outlined services and infrastructure (trusted third party, and safe haven) which could be put in place and which would facilitate safe access to, and use of, sensitive health data. Were these to be put in place, growth in the type and extent of health research carried out could be anticipated; in addition, benefits to population well-being and healthcare could be expected. We have already noted that agreement on the need for the proposed services and infrastructure is shared by a range of players in the broader data ecosystem in Ireland, and that a solution, while serving the health area, would also serve broader economic, social and governmental agendas.

We shall build the enabling infrastructure proposed into a model, and discuss it in more detail, in Chapter 4.

2.6 Special Legislation

There is a considerable body of extant legislation relevant to national health and social care data collections in Ireland. HIQA 2013 provides a non-definitive list of 12 pieces of such legislation.⁶⁸ In a related vein, Sethi and Laurie 2013⁶⁹ report that the UK Department of Health while reporting on the research framework within the UK ‘identified 43 pieces of relevant legislation, 12 sets of relevant standards and 8 professional codes of conduct...’, concluding that what ‘this has bred is a culture of caution, confusion, uncertainty and inconsistency’.

While the legislative environment governing data use for research has been much criticised,⁷⁰ some pieces of special legislation have indeed facilitated access to, and use of, health data in particular circumstances. While special legislation might be usefully considered to underpin the operation of the infrastructure described in Section 2.5, some countries (e.g. Scotland) have developed alternative approaches. For example, Scotland has developed authorisation structures such as the Public Benefits and Privacy Panel. This panel uses templates and frameworks which have been developed to assess adherence to legislative requirements, but also to guide deliberation on the grey areas between legislative regimes towards a rounded risk assessment and, ultimately, to good governance in the conduct of research. In this context, Sethi and Laurie (2015)⁷¹ point out that good governance is more than just compliance with the law. They argue that legislative reform is not necessarily the most effective way to produce good governance: ‘the argument in this article is that legal reform is not required. It is, in many senses, a distraction. Instead, we posit that the broad parameters for delivering good governance are already laid down in the legal architecture and that more law is not the answer. What is required, however, is a deeper understanding of how to operate within those parameters and in keeping with the established data protection principles in both a robust and effective manner to give effect to the twin purpose of the law to promote responsible sharing whilst adequately protecting privacy’. We will return to these matters later.

In the Irish context, three pieces of legislation – two forthcoming and one existing – have potential to facilitate the conduct of research.

Health Information and Patient Safety Bill

The forthcoming Health Information and Patient Safety Bill (due to be published by the end of the second quarter of 2016) contains a number of information elements that may benefit research. The Bill provides for the Minister to prescribe, subject to controls, national health information resources (effectively population registers, indexes, databases). Once the resource is prescribed, the data controller involved can require those holding relevant information (whether in the public and/

⁶⁸ <http://www.hiqa.ie/publications/guiding-principles-national-health-and-social-care-data-collections>, p.33

⁶⁹ Sethi and Laurie (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

⁷⁰ Thomas R and Walport M *Data Sharing Review*, 11 July 2008, para 2.28. <http://webarchive.nationalarchives.gov.uk/+/http://www.justice.gov.uk/docs/data-sharing-review.pdf>

⁷¹ <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr> Sethi and Laurie (2015), p.178

or private side of the health system) to provide it, and they must comply. In addition, the Bill also provides, subject to the same controls, that the Minister will be able to prescribe data linkage or data-matching programmes. It will be for the Minister to decide what information resources and what data-matching programmes will be prescribed. Given the far-reaching nature of these powers, it is anticipated that they will be used sparingly and, in all cases, after consultation with the DPC. The provisions on health information resources and data-matching programmes are designed primarily to support health service management and disease prevention, monitoring and control. It is envisaged that only anonymised and aggregated data will be accessible by third party researchers. For example, in relation to linkage, researchers might provide their data to a data controller operating a prescribed health information resource for linkage purposes, but only anonymised data will be returned to the researchers.

It is anticipated that the Bill will result in an expansion of the amount and variety of health information available for research. The inclusion of data collected in the private healthcare system under the Act is particularly welcome, since many existing data collections do not cover the private health sector and the resultant limitations in coverage can result in biases, and limit the value and relevance of research findings and policy analyses.

Data Sharing and Governance Bill

The proposed Data Sharing and Governance Bill being put forward by the Department of Public Expenditure and Reform may also facilitate researchers, particularly in accessing administrative datasets. The Bill will provide a general legal basis for data sharing between public bodies. Among the suggestions that were the subject of public consultation around the Bill were the establishment of 'honest broker' or 'trusted third party' services⁷² – a development that would align well with the proposals made in Section 2.5 above. However, it appears that these will not feature in the forthcoming Bill. The Bill will require a screening assessment and a contingent privacy impact assessment to be carried out on the proposed data sharing, thus facilitating good governance and good practice in ensuring privacy and confidentiality.

Health (Provision of Information) Act, 1997

Access to, and use of, National Cancer Registry in Ireland data is specifically facilitated under the Health (Provision of Information) Act, 1997, where an 'overriding public concern', i.e., cancer prevention, was deemed to exist. Thus, Barron *et al.*'s (2014) research, which was covered by this Act, was able to show that recent pre-diagnostic aspirin use is protective against lymph node-positive breast cancer; the provision of such evidence was made possible by linking patient records in the National Cancer Registry Ireland database to prescription dispensing data and mammographic screening through BreastCheck.⁷³

2.7

Striking a Balance – a Role for Guidance and Codes of Practice

The above discussion shows that access to research data for re-use is not straightforward, and can be cumbersome. In the first instance, the agreement of the data controller or holder of the data to provide or share their data with the researcher is required. Patient privacy and confidentiality must be protected, and processing of data has to be carried out in compliance with the Data Protection Acts 1988 and 2003 and other relevant legislation.

What is critical – in order for health research to be facilitated to contribute maximally to patient well-being and safety, service efficiency, policy and planning etc. – is that a balance is struck between patient privacy and confidentiality on the one hand and the conduct of publicly important research on the other.⁷⁴ Internationally, there is increased recognition that this cannot be achieved if a wholly risk-averse culture predominates.⁷⁵ Research ethics committees, which of course have to operate within the law, have an important role to play in ensuring ethical conduct of research, and in this they are guided by internationally recognised

⁷³ Barron T, Flahavan EM, Sharp L. *et al.* (2014). *Recent Prediagnostic Aspirin Use, Lymph Node Involvement, and 5-Year Mortality in Women with Stage I–III Breast Cancer: A Nationwide Population-Based Cohort Study*. <http://cancerres.aacrjournals.org/content/74/15/4065.abstract?sid=4acdfe4-38c3-4548-a206-16d7b79eed76>

⁷⁴ The current debates around the proposed EU General Data Protection Regulation echo these concerns.

⁷⁵ See discussion of digital security risk management cycle in <http://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf>

⁷² <http://www.per.gov.ie/en/datasharing/>

guidelines such as the Declaration of Helsinki,⁷⁶ the Belmont report,⁷⁷ and the Council for International Organizations of Medical Sciences/World Health Organization guidelines.⁷⁸ It is good practice internationally that, in addition to consideration of matters of ethics by research ethics committees, that a governance or authorisation mechanism(s)⁷⁹ is put in place which will weigh up the risks and benefits related to access and linkage of data *specifically*, and make a determination in relation to each particular research project (e.g. Caldicott Guardians and the Scottish Public Benefits and Privacy Panel). In their deliberations, these authorising entities are guided by principles, codes of conduct, professional guidance and standards etc. The approach leads to what has become known as principled, proportionate, risk-based approach to data governance; this will be discussed in further detail in Chapter 4.

HIQA, in addition to providing *National Guiding Principles for National Health and Social Care Data Collections 2013*,⁸⁰ advises the use of measures such as data-sharing agreements between bodies/entities sharing data⁸¹ and the use of privacy impact assessments (PIAs).⁸² Data-sharing agreements typically require the agency receiving data to adhere to the same privacy, confidentiality and security principles as the data controllers, while the PIA, as advised by HIQA, will identify actual or potential privacy risks and concerns related to a project, including risks related to legislative compliance. In 2010, HIQA issued guidance in relation to PIAs and their conduct which, *inter alia*, provides guidance on how to evaluate the privacy implications of a project and how to mitigate risks. HIQA notes that in some cases when doing a PIA 'it may be necessary to balance risks to privacy of personal information against the public good while having regard to legal requirements in this area'.⁸³

HIQA's *National Guiding Principles 2013*, *inter alia*, also advocate the use of 'statements of information practice', particularly in situations where the data collector does not have direct contact with the data subject. These statements are generic documents which are made available to service users (for example, by being displayed in hospital settings), and which outline the information practices undertaken by particular services. Typically, these statements set out what information is collected, and how it will be used. In addition, they identify with whom the information will be shared and for what purpose, what safeguards are in place to protect the information etc. Thus, as a basic principle, the data subject is aware of how their personal health and social care information is being used.

HIQA's *International Review of Secondary Use of Personal Health Information* (HIQA 2012) identified a consensus among the countries reviewed regarding the need for guidance on the secondary use of information, and noted that in the Irish context the legislative provisions outlining the instances in which information can be used 'are somewhat ambiguous and open to interpretation'. HIQA cites examples of guidance documents such as the British Medical Association's publication *How to respond to requests for data for secondary purposes*.⁸⁴ A related document from Northern Ireland – *Code of Practice on Protecting the Confidentiality of Service User Information* – provides guidance for those involved in health and social care concerning decisions about the protection, use and disclosure of service user information.⁸⁵ In addition, in some countries, bodies have been established to provide guidance and advice regarding the re-use of data, e.g., the Ethics and Confidentiality Committee in England.⁸⁶

The Irish Data Protection Commissioner welcomes interactions with professional bodies and will provide advice in setting codes of practice 'to clarify how data protection rules are to be applied' to particular sectors.⁸⁷ This is an option that should be availed of by the health research community to

76 <http://jama.jamanetwork.com/article.aspx?articleid=1760318>

77 http://www.videocast.nih.gov/pdf/ohrp_belmont_report.pdf

78 http://www.cioms.ch/publications/layout_guide2002.pdf

79 These can be authorising individuals (e.g. Caldicott Guardians) or Committees (e.g. PBPP) or a combination of both.

80 http://hiqa.ie/system/files/Guiding-Principles-Data-Collections_o.pdf

81 <http://www.hiqa.ie/healthcare/health-information/information-governance>

82 <http://www.hiqa.ie/publications/guidance-privacy-impact-assessment-health-and-social-care>

83 HIQA (2010). *Guidance on Privacy Impact Assessment in Health and Social Care*. Dublin. www.hiqa.ie

84 HIQA (2012). *National Guiding Principles for National Health and Social Care Data Collections*, p.49. www.hiqa.ie

85 Department of Health, Social Services and Public Safety (2012). *Code of Practice on Protecting the Confidentiality of Service User Information*. <http://www.dhsspsni.gov.uk/confidentiality-code-of-practice0109.pdf>

86 Many of the functions of the ECC have been transferred to the Health Research Authority.

87 <https://www.dataprotection.ie/viewdoc.asp?m=m&fn=/documents/enforcement/5y.htm>

help provide much needed clarity in this area. The initiative could build on the guidance document for health research from the Office of the Data Protection Commissioner (2007),⁸⁸ the HSE National Consent Policy of 2013,⁸⁹ as well as numerous HIQA publications and the international literature, particularly those which address risk and benefit. It is to be hoped that the forthcoming EU General Data Protection Regulation which is due to enter into effect in 2018 will in its implementation be conducive to research.

2.8 Conclusions and Recommendations

Conclusions

- A rich body of health and related data exists in the Irish context, but there are barriers to the exploitation of this valuable national resource.
- A range of safeguarding provisions and data protection measures can be employed to enable the safe use of data for research and related purposes.
- A need exists to put data infrastructure and services (health research data hub, safe haven, trusted third party and data linkage services) in place to enable health researchers to overcome the barriers they currently encounter in trying to access and use research data.

Recommendations

- The current research points to the value of developing a health research data hub, which would be engineered to meet the stringent governance requirements for storage and access to health data, inter alia. It would provide safe access to agreed, routinely collected health datasets deposited by the HSE, other health data custodians and by health registers.
- There was general agreement among those interviewed that the ISSDA needs to be established on a firm financial footing, so that it can continue to provide access to anonymised social science data (AMFs) in line with emerging

needs and developments. There will be synergies between the ISSDA and the health research data hub.

- The health research community should make a case for inclusion of datasets of value to research in the body of official statistics.
- A comprehensive infrastructure for data linkage needs to be established. This would include trusted third party/honest broker, safe haven and research support facilities.
- Healthcare professionals, researchers and related stakeholders should work with the Office of the Data Protection Commissioner with a view to drawing up guidelines to steer researchers in their practice and towards the safe use of data.
- When obtaining consent from study participants, the consent should include making provision for sharing data and should also take into account any immediate or future uses of data.⁹⁰ In addition, consent forms should promise to not destroy data unnecessarily.⁹¹
- A discussion needs to take place regarding the necessity and utility of the introduction of special legislation to underpin the infrastructure and services identified as needed in the Irish context. Lessons learned from international literature suggest that special legislation will not necessarily result in good or better governance. Rather, what is needed is a robust authorising mechanism which can deliberate not only on the conformity of research projects to existing legislation but also on research projects' conformity to guiding principles and best practice which help navigate the spaces in between legal requirements. Frameworks and templates have been developed to facilitate this process.

Thus, with appropriate safeguards to protect the privacy of patient information, access and re-use of data can be facilitated, allowing for possible patient best interest and scientific advancement.

The work undertaken for this project revealed a number of synergies in the broader data ecosystem. The work suggests that the pieces of infrastructure identified as requirements for a robust health research community were also required by a variety of actors in the broader data

⁸⁸ Data Protection Commission (2007). *Data Protection Guidelines on Research in the Health Sector*. https://www.dataprotection.ie/documents/guidance/Health_research.pdf This is currently being reviewed (personal communication from ODPC, October 2015).

⁸⁹ http://www.hse.ie/eng/about/Who/qualityandpatientsafety/National_Consent_Policy/National%20Consent%20PolicyMay14.pdf


⁹⁰ The TILDA study is a good example. TILDA obtains consent from participants to allow their prescribing PCRS record to be linked anonymously to their TILDA cohort data. This has enabled valuable research work on medication use and compliance.

⁹¹ UK Data Archive <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent>

ecosystem, and should be construed as elements of the national statistical infrastructure i.e., part of the Irish statistical system.

A model which configures the elements of infrastructure needed by researchers, as identified above, is outlined in Chapter 4.

The operationalisation of the model, including the most appropriate structures and processes will require debate and input from many stakeholders; in addition, the most appropriate technical solutions will change as knowledge advances.



3

Attitudes, Practices and Incentives – Data Access, Sharing and Linkage

3.1 Introduction

The project explored interviewees' attitudes and practices around data access, sharing and linkage. Interviewees' views on the way forward were elicited: perceptions of, and attitudes towards, extant and needed infrastructures for sharing were investigated. Respondents were asked what kind of incentives would be most effective in encouraging researchers to share data, and what factors would aid compliance. Thus, questioning aimed to depict the culture around data usage and sharing, as well as respondents' views regarding the measures required to create a robust data environment for health research. The findings and associated recommendations are set out under three headings – attitudes/culture, data management, and data linkage and needed services. With regard to how to incentivise data sharing and encourage compliance, the views of interviewees are reported directly in a final section.

3.2 Attitudes/Culture

Attitudes to data sharing and linkage were found to vary, depending on the types of data involved. It was generally agreed that better access, sharing and linkage of data would facilitate research, thus enabling many valuable research questions to be tackled. It was also noted that use of data through sharing helped to ensure, and possibly improve, data quality. The possibility of re-use of expensively collected information was particularly valued. Caution increased, however, with the level of sensitivity of the data collected – the greater the sensitivity the greater the governance arrangements perceived to be needed in order to ensure safe access and use.

Therefore, while researchers generally did not have an issue with access to anonymised data, for example through the CSO and the ISSDA, they felt that particular safeguards need to be in place in order to eliminate or minimise the possibility of disclosure when providing access to register data or to routinely collected health data. The researchers welcomed the protocols that had been developed by the CSO and around the GUI study, which enabled the re-use of data. Certain existing requirements within protocols for sharing were considered over-elaborate – for example the

CSO requirement to work on Research Micro Files (RMF) files on site; in addition, some researchers perceived the ODPC to be unduly risk adverse.

A number of interviewees expressed the opinion that willingness to share data was inhibited by a fear of contravening data protection requirements. This fear was seen to stem partly from lack of comprehensive knowledge of data protection requirements (for example, none of the interviewees mentioned the 2007 *Data Protection Guidelines on Research in the Health Sector* published by the ODPC⁹² which are currently being reviewed by the ODPC⁹³) and their perceived possible ambiguities. It was also noted that 'data protection' was sometimes used as a pretext to refuse access to data by those unwilling to share data. This unwillingness seemed to stem, in some cases, from reasons such as lack of training regarding the preparation of data for sharing, the time involved, or a lack of a requirement by funders to share data.

The culture around data sharing in the health area was considered to be quite 'closed' in Ireland, with a tendency for 'researchers to hang onto their data' in contrast to the situation that applies in other countries – particularly Scandinavian countries and Australia. A number of interviewees stated, however, that a culture of openness was growing, particularly among younger researchers who had grown up in the 'digital age'. A general apprehension regarding the sharing of health data for commercial purposes was noted.

It was generally agreed that funding bodies had a great deal of power and could play a significant role in promoting a culture of sharing – we shall return to this subject later in the report.

Recommendations

- Professional associations and educators need to promote a culture of openness and good governance in relation to access to, and sharing of, research data.
- Researchers need to be more aware of requirements under the data protection legislation and existing guidelines from the

ODPC. In addition, researchers should enter into discussions with the ODPC when clarifications are needed or when perceived ambiguities exist.

- Access and sharing would be facilitated by researchers obtaining informed consent, which would allow for research use beyond the current study, e.g., obtaining consent for specified anticipated future research uses or to archive data for secondary use.
- Data custodians could facilitate use of their data by providing comprehensive metadata and user-friendly tools that would help to contextualise, visualise and analyse their data.
- Willingness to share data is facilitated by data-sharing agreements which ensure the bona fides of the research and those with whom data are shared.

3.3

Data Management Practices

During the discussions on access to data there were a significant number of references to the value of data standards involving, for example, common definitions and data dictionaries in facilitating the re-use and sharing of data, coupled with references to the Health Information and Quality Authority's (HIQA) work in this area. However, there was far less knowledge among interviewees about the important concept of data management and the value of data management plans (DMPs). DMPs 'help researchers consider, when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community' (UK Data Archive).⁹⁴

Typically, DMPs comprise the following components:

- description of the data generated during the research, and the context of the research
- metadata, standards and quality assurances measures adopted
- data formats, storage, back-up and security
- data management roles and responsibilities
- plans for sharing data
- ethical and legal issues or restrictions on data sharing
- copyright and intellectual property rights of data

⁹² The Office of the Data Protection Commissioner (2007). *Data Protection Guidelines on Research in the Health Sector*. Available at <http://www.dataprotection.ie>

⁹³ Personal communication, October 2015

⁹⁴ <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

- storage, for example, lodging data in a repository
- costing of the resources needed for data management.

A number of agencies have provided guidelines for constructing DMPs and related checklists – for example, UCD Library/ISSDA,⁹⁵ UK Data Archive,⁹⁶ Digital Curation Centre,⁹⁷ JISC,⁹⁸ Wellcome Trust,⁹⁹ and the EU Framework Programme.¹⁰⁰ In addition, in recent years many UK and international research funders (e.g., Medical Research Council, Economic and Social Research Council, Natural Environment Research Council, Wellcome Trust)¹⁰¹ have stipulated a requirement in their data policies that a data management and sharing plan be included in research grant applications. The work of the GUI team in the development of documentation around data management, and its frequent provision of training courses for researchers who wish to use the GUI data, is an interesting example of a group tackling the data sharing issue constructively. However, with the exception of researchers who are involved in large database research in Ireland, such as GUI, The Irish Longitudinal Study on Ageing (TILDA) and Lifeways, few interviewees were familiar with these important developments. The awareness-raising discussion carried out as part of this project was welcomed, as was the related information note provided to interviewees.¹⁰²

A lack of training in data management for researchers was apparent, with interviewees being aware of, or involved in, only a handful of courses or inputs to courses in academia in Ireland. Thus, the finding by HIQA that most health service data in this country are not in good shape for sharing should not come as a surprise.

Recommendations

- Adoption of HIQA and related guiding principles and practices for health and social care data collection would support the re-use of expensively collected health and health-related data, as the adoption of common data standards would greatly facilitate data access, sharing and linkage.
- Development and upkeep of DMPs should be an essential part of the research process, where data collection represents a significant part of the project. Such plans should be externally monitored, so as to ensure compliance.
- The HRB, perhaps in collaboration with other research funders, should consider the commissioning of work to provide guidance on the type of DMP that should be required of grant applicants. There is a considerable literature available which could guide this development, including work undertaken by the UCD Library/ISSDA.¹⁰³ Alternatively, suitable guidelines from another agency or funder could be used, with any necessary amendments made to suit the Irish environment.¹⁰⁴ This work could link in with work already under way in the HRB, where international developments in relation to data policies are being monitored.
- On foot of this work, funders could consider requiring researchers to include a DMP, where relevant, as part of a grant application. This DMP would include, inter alia, specification of where the data generated by the research would be repositied and how it would be set up and maintained for the purposes of sharing.
- The costing of DMPs should be an allowable expense in relevant grant applications.¹⁰⁵ The UK Data Archive has developed a costing tool that can be used for costing data management in the social sciences; this tool could also be used in the preparation of research grant applications.
- The HRB along with other funders should consider funding the development and roll-out of courses on DMPs, which could be delivered as part of undergraduate and postgraduate research methodology training programmes. An online format could be considered, for example, MANTRA.¹⁰⁶ Training materials and the delivery of training for data custodians is also recommended.

⁹⁵ <http://libguides.ucd.ie/data>

⁹⁶ <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

⁹⁷ <http://www.dcc.ac.uk/resources/data-management-plans/checklist>

⁹⁸ <http://www.jisc.ac.uk/guides/research-data-management>

⁹⁹ <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Data-management-and-sharing/index.htm>

¹⁰⁰ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/h2020-hi-oa-data-mgt_en.pdf *Guidelines on Data Management in Horizon 2020*, Version 1, 11 December 2013. The EU Framework Programme for Research and Innovation, EC.

¹⁰¹ Science Europe recently conducted a research data management survey among member organisations. Preliminary results show that the majority had research data management [RDM] policies, 25% require Research Data Management (RDM) plans, and almost 50% compensate researchers for RDM activities.

¹⁰² Drafted by Patricia Clarke – see Chapter 1 and Appendix 3.

¹⁰³ <http://www.ucd.ie/issda/data/>

¹⁰⁴ UCD Library provides examples of funding bodies 'example' DMPs. See http://libguides.ucd.ie/data/checklist_plans

¹⁰⁵ Costs which could be considered include those relating to the preparation of data for deposit and ingestion, data storage, ongoing digital preservation and curation after the project.

¹⁰⁶ <http://datalib.edina.ac.uk/mantra/>

3.4

Data Linkage and Needed Services

Linkage of data is not confined to the health area. Linkage of health to administrative and other research datasets opens up major opportunities for research. It can help to contextualise health research questions within a broader social and cultural framework; for example, linking health data to data available through the census can provide a rich picture of phenomena at very little cost.

A number of interviewees had had experience of using data linkage while carrying out health research both in Ireland and abroad. The experiences did not compare favourably. Researchers who had used the CSO to provide linking services reported long and complex negotiations involving, typically, the data controllers of the datasets they wished to link, the CSO and the ODPC. The negotiation process often took up to nine months to complete, which had knock-on effects on deliverable dates, draw-down of funding and reputational damage. The difficulties researchers experienced included:

- the number of entities involved in authorisation for linkage – e.g., data controllers, the CSO, the ODPC and ethics committees – where roles and mandates in the area were often perceived to be unclear
- lack of clarity regarding the order in which the entities should be contacted
- delays encountered, due to the time taken to complete authorisations
- perception by a number of researchers that some of the data protection issues raised by research were subject to interpretation, coupled with a feeling that the ODPC was overcautious in its approach
- lack of readiness of datasets for linking, due to lack of common standards, adequate metadata, or a data management plan
- the lack of availability of secure remote access facilities to linked data by the CSO, thus requiring researchers to work in the CSO premises, raises the issue of geographic inequality for researchers in the regions
- the lack of a unique identifier was cited by almost all interviewees as an impediment to health service delivery and data processing in the health area. The new Health Identifiers Act 2014 provides for unique health identifiers

for individuals, health practitioners and organisations, and its implementation was eagerly anticipated.

Recommendations

- A properly resourced national infrastructure is needed for secure linkage of data (e.g., trusted third party, safe haven) with related supports for researchers to facilitate not only health and well-being but also the economic and enterprise agendas.
- The NSB has championed the idea of the Irish statistical system (ISS) to describe the resource base for official statistics, and has identified the need for a national data infrastructure to address the issues which inhibit the linking and sharing of data.^{107, 108} It would be desirable that the NSB review and engage with relevant stakeholders regarding the recommendations set out in this report and their implications for the production of official statistics in Ireland.
- A balance needs to be struck between data protection considerations and the public good when decisions regarding data linkage are being taken and a risk management approach needs to be adopted, with due cognisance taken of safeguarding measures employed by researchers to ensure safe data practices.
- Guidance is needed from the ODPC and/or from professional bodies and other stakeholders in collaboration with the ODPC, in relation to data sharing and linkage.
- Training for data sharing is needed for data custodians and researchers who wish to re-use data.
- Capacity building around the statistical expertise required for linking and sharing needs to be developed. Such training programmes for researchers are regularly advertised in the UK, in other EU countries, and internationally.
- The desirability and implications of cost recovery in relation to data access and data linkage services needs to be examined.

¹⁰⁷ <http://www.nsb.ie/media/nsbie/pdffdocs/StrategyforStatistics2009-2014.pdf>

¹⁰⁸ O'Hara, P. 2013. The Irish Statistical System - Preserving Trust through Quality Standards'. Paper presented at SSISI Symposium Safeguarding Trust in Official Statistics. See <http://www.tara.tcd.ie/bitstream/handle/2262/72772/2a%200%27hara.pdf?sequence=1&isAllowed=y>

3.5

Incentives to Facilitate Data Sharing and Encourage Compliance

Interviewees were asked to suggest measures and incentives to facilitate data sharing and encourage compliance. Their views on the role that the HRB and other agencies/entities might play were explored.

Funders were seen to have a lot of power to change research culture and researcher behaviour, and it was felt that they could use this power to support better access and sharing of data. Interviewees felt that leadership was needed to harness this power.

Interviewees felt that it would be of benefit if funders developed a common set of principles regarding the re-use and sharing of data. It was suggested that the HRB, Irish Research Council (IRC) and SFI should agree a common approach so that, as one interviewee remarked, 'there would be nowhere to hide'.

Many interviewees saw a requirement by funders for researchers to deposit publicly funded research data (where copyright did not apply) into a data repository for use by other researchers as a major incentive. It is interesting to note that the Economic and Social Research Council will not provide the final payment of a grant unless the researcher deposits their data in the UK Data Service within three months of the end of the project.

It was generally agreed that preparing data for sharing involves time and money, and that funders should allow sharing and related activities, such as the formulation of a DMP, as an allowable expense. It was further suggested that the funding bodies could pay an annual subscription for a body such as the Digital Repository of Ireland to archive data.

Interviewees noted that the existing research culture does not recognise or reward data sharing or preparatory data management activities. There is a need for a system of authorship credits for members of a research team involved in managing, analysing and curating data. In addition, a proper career structure for data scientists and data

management professionals needs to be developed. The literature points to these as long-noted deficiencies that have not been addressed.¹⁰⁹

Academics should promote a requirement to consider data access and sharing in criteria used for academic selection and promotion. For example, application forms for posts could feature a section requesting applicants to specify whether they have reposit data; if so, where the data are deposited; whether they are openly available; whether the applicant's data have been used by other researchers, and if so, when and how often.

Research funding for the analysis of secondary data should be promoted and increased.

3.6

Conclusions

The HRB Data Project found that researchers are generally willing to share data. However, willingness is dependent on the type of data involved. Moreover, ambiguities in relation to the governance arrangements around sharing, and the lack of research and statistical infrastructure, are inhibiting the progress of health research in Ireland. Exploration of these issues confirmed the findings presented in the previous chapter regarding the need for infrastructure and third party/honest broker services etc. – a need shared by players in the broader Irish data ecosystem as well as those in the health area.

The next chapter configures the pieces of infrastructure identified as needed to support health research, and proposes a model for discussion and comment that is based on experience of similar models in use, mainly in Northern Ireland, Scotland and France.

¹⁰⁹ See, for example, the Expert Advisory Group on Data Access [EAGDA] report, *Establishing incentives and changing cultures to support data access*. May 2014

4

The DASSL Model and the Proposed Research Data Trust

A proposal for a data access, storage, sharing and linkage infrastructure

4.1 Introduction

The research reported in chapters 2 and 3 has identified the need among health researchers in Ireland for infrastructure to facilitate data access, storing, sharing and linkage. The lack of infrastructure has inhibited and frustrated the conduct of important health research that could inform policy, planning, practice and treatments. The need for certain new pieces of infrastructure which would provide secure storage and access to data, and an associated research support unit to support researchers, was identified. In addition, a need to augment or strengthen certain services and facilities, including third party and safe haven services was noted. The data project has established that these pieces of infrastructure would serve not only the needs of the health research community but also the needs of players in the wider data ecosystem in Ireland. Interviews conducted with personnel from government, enterprise, academia and the broader research area revealed that they require access and linkage facilities for their own purposes. The proposed

infrastructure is seen as a necessary addition to the national data infrastructure and also to the Irish statistical system described in O'Hara 2013.¹¹⁰

In this chapter, the requirements identified are configured into a model which allows for the safe and efficient use of data for research purposes – a data access, storing, sharing and linkage model, i.e., the DASSL model. Good governance, safe data outputs and public engagement are integral parts of this proposed model.

The previous chapters have shown that variations of this type of infrastructure have been developed in many countries, including the UK, France, Australia and Canada. In some cases, over 10 years of experience in implementation have been gained. Ireland is in a good position to benefit from this experience.

The DASSL model has been developed on foot of learning from the experiences of those who implemented the solutions developed in Northern Ireland (Northern Ireland Statistics and Research Agency (NISRA),¹¹¹ and the Administrative Data

¹¹⁰ O'Hara, P. (2013) 'The Irish Statistical System – Preserving Trust through Quality Standards'. Paper presented at SSISI Symposium Safeguarding Trust in Official Statistics. See <http://www.tara.tcd.ie/bitstream/handle/2262/72772/2a%20O%27hara.pdf?sequence=1&isAllowed=y>

¹¹¹ Northern Ireland Statistics and Research Agency. See www.nisra.gov.uk

Research Centre (ADRC)-NI/ADR Network,¹¹² Health and Social Care Business Services Organisation (HSC BSO¹¹³) and in France (Centre d'Accès Sécurisé Distant aux Données (CASD), the Group of National Economics and Statistics Schools (GENES), National Institute for Statistics and Economic Studies, France (INSEE)). Work carried out in Scotland (Scottish Health Informatics Programme (SHIP); Farr Institute) and Wales (SAIL) also influenced the model, as did ongoing conversations with colleagues in the CSO. It is interesting to note that the infrastructure based in NISRA has developed on foot of data access and linkage requirements emerging from researchers working on the Northern Ireland Longitudinal Study (NILS), inter alia, including their need to link research data to census and administrative datasets. In addition, the experience of providing linkage and honest broker services and setting up a 'safe haven' specifically for health-related data in Northern Ireland (Health and Social Care Business Services Organisation in collaboration with NISRA) informed the DASSL model. The proposals presented below were further informed by the international literature and insights from interviews carried out as part of the HRB Data Project with senior staff in government departments, agencies and academia. The operationalisation of the model, as well as the most appropriate structures and processes etc., will require debate and input from many stakeholders; in addition, the most appropriate technical solutions will change as knowledge advances.

4.2 The DASSL Model and the Research Data Trust

It is proposed that an entity – the Research Data Trust (RDT) – would be established (either on a stand alone basis or as part of an existing facility/facilities). The role of this entity would be to provide the institutional and technical environment to respond in a concerted manner to the growing needs for data access, storage, sharing and linkage within the Irish research and broader data ecosystem. The operationalisation of the DASSL model would take place within this context. The RDT would be a valued addition to the Irish statistical system.

A phased development of the DASSL model is likely to be required, and it is suggested (given extant work on the governance issues which apply) that the health area should be prioritised. Application of the model to the requirements of the broader data ecosystem would be likely to be implemented relatively smoothly.

The DASSL model comprises seven main elements, designed to facilitate the conduct of research, as follows:

1. governance
2. health research data hub
3. third party and data linkage service
4. safe setting/safe haven
5. research support unit (data access, sharing and linkage)
6. output checking and disclosure control
7. public engagement and communications

The model responds to the five safety principles identified by ADNR:¹¹⁴

- safe projects (valid research purpose)
- safe people (trusted researchers)
- safe data (protection of data)
- safe setting (security controls)
- safe outputs (disclosure control of outputs)

The main elements of the DASSL model are described below. This is followed by a high-level walk-through of the model in use by a researcher. The presentation aims to stimulate discussion and debate regarding options going forward.

Governance

Since health research typically involves sensitive data and is subject to data protection and related legislation, ensuring sound governance is critical to the operation of the DASSL model. HIQA asserts that 'good information governance allows organisations and individuals to ensure that personal information, such as that contained in a healthcare record, is handled legally, securely, efficiently and effectively...'.¹¹⁵ Good governance requires compliance with legislation and proportionality when weighing up public benefits and the protection of privacy and other interests of citizens. Without good governance, securing the confidence of professionals, as well as securing

¹¹² ADNR see www.adnr.ac.uk

¹¹³ <http://www.nidirect.gov.uk/health-and-social-care-business-services-organisation>

¹¹⁴ Slides presented at NILS 2011 Census Launch event. NISRA, Belfast June 2014.

¹¹⁵ <http://www.hiqa.ie/healthcare/health-information/information-governance>

public confidence and trust, cannot be expected. Given that the model is designed to address the rigorous governance requirements necessary for handling sensitive health data, it should certainly be robust enough to respond to the needs of the broader data ecosystem.

A principled, proportionate governance approach

Following the excellent work carried out in Scotland (e.g. SHIP, Scottish Government,¹¹⁶ Laurie and Sethi's work at the Mason Institute¹¹⁷ etc.), it is suggested that optimal governance can be achieved through adopting a principled, proportionate, risk-based approach to governance. This approach uses properly trained and suitable appointed authorising entities who adjudicate on data sharing and linkage proposals in the first instance with reference to regulatory requirements which place much emphasis on consent and anonymity. No legal framework can address each and every circumstance thrown up by complex research studies in a changing social and technical environment, and it is foolhardy to hold such expectations. Thus, importantly, the principled, proportionate governance (PPG) approach deals with the 'spaces in between legal provisions' (Sethi and Laurie 2013¹¹⁸) where judgement calls are required in order to address ambiguities and grey areas which research has shown has resulted in a 'culture of caution' and lassitude in sharing. In navigating the grey areas, the Scottish (and other OECD countries) approach is guided by key principles (e.g. privacy, consent, anonymisation, public interest), and by robust and transparent policies, processes and procedures for holistic risk assessment which are informed by stakeholder engagement.¹¹⁹

This principled, proportionate approach to achieving balance between the protection of privacy and individual interests and public benefit in the conduct of research is receiving growing acceptance internationally. It allows for an adaptable and flexible governance mechanism which can deal responsibly and proportionately

with the complexities that modern health research poses, and can inform the interpretations required by the current legislative environment or, as Sethi and Laurie (2013, p.178)¹²⁰ describe it, to assist 'data controllers and decision makers who (are required to) operate in the regulatory spaces in between the legal architecture'.

Sethi and Laurie (2013) provide a 'template for optimal governance' which identifies the pertinent issues including risks, options and opportunities that need to be weighed up in producing good governance decision-making. Experience gained in applying the template led to the construction of a principled, proportionate governance model with the following four key elements:

1. guiding principles and best practice
2. safe, effective and proportionate governance
3. an articulation of the roles and responsibilities of data controllers and data processors and
4. the development of a researcher (and stakeholder) training programme, including vetting procedures prior to data sharing.

The English Care Act 2014 reflects a similar approach by requiring the Health Research Authority to publish guidance on '...(a) principles of good practice in the management and conduct of health and social care research; [and] (b) requirements, whether imposed by enactments of otherwise, to which persons conducting health and social care research are subject' Laurie *et al.* (2015).^{121, 122}

The PPG approach is in harmony with many of the conclusions emerging from the OECD (2015) report *Health Data Governance: Privacy, Monitoring and Research*¹²³ which reviewed data governance and privacy-protective mechanisms across 22 states. Among the key mechanisms identified to enhance privacy-protective data use are: fair and transparent project approval processes, certification/accreditation processes,

116 <http://www.gov.scot/Topics/Statistics/datalinkageframework> and <http://www.gov.scot/Topics/Statistics/datalinkageframework/GuidingPrinciples>

117 The Mason Institute (School of Law, University of Edinburgh) is an interdisciplinary network aimed at investigating the ethical, legal, social and political issues at the interface between medicine, life sciences and the law. <http://masoninstitute.org/>. G. Laurie is a professor of medical jurisprudence and N. Sethi is a doctoral scholar at the Institute.

118 Sethi, N. and Laurie, G. (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

119 See Chapter 2 for discussion of principles.

120 Sethi, N. and Laurie, G. (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

121 <http://www.sciencedirect.com/science/article/pii/S1386505615300356>

122 The Health Research Authority in England has been given power by the Care Act 2014 'to authorise the processing of confidential medical information for medical research, subject to approval by an ethics committee (section 117), and requires the HRA to put '...in place a system for reviewing decisions' – see Laurie *et al.* (2015) at <http://www.sciencedirect.com/science/article/pii/S1386505615300356>

123 <http://www.oecd.org/publications/health-data-governance-9789264244566-en.htm>

best practice in data de-identification, data security and data management. They advocate that 'governance mechanisms are periodically reviewed at an international level to maximise societal benefit and minimise societal risks as new data sources and new technologies are introduced', OECD 2015.

The suggestions regarding structures, processes and procedures put forward in this chapter towards the proposed implementation of the DASSL model have been informed by these international developments. Planning and implementation of the model will require debate among all the relevant stakeholders in the Irish context e.g. data custodians, regulators, standards body, researchers, and the public. Debate regarding principles, decision-making frameworks, authorisation, certification, and training, *inter alia*, will need to take place. Here we can be guided by several of the examples of good practice referred to throughout this document and the forthcoming Health Information and Patient Safety Bill, *inter alia*.

Thus, in the Irish context, in addition to scientific reviews of health research proposals, it is proposed that research involving access to, and linkage of, sensitive health data should be subject to adjudication which, as well as making deliberations regarding conformance to regulatory requirements would weigh up public benefits and private interests in sanctioning particular research projects or not, while being guided by a principled proportionate approach to governance.

Helpfully, the operation of principled proportionate governance can be guided by work carried out by SHIP/Farr Scotland (involving NHS Scotland and other stakeholders). In this context, Sethi and Laurie (2013)¹²⁴ provide a means of categorising research projects into one of four levels of risk following scrutiny with respect to benchmarks relating to safe data, people and environment and the conduct of a privacy risk assessment. Low-level risk projects typically would involve use of data in the public domain where outputs are non-disclosive and non-sensitive. High-risk projects would include those which do not satisfy criteria such as demonstrating public interest, adherence to safeguarding criteria relating

to safe environments (e.g. use of safe haven, where indicated), or which would involve very sensitive data or data with a high risk of disclosure.

Scotland has developed an authorising body called the Public Benefits and Privacy Panel for Health and Social Care (PBPP) to 'balance safeguarding privacy with fiduciary duty to make best use of (health and social care) data collected' (Murray 2015).¹²⁵ In the interest of proportionate governance and efficiency of the research review process, a two-tier scrutiny process has been put in place: operational Tier 1 assesses technical, security and legal aspects of requests to determine whether access can be approved, or whether scrutiny by Tier 2 is appropriate; Tier 2 considers contentious high-risk applications or proposed policies regarding health and social care data use. Murray 2015 provides an outline of the PBPP's role and remit, assessment criteria, procedures, committee members and their activities. McGrail (forthcoming)¹²⁶ describes a similar two-tier review and approvals process i.e. fast-track reviews and full review using principles and proportionate decision-making which is used by the British Columbia Linked Health Database (1996–2002) and is feeding in to the development of a new framework.

Information governance review panel

It is suggested that a discussion should take place regarding the value of establishing a suitably appointed *information governance review panel* in the Irish context, loosely modelled on the Scottish PBPP described above. The panel would focus on decision-making in relation to governance issues *specifically* related to access, sharing and linkage of data, and would comprise members with related technical, healthcare and governance expertise. The relationship of such a panel with the Office of the Data Protection Commissioner¹²⁷ and the research ethics committee structures¹²⁸ in the light of the Health Information and Patient Safety Bill would need to be the subject of discussion.

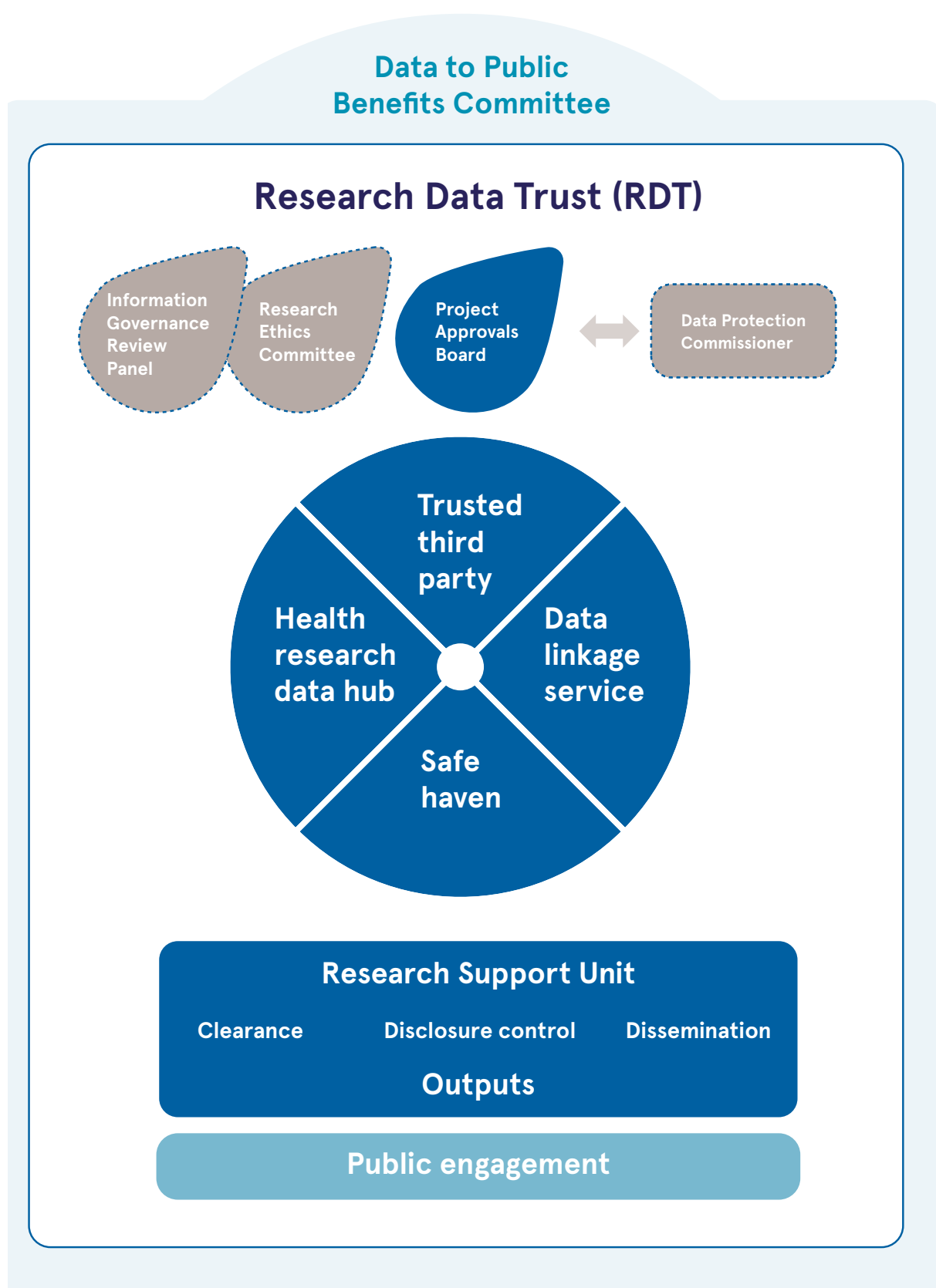
¹²⁴ Sethi and Laurie (2013). *Delivering proportionate governance in the era of eHealth: making linkage and privacy work together*. <http://mli.sagepub.com/content/13/2-3/168.short?rss=1&ssource=mfr>

¹²⁵ Dr Jane Murray 2015. From slides for paper entitled 'Public benefits and privacy panel – Proportionate governance of health data' presented at the FARR conference, Scotland 2015

¹²⁶ McGrail, K.M. *et al.* forthcoming. 'Building on principles: The case for comprehensive, proportionate governance of data access'. Chapter to appear in <http://www.springer.com/us/book/9783319236322#aboutAuthors>

¹²⁷ The panel might act in a consultative capacity (see 33.5 of Revised General Scheme, Health Information and Patient Safety Bill).

¹²⁸ The panel might act as a sub-group of a research ethics committee (see 25.1 of Revised General Scheme, Health Information and Patient Safety Bill).

Figure 2: The DASSL model

The key elements of the DASSL model are outlined here. Their configuration and operationalisation will require further discussion and agreement.

This authorising entity would adopt a principled and proportionate governance approach to responsible decision-making regarding what data can be accessed, shared or linked in accordance with legal requirements and good governance principles, which include deliberations regarding anonymisation, consent, privacy, public interest etc. (see Sethi and Laurie, 2013, p.188). A two-tier structure which would address different levels of risk along the lines outlined above is suggested, in order to allow for proportionality and efficiency in the review process. This format has the advantage of weighing up each research project on its merits and moves away from a one-size-fits-all approach. Procedures to ensure formal accountability for decisions made would have to be in place.

Discussions need to take place involving all the relevant stakeholders including, in particular, the research ethics community and the ODPC regarding the most appropriate and proportionate approach to achieving good data governance. Such discussions were held in Scotland, and led to a decision whereby an agreed Memorandum of Understanding frames the work of the research ethics committees and the PBPP. The present project noted in the literature that the workings of such panels are audited, and adjustments are made to structure, functions and relationships with other aspects of the governance framework as needed. Whatever solution is put in place, it has to work for all the different stakeholders: legal, administrative, public and researchers.

Research ethics committee

While the information governance review panel authorisation above focuses on good information governance relating to data access, sharing and linkage, health research projects will also require approval by a recognised research ethics committee (REC). Typically, RECs work to maintain ethical standards of practice in research, and to protect the rights and interests of human subjects. Structures and processes relating to the operation of RECs in Ireland are being reviewed, and the forthcoming Health Information and Patient Safety Bill will provide direction on the workings of RECs and related sub-committees.

Project approvals board

Typically, within an entity such as the proposed research data trust (RDT), a project approvals board would consider operational issues including the feasibility, practicality and resource implications of proposals involving data sharing and linkage, and would assess and adjudicate on compliance with

standing operating procedures.¹²⁹ Such a board could provide advice on how projects might be improved, or they could stipulate changes to be made. Typically, approval from this type of board is conditional on receipt of research ethics committee approval.

The following sections outline the core elements of the DASSL model as shown in Figure 2. They represent the core data services provided by the RDT: health research data hub, trusted third party service, data linkage and safe haven. This is followed by a description of the research support unit and its proposed disclosure control and public engagement functions.

Health research data hub and use of trusted third party service

The preceding chapter identified the need for an environment that can accept health and related research data, and make it available for use in a safe and secure manner by other researchers. We saw that the ISSDA¹³⁰ provides this function for social science and related data which it receives in an anonymised form. The HRB Data Project has identified the need to provide safe and secure access to health research data, and to routinely collect health data such as hospital admissions and discharges. Millions of euro are spent annually collecting routine health data in the course of treatment provision and monitoring. Typically, these data are hugely under-exploited. If they were more widely available, they could provide valuable insights for use in policy-making and practice.

The custodians of such routinely collected health information, together with the researchers and administrators interviewed as part of the HRB Data Project, while agreeing that there should be greater access to such data, were of the view that such data, due to its sensitivity, required privacy protecting, anonymisation, storage and access regimes that were not only rigorous but were also widely perceived to be so. A number of data trusts¹³¹ have been developed internationally to exploit the economic and health potential of electronic patient records and research data in a manner that fosters safe data use. Ireland is in a fortunate position to be able to examine the different technical and

¹²⁹ See RSU and Research Advisory Group – RAG in Northern Ireland <http://www.qub.ac.uk/research-centres/NILSResearchSupportUnit/GuidesResources/Access>

¹³⁰ ISSDA receives anonymised data, which it then checks for anonymity.

¹³¹ John Dunne of the CSO 2014 in 'Briefing note: national data trust' personal communication, has outlined a similar concept for national statistics.

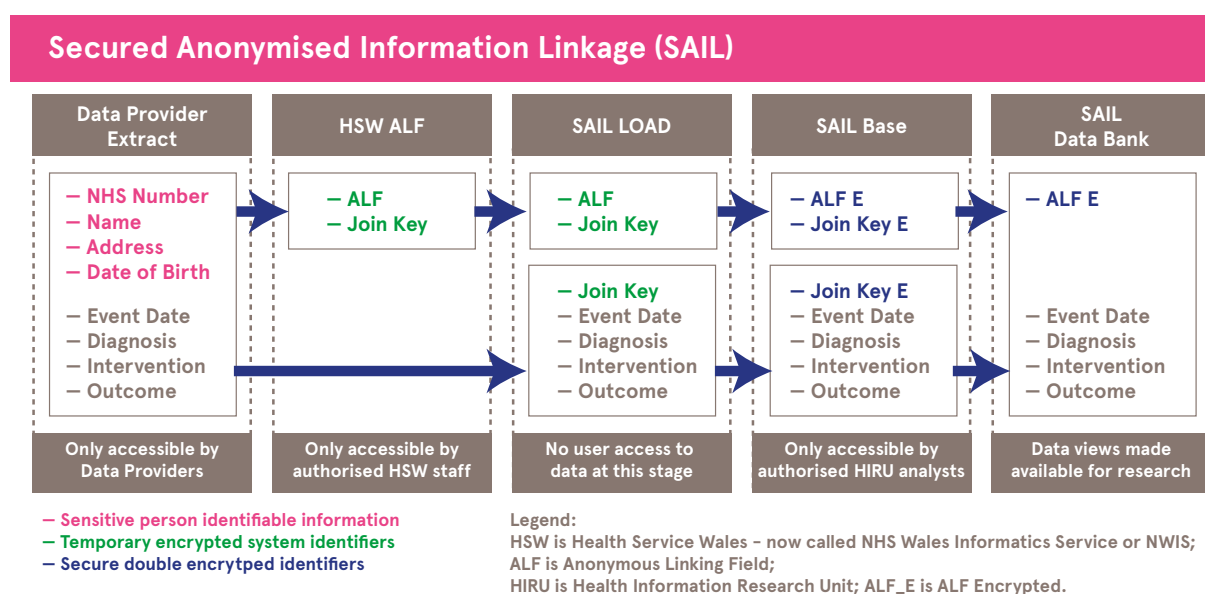
institutional solutions implemented internationally, and to learn from them in constructing a culturally appropriate solution.

In the Irish context, it is proposed that an RDT should be set up (either within an existing or newly established entity); this RDT or its health research data hub would liaise with data custodians to gather (either anonymised or identifiable data) and make such data available to researchers and other data users for safe use. In the case of identifiable data, the health research data hub would use the services of a trusted third party to index or anonymise data and perform disclosure control, and thus would make data available for safe use by researchers^{132,133} using either federated (project-by-project access, data are not held) or centralised (warehousing) type data solutions.¹³⁴ An example of a health research data hub is presented in Figure 3, while the technical details are represented in Box 4 overleaf.

The SAIL¹³⁵ research centre in Swansea, Wales, has a long track record of information system

research using electronic health records: 'It hosts the Welsh SAIL Databank – a large-scale warehouse of person-centred, linked, anonymised data, and it supports multiple observational and interventional research at individual and household level, including linkage to major UK cohort studies. The National Institute for Social Care and Health Research (NISCHR) is part of SAIL and focuses on routinely collected health data to conduct and support high-quality research.' The SAIL databank holds over 500 million anonymised and encrypted individual-level records from a range of sources relevant to health and well-being.¹³⁶ These include data from Welsh national screening programmes and the cancer registry, registers of births and deaths, national community child health services, emergency services, primary care general practices, secondary care hospitals (inpatients, outpatients, day cases, A&E attendances, pathology results) and social care. These datasets are continually being added to.

Figure 3: Secure anonymised data access through SAIL (Ford et al. 2009)^{137,138}



132 Third party indexing can result in the hub never having access to, or sight of, personal identifying data; nor would the researchers who would make use of the data held by the hub – see ADNRN, p.6 <http://www.esrc.ac.uk/files/news-events-and-publications/publications/themed-publications/improving-access-for-research-and-policy/>

133 This is usually accompanied by a requirement for analysis of the protected data by researchers to take place in a secure environment (see safe haven below).

134 See Chapter 2; 2.5 for more detailed discussion.

135 Jones KH, Ford DV, Jones C, Dsilva, R, Thompson, S, Brooks, CJ, Lyons, RA (2014) A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation *Journal of Biomedical Informatics* 50(100), 196–204. doi:10.1016/j.jbi.2014.01.003

136 See Ford et al. 2009. *The SAIL Databank: building a national architecture for e-health research and evaluation*. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744675/>

137 Diagram from Ford DV, Jones KH, Verplanck J-P, Lyons RA, John G, Brown G, Brooks CJ, Thompson S, Bodger O, Couch T and Leake K (2009). The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res*. 2009; 9: 157. Published online 2009 September 4. doi: 10.1186/1472-6963-9-157 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744675/>

138 See the following for more technical details and governance details, as well as examples of other similar data protection techniques, in Jones KH, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ, Heaven ML, Thayer DS, McNerney CL and Lyons RA (2014). A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation <http://www.sciencedirect.com/science/article/pii/S1532046414000045>

BOX 4

The SAIL technique for providing safe access to data¹³⁹

The SAIL technique involves separation of the identifying information from descriptive information. Thus, the commonly recognised identifiers (designated File 1 (for person-level data) and File 1R (for address-level data)) are separated from the clinical or event-based descriptive data (designated File 2), as part of standard SAIL methods.

The files holding the identifiable data (File 1/1R) are sent to the Trusted Third Party for SAIL (i.e. NHS Wales Informatics Service – NWIS, formally known as HSW), which carries out matching and anonymisation, whereby the commonly recognised identifiers are replaced with an Anonymous Linking Field (ALF) assigned to each person represented in the File 1 dataset (thereby creating File 3), or a Residential Anonymous Linking Field (RALF) assigned to each address in the File 1R dataset (similarly creating File 3R). File 3/3R with the ALF are then sent to SAIL for recombination with File 2, which holds the descriptive data which are sent directly to SAIL.

The de-identified data are then accessible by researchers in accordance with detailed privacy-protecting policies and procedures.

Figure 3 above provides a diagrammatic representation of the secure anonymisation service, SAIL.

The establishment and use of the health research data hub will require agreements between all the relevant stakeholders, such as data providers, regulators etc. In addition, agreements will need to be put in place with regard to protocols and procedures for safe access, use, disclosure control, checking, retention, destruction etc. – all informed by good practice internationally. Such agreements would involve inputs from the Data Protection Commissioner's Office, HIQA, HSE, HRB and other relevant stakeholders. The forthcoming Health Information and Patient Safety Bill would also inform such initiatives.

An up-and-running health research data hub, which would gather (and could store) and make available safe health and related administrative and other data for research and evidence purposes, would make a major contribution to the health research community and the broader data ecosystem. It would facilitate the conduct of research that can inform health and well-being, health service delivery and the effective management of services.

Trusted third party data linkage service

Chapter 2 showed that linking different datasets provides researchers with inexpensive access to larger datasets and widens the variable range for particular projects, thus allowing for powerful hypothesis testing. Many developments in modern medicine have relied on linked datasets. Chapter 2 also showed that safe linkage is facilitated by the services of a trusted third party, since researchers cannot directly access personal information collected by another researcher.

There are a number of recognised techniques for carrying out data linkage for datasets holding personal information.^{140,141} Many, in the first instance, use a trusted third party technique for data indexing similar to that described above for SAIL; this is then followed by the use of a separate linkage service. The Administrative Data Research

¹⁴⁰ See Jones KH, Ford DV, Jones C, Dsilva, R, Thompson, S, Brooks, CJ, Lyons, RA (2014) A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation *Journal of Biomedical Informatics* 50(100), 196–204. doi:10.1016/j.jbi.2014.01.003; Note that the Health Intelligence Unit, Health and Wellbeing Directorate, HSE is working on a similar concept to respond to information demands and auditing requirements within the HSE. Personal communication from Howard Johnson, 29/04/15 'Health Research Hub'.

¹⁴¹ <http://www.esrc.ac.uk/files/news-events-and-publications/publications/themed-publications/improving-access-for-research-and-policy/> i.e. ADRN, UK, Appendix 2

¹³⁹ Text adapted from Secure Anonymised Information Linkage (SAIL) – Data Management Policy (2013), p.5. <http://www.saildatabank.com/media/20689/SAIL%20data%20management%20policy%20v1%200.pdf>

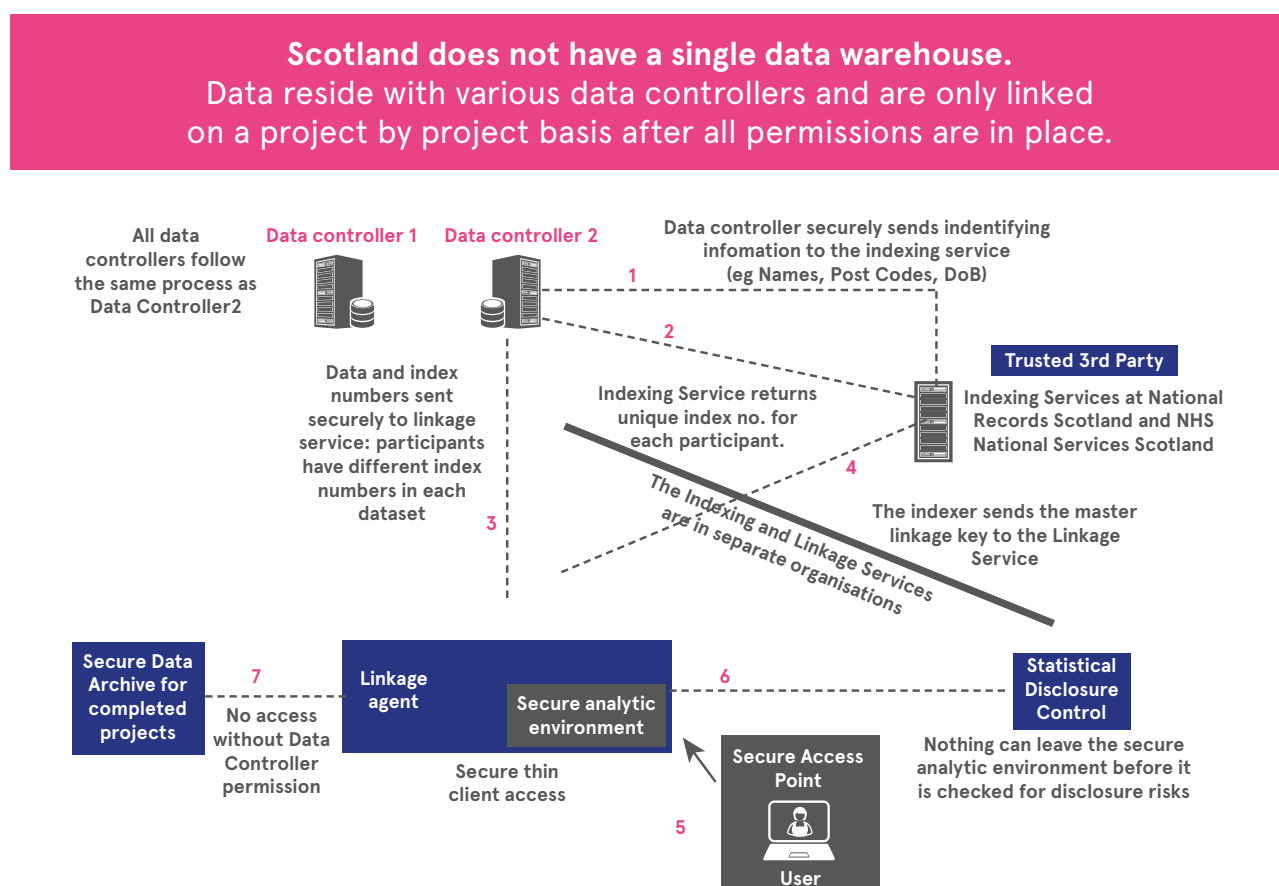
Network (ADRN) has produced a short video which provides a succinct account of the privacy-protecting approach used by them.¹⁴² The ADRN approach has been implemented in relation to administrative data in a number of locations in the UK, including Northern Ireland. Similar approaches have been developed in relation to health data – see, for example, the approach adopted by SHIP, which is shown diagrammatically in Figure 4.

Safe linkage is achieved by using advanced anonymisation and encryption to minimise re-identification of individuals; in many cases the linkage service is separated from the indexing service, in order to augment privacy and confidentiality. Furthermore, the data linkage is carried out in a secure environment to which the researcher does not have access.

To date, the CSO has undertaken data linkage for projects subject to available resources. However, the service is not resourced to respond to the broader needs of health researchers where these needs go beyond the realm of official statistics.¹⁴³ Thus a data linkage solution needs to be put in place that is capable of responding to researchers' needs in a timely and comprehensive manner. It is proposed that this data linkage service would be provided through the RDT. Discussion needs to take place regarding the most appropriate techniques to be used for particular projects in the Irish context.

It should be noted that many countries have identified data linkage services as an essential element of their statistical infrastructure; moreover, the International Population Health Data Linkage Network supports such

Figure 4: Trusted third party data linkage service, Scotland (Pavis 2015)¹⁴⁴



¹⁴³ For the CSO to provide such a services legislative change, is likely to be required, inter alia.

¹⁴⁴ Diagram taken/adapted from slides presented by Dr Stephen Pavis, NHS Scotland, from a presentation entitled 'Scotland Health Data and its Governance' at the FARR International Conference 2015: Data Intensive Health Research and Care, St Andrews, Scotland. see <http://www.abdn.ac.uk/events/6475/>

¹⁴² ADRN video link https://www.youtube.com/watch?v=nnxz3_XGMAE

endeavours.¹⁴⁵ In the UK, the Farr Institute for Health Informatics Research, which comprises four centres focused on electronic health records research, aims to position the UK at the forefront of research using linked electronic health records, together with further linking of such data to other forms of research and routinely collected data. The centres 'will support innovation in the public sector and industry, leading to advances in preventive medicine, improvements in healthcare delivery, and better development of commercial drugs and diagnostics'.¹⁴⁶

Safe setting/safe haven

The next piece of infrastructure identified as needed by the health research community – and more widely identified as needed to advance the national economic and social agenda – is a safe setting or safe haven. In essence, this is an environment designed to ensure safe use of sensitive data.

Typically, safe havens provide a locked-down environment with highly controlled and restricted access, including supervised use and sealed work folders with no Internet connection. Internet, USB ports and CD drives are disabled, as are the printing and the taking of screenshots, and any paper used for note-taking is handed in on exiting the safe haven premises. Data cannot be downloaded or transferred. For information on 15 UK safe havens, see <https://adrn.ac.uk/protecting-privacy/secure-environment/safe-centres>

There are three main types of safe havens: (i) an on-site safe haven in an entity such as the proposed RDT; (ii) an external, secure research laboratory or pod placed in a university or similar environment;¹⁴⁷ (iii) the provision of other remote access solutions. The external and remote access solutions need to be privacy-engineered to the same high standard as the on-site solution. In some jurisdictions remote access is only allowed to less sensitive datasets.

Various technical solutions are possible to create a locked-down and leak-proof environment; one such example is the Bubble and Box in France.¹⁴⁸ Organisations in a number of countries are developing solutions for secure remote access, as

is the CSO in Ireland.¹⁴⁹ In the UK, the intention has been to spend part of the £64 million investment by the Economic and Social Research Council (ESRC) in four new administrative research centres (ADRCs) on developing SafePods, which are secure rooms allowing for secure remote access to sensitive data held by the ADRCs (and possibly also other sensitive data that need to be accessed from a secure setting). The ADRCs have been designed to strengthen the UK's competitive advantage in the era of big data government; in addition, the Farr Institute of Health Informatics Research in the UK,¹⁵⁰ funded by a consortium of research bodies including the Medical Research Council, ESRC, National Institute for Health Research and Wellcome Trust, have been developed with the aim of establishing a coordinated approach to safe havens. In the UK, following the Caldicott Review 2013,¹⁵¹ accredited safe havens (ASHs) were established. ASHs are accredited organisations, or a designated part of an organisation, which are contractually and legally bound to process data in ways that prevent the identity of individuals from being disclosed.¹⁵² The HSCIC (Health and Social Care Information Centre), which is the UK national provider of information, data and IT systems for health and social care, is itself an accredited safe haven, but also is expected to have a role in auditing UK safe havens.

It is proposed that a safe haven is developed within the RDT, providing a safe environment for data access, linkage and analysis which will be able to respond to a broad range of research and public service demands. As already noted, the CSO operates an embryonic safe haven. However, as already noted the facility is not resourced to respond to the broader needs of health researchers where these needs go beyond the realm of official statistics.

The anticipated expansion of demand from researchers as a consequence of developing the proposed data hub, coupled with growth in secondary data analysis, *inter alia*, will require investment in this critical piece of the national statistical infrastructure. It is expected that the facility will be used by a wide range of actors in the

¹⁴⁵ <http://www.ipdln.org/>

¹⁴⁶ <http://www.farrinstitute.org>

¹⁴⁷ See <http://www.ucl.ac.uk/jdi/data-lab> for research laboratory at UCL

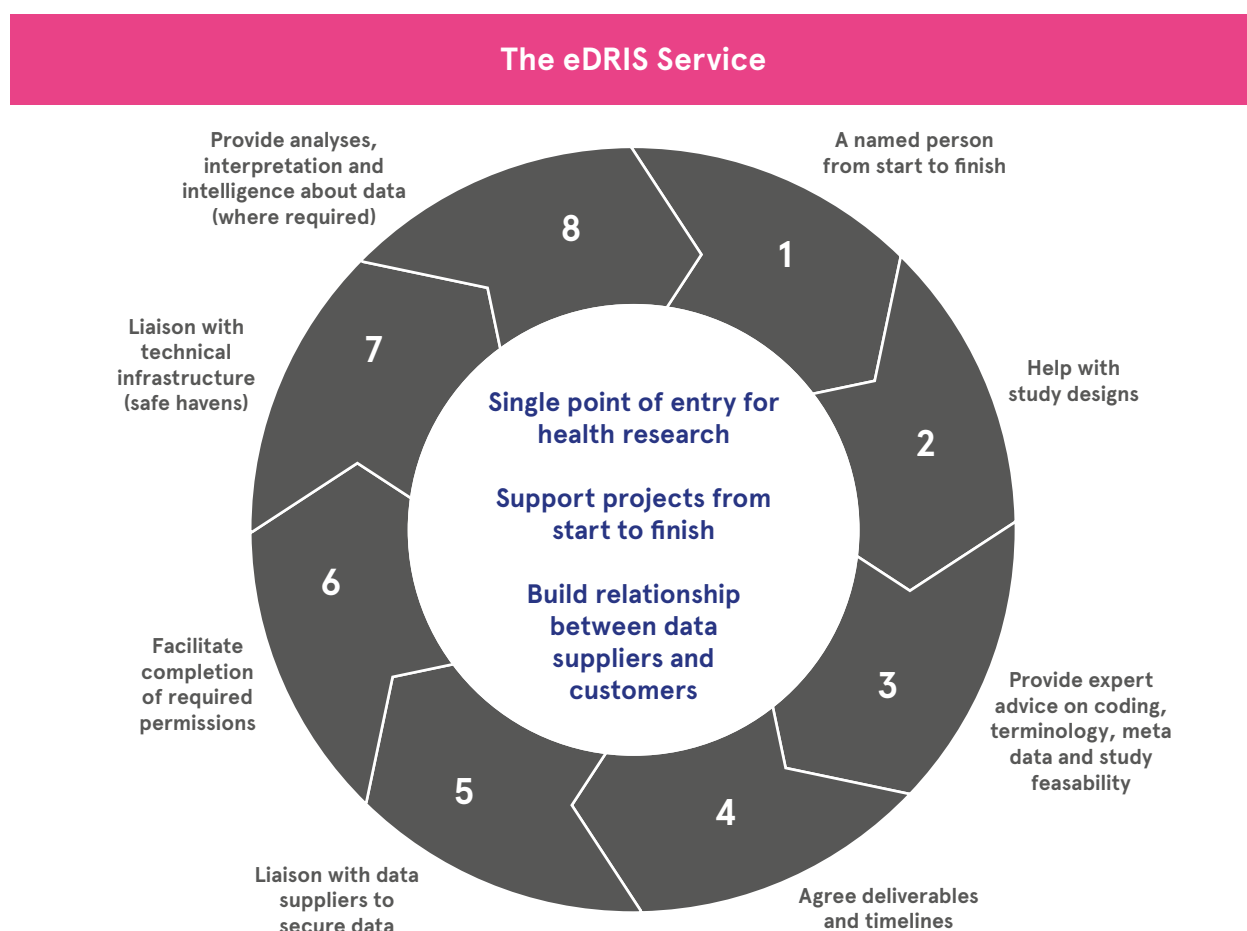
¹⁴⁸ http://doku.iab.de/fdz/events/2012/WDA/presentations/4thWDA_CASD.pdf

¹⁴⁹ The CSO is affiliated with The EU framework Data without Boundaries Project (DwB) – DwB is establishing a network of remote access centres.

¹⁵⁰ The Farr Institute was established in March 2013. See <http://www.farrinstitute.org/>

¹⁵¹ <https://www.gov.uk/government/publications/the-information-governance-review>

¹⁵² <http://www.hscic.gov.uk/article/3638/Personal-data-access-FAQs#ASH-questions>

Figure 5: The eDRIS service (Pavis 2015)¹⁵³

wider data ecosystem. One might expect growth in demand for the facility similar to the growth in demand for ISSDA data services, as shown in the graph in Appendix 7.

Research support unit (data access, sharing and linkage)

Both the facilities visited (i.e. Northern Ireland and France) operate a research support service (RSU) – as do the majority of similar data services reviewed here and in the British Columbian report on ‘data innovation’.¹⁵⁴ Scotland’s Information Security Division’s eDRIS or Electronic Data Research and Innovation Service ‘provides a single point of contact to help researchers in study design, approvals and data access in a secure environment related to patient and non-patient identifiable

data. eDRIS aims to make conducting research easier, more efficient and more convenient. eDRIS encompasses Information Security Division’s Record Linkage services’.¹⁵⁵ (See Figure 5 above, which summarises the services offered.)

A research support unit (RSU) is a vital part of the proposed infrastructure for Ireland also. The RSU would support researchers and data custodians in the processes and procedures required in relation to data access and linkage, so as to ensure maximum and safe use of the infrastructure and services provided in the proposed research data trust. The RSU would provide assistance to researchers and data providers throughout the research process, thus maximising efficiencies and minimising delays.

The RSU would manage and control users and access to data (input and output) and would ensure that each step in the governance process, as set

¹⁵³ Diagram taken from slides presented by Dr Stephen Pavis, NHS Scotland from a presentation entitled ‘Scotland Health Data and its Governance’ at the FARR International Conference 2015: Data Intensive Health Research and Care, August 2015 St Andrews, Scotland

¹⁵⁴ http://www.gov.bc.ca/citz/download/BC_Centre_for_Data_Innovation-FINAL.pdf. Customer functions included – access requests and approvals, data analysis methods, data navigation, collaboration forums, onboarding services, knowledge translation, Q&A services and training.

¹⁵⁵ eDRIS is based at the Farr Institute in Edinburgh, Scotland and contributes to the ScottishH Informatics Programme (SHIP).

out in the various protocols, is adhered to. The RSU would be involved in all elements of the DASSL model. The specific functions of the RSU would be to:

- give support to data providers to provide quality, well-documented data
- put data-sharing agreements in place between data providers and the RSU
- provide guidance to researchers regarding the feasibility of their research projects
- provide advice and comment prior to submission to the research ethics committee and other governance boards
- manage and deliver safe researcher training and researcher certification as an ‘approved researcher’¹⁵⁶
- put licence agreement in place between researchers and the RSU
- oversee Garda vetting of researchers (if required)
- provide access to the safe haven
- supervise use of the safe haven
- facilitate researchers to access study data and undertake analysis
- create project-specific data extracts for researchers
- manage data disclosure process and output controls
- close off projects
- maintain web pages for DASSL elements
- manage public engagement, communications and dissemination.

Detailed protocols and procedures in relation to the operation of an RSU are provided on the NISRA website www.nisra.gov.uk

Software to manage the entire workstream has been developed. It should be noted that an RSU would not provide statistical advice. In addition to data quality assurance, management and administration, the skill sets of people working in the RSU would include statistics, with particular expertise in privacy protection, security and disclosure control. Technical staff would also require high-level database and IT expertise. It is anticipated that there would be synergies between

the RSU, the ISSDA, Health Atlas Ireland and the Centre for Support and Training in Analysis and Research (CSTAR).¹⁵⁷

Output checking and disclosure control

The different pieces of infrastructure described above would be involved in the output of data, which would be released through the RSU. These data, as required, would have to be thoroughly checked by highly trained statisticians with expertise in disclosure control, in order to ensure that individuals or entities cannot be identified. Checks carried out would, for instance, ensure that cell counts in tables have a minimum frequency of 10.

Guidelines for different types of statistical output are available in the literature. For example, the Network of Excellence in the European Statistical System in the field of Statistical Disclosure Control (ESSNet SDC) provides detailed guidance, along with recommendations and best practice, on the organisational and procedural aspects of output checking.¹⁵⁸ Typically, checking is very time-consuming. Turnaround times are usually specified, and in some countries there is a charge for the checking service. When checking has been completed, an RSU team member can transfer the checked data files to those authorised to receive outputs, i.e. licence agreement holders.

Data providers may stipulate that final outputs such as working papers, publications and presentations are checked by themselves, so as to ensure that the data are correctly described, and that the approved acknowledgement has been used. Obviously, this additional step will increase the time input. Researchers may also be required to notify the RSU when a presentation/publication has been delivered or published. In this way, a listing of all such presentations and publications could be maintained on the RSU’s website, which would contribute to transparency and would also serve as an information resource for researchers and policy-makers.

A disclosure control policy typically guides practice in relation to the release of outputs. Practice varies with regard to the destruction/conservation of data. In the Irish context, as outlined above, it is recommended that data are stored safely and preserved into the future.

¹⁵⁶ Accreditation as an Approved Researcher: as of 1 July 2014, NILS/RSU of NISRA requires new and existing researchers to attain Approved Researcher status by completing the Application for Accreditation as an Approved Researcher form. The application form asks for details of experience and professionalism, and will be used by senior NISRA staff to approve, or not, the researcher as a ‘fit and proper’ person.

¹⁵⁷ CSTAR is the acronym for Centre for Support and Training in Analysis and Research. See <http://www.cstar.ie/>

¹⁵⁸ http://neon.vb.cbs.nl/casc/ESSnet/guidelines_on_outputchecking.pdf

Public engagement and communication

Securing the trust of the general public is essential for conducting health research. People who provide their personal information for research purposes need to be confident that their data will be used in the public interest; that the data will be held securely, and that their privacy and confidentiality will be respected. There is ongoing public discourse regarding potential breaches of privacy, particularly in relation to social media, which has a corrosive effect on trust in institutions of all types to manage data securely. In order to help ensure public confidence, a high level of transparency in relation to activities is needed.

In recognition of the importance of public acceptance and legitimacy, programmes for public engagement have been developed across a wide range of science and policy initiatives. The major centres providing data access and linkage services internationally have been quick to establish a public engagement stream in their work; such centres include SAIL in Wales,¹⁵⁹ SHIP in Scotland,¹⁶⁰ the Western Australia Data Linkage Unit¹⁶¹ and the UK network of e-health record centres (Centre for the Improvement of Population Health through E-records Research – CIPHER). Also in the UK, the Administrative Data Research Network which reports to the ESRC has a workstream on public engagement in data linkage research.

Typical activities undertaken as part of public engagement include information provision, awareness raising and public education, involvement of the public in policy and projects through consumer or user panels, public representation on research ethics committees and other governance structures around data linkage and research, monitoring of public sentiment, and attitude research. The provision of a discussion forum on data centre and RSU websites, such as the DASSL, health research data hub and RSU described above, would allow people who wish to engage, and have an input to the work of such centres, to do so.

A public engagement policy needs to be developed as part of the DASSL model and could be guided by the principles for public engagement put forward by the Policy, Ethics and Life Science Research

Centre in the UK,¹⁶² and applied by SAIL in setting up its consumer panels (Jones *et al.* 2014).¹⁶³ A high level of transparency in activities carried out is essential.

4.3

A Walk-Through of the Proposed New Infrastructure

The following walk-through describes the steps a researcher might take using data from the health research data hub, linking data using the third party or broker service, and analysing data using the safe setting. It is largely based on the procedures and protocols operated by NISRA, which are outlined in detail on the agency's website.¹⁶⁴ Some adaptations have been made to suit the proposed implementation of the DASSL model and more are likely once all stakeholders have given detailed consideration to, and debated, the DASSL model and related governance policies. The order in which the different approvals are sought requires particular consideration and debate. Supporting documents, such as the application form and other forms that NISRA requires to be completed, are all available on the NISRA website. The SAIL data user journey is shown in Appendix 9.

- A researcher may first discuss their research proposal with personnel in the research support unit (RSU). Once quality metadata have been provided to the RSU, support staff will help confirm the suitability of the data for use in the proposed research. It may be necessary for the researcher to contact the data producer to confirm that the requested data match his or her needs.
- The researcher who wishes to access or link data then completes a project application form, providing details of the study aims, investigators, start/finish dates, planned publication and dissemination of findings. This form will also specify what data extracts are needed. Where linkage is involved, details of the methodology proposed and the legal considerations considered will be

¹⁵⁹ <http://www.saildatabank.com/governance>

¹⁶⁰ <http://www.scot-ship.ac.uk/c4.html>

¹⁶¹ <http://www.data-linkage-wa.org.au/>

¹⁶² <http://www.ncl.ac.uk/peals/research/>

¹⁶³ Jones, K, McNeerney C and Ford D (2014). Involving consumers in the work of a data linkage research unit. *International Journal of Consumer Studies* 38 [2014] 45-51. See also www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/7489.pdf

¹⁶⁴ <http://www.qub.ac.uk/research-centres/NILSRResearchSupportUnit/GuidesResources/Access>

outlined. The application form will also include a commitment in principle from the data controller that they will sign the data access/transfer agreement where required, should the project receive all approvals. A formal approval of the information governance aspects of the project by the *information governance review panel* (which might involve a noting of the project by the Data Protection Commissioner, without raising significant data protection issues), will be required.

- The project application form is considered by the *projects approvals board* that assesses and adjudicates on compliance with standard operating procedures (SOPs) and data management policies and resource requirements. Approval is granted by this board, but is conditional on receipt of approval by the research ethics committee.
- On receipt of provisional approval, the project is sent to a research ethics committee for ethical review.
- Following the approvals, a licence agreement is drawn up between the research data trust, and signed by all project team members and the chief investigator. An institutional signatory will be required; this signatory will have ultimate responsibility for the research team members.
- All researchers named on a project will be required to undertake safe researcher training, which typically lasts one day and covers legal and computer system aspects and statistical confidentiality. In addition, Garda clearance may be required.
- Next, the research data agency (most likely the RSU division) generates a project-specific dataset based on the details supplied in the application form. This dataset will only be made available to the project team members named on the application form who have signed the licence agreement. Agreed external data can also be linked at this point.
- All primary data analysis involving potentially identifiable data will be carried out within the safe haven. Researchers using the safe setting will be supervised at all times. In some circumstances, arrangements may be made to run analyses remotely.
- Outputs from these analyses will normally be released from the safe haven, once they have been checked for disclosure risk and cleared by the RSU.
- Closing-off of a project is marked by termination of the contract or agreement, and the status of the project is updated on the RSU database. Closure documentation will include information regarding storage and access and destruction if relevant.

4.4

Conclusions and Recommendations

Conclusions

The HRB Data Project has, with inputs from a wide range of stakeholders and with reference to international experience, identified the infrastructure and services required to create a robust data sharing environment in Ireland. The DASSL model developed here to address the requirements for data access, sharing and linkage experienced by Irish researchers needs to be incorporated into the Irish data infrastructure as part of the Irish statistical system. Implementation of DASSL will benefit not only the health research community but analysts in the wider data ecosystem, allowing for the exploitation of data to serve social and economic agendas.

Recommendations

- The proposals and practical examples presented here in relation to establishing the seven elements of the data access, storage, sharing and linkage model i.e. DASSL model, need to be debated by all relevant stakeholder groups (e.g. data custodians, researchers, professional bodies, regulators, general public etc.).
- The research carried out would suggest that all seven elements of the DASSL model should be implemented within a research data trust (RDT) environment in order to ensure maximum efficiency and quality and exploit synergies. Consideration needs to be given, however, to the urgent need to put in place trusted third party/data linkage and safe haven facilities within a robust, principled, proportionate governance framework.
- Public engagement, transparency and the development of trust is recommended as essential for the successful implementation of the DASSL model.
- A high-level 'Data to Benefits Committee' needs to be established to drive discussions towards the implementation of the DASSL model components.
- It would be desirable that the National Statistics Board (NSB), HSE, inter alia, review and engage with relevant stakeholders regarding the recommendations set out in the report and their implications for the data environment in Ireland.

5

Recommendations and Conclusions

5.1

Recommendations and Conclusions

The recommendations based on the research carried out are brought together here¹⁶⁵ broadly in chapter sequence. These are presented for discussion by the relevant stakeholders. A conclusions section follows.

Chapter 2: Health and related data – safeguarding provisions for safe access and use

1. A comprehensive infrastructure for data access and linkage needs to be established in Ireland. This would include trusted third party/honest broker, safe haven and research support facilities. Such infrastructure would facilitate not only health and well-being but also the economic and enterprise agendas.
2. Based on the research carried out, it is recommended that consideration be given to the development of a health research data hub, which would be engineered to meet stringent governance requirements for storage and access to health data, inter alia. It would provide safe access to agreed routinely collected health datasets deposited by the HSE, other health data custodians and by health registers.
3. A discussion needs to take place regarding the necessity and utility of the introduction of special legislation to underpin the infrastructure and services identified here as required in the Irish context. Lessons learned from international literature suggest that such legislation will not necessarily result in

good or better governance. Rather, what is needed is a robust authorising mechanism which can deliberate not only on conformity of research projects to existing legislation but also on research projects' conformity to guiding principles and best practice which help navigate the spaces in between legal requirements. Frameworks and templates have been developed to facilitate this process.

4. It is recommended that a principled, proportionate governance approach be adopted in order to guide decision-makers in navigating the ambiguities posed by existing legislative regimes and complex research questions. Good governance in the sharing and linkage of data is critical.
5. Healthcare professionals, researchers and related stakeholders should work with the Office of the Data Protection Commissioner (ODPC), with a view to cooperation in the preparation of guidelines to steer researchers in their practice and towards the safe use of data.
6. The ISSDA is a trusted repository which provides access to anonymised social science and related data. It is recommended that the ISSDA be established on a firm financial footing, so that it can continue to respond to emerging needs and developments.
7. The health research community should make a case for inclusion of datasets of value to research in the body of official statistics.

Chapter 3: Attitudes and practices – data access, sharing and linkage

Attitudes and culture

8. Professional associations and educators need to promote a culture of openness and good governance in relation to access to, and sharing of, research data.

¹⁶⁵ The recommendations are edited, reordered etc. in some cases.

9. When obtaining consent from study participants, the consent should take into account any immediate or future uses of the data, including archiving of data for secondary use. In addition, consent forms should promise to not destroy data unnecessarily.¹⁶⁶
10. In order to facilitate re-use of data, it is recommended that data custodians provide comprehensive metadata and user-friendly tools that would help to contextualise, visualise and analyse their data.
11. Researchers need to be more aware of requirements under the data protection legislation and existing guidelines from the ODPC.¹⁶⁷

Data management practices (DMPs)

12. HIQA and related guiding principles and practices for health data collection should be adopted to facilitate data re-use. In addition, common data standards should be adopted which would greatly support data access, sharing and linkage.
13. Development and upkeep of data management plans (DMPs) should be an essential part of the research process, where data collection represents a significant element of the project. Such plans should be externally monitored, so as to ensure compliance.
14. The HRB, perhaps in collaboration with other research funders, should consider the commissioning of work to provide guidance on the type of DMP that should be required of grant applicants. Alternatively, suitable guidelines from another agency or funder could be used, with any necessary amendments made to suit the Irish environment.
15. On foot of this work, funders should consider requiring researchers to include a DMP, where relevant, as part of a grant application. This DMP would include, inter alia, specification of where the data generated by the research would be repositied and how it would be set up and maintained for the purposes of sharing.
16. The costing of DMPs should be an allowable expense in relevant grant applications.
17. The HRB, along with other funders, should consider funding the development and roll-out of training courses on DMPs, which could be delivered as part of research methodology training programmes. An online format could be considered, for example, MANTRA¹⁶⁸. The delivery of training for data custodians is also recommended.

Data linkage

18. A balance needs to be struck between data protection considerations and the public good when decisions regarding data linkage are being taken. Moreover, a risk management approach needs to be adopted, with due cognisance taken of safeguarding measures employed by researchers to ensure safe data practices.
19. Guidance is needed from the ODPC and/or from professional bodies and other stakeholders in collaboration with the ODPC, in relation to data sharing and linkage.
20. Training for data sharing is needed for data custodians and researchers who wish to re-use and/or link data.
21. Capacity building around the statistical expertise required for linking and sharing needs to be developed.
22. The desirability and implications of cost recovery in relation to data access and data linkage services needs be to be examined.

Incentives for data sharing

23. Funders were regarded to have a lot of power to change research culture and behaviour, and this should be used to support better access to, and sharing of, data.
24. Funders should require researchers to deposit publicly funded research data (where copyright does not apply) into a repository for use by other researchers.
25. It was generally agreed that preparing data for sharing involves time and money, and that funders should allow sharing and related activities, such as the formulation of a DMP, as an allowable expense.
26. Since existing research culture does not recognise or reward data sharing or preparatory data management activities, it is recommended that a system of authorship credits for members of a research team involved in managing, analysing and curating data be developed. In addition, a proper career structure for data scientists and data management professionals needs to be developed.

¹⁶⁶ UK Data Archive <http://www.data-archive.ac.uk/create-manage/consent-ethics/consent>

¹⁶⁷ The latter are currently being reviewed by the ODPC (personal communication, October 2015).

¹⁶⁸ <http://datalib.edina.ac.uk/mantra/>

27. Academics should promote a requirement to consider data access and sharing in criteria used for academic selection and promotion.
28. Research funding for the analysis of secondary data should be promoted and increased.

Chapter 4: The DASSL model and the proposed research data trust

29. The proposals and practical examples presented here in relation to establishing the seven elements of the DASSL (data access, storage, sharing and linkage) model need to be debated by all relevant stakeholder groups. The relevant elements are as follows:
 - governance (the principled, proportionate governance approach is recommended)
 - health research data hub
 - trusted third party and data linkage service
 - safe setting/safe haven
 - research support unit for data access, sharing and linkage
 - output checking and disclosure control
 - public engagement and communications

These seven elements are designed to ensure safe projects, safe researchers, safe data, safe settings and safe outputs. These debates should aim to develop shared understandings towards consensus on the implementation of the model.

30. The research carried out would suggest that all seven elements of the DASSL model should be implemented within a research data trust (RDT) environment in order to ensure maximum efficiency and quality and exploit synergies. Consideration needs to be given, however, to the urgent need to put in place trusted third party/data linkage and safe haven facilities within a robust, principled, proportionate governance framework.
31. It would be desirable that the National Statistics Board (NSB), HSE and other relevant agencies review the recommendations set out in the report and their implications for the use, sharing and linkage of data in Ireland.
32. Public engagement, transparency and the development of trust should be a priority in the implementation of the DASSL model to ensure its successful adoption.
33. A high-level, possibly cross-governmental, 'Data to Benefits Committee' needs to be established to drive discussions towards the implementation of the DASSL model components.

5.2 Conclusions

The research carried out as part of the HRB Data Project identified a need to put data infrastructure and services (research data trust, health research data hub, safe haven, trusted third party and data linkage services, and a research support unit) in place to enable health researchers to overcome the barriers they currently encounter in trying to access and use research data. The current environment impedes opportunities for research, and creates barriers and delays that are not conducive to establishing a world-class health research environment in Ireland.

This report proposes a model – the DASSL (data access, storage, sharing and linkage) model – to address the deficits identified in the course of the HRB Data Project. These deficits were found to be shared by many players in the Irish data ecosystem. The model and its proposed implementation were informed by an investigation of the legal, governance, social and cultural factors involved in establishing an enabling environment for health research. Given the sensitivity of most health data, it is expected that the model will not only respond well to the governance requirements of health researchers but will also respond well to the requirements of researchers in the broader data ecosystem, and thus can serve economic, social as well as health agendas in the successful exploitation of a core asset in the economy – i.e., data. It is proposed that the model elements be incorporated into the Irish data infrastructure as part of the Irish statistical system. A whole-of-government approach is desirable in order to ensure successful implementation of the model.

The infrastructure and services required to create an enabling research environment outlined above will not happen automatically and will require concerted and committed engagement by a large number of stakeholders. It is proposed that a Data to Benefits Committee be established to examine the proposals presented here. The committee would include representatives of the major stakeholders and would spearhead discussions about the development of the governance, infrastructure and services required to create a welcoming environment for health research in Ireland.

There is timeliness to the presentation of the model. The call for policy and practice to be informed by quality data and research is widespread, and strongly underpins developments planned under the government's Healthy Ireland framework, the new clinical care programmes in the HSE, as well as healthcare delivery generally. The opening up of data and its exploitation for enterprise, public service and job creation are central to government policy. In the emerging modern data environment, initiatives under the eHealth Strategy, Connected Health and big data would be facilitated by the DASSL infrastructure. The forthcoming Health Information and Patient Safety Bill and the Data Sharing and Governance Bill legislation demonstrate the Government's support for enhanced exploitation of data.

Experience from other countries where DASSL-type infrastructure and services have been put in place suggests that implementation of the model would undoubtedly expand the type and extent of research carried out in Ireland. The HRB Data Project found that health researchers are anxious to avail of the opportunities that would be presented to conduct research, which to date has been either impossible or inordinately burdensome to carry out. The DASSL model would enable the health research community to explore new avenues that would inform policy and practice, and help leverage EU and international funding for health research. Stakeholders in the broader data ecosystem would be able to avail of the new data infrastructure for business and enterprise purposes, with significant consequences for economic development and job creation.

If we want to have a safe and trusted modern infrastructure that will enable researchers to unlock the significant value of currently under-exploited data for the public good, then a DASSL or similar model needs to be implemented in Ireland.

If we want to have a safe and trusted modern infrastructure that will enable researchers to unlock the significant value of currently underexploited data for the public good, then the DASSL model or a similar model needs to be implemented, in order to facilitate not only health research but other research that serves national economic and social agendas.

Appendices

An abstract network diagram is overlaid on the page. It consists of several circles of varying sizes connected by thin lines. The circles are distributed across the page, with a larger circle on the right side and several smaller ones on the left and bottom. The lines connect these circles, creating a web-like structure that suggests interconnectedness or a data network.

Appendix 1:

Examples of Research Which Used Data Access, Storage, Sharing and Linkage

i.e. DASSL-Type Supportive Infrastructure and Services for Health, Economic and Social Benefits

There is a large international body of literature available which demonstrates the health, social and economic benefits of re-using data that already exist and also demonstrates the advantages of linking data from different datasets. Sharing and linking data optimises the use of expensively collected data and makes possible other types of research – for example, longitudinal studies or research that requires large amounts of data, as is the case when investigating infrequently occurring phenomena (e.g., rare conditions, suicide). Sharing and linkage of health, administrative and other datasets allows for investigation of health issues within their wider societal, economic contexts, thus providing a powerful means to increase understanding of the causal and preventive factors involved in health, and to identify potentially modifiable protective and risk factors.

For convenience, the following examples are all taken from the Farr Institute of Health Informatics Research website. However, this is just one of many websites providing examples of published research that was made possible through the use of infrastructure and services as proposed here in the DASSL model (see, for example, research which used the Secure Anonymised Information Linkage (SAIL) data¹⁶⁹, or the special edition of *Scandinavian Journal of Public Health* 2011, which provides examples based on Danish social and health datasets, including disease registers). These research infrastructural supports are well developed and advanced in other countries. Through implementation of the DASSL model, similar valuable research can be carried out and facilitated in Ireland, thus providing knowledge to guide Irish health, economic and social policy and practice, as well as allowing Irish researchers to compete on the international stage. The chosen examples below are presented using the headings featured on the Farr Institute website http://www.farrinstitute.org/188_Case-studies.html

Disease treatment

The economic and health resource impact of statin use for the primary prevention of cardiovascular disease

High cholesterol levels increase the risk of cardiovascular disease. Statins lower cholesterol and reduce the risk of heart problems, but the long-term economic effect of treating healthy people with these drugs, and the impact on health resource usage, have been unclear. The West of Scotland Coronary Prevention Study (WOSCOPS) recruited over 6,500 middle-aged men with high cholesterol and no history of heart attack between 1989 and 1991. Over a five-year period, half were treated with the cholesterol-lowering drug pravastatin and half were treated with a placebo. In order to explore long-term outcomes, patients were linked to electronic hospital discharge records, the cancer registry and death records, providing an extra 10 years of follow-up. Treatment with pravastatin was found to be cost-saving in the long term. Over the 15-year period, it saved the NHS £710,000 for every 1,000 patients treated for five years. There was also a significant reduction in the number of days spent in hospital and a previously unseen drop in hospital admissions due to heart failure. It is likely that a similar outcome would be observed for other cholesterol-lowering drugs.

This study was carried out by researchers at the Farr Institute Scotland.

Comment: This study shows the benefits of linking data collected on cholesterol levels to a number of existing datasets (i.e., hospital discharge data, cancer registry, death records), in order to explore long-term outcomes of statin use, the cost-benefit of such use, impact on admissions to hospital and length of stay in hospital. The DASSL-type infrastructure would enable Irish researchers to conduct such research by providing data access and linkage services within an appropriate governance framework.

Health services delivery

The risk of special educational needs among children born pre-term and early term

Pre-term babies born more than three weeks early are known to have a higher risk of neurodevelopmental problems, which can lead to special educational needs (SEN) later in childhood. However, the risk of SEN associated with early-term births – when mothers choose to deliver their babies just 12 weeks early – is not as clear.

¹⁶⁹ http://www.saildatabank.com/media/13007/130522_sail_projects_list_-_master.pdf

A study in Scotland linked health and education data for more than 400,000 school-aged children to determine the risk of SEN among children born 1-16 weeks early. It found that this risk increases steadily with increasing prematurity. Even babies born just 1-3 weeks early are more likely to have SEN than full-term babies. This suggests that choosing to give birth early is not a risk-free choice. Mothers should be made aware of the risks and benefits relevant to their circumstances, so that they can make an informed decision.

This study was carried out by researchers at the Farr Institute Scotland.

Comment: This study illustrates the value of linking health data to administrative data – in this case linking health data to education data. Again, the DASSL infrastructure could support this type of linkage of datasets in a secure environment.

Drug safety

Selective serotonin reuptake inhibitors (SSRIs) and the risk of self-harm and suicide

SSRIs are the most common class of drugs prescribed for depression. However, there have been concerns that SSRI use may promote self-harm and suicidal thoughts, particularly among adolescents. Previous studies investigating this link have been limited by their small size, as large amounts of data are needed to explore a rare event such as suicide. In light of this, the Medicines & Healthcare products Regulatory Agency (MHRA) – in collaboration with several universities – initiated a study in 2005 of 146,000 patients aged 10 years and older, using prescribing and diagnosis data from the General Practice Research Database. Its aim was to compare the risk of self-harm and suicide between people taking SSRIs for depression and those taking alternative antidepressant drugs. SSRI use in adults was not associated with a greater risk of self-harm or suicide. However, patients aged 18 years and younger taking SSRIs were found to have a slightly higher risk of self-harm. This finding confirmed existing regulatory advice that SSRIs should be prescribed cautiously to this age group.

This study was carried out by researchers at the Farr Institute CIPHER.

Comment: Access to a large dataset is necessary in order to test the anecdotal relationship between anti-depression SSRI use and relatively rare events such as suicide and self-harm. The results of this work confirmed regulatory advice, the adoption

of which should lower the incidence of self-harm. DASSL could support the conduct of such research in Ireland through secure linking of existing Irish self-harm and suicide data to prescribing data and other possible clinical data from an expanded Irish Primary Care Research Network (IPCRN).

Public health monitoring

Environmental triggers of myocardial infarction

Temperature and pollution are known to affect mortality, but the influence of these environmental factors on heart attack risk has been less clear. In order to explore this association, researchers at the London School of Hygiene & Tropical Medicine linked data from the Met Office and the UK National Air Quality Archive to data from the Myocardial Ischaemia National Audit Project (MINAP) for 15 urban areas across England and Wales between 2003 and 2006. They discovered that each 1°C drop in temperature was associated with a 2% short-term increase in the risk of suffering a heart attack, particularly among older people and those with previous coronary heart disease. Two products of traffic pollution – small airborne particles and nitrogen dioxide – were also associated with an increased risk of heart attack in high-risk groups. The pollutants appeared to trigger heart attacks early in these people, effectively bringing forward attacks that would have happened at a later time.

This study was carried out by researchers at the Farr Institute London.

Comment: This study demonstrates the value of using existing environmental data (collected in this case by the UK Met Office) and relating it to routinely collected health data which are stored in the National Institute for Cardiovascular Outcomes Research (NICOR) in University College London, where data can be accessed by researchers in accordance with good governance protocols. NICOR provides a linking service to other national datasets, thus enabling a large variety of studies. It is anticipated that the health research data hub, along with the proposed linkage and safe haven functions, would enable the development of a body of similar research in the Irish context.

Disease risk

The impact of pre-term birth on childhood respiratory disease

The number of pre-term babies has grown over the last 10 years. Very pre-term infants are more likely to have poor lung function in childhood, but it is not clear whether this is the case for moderately and late pre-term babies. In a study partly funded

by the Medical Research Council, a team of Welsh researchers used the Wales Electronic Cohort for Children (WECC) to explore the effect of different degrees of prematurity, from very pre-term to late pre-term birth, on the risk of emergency respiratory hospital admissions in early childhood. Using the WECC enabled the team to link and analyse the health records of over 300,000 children born in Wales between 1998 and 2008. These records included birth records, inpatient admissions and mortality data. Their study found that the more premature a baby is, the greater its risk of being admitted to hospital with respiratory disease in childhood. Even babies born at 39 weeks have a 10% greater risk when compared with babies born at 40–42 weeks. Given the growing number of pre-term infants, this finding suggests a potentially significant future impact on paediatric healthcare services.

This study was carried out by researchers at the Farr Institute CIPHER.

Comment: The WECC anonymously links routinely collected datasets to create the first complete population e-cohort in the UK. Anonymous linking fields are used to link records for the same child with both their mothers and their local environment data (including datasets from the Secure Anonymised Information Linkage (SAIL) databank). Such linkage enables the conduct of research into a really broad range of social, economic and environmental determinants of child health, well-being and social outcomes.¹⁷⁰ The possibility of creating such national cohorts would be enabled by the implementation of DASSL infrastructure and services.

Disease mechanism

Wider health benefits of diabetes drug

Metformin is a drug widely used by patients with type 2 diabetes. Following an observed reduction in cancer incidence in these patients, record linkage was used to test the hypothesis that metformin decreases the risk of developing cancer in diabetic patients. Researchers linked several Scottish health datasets, including the cancer and diabetes registers and the database of dispensed prescriptions. They found that metformin is associated with a reduction in cancer incidence. Subsequent biochemical and genetic studies have shown that the potential mechanism for this protective effect is through a cancer suppressor gene associated with the response to metformin. Similar record linkage studies will likely play an important role in stratifying patients for the most effective treatments and also in discovering new drug targets.

This study was carried out by researchers at the Farr Institute Scotland.

Comment: This study shows that easy access to health and related datasets allows for exploration of many diverse hypotheses, while the ability to link different datasets further expands the potential for discovery of health-related outcomes.

¹⁷⁰ http://fn.bmj.com/content/96/Suppl_1/Fa18.1.abstract

Appendix 2:

Summary of Anticipated Benefits Associated with the British Columbia Centre for Data Innovation by Stakeholder Group¹⁷¹



Citizens	Researchers	Private Sector	Government
Enhanced transparency and trust	Access to one of the most robust, linked data resources in Canada	Better understanding of needs of citizens & government	Improved understanding of citizen needs
Personalized services and products	Predictable, timely turnaround on access requests	Ability to bring new/enhanced data products & services to market	Improved policy development & decision making
More employment opportunities associated with economic growth	Increased productivity arising from enabling services and reduced transition costs for data requests	Enhanced innovation and productivity through the 'cluster effect'	Time and resource savings due to more efficient service delivery
Improved outcomes and efficiency of government services (healthcare, social services, environment etc.)	Improved access to research funding.	Improved global competitiveness	Improved data quality
Improved security and privacy of citizen data.		Improved competitiveness for funding	Enhanced innovation through partnerships with citizens and the private sector (co-creation)

It should be noted that the reference to "Government" in this figure includes the broader public sector within the province (e.g., school districts, health authorities and other public sector entities)

¹⁷¹ BC Centre for Data Innovation 2014. 'BC Centre for Data Innovation Final Report of Working Group'. p.7 at http://www.gov.bc.ca/citz/down/BC_Centre_for_Data_Innovation-FINAL.pdf

Appendix 3: The Research Data Agenda – Communication to Participants in the HRB Data Project¹⁷²

Background – developments in relation to research data

Research is being transformed by the development of extraordinary new ways of collecting, storing, manipulating and transmitting data and information. Good data management and sharing practice allows reliable verification of results, and permits new and innovative research built on existing information. The boundaries between previously distinct fields are blurring as ideas and tools are exported from one discipline to another. Informatics and the ability to mine large datasets and combine them with information from many other sources offer huge potential to advance research. This is important if the full value of public investment in research is to be realised.

International developments – managing and sharing research data

- There is growing international support for managing and sharing research data:
- In 1997, the US National Research Council argued that ‘full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.’
- In 2007, the OECD published a set of *Principles and Guidelines for Access to Research Data from Public Funding*.¹⁷³
- A 2010 report by the European Commission’s High Level Expert Group on Scientific Data, *Riding the Wave*, called on the Commission to accelerate moves towards a common data infrastructure.¹⁷⁴
- In 2010, the US National Science Foundation announced that it would alter its data sharing policy to require data management plans in future grant proposals to the agency.
- In 2011, a declaration on *Sharing Research Data to Improve Public Health* was signed by 17 major international public health research funders.¹⁷⁵
- In 2012, the UK Royal Society landmark report, *Science as an Open Enterprise*, stressed the potential for data re-use and a need for rapid data sharing, so that we can respond to global challenges, such as flu epidemics or disaster risks.¹⁷⁶
- In 2012, the Research Data Alliance (RDA), involving Australia, the US and the EU, was formed to accelerate and facilitate research data sharing and exchange by promoting and encouraging both a bottom-up and interdisciplinary approach.¹⁷⁷ In March 2014, the next plenary meeting of the RDA will be held in Dublin. In 2012, the EU Commission issued two communications, one outlining its intention to promote open access to research data and to set a pilot framework in Horizon 2020¹⁷⁸ and a second setting out key priorities for completing the European Research Area, including the optimal circulation, access to, and transfer of scientific knowledge, including research data.¹⁷⁹ In July 2013, the European Commission held a public consultation on open research data to inform its approach to Horizon 2020. In 2013, the US Office of Science and Technology Policy Memorandum, *Expanding Public Access to the Results of Federally Funded Research*, called on funders ‘to develop a plan to support increased public access to the results of research’, including data.¹⁸⁰ Funding agencies were asked to provide plans by September 2013.
- In 2013, G8 leaders signed an Open Data Charter following a G8 science ministers’ statement of principles for more open data, including that publicly funded research data should be made open, and that open data, by definition, should be easily discoverable, accessible and assessable.

¹⁷² Note compiled by P. Clarke for participants in the HRB Data Project.

¹⁷³ <http://www.oecd.org/sti/sci-tech/38500813.pdf>

¹⁷⁴ <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

¹⁷⁵ <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>

¹⁷⁶ <http://royalsociety.org/policy/projects/science-public-enterprise/report/>

¹⁷⁷ <https://rd-alliance.org/about.html>

¹⁷⁸ *Scientific data: open access to research results will boost Europe’s innovation capacity.*
Go to http://europa.eu/rapid/press-release_IP-12-790_en.htm

¹⁷⁹ *A Reinforced European Research Area Partnership for Excellence and Growth.*

Go to http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-partnership-excellence-growth_en.pdf

¹⁸⁰ <http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>

- Recently, within the UK, the Research Councils UK,¹⁸¹ the Medical Research Council¹⁸² and the Wellcome Trust¹⁸³ have all developed new, more comprehensive guidelines to govern management and sharing of research data.
- At EU level, a Science Europe working group has been convened to consider research data as a key policy priority area. The HRB is represented on this group.

Any new HRB policy will need to consider the different types of data, the availability of repositories, the level of established good practice for data sharing, and the costs associated with new practices.

In some research fields – such as genetics and physics – data sharing is well established. In health research, however, while research collaborations are growing more common, the sharing of data is not yet the norm. The HRB will work with the research community and other relevant stakeholders to shape and apply good practice with regard to the effective management of research data arising from HRB-funded research throughout and beyond the research life cycle.

Possible implications for researchers

This changing agenda will have implications for our researchers. In pursuing the shift away from a research culture where data are viewed as a private preserve, researchers are typically required to submit a data management and sharing plan (DMP) as part of their research proposal. This DMP should explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication issues. Other changes that may be forthcoming include:

- developing necessary accompanying measures and incentives for scientists to share their data
- the development and use of common standards for communicating data
- data that underpin a journal article being made concurrently available in an accessible database
- expanding the criteria used to evaluate research, so as to give credit for useful data communication and novel ways of collaborating
- strengthening the cohort of data scientists needed to manage and support the use of digital data, and facilitating the interface between computer scientists and disciplinary scientists.

¹⁸¹ <http://www.rcuk.ac.uk/research/datapolicy/>

¹⁸² <http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/>

¹⁸³ <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

Appendix 4:

HRB Data Project – List of Informants

Title	First name	Surname	Organisation
Dr	Ella	Arensman	National Suicide Research Foundation
Dr	Kevin	Ballinda	Institute of Public Health in Ireland (IPH)
Ms	Julia	Barrett	Irish Social Science Data Archive (ISSDA)
Mr	William	Beausang	Department of Public Expenditure and Reform (DPER)
Ms	Jasmina	Behan	SOLAS
Mr	John	Brazil*	Health Protection Surveillance Centre
Mr	Marcus	Breathnach	Enterprise Ireland (EI)
Mr	Alan	Cahill*	Department of Health
Mr	Aidan	Clancy*	Department of Health
Dr	Sandra	Collins	Digital Repository of Ireland (DRI)
Mr	Cathal	Daly*	Assistant Data Protection Commissioner DPC
Prof	Stefan	Decker*	NUI Galway
Ms	Rhona	Dempsey	Quality and Qualifications Ireland
Dr	J-C	Desplat	Irish Centre for High-End Computing (ICHEC)
Ms	Mary	Doyle	Department of Education and Skills
Mr	John	Dunne	Administrative Data Centre and Methodology, CSO
Mr	Garry	Dunphy*	ICT Service Delivery, CSO
Dr	Tom	Fahey	Royal College of Surgeons in Ireland (RCSI)
Ms	Clare	Finn	Department of Children and Youth Affairs (DCYA)
Mr	Geoffrey	Fletcher*	Cystic Fibrosis Registry of Ireland
Ms	Barbara	Foley*	Health Information and Quality Authority (HIQA)
Dr	Margaret	Foley*	Data Manager, TILDA, TCD
Dr	Jane	Gray	Irish Qualitative Data Archive (IQDA)
Prof	Jane	Grimson	Health Information and Quality Authority (HIQA)
Dr	Sinead	Hanafin	Research Matters
Dr	Ann	Hever*	TILDA, TCD
Mr	Richard	Howell*	Department of Agriculture, Food and the Marine
Ms	Abie	Jackson	UCD School of Public Health
Mr	Howard	Johnson	Health Atlas Ireland, Health Intelligence Unit, HSE
Prof	Cecily	Kelleher	UCD
Mr	Raymond	Kelly*	Teagasc
Ms	Aoife	Lawlor	HSE
Prof	Richard	Layte*	TCD, Department of Sociology
Ms	Deirdre	Lee*	Insight Centre for Data Analytics, NUI Galway, Ireland
Mr	Peter	Lennon	Department of Health
Dr	Nancy	Meagher*	International Population Data Linkage Network
Mr	Tom	Meagher	St Patrick's Hospital

Title	First name	Surname	Organisation
Dr	Patrick	Moore*	TILDA, TCD
Ms	Mary	Morrissey*	HSE
Ms	Eibhlin	Mulroe*	Irish Platform for Patients' Organisations, Science and Industry (IPPOSI)
Mr	Colm	Moroney	Dublin Region Homeless Executive
Dr	Sean	Mulvany	Enterprise Ireland (EI)
Dr	Tracy	O'Carroll	HIQA
Ms	Evelyn	O'Connor*	Department of Public Expenditure and Reform (DPER)
Dr	Bernie	O'Donoghue-Hynes	Dublin Region Homeless Executive
Dr	Stephanie	O'Keefe*	HSE
Ms	Brittany	O'Neill	School of Information & Library Studies of UCD
Dr	Pat	O'Hara	National Statistics Board
Mr	Michael	Perkins*	Department of Public Expenditure and Reform (DPER)
Dr	Patrick	Redmond*	Royal College of Surgeons in Ireland (RCSI)
Ms	Bennery	Rickard	HSE
Prof	Frances	Ruane*	Economic and Social Research Institute (ESRI)
Dr	Ratnesh	Sahay*	INSIGHT Centre for Data Analytics
Dr	Kalpana	Shanker*	School of Information & Library Studies of UCD
Ms	Frances	Spillane	Department of Health
Prof	Tony	Staines	DCU, School of Nursing
Dr	Victoria	Tsoukala*	EU Recode Project, Greece
Dr	Mark	Ward*	Department of Sociology, TCD
Prof	James	Williams*	Economic and Social Research Institute (ESRI)

* Discussion held versus full interview

Appendix 5:

Listing of Health and Social Care Data Collections from HIQA 2014¹⁸⁴

1. National Data Collections of Health and Social Care in Ireland (n=74)

- 1 Alcohol Hand Rub Consumption Monitoring
- 2 Alpha One Patient Registry
- 3 Blood Donor Database
- 4 BreastCheck (The National Breast Screening Programme)
- 5 Central Treatment List (CTL)
- 6 CervicalCheck: Cervical Screening Register
- 7 Clostridium difficile enhanced surveillance
- 8 Computerised Infectious Disease Reporting (CIDR)
- 9 Cystic Fibrosis Registry of Ireland
- 10 Enhanced Bacteraemia (Bloodstream infections) Surveillance in Ireland
- 11 Fatalities and other Traffic Statistics
- 12 Hand Hygiene Compliance Monitoring
- 13 Heart Rhythm Ireland (Irish National Pacemaker Register)
- 14 HIV Antenatal Testing
- 15 Hospital In-Patient Enquiry
- 16 HSE Performance Reports – Acute Hospitals including Clinical Programmes, National Ambulance Service and National Cancer Control Programme
- 17 HSE Performance Reports – Child Protection and Welfare Services
- 18 HSE Performance Reports – Disability Services
- 19 HSE Performance Reports – Health and Wellbeing and Governance
- 20 HSE Performance Reports – Mental Health Services
- 21 HSE Performance Reports – Older People Services
- 22 HSE Performance Reports – Primary Care and Social Inclusion and Palliative Care
- 23 Immunisation Uptake Statistics
- 24 Irish Audit of Surgical Mortality (IASM)
- 25 Irish Biologic Therapies Register
- 26 Irish Childhood Diabetes National Register
- 27 Irish Epilepsy and Pregnancy Register
- 28 Irish Heart Valve Bank Register
- 29 Irish Hip Fracture Database
- 30 Irish Motor Neurone Disease Register
- 31 Irish National Orthopaedic Register (INOR)
- 32 Irish Unrelated Bone Marrow Registry (IUBMR)
- 33 Major Trauma Audit (MTA)
- 34 MHC – Admissions of Children to Approved Centres
- 35 MHC – Deaths relating to all residents in Approved Centres
- 36 MHC – Involuntary Admission Activity
- 37 MHC – Administration of Electro-convulsive Therapy (ECT) in approved centres
- 38 MHC – Use of Seclusion, Mechanical Restraint and Physical Restraint in approved centres
- 39 National Adverse Event Management System (NAEMS)

¹⁸⁴ www.hiqa.ie/system/files/Data-Catalogue-2014.pdf

- 40 National Antimicrobial Resistance Surveillance
- 41 National Cancer Control Programme (NCCP)
- 42 National Cancer Drug Management Programme
- 43 National Cancer Registry Ireland
- 44 National Cleft Database
- 45 National Drug-Related Deaths Index (NDRDI)
- 46 National Drug Treatment Reporting System (NDTRS)
- 47 National Haemophilia Register
- 48 National Health Schemes Data (Primary Care Reimbursement Service)
- 49 National Hepatitis C Database
- 50 National Intellectual Disability Database (NIDD)
- 51 National Intensive Care Audit (ICU Audit)
- 52 National Organ Procurement Service Statistics
- 53 National Paediatric Haemopoietic Stem Cell Transplantation
- 54 National Paediatric Mortality Register
- 55 National Perinatal Epidemiology databases: Perinatal Mortality, Severe Maternal Morbidity and Homebirth Databases
- 56 National Perinatal Reporting System
- 57 National Physical and Sensory Disability Database (NPSDD)
- 58 National Poisons Information Centre Database
- 59 National Psychiatric Inpatient Reporting System (NPIRS)
- 60 National Registry of Deliberate Self Harm Ireland
- 61 National Renal Transplant Registry
- 62 National Spinal Injuries Unit
- 63 NHS Blood and Transplant Audit UK & Ireland
- 64 Out-of-Hospital Cardiac Arrest Register (OHCAR)
- 65 Patient Treatment Register (PTR)
- 66 Pre-Hospital Emergency Care Council (PH)– Cardiac First Response Report (CFRR)
- 67 PHECC – Patient Care Report (PCR)
- 68 PHECC – Patient Transport Report (PTR)
- 69 Sentinel Flu Surveillance
- 70 Surveillance of antimicrobial consumption in Ireland
- 71 Vital Statistics – Deaths Registration
- 72 Vital Statistics – Live Births Registration
- 73 Work-Related Injuries Database
- 74 Workplace Fatalities Database

2. National Censuses (n=3)

- 1 Census of Population and other Population Data
- 2 Health Service Personnel Census
- 3 Irish Psychiatric Units and Hospitals Census

3. Data collections without national coverage/regional collections (n=6)

- 1 Cardiac surgery registers
- 2 Eastern Region Cerebral Palsy Register
- 3 Coronary Heart Attack Ireland Register (CHAIR)
- 4 e-Heartbeat
- 5 Heartwatch
- 6 EUROCAT South, East, and South-East (Congenital Anomaly Register)

4. Systems that collated data from a number of different sources (n=15)

National

- 1 Ageing in Ireland
- 2 Community Profiles Tool
- 3 CompStat
- 4 Drug Situation Ireland
- 5 Health Atlas
- 6 Health in Ireland – Key Trends
- 7 Irish Casemix Programme
- 8 Public Health Information System (PHIS)
- 9 State of the Nation's Children
- 10 Statistical Information on Social Welfare Services
- 11 Women and Men in Ireland

European/International

- 12 European Health for all Database (HFA-DB)
- 13 Eurostat Health Statistics
- 14 OECD Health Data
- 15 World Health Organization Global Health Observatory

5. National Surveys (n=9).

- 1 European Social Survey
- 2 Growing Up In Ireland
- 3 Health Behaviour in School-aged Children
- 4 Lifeways
- 5 SHARE – Survey of Health, Ageing and Retirement in Europe (SHARE)
- 6 SILC – Survey on Income and Living Conditions
- 7 SLÁN – Survey of Lifestyle, Attitudes, and Nutrition
- 8 QNHS – Quarterly National Household Survey – Health Module
- 9 TILDA – The Irish Longitudinal Study on Ageing

Appendix 6:

Listing of Datasets in ISSDA¹⁸⁵

Study	Datasets	Data Provider	Availability	Purpose	Formats
Adapting to Diversity: Irish Schools and newcomer students		ESRI and Department of Education	Apply to ISSDA	Research	SPSS
Ageism and Ageing	2003 2008		Download		SPSS
All Ireland Traveller Health Survey (AITHS)		School of Public Health & Population Science, University College Dublin	Apply to ISSDA	Research Teaching	SPSS
Census of Population	COPSAR 1996 2002 2006	CSO	Apply to ISSDA	Research Only	SPSS
	POWSAR 2002				SPSS
	SAPS 1981 1986 1991 1996 2002				CSV
Children's Sport Participation and Physical Activity (CSPPA)	2009	Dublin City University, University of Limerick, University College Cork, and The Irish Sports Council	Apply to ISSDA	Research Teaching	SPSS
CoHeart		Health Research Board	Apply to ISSDA	Research	SPSS
Commission for Energy Regulation	Electricity Smart Meter CBT Gas Smart Meter CBT	Commission for Energy Regulation	Apply to ISSDA	Research	CSV
Cross National Comparative Data	Various	Various			
Doodle Den	2008–2011	Childhood Development Initiative (Tallaght West) and Centre for Effective Education, School of Education, Queen's University Belfast	Apply to ISSDA	Research	PSS

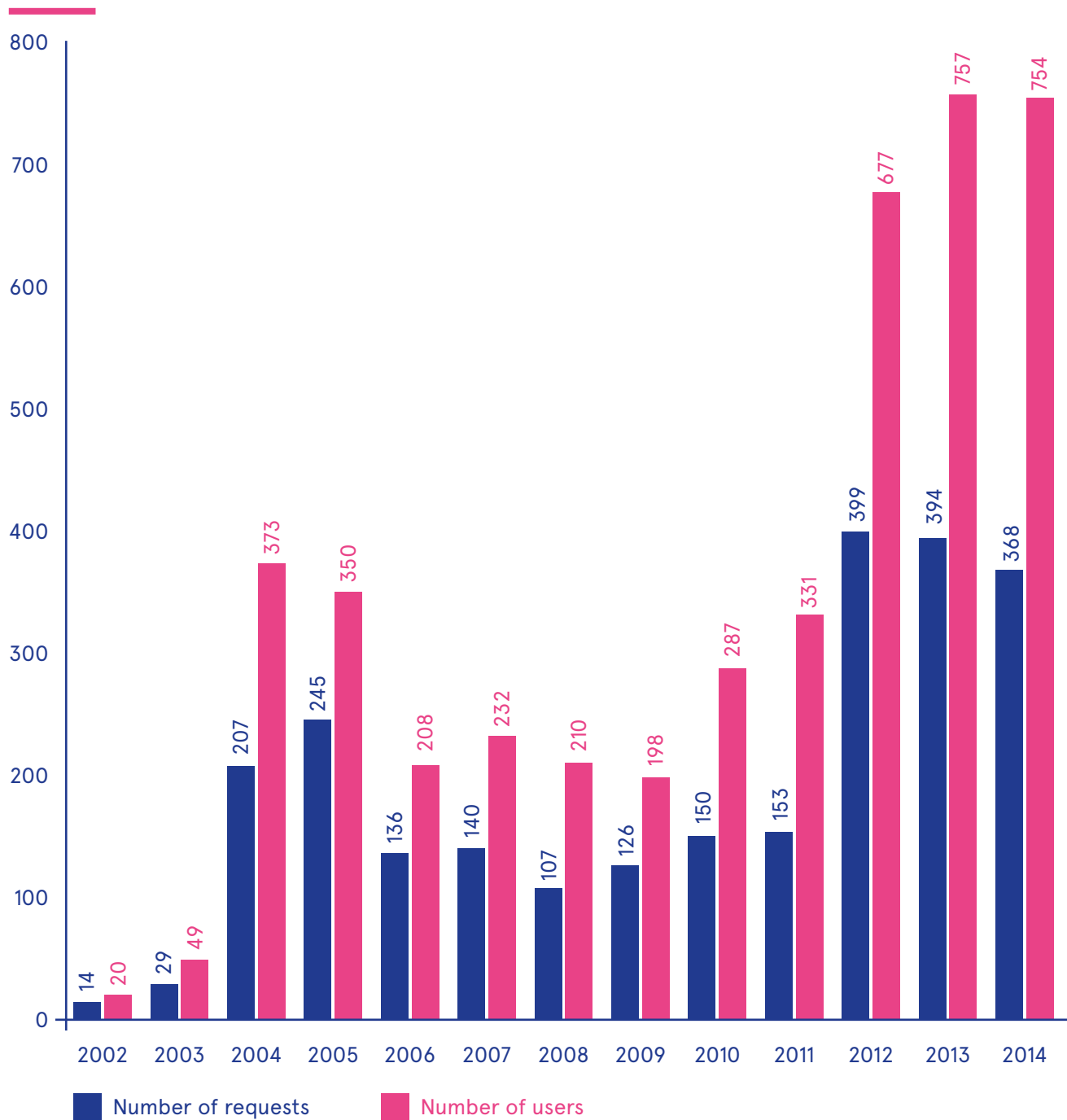
¹⁸⁵ <http://www.ucd.ie/issda/data/>

Study	Datasets	Data Provider	Availability	Purpose	Formats
EU Referendums	Lisbon 2008 2009	Department of Foreign Affairs	Download		SPSS
EU Survey of Income and Living Conditions (EU-SILC)	2003–2012	CSO	Apply to ISSDA	Research	SAS, CSV (SPSS and Stata are also available for 2003–2009)
Eurostudent	Round III Round IV Round V	HEA	Apply to ISSDA	Research	Stata (III) SPSS (IV)
Funded Project Data	Various	Various	Download		
Growing Up in Ireland (GUI)	Infant Cohort Wave1 (9mo) Wave2 (3yrs) Child Cohort Wave1 (9yrs) Wave1 (Time Use Data) Wave2 (13yrs)	Department of Children and Youth Affairs, Department of Social Protection, CSO, ESRI and TCD	Apply to ISSDA	Research Teaching	SPSS
Household Budget Survey (HSB)	1987 1994/95 1999/2000 2004/05 2009/10	CSO	Apply to ISSDA	Research	SPSS, Stata, SAS, CSV (1987, 1994/95, 1999/2000) SPSS, Stata (2004/05) SAS, CSV (2009/10)
Insight '07 (HSE)		HSE and the School of Public Health Physiotherapy and Population Science, University College Dublin.	Apply to ISSDA	Research	SPSS
Irish National Election Study (INES)	2002–2007	ESRI, TCD and UCD	Download	Research	SPSS, Stata
Irish National Survey of Housing Quality (NSHQ)	2001/02	Department of the Environment and ESRI	Apply to ISSDA	Research	SPSS
Irish National Time Use Survey	2005	ESRI and NDP Gender Equality Unit of the Department of Justice, Equality, and Law Reform	Apply to ISSDA	Research	SPSS

Study	Datasets	Data Provider	Availability	Purpose	Formats
Irish Social and Political Attitudes Survey (ISPAS)	2001	ESRI, TCD and UCD	Apply to ISSDA	Research	SPSS
Irish Study of Sexual Health and Relationships (ISSHR)	2006	ESRI, HSE Crisis Pregnancy Programme	Apply to ISSDA	Research Teaching	SPSS, Stata
Irish Survey of Contraception and Crisis Pregnancy (ICCP)	2004 2010	ESRI, HSE Crisis Pregnancy Programme	Apply to ISSDA	Research	SPSS
Labour Force Survey	1995-1997	CSO	Apply to ISSDA	Research	
Living in Ireland (LII)	1994-2001	ESRI, European Community Household Panel (ECHP)	Apply to ISSDA	Research	SPSS
National Psychological Wellbeing and Distress Survey	2006 2008	ESRI and HRB	Apply to ISSDA	Research	SPSS
National Travel Survey (NTS)	2009	CSO and the Department of Transport, Tourism and Sport	Apply to ISSDA	Research	SPSS, Stata, SAS and CSV
NCPP Surveys of Employee Attitudes	2003 2009	NCPP, ESRI and Amarach Research	Apply to ISSDA	Research	SPSS
Opinion Poll Data	2001-2008 2002-2005	Taylor Nelson Sofres MRBI Millward Brown	Apply to ISSDA	Research	SPSS
Quarterly National Household Survey (QNHS) - Microdata	1998-2012	CSO	Apply to ISSDA	Research	SAS

Study	Datasets	Data Provider	Availability	Purpose	Formats
QNHS Special Modules Health data Q3-2001, Q3-2007, Q3-2010 Pensions data Q1-2002, 2005-2008, Q4-2009 Disability data Q2-2002 Childcare data Q4-2002, Q4-2007 Lifelong learning data Q2-2003 Crime and victimisation Q4-1998, Q4-2003 Work organisation and working time data Q2-2004 Equality data Q4-2004, Q4-2010 Childcare data Q2-2004 Voter Participation/Abstention data Q3-2002		CSO	Apply to ISSDA	Research	Various
School Leavers Survey	1980-2007	ESRI and the Department of Education	Apply to ISSDA	Research	SPSS
SPHERE		HRB, Irish Heart Foundation, NUIG, QUB, RCSI	Apply to ISSDA	Research Teaching	SPSS
Survey of Public Attitudes to Disability	2006 2011	NDA	Apply to ISSDA	Research	SPSS
Survey of Public Attitudes Towards Forestry in Ireland	1998	ESRI and the Department of Communications, Energy and Natural Resources	Apply to ISSDA	Research	SPSS, Stata
Survey on Lifestyle and Attitudes to Nutrition (SLÁN)	1998, 2002, 2007	ESRI, Department of Health & Children, and others	Apply to ISSDA	Research Teaching	SPSS (all three), and Stata and SAS (1998 and 2002 only)
Teagasc National Farm Survey (NFS)	2005-2012	Teagasc	Apply to ISSDA	Research	Excel
The Irish Longitudinal Study on Ageing (TILDA)	Wave 1	TCD, Irish Life, Atlantic Philanthropies, and the Department of Health	Apply to ISSDA	Research Teaching	

Appendix 7: Demand for ISSDA Data, 2002–14¹⁸⁶



¹⁸⁶ Received from Julia Barrett, UCD library, 2/3/2015

Appendix 8:

Guidelines for Data Sharing in the Public Sector – Office of the Data Protection Commissioner¹⁸⁷

Compliance with the following guidelines can provide a basis for a general approach to data sharing within the public sector based on the principles set out below. These principles should ensure that such data sharing is proportionate and in accordance with the Data Protection Acts.

1. Demonstrable justification

The public policy objective being pursued by a particular data sharing arrangement, without the consent of the individual(s) whose data are involved, should be explicit. An assessment should be made as to whether the likely benefits of the sharing justify the overriding of the individual's data protection rights. The assessment should represent a careful balancing of these factors. It should take account of the fact that such sharing could increase the reluctance of individuals to provide accurate personal data to State authorities. It should also take account of any disproportionately negative impact on particular sections of society.

2. Explicit legal basis

The legal basis for data sharing, including the conditions under which such sharing is permitted, should be set out in primary legislation.

3. Authorisation

Any decision to share personal data between public bodies (and thereby to set aside a person's right to privacy) must not be taken lightly. This is especially the case when bulk data are shared. Such decisions should only be taken following due consideration at senior management level.

4. Transparency

If relevant, it should be made clear to individuals when they give personal data to a State body that this information may be shared with other State bodies. The reason for such sharing should be stated clearly. Under the Data Protection Acts, State bodies are legally required to include such disclosures in their public registration with our

Office. In addition, it is good practice for a public body to regularly publish a list of its data sharing arrangements.

5. Data minimisation

Only the minimum amount of personal data should be shared. In many cases, all that is required is a 'yes' or 'no' with regard to whether an individual is, for example, a holder of a permit or a licence.

6. Data access and security

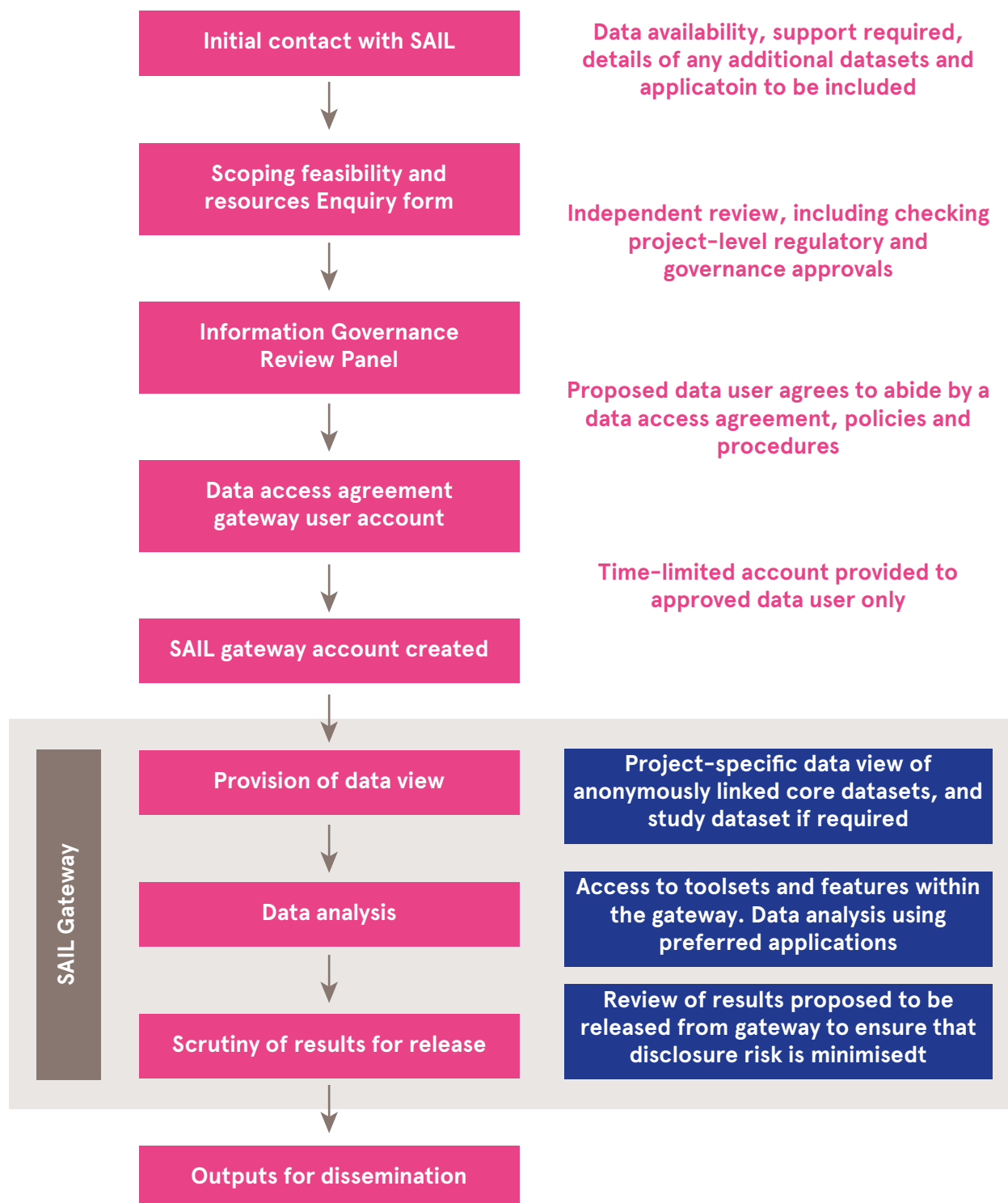
Enhanced access and security requirements should apply to personal data received as part of an approved data sharing arrangement. Access to such data should be limited to a very small number of officials, and security measures should rule out any possibility of data leakage (bearing in mind the increased emphasis on the State's responsibility to prevent data breaches and the reputational damage that would result from failure to protect shared personal data).

7. Data retention

Personal data provided as part of an approved data sharing arrangement should be securely destroyed when no longer required.

¹⁸⁷ https://www.dataprotection.ie/viewdoc.asp?m=m&fn=/documents/guidance/Data_Sharing_in_the_Public_Sector.htm

Appendix 9: The Sail Data User Journey¹⁸⁸



Appendix 10:

Abbreviations and Acronyms

ADRC-NI	Administrative Data Research Centre – Northern Ireland
ADRN	Administrative Data Research Network
CASD	Centre d’Accès Sécurisé Distant aux Données, France
CSO	Central Statistics Office
CSTAR	Centre for Support and Training in Analysis and Research
DAFM	Department of Agriculture, Food and the Marine
DASSL	Data access, storage, sharing and linkage
DERI	Digital Enterprise Research Unit
DES	Department of Education and Skills
DJEI	Department of Jobs, Enterprise and Innovation
DMP	data management plan
DoH	Department of Health
DPER	Department of Public Expenditure and Reform
DRI	Digital Repository Ireland
EI	Enterprise Ireland
ESRI	Economic and Social Research Institute
EU	European Union
EU SILC	EU Survey on Income and Living Conditions
GENES	The Group of National Economics and Statistics Schools, France
GUI	Growing Up in Ireland
HBSC	health behaviours in school-aged children
HEA	Higher Education Authority
HIPE	Hospital In-Patient Enquiry
HIQA	Health Information and Quality Authority
Horizon 2020	Horizon 2020 Framework Programme of EU (2014–2020)
HRB	Health Research Board
HSE	Health Service Executive
HSE BSO	Health and Social Care Business Services Organisation
HSR	health services research
INSEE	National Institute for Statistics and Economic Studies, France
IPCRN	Irish Primary Care Research Network
IPH	Institute of Public Health in Ireland
IPPOSI	Irish Platform for Patients’ Organisations, Science and Industry
IQDA	Irish Qualitative Data Archive
IRC	Irish Research Council
ISS	Irish statistical system
ISSDA	Irish Social Science Data Archive
MHC	Mental Health Commission
MHPRA	Medicines & Healthcare products Regulatory Agency
MINAP	Myocardial Ischaemia National Audit Project
MOU	Memorandum of Understanding
MRC	Medical Research Council (UK)
NCR	National Cancer Registry
NDP	National Development Plan

NDRDI	National Drug-Related Deaths Index
NDTRS	National Drug Treatment Reporting System
NHIS	National Health Information Systems
NI	Northern Ireland
NICOR	National Institute for Cardiovascular Outcomes Research
NIDD	National Intellectual Disability Database
NIHR	National Institute for Health Research
NILS	Northern Ireland Longitudinal Study
NISRA	Northern Ireland Statistics and Research Agency
NPIRS	National Psychiatric In-Patient Reporting System
NRPE	National Research Prioritisation Exercise
ODPC	Office of the Data Protection Commissioner
OECD	Organisation for Economic Co-operation and Development
PEER	Policy, Evaluation and External Relations, HRB
PH	public health
PHHSR	population health and health services research
PRTLl	Programme for Research in Third-Level Institutions
R&D	research and development
RCSI	Royal College of Surgeons in Ireland
RDA	Research Data Alliance
SAIL	Secure Anonymised Information Linkage
SEN	special educational needs
SFI	Science Foundation Ireland
SILC	EU Survey on Income and Living Conditions
SLÁN	Survey of Lifestyles, Attitudes and Nutrition
SSRI	selective serotonin reuptake inhibitor
TILDA	The Irish Longitudinal Study on Ageing
UCD SILS	School of Information and Library Studies, UCD
UHI	unique health identifier
WECC	Wales Electronic Cohort for Children
WHO	World Health Organization
WOSCOPS	West of Scotland Coronary Prevention Study
WT	Wellcome Trust

Notes

Notes

An Bord Taighde Sláinte
Teach Grattan
67-72 Sráid an Mhóta Íochtarach
Baile Átha Cliath 2
D02 H638
Éire

Health Research Board
Grattan House
67-72 Lower Mount Street
Dublin 2
D02 H638
Ireland

t +353 1 234 5000
f +353 1 661 2335
e hrb@hrb.ie
w www.hrb.ie

 [@hrbireland](https://twitter.com/hrbireland)