PRIVITAR
Research & Policy

Health Data Sharing Case Studies

2021

www.privitar.com

Contents

Acknowledgements 4

Abbreviations 5

About Privitar 6

Executive summary 7

1. Introduction 9
   - 1.1 Improving data-driven health research 9
   - 1.2 The case study organisations 10
   - 1.3 Why case studies? 10
   - 1.4 Our methodology 11

2. Regulatory requirements for health data 13
   - 2.1 Legislation and common law 13
   - 2.2 Policy and guidance 14
   - 2.3 Oversight bodies 15

3. Summary of findings 17
   - 3.1 Challenges and mitigations 17
   - 3.2 Best practice – Set up an information governance framework and a data sharing process 22
   - 3.3 Best practice - define roles and responsibilities 25
   - 3.4 Best practice - controls 26

4. Cambridge University Hospitals (CUH) 29
   - 4.1 Background on CUH 29
   - 4.2 CUH's data sharing context 29
   - 4.3 Roles and responsibilities 30
   - 4.4 CUH's data sharing process 31
   - 4.5 CUH simplified process diagram 36

5 Centre for Epidemiology Versus Arthritis (CfE) 37
   - 5.1 Background on CfE 37
   - 5.2 CfE's data sharing context 37
   - 5.3 Roles and responsibilities 39
   - 5.4 CfE's data sharing process 41
   - 5.5 CfE simplified process diagram 47
   - 5.6 Data Safe Haven Plus (DSH+) 48
   - 5.7 CfE - DSH+ simplified process diagram 52

Copyright 2021, Privitar LTD

Where do we go from here? 53

About the authors 54

Appendix A – Glossary and legal terminology 55

Appendix B – Common data controls 57

Appendix C – Common contract terms governing data use 58

Copyright 2021, Privitar LTD

Acknowledgements

We would like to thank the following.

- The Research & Development Department, the Information Governance, the Clinical Informatics (part of the NIHR Cambridge Biomedical Research Centre) and the eHospital teams at Cambridge University Hospitals.
- The Centre for Epidemiology, the Information Governance Office, and Research Lifecycle Programme: Project S at the University of Manchester and in particular – Elaine Mackey, Laurence Malbeaux, William Dixon, John Mcbeth, Alex Daybank, Gary Leeming, Mary Mcderby and Rachel Thompson.
- Our panel of four independent reviewers, including the Office of the National Data Guardian and the Medical Research Council Regulatory Support Centre. We're grateful to all our reviewers for their helpful, expert advice.

CONTENTS

Copyright 2021, Privitar LTD

Abbreviations

Anonymisation Decision-making Framework	ADF
Confidential Patient Information	CPI
Public Health England	PHE
National Health Service	NHS
Department for Health and Social Care	DHSC
Health Research Authority	HRA
Medical Research Council	MRC
Information Commissioner’s Office	ICO
National Data Guardian	NDG
Office for National Statistics	ONS
General Data Protection Regulation	GDPR
Privacy Enhancing Technologies	PETs
Data Protection Impact Assessment	DPIA
Cambridge University Hospital Trust	CUH
The Centre for Epidemiology Versus Arthritis	CfE
Data Protection Act 2018	DPA 2018
Medicines and Healthcare products Regulatory Agency	MHRA
Integrated Research Application System	IRAS
Clinical Practice Research Datalink	CPRD

CONTENTS

Copyright 2021, Privitar LTD

About Privitar

Privitar is the leader in modern data provisioning. We empower organisations to use data safely, quickly and at scale. Our clients use the Privitar Data Provisioning Platform to share data, unlock insights, keep data safe and support regulatory compliance. Our platform includes state-of-the-art privacy enhancing technologies, and our experts help customers use them effectively. Only Privitar has the right combination of technology and expertise to create a safe data provisioning ecosystem.

Privitar's Research and Policy teams work with world leading academics, policy makers, regulators, and other experts to investigate how technology can help to preserve privacy while utilising data. Privacy is context specific. Privacy risk to individuals varies depending on factors including what data is being used, by whom and for what purposes. We work to understand the context and the role that technology can play in helping organisations to manage privacy risk.

We publish authoritative reports, including on differential privacy for the UK Government Statistical Service, on de-identification, and on anonymisation for the Anonymisation Decision-Making Framework (ADF). We advise and collaborate on a range of projects on privacy enhancing technologies including with RUSI and the Royal Society. We organise events, including In:Confidence (our flagship event for data professionals) and the Data Policy Network.

CONTENTS

Copyright 2021, Privitar LTD

Executive summary

The benefits of data sharing and the need to make use of them have never been greater. Sharing health data for research purposes saves lives. Advances in technology, from wearables to cloud computing, allow more high-quality data to be collected and analysed. However, this also raises significant issues around privacy and trust. The stakes are high, and the way forward is unclear.

Organisations holding health data are unsure what data they can - or should - share, when, with whom and under what conditions. This means that researchers can find it challenging to get timely access to the data they need. Regulators and standard setters have issued guidance, but recent changes and different approaches can make it challenging for organisations to apply the guidance in practice.

Policymakers recognise this challenge. The Secretary of State for Health and Social Care has launched a review, led by Dr Ben Goldacre, into the efficient and safe use of health data for research.

This report aims to provide a practical description of how organisations can share data responsibly and compliantly. It is informed by our work over several years with organisations sharing health data. During that time we have collaborated with Reform on their report on access to and use of NHS data, interacted with data leaders in healthcare organisations through our events, and held numerous informal discussions with researchers, data scientists, patients' groups and other health data stakeholders. We found that organisations would welcome detailed examples showing how they could translate legal and regulatory guidance into operational processes.

In response to that need, we launched a project to produce case studies documenting the data sharing process at leading organisations, with commentary and analysis. We believe that case studies can work alongside guidance and help organisations manage their own data sharing by providing worked examples.

This report accompanies the first two case studies in our series. Each case study is based on extensive interviews and document reviews with teams at the case study organisations: Cambridge University Hospitals Trust (CUH) and the Centre for Epidemiology Versus Arthritis (CfE) at the University of Manchester. Independent experts, including the Office of the National Data Guardian and the MRC Regulatory Support Centre" reviewed and provided feedback on the report.

We completed the case studies in 2020, before the end of the Brexit transition period. However, given the UK GDPR remains in line with the EU GDPR and that the European Commission has indicated that the UK will be considered 'adequate' for international data transfer purposes, our findings remain relevant after 1 January 2021.

This report is divided into five sections.

1. The Introduction describes the case study organisations and our methodology.
2. The Regulatory ecosystem section describes the legal and regulatory requirements and the key oversight and standards bodies in England relevant to sharing health data for research.
3. The Case studies section presents the case studies in the same format. Each case study documents the key decision points in the data-sharing process. For each decision point, we describe who is responsible for making what decision, the factors they consider, any technical or organisational controls applied, and legal considerations.
4. Our Summary of findings section sets out analysis and commentary on the case studies. It highlights similarities and best practice that could apply to other organisations.
5. Our conclusion, 'Where do we go from here?' sets out possible next steps.

CONTENTS

Copyright 2021, Privitar LTD

Our key findings include the following.

1. We found six common challenges:
    - managing competing priorities;
    - co-ordinating and carrying out the data sharing;
    - managing re-identification risk;
    - managing the risk of data misuse;
    - providing the right amount of information about the data; and
    - continuously improving the data sharing process.
2. While there were differences, both organisations followed fairly similar processes and applied a broad set of controls to:
    - the data (such as pseudonymisation);
    - the users (such as vetting and contracts);
    - the environment the data is accessed in (such as access controls); and
    - any research outputs (such as statistical disclosure controls).
3. Both organisations aimed to decrease the ‘time to data’ - the time it took from receiving a request to providing the data. We saw the importance of precedent-based judgments about the risk associated with a request to access data. This means using examples from previous, similar data sharing requests to speed up decision-making without increasing risk.
4. Anonymisation is a concern, as existing guidance is challenging to apply in practice. Specifically, it is not clear what factors organisations can consider when assessing if data is anonymous or not, nor how to evaluate the threshold for anonymisation.

We are grateful to the case study owners, reviewers and others who have given their time and insights to support this project. We welcome your thoughts on these case studies, and on if they are in line with or different from your experience. We are also open to carrying out further case studies as we build a library of best practice. You can contact us on policy@privitar.com.

CONTENTS

Copyright 2021, Privitar LTD

1. Introduction

1.1 Improving data-driven health research

The United Kingdom is in a good position to get the benefits from data-driven health research. The government's Life Sciences Industrial Strategy highlights the UK's large single payer system, population-level electronic patient records, and strong academic and research skills. The MRC notes that using that data responsibly leads to research that delivers more effective treatment, helps to identify and track public health risks, and improves service delivery.

However, the National Data Strategy cautions that: "used badly, data could harm people or communities, or have its overwhelming benefits overshadowed by public mistrust." Health data is particularly sensitive, significantly increasing the risks associated with irresponsible data use. Designing and implementing a clear and effective health data sharing process helps organisations to share data responsibly.

The data sharing process is part of an organisation's information governance framework which in turn reflects wider regulatory requirements. The high-level requirements for working with data are set out in terms of general principles (for example, the data minimisation principle in the GDPR), supported by guidance (for example, NHS codes of conduct) and overseen by regulatory bodies (for example, the ICO). Section 2 below looks at this in more detail.

An organisation's information governance framework is the internal system of rules and procedures that define how the organisation should handle data. The information governance framework makes sure the organisation complies with regulatory requirements, often by translating general principles into actionable processes. For example, a transparency principle may translate into an information governance requirement to publish summaries of data access requests. The information governance framework is influenced by more than just compliance. For example, it will take account of an organisation's risk appetite, priorities and available resources.[1]

The information governance framework includes the data sharing process. The two are closely related. For example, the information governance framework may include a process to get consent from research participants. Organisations may only be able to share data if it is within the participant's reasonable expectations, as set out in the consent form they completed. The data sharing process complies with the information governance framework, and takes account of broader considerations including scientific merit, ethics, public benefit or a desire for international collaboration. This report focuses on the data sharing processes, looking at broader information governance issues where those are relevant to understanding the data sharing process.

We believe that a good health data sharing process should comply with regulation and information governance rules, and allow fast, safe access to data for researchers. The National Data Strategy calls this "data availability." A process that can show it meets these requirements builds trust and confidence in data sharing and data use. We also believe that highlighting data sharing best practice will increase data availability and improve UK health research.

[1] We've chosen the term 'information governance' because it is generally understood to cover a range of issues including legal or compliance and ethical considerations as well as information security, audit and operational issues. Some organisations prefer the term 'data governance'.

CONTENTS

Copyright 2021, Privitar LTD

1.2 The case study organisations

This report presents case studies setting out the data sharing processes at two leading organisations: Cambridge University Hospital Trust (CUH) and the Centre for Epidemiology Versus Arthritis (CfE) at the University of Manchester. This section briefly introduces each organisation, highlighting similarities and differences between them. Each case study provides more background and details on the specific context in which each organisation operates.

CUH is one of the largest NHS Trusts in the UK. It delivers healthcare, including specialist treatment. CUH is also a government-designated biomedical research centre and a university teaching hospital. CUH accepts applications from researchers to access data in the Trust's electronic patient records. Data sharing supports research and teaching but is not CUH's main activity. The current formal data sharing arrangements have been in place for five years and CUH approves data sharing arrangements with roughly 60 research projects a year. CUH shares data with Trust or University-affiliated researchers under a framework agreement or with NHS-affiliated researchers under the NHS Research Passport scheme.

CfE is an academic research centre at the University of Manchester (UoM), focused on the epidemiology of arthritis and other musculoskeletal disorders.[2] CfE collects data directly from study participants in observational studies and randomised clinical trials, as well as analysing de-identified routinely collected health data, such as data provided by the Clinical Practice Research Datalink (CPRD). It works and shares data mainly with other academic institutions in the UK and internationally, with NHS Trusts, charity sector organisations, and industry partners. CfE data sharing is currently small scale; it approves roughly 25 requests a year. However, the UoM is developing a new, cloud-based data service that will provide a new way for researchers to access CfE-held data, which avoids sending data outside the organisation. CfE anticipates that its data sharing activities will increase once the new service starts.

Those brief descriptions highlight three key differences between the organisations. First, the scale of their data sharing activity. We focus on data sharing for research purposes, which leaves some of CUH's activity (for example, sharing data for clinical audit or service delivery purposes) out of scope. Second, the nature of their systems. CfE is exploring a cloud-based solution, while CUH relies on an on-premises IT system. Finally, the range of eligible researchers. CUH limits data sharing to NHS-affiliated researchers while CfE is open to sharing data with a wider range of organisations.

1.3 Why case studies?

This report builds on existing guidance. Feedback in the scoping stage suggested that information governance professionals find existing guidance useful but would welcome further support in the form of worked examples. This would help them to translate general principles in the guidance into specific, actionable processes in their organisations.

Case studies also allow us to consider each stakeholder's interaction with the decision-making process in relation to data sharing. Considering stakeholders in detail allows us to be more specific about the risks and benefits of data sharing and move from a high-level discussion towards answering the questions 'risks to whom?' and 'benefits for whom?'.

[2] Epidemiology is the study of how often diseases occur in different groups of people and why. Musculoskeletal disorders affect the joints, bones and muscles, and also include rarer autoimmune diseases and back pain.

CONTENTS

Copyright 2021, Privitar LTD

We believe that case studies will do the following.

1. Support organisations translating guidance and legal requirements into actionable processes. The case studies show how CUH and CfE have interpreted and applied guidance and best practice, using it to support their processes in their specific context. The differences between the two organisations allow us to look at a broad range of considerations. This makes sure this report is relevant to a wide audience.
2. Highlight common approaches and different solutions to common challenges. We found that the case study organisations face common challenges. In some cases, they have developed similar responses. Studying these responses provides practical examples that may be used in other contexts.
3. Provide a detailed, operational description of each organisation's approach to data sharing. This level of detail allows readers to compare or benchmark their own processes, helps to spread best practice, and may serve as a model for other organisations wanting to share data. Crucially, the case studies show how practitioners make judgments around risk: assessing different types of risk (for example, re-identification or data misuse), and balancing residual risk against allowing research to continue.

This report focuses on sharing health data for research purposes. We believe the findings will be relevant to data sharing challenges in other sectors.

1.4 Our methodology

The case studies describe each organisation's data sharing process in detail, focusing on how they manage the risks of sharing data for research purposes.[3] The research focus is deliberate; we exclude data sharing for other purposes (for example, clinical audit or direct patient care) because they operate under different regulatory and information governance requirements.

The case study organisations are based in England. This report does not cover variations in the legal and regulatory regimes for sharing data in the devolved administrations.

Each case study provides the following.

1. An overview of the organisation, including information on the context in which it operates.
2. A description of the roles and responsibilities involved in the data sharing process.
3. The key decision points in the data sharing process, including a summary of what decision is taken, who is involved, and the issues considered.
4. A flow diagram showing a simplified version of the data sharing process.

The case studies are based on extensive, semi-structured interviews with teams at the case study organisations. We carried out the interviews between March and September 2020. In some cases, the organisation shared internal documents to help us understand their processes. Each organisation reviewed and commented on the case study describing their process.

The fact that UK GDPR remains in line with EU GDPR and that the European Commission has indicated that the UK will be considered 'adequate' for international data transfer purposes both mean our findings remain relevant after 1 January 2021.

[3] We'll use 'manage' as shorthand for a comprehensive approach to risk including avoiding, transferring, mitigating, or accepting risk.

CONTENTS

We also worked with four reviewers, including the Office of the National Data Guardian and the MRC Regulatory Support Centre. We are grateful to the case study owners, interviewees and reviewers for the significant time and effort that they dedicated to this project.

We took steps to make sure the case studies were consistent. We chose to use one common set of terminology where the different organisations use different internal terminology or where individuals or departments with different names have a similar role.

The case studies all describe sharing tabular, row level data with recipients based mainly in the UK or European Economic Area (EEA). The geographic scope means that, in most cases, the organisations did not need to comply with the international data transfer requirements in the GDPR.

CONTENTS

Copyright 2021, Privitar LTD

2. Regulatory requirements for health data

Health data use must meet requirements from legislation and common law, policy and guidance, and oversight bodies. We use ‘regulatory requirements’ as a collective term to refer to all these. This section summarises some elements of the regulatory requirements in England; we recommend the joint British Academy and Royal Society data governance landscape review for more detail.

2.1 Legislation and common law

We can put legislation in two broad categories: data protection law in general, and specific legal requirements for processing health data in England. The General Data Protection Regulation (GDPR), implemented by the Data Protection Act 2018, sets out the general requirements for data processing. The GDPR was incorporated in UK law, becoming the UK GDPR. The rights, principles and obligations remain the same.

Common law is also relevant to health data sharing. It establishes the common law duty of confidentiality and provides a set of precedents in case law. Common law provides a way to disclose confidential patient information in circumstances where the disclosure would be in the substantial public interest. However, there are also ways to disclose data set out in law, as described below.

The GDPR includes data relating to an individuals’ health in its definition of ‘special category’ data, meaning that it needs additional protection. The GDPR prohibits the processing of special category data unless one of ten exemptions applies. The case study organisations rely on the exemptions in Article 9(2)(i) and (j) of the GDPR, allowing processing necessary for the public interest for public health or scientific research.

The National Health Service Act 2006 section 251(11) defines “confidential patient information” with a three part test. Data is confidential patient information if it:

1 is identifiable or likely to be identifiable;

2 was given in circumstances where the individual is owed an obligation of confidence; and

3 relates to the physical or mental health of an individual or is derived from that information.

The Health Service (Control of Patient Information) Regulations 2002, as amended by section 117 of the Care Act 2014, allows confidential patient information to be processed without consent for medical purposes, including medical research, in some circumstances and with approval. The Confidentiality Advisory Group (CAG) considers applications to process confidential patient information without consent for both research and non-research purposes. The CAG advises the Health Research Authority (HRA), which approves applications for research purposes, and the Secretary of State for Health and Social Care who approves applications for other purposes.

The Health Service (Control of Patient Information) Regulations 2002 also allow the Secretary of State for Health and Social Care to require organisations to process confidential patient information. The Secretary of State used this power in July 2020 to require organisations to process data for purposes relating to COVID-19. The power to require processing is used rarely and only in the context of responding to threats to public health (for example, from infectious diseases).

Taken together, these legal requirements set out effective protection for identifiable health data and provide a legal way to process this data, in an identifiable form if necessary, for health research purposes. Organisations rely on the policy and guidance documents described in the following section to help them understand and interpret the legal requirements.

CONTENTS

Copyright 2021, Privitar LTD

2.2 Policy and guidance

This section summarises the policy and guidance documents most frequently mentioned in interviews. This is not an exhaustive list. We have grouped guidance into three broad categories.

1. General guidance on data protection covering lawful bases for data use and risk assessments. This includes the ICO guidance on data protection impact assessments and on interpreting the GDPR. Some research methodologies may need to follow other types of general guidance, for example the ICO's guidance on AI and data protection, if researchers intend to use machine learning.
2. Guidance, best practice recommendations, and decision-making tools on anonymisation. This includes the Anonymisation Decision-Making Framework (ADF), the Five Safes, and the ICO's 2012 Anonymisation code of practice. An updated version of the ADF was published in October 2020, after we carried out our interviews. The ICO website states that work is ongoing to update the 2012 guidance. However, there is no target date for publishing the update and the ICO refers to the existing guidance as a "good starting point".
3. Guidance on using health data for research. This includes the NHS's Codes of Practice for handling information, NHS Digital's guidance on data security and information governance, the HRA's UK Policy Framework for Health and Social Care Research and guidance on using patient data without consent, and the Medical Research Council's policy and guidance on data sharing for researchers and Guidance Note 5 on identifiability, anonymisation and pseudonymisation.

Anonymisation

- The GDPR distinguishes between personal data and anonymous data. Personal data includes identified and identifiable or pseudonymised information.
- Anonymous information is not covered by data protection law.
- However, anonymisation can be challenging for three main reasons.
  - The legal and policy position is unclear, making it difficult to be certain that data is anonymous.
  - The unclear legal position and the fact that anonymisation varies depending on context makes it challenging to communicate to patients and the public, which can lead to confusion.
  - The controls required to achieve anonymisation can undermine the usefulness of the data for research purposes. The case studies show that organisations, broadly speaking, prefer environmental controls (for example, contract restrictions on data use) over making changes to the data (for example, perturbation or masking) to keep its usefulness.
- We set out the challenges for anonymisation policy in this paper.

CONTENTS

Copyright 2021, Privitar LTD

2.3 Oversight bodies

These organisations develop policy, issue guidance, and in some cases provide external checks and balances on data use. For example, researchers must get HRA approval for their proposed research in some circumstances. As with the policy and regulatory section above, some of these bodies focus on general requirements applying to all data use (for example, the ICO) and others focus specifically on using data in the health research context (for example, the HRA, MRC, and Caldicott Guardians).

1. The Information Commissioner's Office (ICO). The ICO is the UK's independent authority set up to uphold information rights in the public interest, promote openness by public bodies, and data privacy for individuals. It oversees compliance with a number of laws, including the UK GDPR and the Data Protection Act 2018.

2. The Health Research Authority (HRA). The HRA is an executive non-departmental public body of the Department for Health and Social Care (DHSC). It has a legal responsibility to provide guidance about health research. It also reviews health research proposals and provides recommendations on the processing of identifiable patient information for research and non-research projects. For data sharing proposals, HRA approval will likely be made up of three elements.

    - HRA's assessment of governance and legal compliance.

    - An independent ethical opinion from the Research Ethics Committee (REC). In many cases, REC review is a part of the overall HRA review process.

    - A recommendation from the Confidentiality Advisory Group (CAG). The CAG is an independent body, established by the Care Act 2014. It reviews requests to use confidential patient information without consent under section 251 of the NHS Act 2006 (we'll refer to this as s251 for brevity) on behalf of the HRA and the Secretary of State for Health and Social Care. The CAG advises the HRA, which makes the final decision to allow or deny access to data. It can be complicated to decide if CAG approval is needed or not. The CAG publishes guidance to help researchers decide if they need CAG approval.

3. The National Data Guardian (NDG). The NDG aims to build public trust that health data is protected and used appropriately. The role was set up in 2014 and put into law in 2018. The NDG acts as an independent champion for patients and the public, and issues guidance about processing health and social care data. The NDG asks an independent expert group, the NDG Panel, to advise and support its work.

4. The Caldicott Guardians. The UK Caldicott Guardian Council is a subset of the NDG Panel. The chair of the Council sits on the NDG Panel. The Council is the national body for the Caldicott Guardians. In 1997, Dame Fiona Caldicott's review on the use of patient identifiable data recommended a set of principles, which became known as the Caldicott principles. The principles are available on the NDG's website. The review also recommended that "a senior person, preferably a health professional, should be nominated in each health organisation to act as a guardian, responsible for safeguarding the confidentiality of patient information." These individuals are known as the Caldicott Guardians.

5. The Medical Research Council (MRC). The MRC is part of UK Research and Innovation, an independent body mainly funded by the Department for Business, Energy and Industrial Strategy (BEIS). The MRC invests in research on behalf of the UK taxpayer and has an important policy role in developing guidance for researchers on data use.

Copyright 2021, Privitar LTD

Other research funding bodies, such as the National Institute for Health Research (NIHR) and the Wellcome Trust, also play an important role in developing best practice for researchers using health data.

6 Patient groups and advocacy bodies. These groups play an important role in building public and patient trust in how data is used. Organisations such as Understanding Patient Data carry out research, support balanced media reporting, and develop policy insights. Others, such as the Association of Medical Research Charities (AMRC), develop guidance and help to share best practice.

The limited overview in this section shows that the regulatory requirements can be complicated. We chose to focus on the most commonly mentioned requirements and left out specific requirements governing genomic data, data about sexual and reproductive health, and protections from gender recognition law. There is a substantial amount of general guidance on the duties, rights and responsibilities of the stakeholders involved in health data sharing. The following section describes our key findings from the case studies, showing how the case study organisations translate the guidance into actionable processes.

CONTENTS

Copyright 2021, Privitar LTD

3. Summary of findings

This section summarises challenges and mitigations, and best practices that we observed, setting out similarities between the case studies. We initially intended to structure the case studies by grouping issues into legal, risk, and mitigations. However, we found that these neat boundaries did not reflect real world processes. Instead, we found that organisations address issues through roles and responsibilities, and processes.

Section 3.1 describes the challenges the organisations had when sharing data and sets out their approaches to overcome or mitigate the impact of the challenges. Sections 3.2 to 3.4 describe our view of best practice for sharing data. They focus on the:

- information governance framework and data sharing process (Section 3.2);
- roles and responsibilities involved (Section 3.3); and
- controls operating at a number of levels, which minimise risk, preserve the usefulness of data and are easy to understand and audit (Section 3.4).

Taken together, the three sections may provide a useful starting point for organisations developing their own data sharing processes.

We also found a lot of similarity between the data sharing processes. Broadly speaking, they consider similar issues and approach those in similar ways. This is unsurprising, as the case study organisations operate with the same regulatory requirements. The main differences relate to the different contexts in which they operate, as described in the introduction to each case study.

3.1 Challenges and mitigations

We have grouped the challenges into six categories:

1. managing competing priorities;
2. co-ordinating and carrying out the data sharing;
3. managing re-identification risk;
4. managing the risk of data misuse;
5. providing the right amount of information about the data; and
6. continuously improving the data sharing process.

The categories overlap or relate closely to each other. For example, the challenges around competing priorities and co-ordinating stakeholders both come from wide-ranging data sharing processes that looked at many aspects of the request to share data. For each category we describe the challenge, provide examples and set out mitigations we observed.

3.1.1 Managing competing priorities

We observed that organisations look at the privacy and data protection aspects of the proposed data sharing arrangement with a range of other factors (for example, scientific merit or resource constraints). This can lead to different views between stakeholders, making it difficult to find a way forward. The organisation has to consider decisions which, for example, advance one set of interests (for example, increasing privacy protection) but limit others (for example, reducing the usefulness of the data).

CONTENTS

Copyright 2021, Privitar LTD

Responses to the optimisation problem included the following.

- Clearly defining organisational priorities and risk appetite. In both cases the senior leadership teams defined organisational priorities to guide decisions on the best way forward.

3.1.2 Co-ordinating and carrying out the data sharing

This challenge comes from the fact that the data sharing process needs many internal and external stakeholders to work together, as well as stages in the process only happening when other stages are complete. It can be difficult for stakeholders, including the researcher requesting data, to follow the process and to understand their role at each stage.

This is distinct from the challenge described above. Any process involving many stakeholders can have co-ordination challenges, even where all of the stakeholders agree on the best way forward.

If they are not dealt with, co-ordination challenges will lead to longer ‘time to data’ (the period of time between a researcher making a request and receiving the requested data). It will also need more resources to make the data sharing process work. We found examples of projects where the time to data was measured in years.

The case study organisations took various steps to manage the co-ordination challenge, including the following.

- Clearly document the data sharing process. This improves internal co-ordination by making sure all the internal stakeholders in the process are able to follow it, understand their role, and what happens at each stage. Both organisations document their processes.
- Have a process to assess requests based on previous experience. Both organisations distinguished between ‘routine’ and ‘non-routine’ data sharing requests. Routine requests are those similar to ones reviewed in the past. These can be dealt with more quickly, freeing up resources to focus on riskier or more challenging requests. The case study organisations used different approaches. For example, CfE considers if the processing is high risk as described in the ICO and A29WP guidance. For projects that are potentially high risk the UoM requires the researcher to carry out a screening IG risk review.
- Make decisions without calling a meeting. This allows organisations to consider requests as they arrive rather than waiting for an intermittent committee to meet (for instance, once a month). It can be combined with the process for assessing requests where routine requests are assessed on paper and meetings kept for the most complicated requests.
- Avoid the need to ask the requester for more information. Checks early in the process to make sure only viable requests continue to the next stage reduces the number of requests for extra information from the requester or internal stakeholders. Both case studies include a validity check before the request is passed to the committee stage for consideration.

Copyright 2021, Privitar LTD

3.1.3 Managing re-identification risk

Health research aims to produce general conclusions about groups of individuals. It's about what a group has in common that is of interest, rather than what distinguishes them as individuals.

De-identification is a way to remove privacy risk while potentially keeping the information needed to spot shared trends. Both organisations face a common challenge in managing re-identification risk. Being too risk-averse can prevent innovation, while a high tolerance for risk may not be shared by all stakeholders (for example, patients) and could undermine trust in data sharing.

We can consider the challenges relating to re-identification risk from three perspectives. First, the need to balance the usefulness of the data with the need to protect individuals to avoid losing any key information unnecessarily. This balance will depend on the specific research in question. For example, in many research projects the researcher does not need more detail than the patient's age, but a perinatal study may need a date of birth.[4]

Second, the lawful basis for processing data under the GDPR. The exemptions for processing health data under Article 9 of the GDPR require "suitable and specific measures" to protect individuals. Both organisations rely on Article 9(i) or (j) of the GDPR to provide a lawful basis for processing health data in the public interest for public health or scientific research purposes.

Third, the legal status of the data. First, if a common law duty of confidentiality applies and second, if the data is anonymous and so not covered by the GDPR. The legal status of the data determines the regulatory requirements for processing it. Current guidance comprehensively sets out how to achieve anonymity in terms of which actions to take, but does not provide a rigorous methodology to determine when data is anonymous. This makes it difficult to know when enough has been done to cross the threshold into anonymity and difficult to communicate to stakeholders what has been done to achieve anonymity and why.

In some cases, the case study organisations rely on the concept of 'functional anonymisation'. This means the data may be considered anonymous for the recipient, given the environment they access the data in and the conditions on which they access it, but could be pseudonymous if it was accessed in a different environment and context.

For example, the CfE considers functional anonymisation on a case-by-case basis taking into account:

- the data to be shared;
- recipients;
- other data available to the recipient;
- the destination environment, including if statistical disclosure controls are applied to any outputs; and
- the use of agreements, including contracts.

[4] According to the WHO definition, the perinatal period starts at the 22nd week of pregnancy and ends seven days after birth.

Copyright 2021, Privitar LTD

Both organisations apply measures to address these challenges, including the following.

- Take a 'belt and braces' approach to anonymisation. We found that, in some instances, the organisations take steps to anonymise the data, but still treat the result as if it were personal. Applying controls carries a cost (for example, in terms of the usefulness of the data or time to data). We believe that a better and more practical approach to anonymisation from regulators would be helpful to make data sharing processes more efficient. Our recommendations on anonymisation policy set out the key challenges and suggest how to overcome them by improving regulation.
- Actively involve patients and the public to explain the risks and benefits. Both organisations have mechanisms in place to involve patients and the public in the data sharing process. This does not mean putting the responsibility on patients and the public through consent. It means making sure they are informed (for example, about what data is shared, with whom and for what purposes) and that their views are reflected in the decision-making process for data sharing. For example, CfE has a dedicated Patient and Public Involvement and Engagement (PPIE) group, representing the views of patients and the public on data sharing.
- Pseudonymise data and apply data minimisation. Both organisations remove direct identifiers and apply controls to pseudonymise data. (For example, contracts requiring data recipients not to carry out unauthorised linking of data.) Data minimisation means only sharing data that is adequate, relevant and necessary for the specific purposes. Both organisations involve requesters in decisions about controls where appropriate to get the balance right between pseudonymisation and the usefulness of the data. (For example, when the control involves reducing how detailed the data is.) CUH described an instance where a requester asked for dates of birth, but after a discussion agreed that they could carry out their proposed research using age bands instead.

3.1.4 Managing the risk of data misuse

Both organisations currently share data by providing an extract to the data recipient. This creates a challenge around managing the risk of data misuse by the recipient (for example, using the data for an unauthorised purpose or access by unauthorised individuals). Data misuse undermines trust in data sharing and causes reputational damage. The fact that health data is particularly sensitive increases this risk.

Both organisations take steps to address this challenge, including the following.

- Take account of the culture of data use in the requester's organisation when deciding whether to share data. For example, interviewees noted that the NHS and clinical settings have a strong culture of confidentiality and data protection, which may not exist to the same degree in other, non-clinical, settings. Both organisations consider the culture of data use when assessing the risk associated with a data sharing request. For example, CUH only shares data with researchers affiliated with the University or the NHS (via the NHS Research Passport scheme).
- Have contract restrictions or conditions on data use. Both organisations require some form of contract control on data use. For example, when CfE shares data with a researcher employed by the University of Manchester that researcher is bound by their employment contract and the University's policies on data use. Those policies include:
  - IT security requirements (for example, accessing data only on managed devices);
  - purpose limitation;

CONTENTS

Copyright 2021, Privitar LTD

  - only accessing the data from specific locations (for example, only when on campus or from within the UK); and
  - limits on data retention.

We provide more detail on common contract terms in Appendix C.

- Restrict data storage to specific data environments. Both organisations share the data extract with the recipient using an IT environment managed by the hospital Trust or the University. This supports auditing, as they can monitor the environment to confirm compliance with contract restrictions on data use.

3.1.5 Providing the right amount of information about the data

The organisations face a challenge in balancing the risks and benefits of providing information about the data to researchers (for example, how it is set out and what it contains). Providing detailed information about the data at an early stage in the process means the researcher does not need to spend the time and effort to make a formal data access request.

If the researcher has more information about the data, they can make more targeted requests for access, speeding up time to data and supporting data minimisation. On the other hand, providing detailed information about the data risks leaking information about the individuals in the dataset. Initiatives in the health sector, including the Health Data Research Innovation Gateway, play an important role in managing this challenge.

3.1.6 Continuously improving the data sharing process

Once the data sharing process is set up, organisations face a challenge around continuous improvement. We see two main elements to this challenge: (1) monitoring the process to make sure it works as expected and, if not, (2) deciding what is needed to improve it.

Monitoring the process also helps organisations demonstrate the benefits of data sharing. The specific challenge relates to developing performance indicators and collecting metrics which are relevant to different stakeholders. For example, demonstrating the benefits internally (to justify the resources allocated to data sharing activity) or to patients (to encourage them to take part in research).

Organisations also need to demonstrate the benefits. The National Data Guardian's 2016 review of data security, consent, and opt-outs concluded that the public "should be made aware of the use of their data and the benefits."

We recommend organisations focus on collecting metrics about the data sharing process itself. For example:

- the number of requests;
- the average time to data; and
- feedback from requesters and recipients (recognising that not all requests will result in data being shared) on the process.

These may all provide useful indicators about how the process is working and can act as indicators for the benefits of sharing data.

We saw creative approaches to this challenge, including:

- Develop proxy indicators, in other words indicators that act as the best alternative for something that is difficult to measure directly. For example, as CfE shares data mainly

Copyright 2021, Privitar LTD

with academic researchers, it could track research papers based on data it has provided. The number of published papers acts as a proxy measure of the usefulness of the data.

- Collect feedback from stakeholders. This could be a formal process, for example, asking stakeholders to fill in a feedback form.
- Keeping the data sharing process under active review. For example, both organisations described upcoming changes to their data sharing processes. In some cases these involved trialling new technical approaches, such as synthetic data.
- Providing feedback to stakeholders, including patients. For example, CfE is considering options for how to alert patients when data about them is used for health research purposes.

Longer time to data for researchers is a common theme in the challenges set out above. We've heard a number of anecdotal comments about the importance of timely access to data. Long delays in accessing data prevents innovation and is a factor in those working in data science leaving a project or organisation or moving onto other things. This shows the need for organisations to have efficient data sharing processes.

The following three sections build on the challenges and mitigations we set out above. They take our findings from the case studies and set out best practice for:

1. setting up an information governance framework and data sharing process;
2. defining the roles and responsibilities in that process; and
3. selecting and applying controls.

3.2 Best practice – Set up an information governance framework and a data sharing process

3.2.1 Set up an information governance framework

An organisation's information governance framework is the internal system of rules and procedures defining how the organisation should handle data. A clear, well documented framework helps to manage the challenge of co-ordinating and carrying out the data sharing identified in the preceding section. Setting up a comprehensive information governance framework includes a set of related tasks.

- Identifying applicable legal, regulatory, and policy requirements. In some cases, these vary depending on the data and the context of the proposed data sharing arrangement. For example, the common law duty of confidentiality applies to data relating to deceased individuals, but the GDPR does not. Where the GDPR applies, organisations must be clear about the lawful basis for sharing data. The most appropriate lawful basis is usually one of the public interest exemptions in Articles 9(2)(i) and (j) of the GDPR.
- Define the roles and responsibilities, including how they interact at different stages in the decision-making process. Document the terms of reference of any committee or board. Organisations may need to consider if a mix of internal and external roles and responsibilities are relevant. For example, CfE's process includes roles within CfE itself and within the UoM, CfE's parent organisation. Similarly, CUH's process includes NHS-wide roles and responsibilities. The roles and responsibilities should include independent advice and oversight. Organisations should consider giving roles to champion patient and / or public involvement.

Copyright 2021, Privitar LTD

- Define organisational policies and procedures to support data sharing and guide stakeholders through the process. This could include a data sharing policy, data management plan, creating data request forms, and mapping planned data flows.
- Describe the technical and organisational controls necessary to support data sharing. Technical controls may include cybersecurity requirements for the data processing environment (or setting up a secure environment), access controls, and technical measures to prevent data loss. Organisational controls may include staff training, background checks on data recipients or contract limits on data use, which could be included in a standard data sharing contract.
- Assess and document the risks associated with data sharing, for example, in the organisation's risk register. An organisation may rely on frameworks, such as ISO 27001 or data protection impact assessments (DPIAs) to guide their risk assessment. Defining roles and responsibilities will include deciding who owns specific risks. The risks an organisation is willing to accept may vary over time and may be difficult to measure or describe precisely. The decisions the organisation makes about individual requests for access to data will often reflect the organisation's risk appetite.

3.2.2 Prepare to share data

We have identified three core activities that organisations preparing to share data must carry out.

- Onboard data. This involves building a dataset suitable for sharing. This may be a subset of a 'live' dataset (for example, a copy of the electronic patient records data held by a hospital trust) or a dataset that includes linked data from different sources (for example, linking health data with socio-economic data from the ONS).[5] Deciding what data to make available involves understanding the data itself (for example, what characteristics of the data might be particularly sensitive).

  Data onboarding includes the data engineering work common to all data projects. This includes data cleaning and quality checks, for example, making sure records are complete, de-duplicated or properly formatted. It also includes managing links, such as decisions about what data sources to link together and how to be sure that two records actually relate to the same individual when there is no reliable common identifier, such as an NHS number.
- Onboard data requesters. Data requesters may be internal (for example, in-house researchers) or external (for example, research partners working for another institution). Onboarding includes vetting potential requesters to confirm they are eligible to access the data. Vetting may include looking at their CV, their previous relevant experience specific to the kind of data and project in question, where they are geographically, and what other data assets they can access. Vetting can be in house or outsourced, for example, by requiring the recipient to hold an accreditation validated by a third party such as the ONS Accredited Researcher service or through the NHS Research Passport.
- Help researchers to understand the data. This may include publishing information about the data, for example, its structure, number of records, time series, licensing terms and specific conditions on use (for example, commercial or non-commercial). Educating researchers can allow them to make more specific requests, supporting data minimisation and reducing time to data by avoiding unworkable requests.

[5] Not all data sharing models involve centralising data as described in this paragraph. For example, the Web Sciences Institute at the University of Southampton proposes a decentralised model in their Blueprint for a Social Data Foundation.

Copyright 2021, Privitar LTD

3.2.3 Manage data sharing

At this stage the organisation will accept data access requests, apply the data sharing process to evaluate requests, and decide whether to agree to the request and under what conditions. If the request is agreed, the data is shared with the requester.

- Collect information about the request. Requesters can use the information about the data the organisation has provided to inform their data access request. Organisations must aim to collect at the start of the process all the information about the request that they will need to make a decision.
- Apply checks so incomplete requests do not continue to the next stage. This reduces the need to check back with the researcher to clarify elements of their request. This is also an opportunity to provide information to the requester, for example, through explanations on an online form or guidance documents for paper forms. The guidance could include a worked example of a good request.
- Assess the request. Organisations can speed up the review process by identifying common, low-risk, or simple requests that don't need to go through the full review process. The CAG uses a similar, 'precedent pathway' approach.
- Case-by-case review. Projects considered as higher risk can go through a further review. This may include review by an internal board or by an external body such as the HRA. The review process can focus on different elements of the request. For example, a REC review of the ethical issues or a DPIA covering data protection issues.
- Select controls. This may be based on previous requests or following a case-by-case review. It may involve consulting the requester to make sure proposed controls balance risk reduction with the usefulness of the data needed for the proposed research project. The organisation will need to make a final decision on the data sharing request, including a decision to own the residual risk.
- Apply controls and provide the data. This may involve creating a data extract, linking datasets, and applying data transformations in line with the decisions made on controls. It will also include applying environmental controls (for example, contract restrictions on data use).

3.2.4 Audit and reporting

This includes reporting to internal and external stakeholders and audits of the data sharing process itself (for example, is it generating the 'right' decisions?) and of data recipients. Reporting should include publishing information on what data was shared, with whom and for what purposes to support building public trust in data sharing. An organisation could:

- Collect information to support reporting. The organisation can collect management information, for example, number of requests, time to data, and proxies for the usefulness of the data, including publications that refer to it.
- Audit compliance with environmental controls. The process must include ways to monitor compliance with these controls (for example, contract limits on data retention). This could include requiring the data recipient to evidence compliance (for example, through data deletion certificates).
- Apply statistical disclosure controls, to ensure that the recipients' outputs (for example, publications in academic journals) are not disclosive.

CONTENTS

Copyright 2021, Privitar LTD

3.3 Best practice - define roles and responsibilities

This section describes the roles and responsibilities needed to support the data sharing process. The specific names for these functions vary between organisations. In some cases, the roles may be carried out at different levels, for example, NHS-wide or at a local level of a specific hospital trust.

Role	Responsibilities
Senior Leadership Team	• Owns overall responsibility for data sharing. • Sets organisation's priorities and how important data sharing is for research in the organisation. • Sets the general risk appetite. • Decides whether to accept risk in specific cases (deals with exceptions).
Review Board, supported by a secretariat to vet requests.	• Brings together the roles listed below plus the Caldicott Guardian, patient representative, layperson, and clinician to assess the proposed research and data share request. • Agrees proposed data share and decides the controls to apply. • Escalates to the Senior Leadership Team when necessary.
Information Governance	• Provides the framework to comply with legal and regulatory requirements. • Advises on controls to manage re-identification risk and comply with legal and regulatory requirements.
Research Managers	• Assesses a research proposal's scientific value. • Assesses re-identification risk of the proposed data share. • Advises on controls based on the impact on the usefulness of the data. • Acts as a point of contact for the requester.
Legal	• Negotiates data sharing arrangements and contracts with recipients and data providers.
Data Engineers	• Handles database management. • Applies data transformations (including data minimisation). • Makes the final data extract available to the recipient.
IT and Cybersecurity	• Administers IT systems (for example, user authentication). • Monitors use (for example, logging access, queries). • Takes responsibility for cybersecurity (including assessing the security of the proposed recipient's IT environment). • Builds and maintains data infrastructure (this may be physical, for example, dedicated fibre connections). • Makes sure the set up complies with external standards (for example, NHS Information Security Management Code of Practice).
Research Governance	• Takes legal responsibility for research activity. • Manages the process for external approvals (for example, HRA). • Acts as a bridge between the process to approve data sharing and other processes relating to the proposed research (for example, NHS Research Passport, capacity and capability checks, and so on).

CONTENTS

Copyright 2021, Privitar LTD

Audit	• Vets the data recipient (as part of the user onboarding process described above). • Audits the recipient's compliance with contract controls on data use, for example, if they have deleted the data as described in the contract.
Output Checkers	• Applies controls to the research outputs before anything is published.

3.4 Best practice - controls

The controls most consistently used across the organisations are broad, easily explained, auditable, and manage the balance between the risk of re-identification and the usefulness of the data for the project's aim. Mitigating risk from several perspectives requires thinking about not just the data, the recipient, or the project, but all of these.

This theme of a wide-ranging approach to controls is found in many pieces of guidance which our interviewees mentioned as helpful (for example, the ADF, ICO and MRC guidance on anonymisation). In particular, we found that interviewees tended to use the Five Safes framework developed by the ONS, perhaps because it offers a simple, memorable set of issues to consider, as set out below.

1. Projects, meaning that the data is used for a valid purpose. Both organisations review the purpose to decide on its scientific merit and compatibility with ethics standards. They also review the proposed project methodology:

    - Both organisations have measures in place to verify that proposed research projects are ethical and in line with the organisation's goals. For example, by requiring the researcher to complete an ethics review. At CfE the type of ethics review depends on the nature of the project.

2. People, meaning that researchers can be trusted to use data appropriately and follow procedures. Both organisations carry out vetting to make sure only eligible researchers can request access to data. This can include checking that they have the right skills and qualifications, as part of existing vetting frameworks:

    - Vetting could be an internal process, for example, checking if they have done (and are up to date with) specific training courses, for example, confidentiality or data protection training.
    - Vetting can also be part of existing schemes. For example, CfE requires some researchers to be accredited by the ONS Research Accreditation Service and CUH relies on the NHS Research Passport scheme.
    - Data recipients are subject to contract controls specifying, for example, the purposes for which they can use data, how that data should be used, and the standards that their processing environment must meet (or, alternatively, a requirement to process data in specific, managed environments).

Copyright 2021, Privitar LTD

3 Settings, meaning controls on the research environment to prevent the unauthorised access to or removal of data. In both case studies, the researcher must process the data in a managed IT environment. However, in situations where the researcher uses data in their own environment, the organisation could impose standards or requirements for that environment through the data sharing contract. Controlling the research environment may include the following:

- Limiting what the recipient can bring into the research environment. Preventing the recipient from bringing data in helps to prevent unauthorised data linkage and can mitigate re-identification risk:
- Limiting how the recipient can use data, or monitoring how data is used in the environment. This may include allowing a limited set of queries or operations to be carried out on the data and / or monitoring data use to prevent and detect unauthorised activity.
- Specifying a limited data retention period and a way to verify data has been deleted once that period expires. This can include requesting deletion certificates if the data is not used in an environment managed by the organisation sharing the data.

4 Data, meaning that the data itself is non disclosive. Both organisations apply controls to reduce the re-identification risk. For example, both remove direct identifiers, apply data minimisation and use de-identification techniques (described in detail at Annex B). De-identification can be done at different stages in the data sharing process, as set out below:

- De-identification at the data onboarding stage. The organisation may decide that not all data is suitable for sharing or that it should be transformed before being onboarded. For example, CfE data is only made available for sharing if certain conditions are met.
- De-identification in response to the specific request. Both organisations consider the specific circumstances of the request to inform decisions about further controls.
- If the requester is able to preview the data, the information about the data available should also be non-disclosive. The UoM is building the ability to preview data into their cloud-based data access model.

5 Outputs, meaning that the statistical results produced do not contain any disclosive results:

- If the data recipient publishes statistics derived from the data as a part of their work. The CfE case study describes output checking and statistical disclosure control in detail.

Copyright 2021, Privitar LTD

Case studies

The following section presents the case studies, in the order that we carried out the interviews. We use the same structure for each case study:

- background on the organisation;
- a description of its operating context;
- the roles and responsibilities involved in data sharing; and
- the data sharing process itself.

In cases where the organisations use different internal terminology, we have used consistent terms across the case studies. We believe the added clarity for the reader outweighs the slight loss of accuracy. We define any jargon when it is first used, and in the glossary in Appendix A.

CONTENTS

Copyright 2021, Privitar LTD

4. Cambridge University Hospitals (CUH)

4.1 Background on CUH

CUH is one of the largest NHS trusts in the UK. It delivers healthcare through Addenbrooke's and The Rosie Hospitals. CUH is a leading specialist treatment centre, a government-designated biomedical research centre (one of five in the UK) and a university teaching hospital.

The CUH Research & Development department (R&D) oversees all biomedical research involving NHS patients, their data, or tissues that takes place on the CUH campus. Many of these research projects are led by academic partners on site, including University of Cambridge School of Clinical Medicine, the Medical Research Council, and Cancer Research UK.

CUH manages a dataset containing Electronic Patient Records (EPR) relating to roughly 3.8 million patients, updated every night. EPR data relates to patient care, including operational data (for example, admission date, waiting time or bed moves), clinical documentation (for example, medications, blood test results), and information relating to specific conditions (for example, cancer). The dataset includes free text data (such as notes from a clinician), which can be shared with a researcher if they have the necessary research governance and ethical approvals.

Data sharing arrangements have been in place for four years and R&D approves data sharing arrangements with roughly 60 research projects a year. The majority are one-off requests, where CUH shares a specific data extract with a recipient. A minority of data sharing arrangements are long-term, where a recipient receives an extract then updates at agreed intervals (for example, every month).

An overarching framework agreement covers research data sharing between CUH and the University of Cambridge School of Clinical Medicine. The framework agreement sets out a set of minimum standards to support the data sharing. Contracts with data recipients require them to comply with the framework agreement.

4.2 CUH's data sharing context

CUH operates in the context of NHS-wide information governance arrangements. Depending on the nature of their work, researchers may need approvals from external bodies, such as the Confidentiality Advisory Group (CAG), Health Research Authority (HRA) and / or Research Ethics Committee (REC). R&D supports researchers applying for external approvals through the Integrated Research Application System (IRAS).

CUH shares data for a number of purposes, including research, clinical audit, and service evaluation. The HRA has published guidance on the differences between these uses of NHS data, which carry different approval requirements. This case study focuses on data sharing for research purposes. The Cambridge Biomedical Research Centre lists research publications showing the range of research carried out on the CUH campus.

CUH routinely shares data for research projects led by researchers employed by CUH or by the University of Cambridge. A framework agreement governs data sharing between the two institutions. The researcher must have either a clinical honorary contract or apply for a Letter of Access.

CUH employees will generally apply for a clinical honorary contract at the same time as their employment contract. University-affiliated researchers apply for a Letter of Access through the Research Passport system covering the specific research project they wish to carry out. The majority (70 - 85%) of data access requests are from Trust or University staff, with the remaining (15 - 30%) of requests from requesters using the Research Passport system.

CONTENTS

Copyright 2021, Privitar LTD

4.3 Roles and responsibilities

This list only covers roles at CUH. Research requiring external approvals may also interact with NHS-wide research governance bodies, for example the CAG.

Role	Responsibilities and background
Information Governance (IG)	• Reviews and approves all studies involving patient data. • Makes sure personal information is managed legally, securely, efficiently, and effectively to appropriate ethical and quality standards. Relevant standards in a research context include the NHS Code of Practice on confidentiality, the GDPR and ISO 27001. A full list of IG standards is available on the CUH website.
Research & Development (R&D) Manager, part of the R&D Department	• Chairs the R&D Committee. • Responsible for overall risk management. • Supports negotiations, led by legal advisers, on contracts governing data use. • Assesses the trade-off between risk and benefits. This means working with the researcher to find a de-identified version of the dataset that will strike the right balance between reducing privacy risk and still being useful. For example, researchers carrying out a large study of admissions for people over 65 years old may agree to receive five-year age bands (for example, 65 - 70) and an upper age band of 'over XX years old.'
R&D Committee	• Considers legal, ethical, and re-identification risk and the benefits of the proposed research using the Fives Safes model to guide deliberations. • The committee is made up of the Caldicott Guardian, a layperson, a legal representative, a clinician, research governance representatives from the University of Cambridge Clinical School and CUH, and a staff representative from the local mental health trust (as an 'external' member). • Some committee members have experience sitting on NHS Research Ethics Committees.
R&D Co-ordinators, part of the R&D Department	• Assesses study protocols to determine which need external approvals or escalating to the R&D Manager. • Leads capability and capacity confirmation process, bringing together approvals from other parts of the trust as needed.
Clinical Informatics	• Creates the data extract reflecting the schema and data transformations (for example, applying age bands) approved by the R&D Committee. • Checks for residual re-identification risk in the data.
Technical Security and Information Architecture	• Responsible for the technical security of the Trust's architecture, including networks, servers, firewalls, and other equipment. • This role is outsourced to a commercial provider and overseen by a CUH liaison with technical expertise.
Legal	• Negotiates contracts governing data use.

CONTENTS

Copyright 2021, Privitar LTD

4.4 CUH’s data sharing process

4.41 Data onboarding

CUH maintains a dataset, separate to the live hospital system, for sharing purposes. The dataset was created from the hospital trust legacy system in 2014, as that legacy system was being shut down.

Roles and responsibilities

1. The dataset is updated through transfers every night from the hospital’s current Electronic Patient Record (EPR) system. Clinical Informatics is responsible for the automated update.

What controls are applied?

1. Records are checked for duplication.
2. Data minimisation. The most frequently requested fields from the EPR are transferred. More specialist content is not included unless specifically requested.

Legal considerations

1. The lawful basis for the processing is Article 9(i) or (j). The processing is necessary for scientific research or public health work in the public interest.
2. The Trust’s standard information notice provided to all patients sets out the fact that patient data collected during routine healthcare is used for research purposes.

4.4.2 Requester vetting

Option 1: The requester has an existing relationship with CUH, either as a Trust or University of Cambridge employee with an honorary clinical contract or honorary research contract. In this case the Principal Investigator (PI) leading the proposed study submits their CV to R&D.

Option 2: The requester does not have an existing relationship with CUH (for example, they are from another trust), so they submit a Research Passport application to R&D Human Resources department (R&D HR).

Roles and responsibilities

1. Option 1: R&D Co-ordinator flags any specific risks associated with the requester. For example, clinicians and requesters from the clinical school may be more familiar with NHS confidentiality requirements than requesters from other, non-clinical schools in the University.
2. Option 2: R&D HR department carries out checks to make sure the requester has appropriate qualifications and training and is appropriately vetted. R&D Human Resources issue a Letter of Access for a specific study.

What controls are applied?

1. Option 1: CV check to make sure the requester is suitably qualified to use CUH data. The existing relationship allows CUH to intervene in case of data misuse. CUH reviews the requester’s suitability to access data, including their Letter of Access, and may require the requester to take refresher Information Governance training.
2. Option 2: The checks minimise the risk of data misuse. The Letter of Access creates a way to legally enforce penalties for data misuse.

Legal considerations?

1. Legal negotiates a data sharing agreement if the requester is from another trust or part of a multi-site study involving CUH.

CONTENTS

Copyright 2021, Privitar LTD

4.4.3 Initial R&D review

The R&D Co-ordinator reviews and assesses research study protocols based on guidance from the R&D Manager. A study protocol is a full description of the proposed research. The assessment identifies protocols that need external approval. If necessary, R&D supports the project through the external HRA / REC approvals process through IRAS.

Roles and responsibilities

1 The R&D Co-ordinator completes a risk assessment informed by the following:

i Requester profile. If they are a clinician, academic or student, their level of seniority and experience, and if they have access to other patient data that may increase re-identification risk.

ii Nature of the patient group. Patient groups may be sensitive when the total number of records is low or when the study intends to investigate a rare condition affecting a small number of patients.

iii Nature of the data. Certain types of information are particularly sensitive (for example, genomic data, data relating to HIV or mental health status, reproductive health, or sexual development disorders).

Legal considerations?

1 Is the data Confidential Patient Information?

2 Is a data sharing contract needed?

4.4.4 External approval

Some studies require HRA, CAG, or REC approval. The HRA has published a decision tool to help researchers understand if they need approval.

Roles and responsibilities

1 R&D makes sure HRA and / or REC approvals are in place for the study, if applicable. Under the UK Policy Framework for Health and Social Care Research, all research studies must have a sponsor, who takes overall responsibility for the study. For studies sponsored by CUH or CUH / University of Cambridge, R&D supports researchers by confirming sponsorship and offering advice on preparing the documents to submit through the online portal (IRAS).

2 The HRA assesses the study against their approval and assessment criteria, available on the HRA website.

What controls are applied?

1 HRA review requires applications to state if the data is identifiable (including pseudonymous), or anonymous and to describe the measures applied to the data to justify that description.

2 The HRA guidance states that data anonymised by a third party (for example, NHS Digital) before being released to researchers is exempt from REC review, as long as there is a lawful basis for the anonymisation.

Legal considerations?

1 HRA includes an initial assessment of whether the study complies with a range of data protection, patient confidentiality, and information security laws and standards. For example, Human Rights Act 1998, Data Protection Act 2018, NHS Act 2006 and NHS Codes of Practice on information governance, confidentiality, information security management, and records management.

CONTENTS

Copyright 2021, Privitar LTD

4.4.5 R&D Committee review

The committee considers all research protocols requesting access to routinely collected data. The committee also provides ethical and research governance oversight for sponsored studies and protocols that do not need an external review.

Roles and responsibilities

1 The R&D Committee reviews the protocol and recommends mitigations proportionate to the risk. In assessing the risk, the Committee considers the following.

i ICO guidance on anonymisation, the Five Safes model, and the Anonymisation Decision-making Framework (ADF).

ii The culture of the requester's organisation. NHS clinical settings have a strong, ingrained culture of confidentiality. Other organisations in the CUH family, for example, non-medical schools within the University, may not have the same culture.

ii Motivation to attack the data. If any features of the dataset relate to an intruder's motivation, for example, a journalist interested in a trending issue.

What controls are applied?

1 Common data controls such as putting data into bands, small count suppression, truncating postcodes, and record swapping. See Annex B for more details on common data controls.

2 Managing outliers. These are values that differ significantly from other values in the data and can carry re-identification risk. Outliers can be redacted or removed. The requester is involved in decisions on how to treat outliers to make sure the data is still useful for their purposes.

3 Redacting specific records. Records relating to high profile individuals (for example, individuals with a public profile such as politicians or celebrities) may be removed or redacted. This can reduce the motivation for an intruder to attack the data.

4 Substitution. Working with the requester to identify options for substitution, for example, replacing postcodes with values from a social deprivation index.

5 Contract controls, including purpose limitation, re-identification ban, ban on onward sharing, storage according to agreed security standards, and evidence of destruction.

4.4.6 Information Governance review

IG checks compliance with the DPA 2018 and confirms that the common law duty of confidentiality does not apply, either by getting explicit patient consent, through s251 support, or by applying controls so the data is no longer identifiable.

Roles and responsibilities

1 IG receives IRAS referrals from R&D and signs off to confirm compliance with the DPA 2018.

What controls are applied?

1 IG assess the protocol based on:

i the type of data being requested;

ii who will have access;

iii the type of processing;

iv storage arrangements; and

v the lawful basis for processing, whether consent, s251, or the datais anonymous.

CONTENTS

Copyright 2021, Privitar LTD

Legal considerations?

1 In some cases, the requester may already have ethical approval including explicit consent for their proposed study, or s251 support.

4.4.7 Confirm capability and capacity

The R&D Co-ordinator confirms capability and capacity. The capability and capacity check makes sure stakeholders outside the data sharing process (for example, Clinical Directors at the Trust) are happy for the project to go ahead and that CUH has the resources and space to support the research.

4.4.8 Data sharing

The data extract is created and shared with the recipient through a secure IT environment.

Roles and responsibilities

1 Clinical Informatics creates the extract and implements technical controls, including pseudonymisation, in line with the schema agreed by the R&D Committee.

2 Technical Security makes sure the transfer is secure. They work closely with the University and the Trust IT security teams to allow data to be transferred to a secure area on a University or Trust server, only accessible to the approved recipient.

What controls are applied?

1 Data is shared directly to a University or Trust server. Transfers to the University use a secure virtual private network (VPN) link. Transfers to the hospital Trust involve extracting the data from the data warehouse into a network drive accessible by the Principal Investigator named in the research protocol. CUH applies a number of cybersecurity and data loss prevention (DLP) tools to protect the data.

2 Data is de-identified using an open source application (Open Pseudonymiser) from NHS Digital. Pseudonyms are study-specific to manage the risk of data linkage across studies.

3 NHS Digital and CUH Auditors audit the infrastructure to support data sharing (servers, network, databases, and so on), including getting penetration testing from an external provider. They outsource some elements of network security to a third party who submits monthly security reports.

Legal considerations?

1 CUH considers the data anonymous or pseudonymous in the hands of the recipient and it is protected by an appropriate data sharing agreement. In many cases, the combination of changes made to the data and environmental controls has reduced re-identification risk to the point where the data is anonymous from the recipient's perspective.

CONTENTS

4.4.9 Audit and monitoring

Roles and responsibilities

1 The recipient organisation makes sure the research complies with any requirements or limits on data use set out in the data sharing agreement.

What controls are applied?

1 Recipients within the Trust or University access the data in a managed IT environment. CUH can use technical means to check deletion and manage access.

2 CUH is looking at options for secure deletion certificates for external recipients.

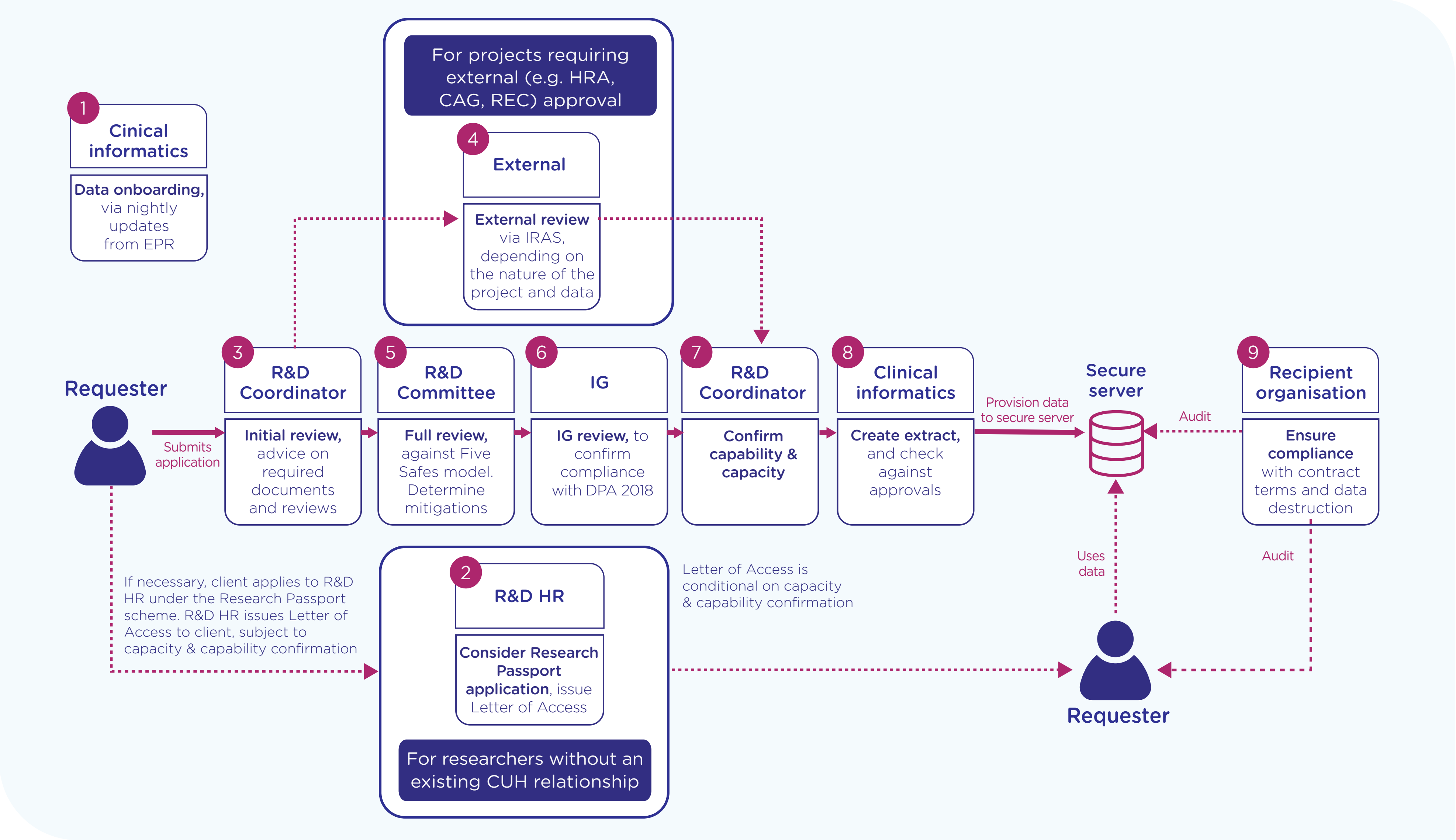
Legal considerations?

1 The fact that the recipient has a contractual link with CUH (for example, a Trust employee) means the Trust has contract and HR options for imposing sanctions for data misuse or non-compliance with contract terms (for example, suspending the researcher's access to data).

CONTENTS

Copyright 2021, Privitar LTD

CUH simplified process diagram



For projects requiring external (e.g. HRA, CAG, REC) approval

4 External

External review via IRAS, depending on the nature of the project and data

1 Cinical informatics

Data onboarding, via nightly updates from EPR

Requester

Submits application

3 R&D Coordinator

Initial review, advice on required documents and reviews

5 R&D Committee

Full review, against Five Safes model. Determine mitigations

6 IG

IG review, to confirm compliance with DPA 2018

7 R&D Coordinator

Confirm capability & capacity

8 Clinical informatics

Create extract, and check against approvals

Provision data to secure server

Secure server

Audit

9 Recipient organisation

Ensure compliance with contract terms and data destruction

If necessary, client applies to R&D HR under the Research Passport scheme. R&D HR issues Letter of Access to client, subject to capacity & capability confirmation

2 R&D HR

Consider Research Passport application, issue Letter of Access

For researchers without an existing CUH relationship

Letter of Access is conditional on capacity & capability confirmation

Uses data

Audit

Requester

CONTENTS

Copyright 2021, Privitar LTD

5. Centre for Epidemiology Versus Arthritis (CfE)

5.1 Background on CFE

CfE is an academic research centre of excellence focused on the epidemiology of arthritis and other musculoskeletal disorders. The CfE has a long history spanning over 65 years. Funding is every five years, and the CfE was most recently funded as a Versus Arthritis centre of excellence in 2018. The CfE is part of the Centre for Musculoskeletal Research at The University of Manchester (UoM).

CfE conducts research in two clinical research areas: (1) how often disease occurs and how it progresses, and (2) the effectiveness and safety of treatment. These are supported by three cross-cutting themes: (1) using digital data, (2) biostatistics[6], and (3) research into practice. The research includes traditional population health studies, clinical trials and long-term registers (for example, the Norfolk Arthritis Register, the British Society for Rheumatology Biologics Register - RA) and digital health studies (for example, Cloudy with a Chance of Pain).

CfE research is developed by a scientific advisory board and, in each theme, a separate Research Advisory Group guides research priorities and helps to publish important findings. CfE also involves patients and members of the public through the research user group made up of patients, carers, and people with an interest in musculoskeletal health.

CfE works mainly with other academic institutions in the UK and internationally and with NHS Trusts, charity-sector organisations and industry partners. As well as Versus Arthritis funding, CfE receives support from the UoM and research funding from bodies such as the Medical Research Council, British Society for Rheumatology, the Nuffield Foundation and the NIHR. More detail on partnerships is available on the CfE website.

Versus Arthritis funds a core infrastructure team in CfE to support its research activities and management. The infrastructure team works across the entire data pipeline and includes roles in data management, data science, information governance and communication and engagement.

CfE works with data in several ways:

1. collecting data directly from study participants in observational studies and randomised clinical trials (through interviews, surveys or digital apps for smartphones and wearable devices like smartwatches);
2. linking data collected with other sources, such as environmental data; and
3. reusing linked data from primary and secondary care settings

5.2 CfE's data sharing context

CfE provides access to data collected or generated through its primary research. This data is collected from study participants, including clinicians and patients. For clarity, this case study will refer to 'primary research' and 'onward sharing.' Primary research refers to the study which collects or generates data from participants. Onward sharing refers to the process by which a researcher can request access to data generated by a primary study for use in their own research.

[6] The development and application of statistical methods to a wide range of topics in biology. It includes designing biological experiments, collecting and analysing data from those experiments and interpreting the results.

CONTENTS

Copyright 2021, Privitar LTD

Unlike the CUH case study, we will cover the research governance process applied to primary research in detail. Carrying out primary research is how CfE creates data. Controls applied at the primary research stage (for example, getting participant consent for onward sharing) affect whether the data is available for onward sharing.

The Principal Investigator (PI) responsible for the primary research will complete a Data Management Plan (DMP) as part of the research governance process. The DMP, along with other key study documents (including the privacy notice, consent forms, and Participant Information Sheet) inform whether and under what conditions data generated by the primary study is available for onward sharing.

For example, Professor Will Dixon's Cloudy with a Chance of Pain study collected data from 13,000 participants through a smartphone app. The DMP included onward sharing. This was included in the ethics application and, crucially, in the study's consent form and Participant Information Sheet (PIS).

The CfE shares data with:

1. internal researchers already affiliated with the UoM (staff and students, visiting scholars under an honorary contract); or
2. external research collaborators.

The data sharing process varies depending on whether the data recipient is internal or external.

As a part of the UoM, the CfE has to comply with University-wide policies making up the University's information governance framework. The University sets baseline requirements for cyber and information security, legal, ethical, and information governance. CfE works with this policy framework and puts in place additional CfE-specific data sharing requirements.

Examples of CfE-specific documents and processes include:

- registers of data assets and data shared;
- a set of CfE templates (for DPIAs, assessing a data share, and associated action plan); and
- best practice guidance for researchers.

CfE's data sharing activity is on a small scale, processing approximately twenty-five data sharing requests a year. However, the UoM is developing a cloud-based, highly restricted data service, which will be called the 'Data Safe Haven Plus' (DSH+). When complete, it will provide another way for researchers to access CfE data.

The DSH+ will allow recipients to access data in a managed environment, providing an additional control on data use. It will also create a way to report on reuse of data to research participants and the wider community as requesters (under this model) have to provide a plain English summary of their research for the CfE website. This case study will describe both the Centre's current model and the proposed DSH+ governance model.

CONTENTS

5.3 Roles and responsibilities

The list of roles and responsibilities reflects the fact that both the UoM and CfE are involved in the data sharing process.

Role	Responsibilities and background
Principal Investigator (PI)	• Leads the primary research project and is responsible for whether and how the data generated can be shared. • Within the CfE, the PI may lead a relatively large team made up of researchers, PPIE partners, a project manager and members of the infrastructure group. They support data management, IG and communication support, as well as external collaborators
CfE roles	
Research Information Governance Manager (RIGM) and Research Information Governance Office (RIGO)Team	• Usually the first point of contact for researchers wishing to carry out a study or share data. • For current data sharing requests, the RIGM carries out a 'data to be shared' assessment and action plan. The assessment (1) gathers information and (2) determines the best approach for sharing data based on the information gathered in step (1). • Consults stakeholders within the CfE (the data science team) and UoM (the centralised roles listed below) to allow the data to be shared. • For sharing data through the DSH+ model, the RIGM and RIGO will manage CfE's (Data Sharing Review Board) Master File - a management system to comply with University and regulatory requirements. CfE's processes for allowing access to its data are in this file and reported every quarter. The Data Safe Haven Operations Group and the Research Compliance Committee review the quarterly compliance reports.
Data Scientists and the Infrastructure Team	• Works with the RIGM and RIGO on data collection, preparation, quality, discovery, management, and transfer. • Leads output checking in line with best practice on statistical disclosure control.
Communication Manager	• Promotes the CfE's research work and raises awareness of data sharing opportunities.
The Data Sharing Review Board (DSRB)	• Set up by the CfE to provide fair, transparent access to health data while making sure the data is properly safeguarded. • Oversees the approvals process for access to data for which the CfE is the business owner. Evaluates data access requests on the basis of data protection impact, ethics, scientific merit, and public benefit. • Works with CfE PIs to identify data suitable for sharing through the DSRB.
UoM central services	
Library Service	• Provides a research data management service, including a tool for data management planning (DMPonline), workshops, advice, and tutorials for researchers completing a DMP.
The Faculty of Biology, Medicine and Health (FBMH)	• Reviews all research studies involving data obtained in or through the NHS to be submitted for NHS REC and HRA approval.

CONTENTS

Copyright 2021, Privitar LTD

Research Governance Team	• Provides sponsorship review of the document pack to be submitted through IRAS. The study's PI and the Research Governance team both sign off submissions to the NHS REC and HRA.
Contracts Office	• Negotiates the data sharing agreement between the UoM and data recipient. This only applies to 'external' recipients. • Responsible for the standard templates.
Information Governance Office (IGO)	• Checks processing activity (including health research) for compliance with the data protection regime (including whether proposed data sharing agreements are appropriate), effective records management (for example, retention, deletion, and storage requirements) and to identify whether there are further Information Governance requirements. • Research involving health data obtained in or through the NHS does not usually need both FBMH Research Governance and Information Governance approval, because the ethical approvals process considers data protection. • Oversees a research information governance risk review (IGRR) for processing considered high risk and / or new.
Research IT	• Responsible for designing, setting up and delivering the DSH+. • Provides access to research data storage, including the DSH+. • Implements some environmental controls (such as identity verification and geographic restrictions on access) and some disclosure controls.
IT security	• Manages cybersecurity on all platforms across the University including the DSH+.
Research Governance, Ethics and Integrity (RGETI)	• Responsible for research ethics and governance. • Researchers using primary data or secondary data (from an onward share) must consider if their proposed study needs ethical approval. That approval can be at national (for example, the HRA) or University level. Both the HRA and UoM publish decision tools to help researchers identify if an ethical review is necessary and which route is most appropriate.
Research Compliance Committee (RCC)	• Sets standards and makes sure the University meets its obligations to comply with statutory, regulatory and policy requirements.
Data Safe Haven Plus Operations Group	• Made up of representatives from RGETI, IGO, Research IT, IT Security, Ethics and academic champions who work together to make sure the DSH meets its operational requirements. • Considers quarterly compliance reports related to Master Files and reports to the RCC.

CONTENTS

Copyright 2021, Privitar LTD

5.4 CfE's data sharing process

5.4.1 Data onboarding

This section describes some of the information governance considerations related to setting up a primary research study and to storing data from primary studies within CfE. All primary research must have a Data Management Plan (DMP). The DMP describes the nature and purpose of the processing, including how to collect, store (including security requirements), and manage (backup, format, and deletion) data. The DMP also captures legal, data protection and ethical issues and plans for publication and onward sharing.

Roles and responsibilities

1. When designing the primary research, the PI works with the RIGM and RGIO and study team to decide the process for recruitment, data capture, storage, analysis, transfer, publication, and onward sharing. This information is recorded in the study's DPIA (using the CfE's DPIA template) and Data Management Plan (DMP).
2. The University's Research Governance, Ethics and Integrity team provides standard templates for consent and the Participant Information Sheet (PIS), which researchers can adapt in line with their research plans. The PIS and consent documents help decide if data can be available for onward sharing. The relevant questions for onward sharing are:
    1. what did participants consent to?
    2. what are participants' reasonable expectations about onward sharing in terms of the purpose of reuse and the type of recipient?
3. The Information Governance Office (IGO) provides a standard privacy notice, used with the consent and PIS templates. The privacy notice includes information required by data protection law. The standard privacy notice is enough for the vast majority of research projects. In a small minority of projects, for instance where a research collaborator requires specific information to be included or where the data will be used for other purposes (for example, teaching), the privacy notice can be adapted.

What controls are applied?

1. If the IGO considered the primary research 'high risk' and / or new processing, that research went through an Information Governance Risk Review (IGRR). For example, processing may be considered new if it involves previously unapproved technology and high risk if it involves international transfers or sharing with private-sector organisations. The IGO provides guidance on the types of processing considered 'high risk.' The DMP and the CfE's DPIA, which considers the data flow in detail, identifies any areas of potential risk and specific mitigations.
2. If the primary research was carried out in or through the NHS, it went through HRA and / or NHS REC approval. The University's FBMH Governance team reviews and provides feedback on all applications to the HRA before submitting them through IRAS. If NHS REC was not required, the research went through a UoM Ethics approval or an assessment to decide that ethical approval was not required, leading to an ethics exemption certificate.
3. Participants consent to take part in primary research studies. This consent is for ethical purposes; it does not provide the lawful basis for processing under the GDPR and the DPA 2018. This consent, and the PIS, include the intention to make the data available for onward sharing.

Copyright 2021, Privitar LTD

Legal considerations?

1 The University is a data controller for data collected during a primary research study if the PI for that study is a University employee (so the University decides the purpose and means of the processing). A student's supervisor takes responsibility for a student's research. The University would also usually be a data controller in cases where an honorary employee decides the purpose and means of the processing. Data controller responsibilities can be held solely by the University or jointly with a research partner.

2 The University processes personal data for research using the 'public interest task' lawful basis in Article 6(1)(e) GDPR. If that data is 'special category,' it is processed on the basis that it is necessary for research or public health, relying on the exemptions in Article 9(2)(j) and (i) GDPR and the conditions and safeguards in Schedule 1, Part 1 of the DPA 2018.

3 If the primary research involves a partner organisation, both parties must agree to a contract before starting the research. That contract sets out the conditions for onward data sharing. It also covers the nature of the partnership, roles and responsibilities, intellectual property, publication, data confidentiality, and data processing arrangements.

5.4.2 Requester vetting

The request for access to CfE data is most commonly made through the PI for the primary research. The PI submits the request to the RIGM and RIGO for consideration. The request triggers a two-part process:

1 the RIGM and RIGO gather information about the intended data recipient; and

2 review the request.

To maintain consistency with the other case studies, we've split this into two stages. We describe the first stage as 'requester vetting' and the second as an 'internal review.'

Roles and responsibilities

The RIGM and RIGO, working with the PI for the primary study, decide the relationship between the CfE and the requester and if the requester is 'internal' or 'external.' The RIGM and RIGO provide guidance to requesters on the process and the requirements that they must meet.

Internal recipients must:

1 be a UoM member of staff, student, or visiting scholar on a UoM honorary contract;

2 go through the UoM ethics decision tool to decide if their proposed use of the data needs ethical oversight;

3 complete a DMP and, if directed by the IGO, the University's research IGRR if their proposed processing is considered high risk; and

4 provide any other documents associated with the proposed research that are needed to support a request for access to data.

External recipients must:

1 get and evidence appropriate approvals, such as a local ethics review or DPIA; and

2 provide a study protocol and / or other documents associated with the proposed research that are needed to support a request for access to data.

Copyright 2021, Privitar LTD

What controls are applied?

The mitigations at this stage support the ‘safe people’ strand of the Five Safes model.

Internal recipients are:

1 bound by their employment contract with the University and up to date with UoM training on information security and data protection, export controls, and ethics;

2 required to comply with UoM policies, procedures, and technical security standards such as Standard Operating Procedures (SOPs) on Information Security Classification, Ownership and Secure Information Handling, Record Retention Schedule, and Acceptable Use Policy;

3 required to work on UoM managed devices or follow the UoM SOP on Bring Your Own Device (BYOD) and remote working; and

4 required to use two-factor authentication to access central UoM services.

External recipients are:

1 most commonly vouched for by the PI (who may be working on the research with the recipient or has previously worked with them);

2 required to confirm their identity (their institution email address is also used to verify their identity); and

3 required to confirm and (where possible) provide evidence they have got any ethical or research governance approvals that may be needed to access the requested data. The ‘data to be shared’ assessment documents these approvals.

Legal considerations?

1 Data sharing with ‘external’ recipients is governed by a data sharing agreement (DSA). The type of DSA depends on whether the ‘external’ recipient is a data controller and if the data is personal data or anonymous in the recipients’ hands.

5.4.3 Onward sharing data review

The onward data sharing review covers the safe data, safe projects, and safe settings aspects of the Five Safes model. These aspects are particularly important to decide if the data is pseudonymised personal data or functionally anonymised in the recipient’s hands.

The RIGM and RIGO work with the PI, who wants to make research data available for onward sharing, and the potential data recipient to complete the ‘data to be shared assessment and action plan’ document (safe data). The RIGM and RIGO review the consent template and PIS from the primary research, the potential recipient’s study protocol, and / or any other relevant study documents (safe project). The RIGM and RIGO consult the Contracts Office and IGO when a DSA is required (safe data).

Roles and responsibilities

1 The RIGM’s review for onward sharing considers whether:

i the data controller(s) authorise the onward sharing;

ii the proposed onward sharing is consistent with participants’ consent preferences and reasonable expectations;

iii the data requested is pseudonymised or functionally anonymised for the recipient in the proposed destination environment (the data is classified in relation to the environment in which it exists);

CONTENTS

Copyright 2021, Privitar LTD

iv the data requested is necessary and proportionate for the proposed purpose;

v the requester had got any necessary approvals, such as ethics;

vi a DSA is needed and if so what type.

2 If the recipient is internal, the RIGM and RIGO may (where appropriate) consult with the UoM central services to complete the 'data to be shared' action plan, which sets out the conditions of the data sharing.

3 If the recipient is external, the RIGM and RIGO consult UoM central services to complete the 'data to be shared' action plan, which sets out the conditions for the data sharing. These conditions are included in the DSA between the recipient and the UoM.

What controls are applied?

1 Data sharing must be consistent with participants' reasonable expectations and consent preferences (usually evidenced by the Participant Information Sheet, consent form, and privacy notice, but may also be evidenced in any public-facing documents provided to participants when the data was collected).

For example, participants may have consented to the onward sharing of anonymised or pseudonymised data. There may also be restrictions on the types of recipients to whom the data can be shared, for example, other academic researchers or broader research groups approved by the PI.

CfE applies data transformations and environmental controls to make sure data is either functionally anonymised or, where functional anonymisation is not possible and the data remains identifiable, represents a low risk of disclosure. All onward sharing aims to provide highly useful data with a low risk of disclosure.

2 As well as removing direct identifiers, changes to data may include limiting or making less detailed key values in the data, such as location, or dates (including dates of birth).

Legal considerations?

1 The CfE decides if the data to be shared is pseudonymised or functionally anonymised from the recipient's perspective on a case-by-case basis. The ADF guides this decision. The factors taken into account to make the decision include:

i what other data exists in the recipient's environment;

ii who will have access to the shared data;

iii how access is governed;

iv how the data will be managed including plans to publish statistical results; and

v what infrastructure arrangements support data management, including storage, use, retention, and destruction.

2 Anonymisation is a form of processing requiring a lawful basis. CfE relies on the same lawful basis used for primary research. As stated above, the University processes personal data for research using the 'public interest task' legal basis in Article 6(1)(e) GDPR. If that data is 'special category,' it is processed on the basis that it is necessary for research or public health, relying on the exemptions in Article 9(2)(j) and (i) GDPR and the conditions and safeguards in Schedule 1, Part 1 of the DPA 2018.

Copyright 2021, Privitar LTD

5.4.4 Negotiate data sharing agreement

For external recipients, the data sharing agreement is the contract. The UoM Contracts Office leads the negotiation with input from the IGO. If the data is being transferred outside of the UoM infrastructure, the agreement includes the necessary safeguards.

Roles and responsibilities

1. The Contracts Office generate standard templates based on the nature of the sharing arrangement (for example, controller-to-controller or controller-to-processor) and the type of data to be collected or shared. The standard template includes defining the parties involved, their relationship and responsibilities, intellectual property, publication rights, confidentiality, data processing arrangements and so on.
2. The PI and the RIGM, in discussion with the Contracts Manager and with advice from the IGO, decide which type of sharing agreement is most appropriate.

What controls are applied?

CfE adds details of any necessary restrictions and controls on the recipient's environment to the standard template to cover:

1. the nature and purpose of the processing (for example, limits on reuse); and
2. the environment in which the data is processed (for example, data flow, method of transfer, data storage, access, retention).

Legal considerations

1. Where the data to be shared is considered identifiable for the recipient and the recipient is in a country outside the EEA which does not have an 'adequacy' agreement with the European Commission, the data sharing agreement includes standard contractual clauses and other safeguards.

5.4.5 Data sharing

This includes preparing the data extract and sharing it with the data recipient. The data extract in the hands of CfE (the data provider) is most commonly pseudonymised personal data. Once shared with the recipient, that data extract may be classed as either functionally anonymised or pseudonymised personal data. This classification is decided on a case-by-case basis and depends on the data shared and the share environment.

Roles and responsibilities

1. The PI, RIGM and RIGO work with the Centre's statisticians and data scientists to prepare the data to be shared.

What controls are applied?

1. University policy requires internal recipients to store research data either in the UoM's Research Data Storage Service or the Data Safe Haven (for highly restricted data). The data is provisioned directly to the recipient's space in the appropriate service.
2. Internal recipients are required to follow University policies, procedures, and technical standards and are given the Centre's specific SOPs, which provide detailed guidance on working with data. The SOPs require the recipient:

   i not to move, copy, or download data from the managed storage service;

   ii make sure no one can look at their computer screen if in a shared office;

Copyright 2021, Privitar LTD

   iii keep written notes associated with their work in a locked drawer;

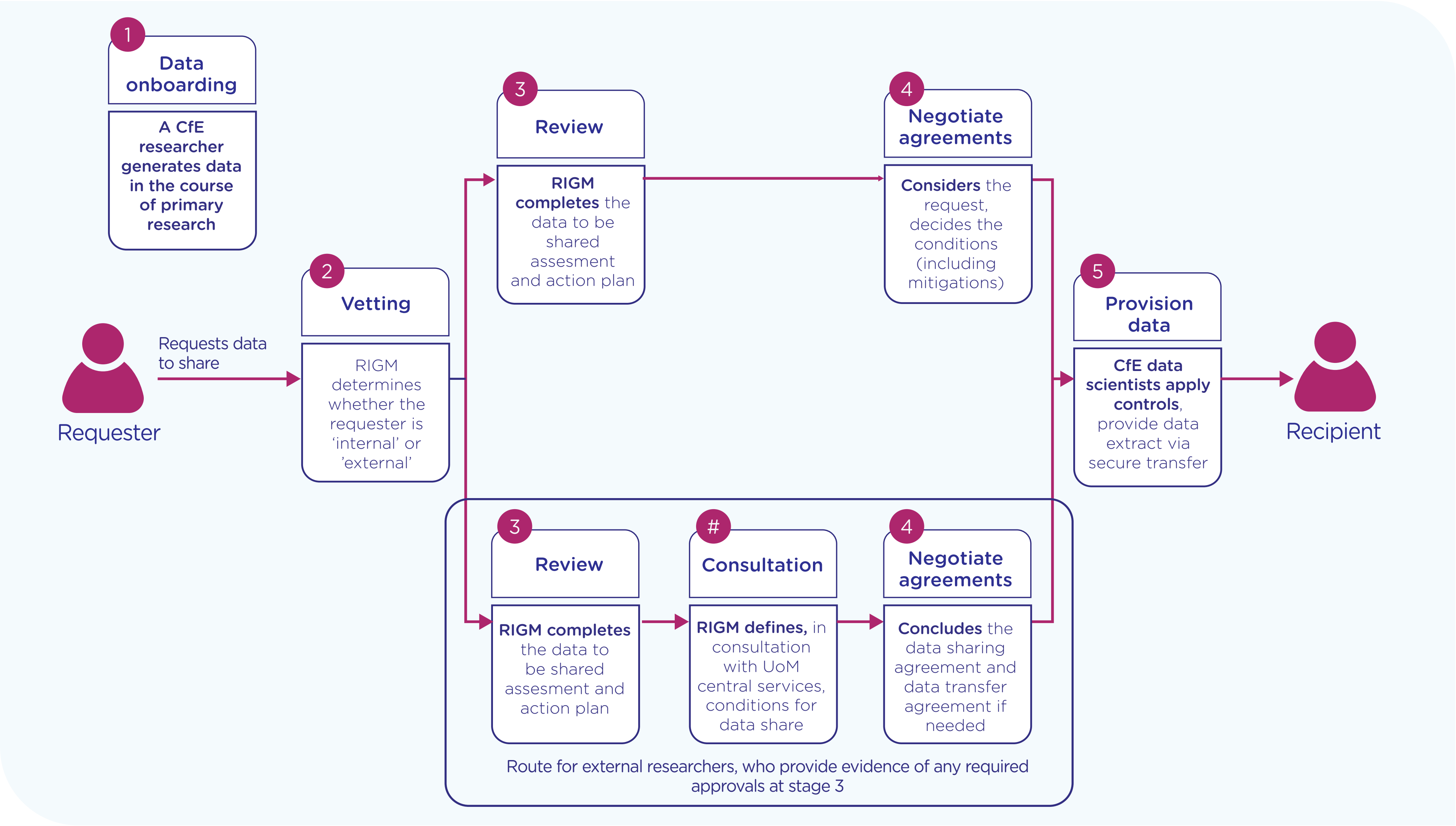
   iv securely destroy written notes at the end of the study; and

   v manage outputs to make sure they are not disclosive.

3 External recipients receive a data extract in line with the security requirements in the UoM policy.

Copyright 2021, Privitar LTD

5.5 CfE simplified process diagram



1 **Data onboarding**

A CfE researcher generates data in the course of primary research

Requester

Requests data to share

2 **Vetting**

RIGM determines whether the requester is 'internal' or 'external'

3 **Review**

RIGM completes the data to be shared assesment and action plan

4 **Negotiate agreements**

Considers the request, decides the conditions (including mitigations)

3 **Review**

RIGM completes the data to be shared assesment and action plan

Consultation

RIGM defines, in consultation with UoM central services, conditions for data share

4 **Negotiate agreements**

Concludes the data sharing agreement and data transfer agreement if needed

Route for external researchers, who provide evidence of any required approvals at stage 3

5 **Provision data**

CfE data scientists apply controls, provide data extract via secure transfer

Recipient

Copyright 2021, Privitar LTD

5.6 Data Safe Haven Plus (DSH+)

The UoM is developing a new, cloud-based research environment called the Data Safe Haven Plus (DSH+). The DSH+ is the next generation of the 'on premises' data safe haven. This section describes the data sharing process for researchers wanting access to CfE data through the DSH+.

The governance model described below has been set up by the CfE to run alongside the UoM's DSH+ infrastructure. It applies to researchers (internal and external) requesting access to data from CfE and provides another way to share data, as well as the process described above.

The CfE's governance model, which uses the UoM's DSH+ infrastructure as a route to sharing data, applies the Five Safes to allow the sharing of data that is, in most cases, functionally anonymised from the recipient's perspective. If the 'functionally anonymised' standard is not possible, other measures may be needed. For example, a legally binding data sharing agreement or a requirement for the recipient to access the data in a secure room.

- Safe projects. Delivered by DPIA on a project-by-project basis and DSRB approval process.
- Safe people. Delivered through the researcher onboarding process to the DSH+, by requiring ONS research accreditation where appropriate, and by the researcher signing the DSRB Terms of Use.
- Safe data. Delivered by the data onboarding process and the DSRB assessment of necessity and proportionality.
- Safe environment. The data stays in the UoM's Data Safe Haven Plus.
- Safe outputs. Delivered through the CfE's output checks against statistical disclosure control principles.

This section describes CfE's governance model as it works in the DSH+ infrastructure. We've identified six decision points, and highlighted differences between this model and the process described in Section 5.4 above.

5.6.1 Data onboarding

The processes for approving primary research are the same. The data onboarding stage in CfE decides if data is suitable for access through the DSH+. This is described in more detail at Section 5.6.3 below. The DSH+ uses a four-tier classification for data and it can accept data classified at tier three and below. Any data classified at tier four stays in the on-premises data safe haven. The general approach of classifying data and environments is described in a paper due to be published soon.

Roles and responsibilities

1 Research IT provides a classification guide on what data can be shared through the DSH+. The guide is based on:

i what other data exists in the recipient's environment;

ii the harm that could result from the data being re-identified; and

iii how the source of the data classifies the data. For example, those responsible for specific data may decide that all data they provide to the DSH+ should be classed at a specific tier.

Legal considerations?

1 The DSH+ is hosted in the UK, avoiding international data transfers under the UK GDPR.

Copyright 2021, Privitar LTD

5.6.2 Data review

Prospective researchers can explore data available for onward sharing through a data previewer, before applying for access to the data through the CfE's data sharing review board. The previewer is currently specific to each study but the CfE aims to produce a general previewer, supported by ongoing CfE work to develop a standard metadata format.

Roles and responsibilities

1 CfE data scientists develop the previewer and manually assess outputs to make sure they meet statistical disclosure control principles.

What controls are applied?

1 The previewer allows a very limited set of queries on the data and returns only aggregated data.

2 Outputs generated by the previewer are checked against statistical disclosure control principles.

5.6.3 Data Protection Impact Assessment (DPIA)

Once the prospective researcher submits a data access request, the RIGM (working with the prospective researcher) completes a DPIA to assess the privacy impact associated with the proposed data share.

Roles and responsibilities

1 The RIGM completes a CfE DPIA template, the data to be shared is classified, taking into account the properties of the data, the share environment (i.e. DSH+) and requirements on the recipient. Where the data to be shared is not considered functionally anonymised for the recipient the RIGM makes recommendations to the DSRB. These could include, for example, reducing detail in the data and / or applying extra controls on access (such as a legally binding data sharing agreement or a requirement to access the DSH+ in a secure room).

Legal considerations?

1 Is sharing consistent with the participant's consent and in line with their reasonable expectations?

2 Is the data necessary to answer the proposed recipient's research question? In other words, does the proposed data share comply with the data minimisation principle?

3 The lawful basis to process data to be shared is the 'public interest' or 'research' basis in Article 9(1)(j), and / or on the 'public interest in the area of public health' basis in Article 9(1)(i).

5.6.4 CfE Data sharing review board (DSRB)

The DSRB is made up of the Chair (UoM lead), a senior UoM academic, an independent academic from a UK institution, the RIGM, and a PPIE representative. The DSRB is supported by a Secretariat.

Roles and responsibilities

1 The DSRB reviews the researcher's application form, accompanying DPIA and any other documents considered necessary for the review. The Board may:

i approve a project;

ii approve a project with minor changes;

CONTENTS

Copyright 2021, Privitar LTD

   iii request substantial changes and a resubmission to a future meeting; or

   iv reject a project.

The Board has an appeals process.

What controls are applied?

1 The DSRB review considers scientific merit, ethics, privacy and public benefit:

   i Scientific merit. The Board decides if the proposed use of the data has enough scientific merit to justify the use of their resources.

   ii Ethical review. The Board assesses applications where ethical approval is needed and is being asked for at the same time. In cases where an applicant's Institution Research Ethics Committee considers a project does not need ethical review, the requestor needs an exemption certificate or relevant evidence of that decision.

   iii Privacy impact. The Board will decide if they are confident that any potential privacy issues have been assessed and can be adequately dealt with by the Centre's RIGM recommendations.

   iv Public benefit. The Board will decide if the proposed use of the data has enough public benefit to justify the use of the Board's resources, and to outweigh any potential privacy impact.

2 The DSRB also considers if the data requested is necessary and proportionate to answer the applicant's proposed research question, in line with the GDPR principle of data minimisation.

5.6.5 Recipient granted access to the DSH+

Once the DSRB approves access, the RIGM and RIGO and central UoM services allow the recipient to access the DSH+. The process varies depending on if the recipient is internal or external to the UoM.

Internal researchers access the DSH+ using their university credentials, with two-factor authentication. Internal researchers must access the DSH+ from a university-managed device, usually on campus. They must comply with the UoM policies relating to research data use described in the previous case study. The process for external researchers is described below.

Roles and responsibilities

1 The RIGM and RIGO submit a request to IT Services.

2 IT Services guides the external recipient through the account set up and access procedures. The researcher will get a UoM account, with two-factor authentication to access to the DSH+.

What controls are applied?

1 Only ONS Accredited Researchers are eligible to access CfE data in the DSH+. The researcher can apply for access before getting research accreditation status but cannot be given access to data until they have it.

2 The researcher agrees to the DSRB Terms of Use, which include controls on data access and use.

3 Researchers access the DSH+ through a web-based virtual desktop that restricts copying and pasting of data, with clearly defined policies on what data can be taken out of or brought in to the DSH+.

Copyright 2021, Privitar LTD

4 Users outside the UoM network use a secure VPN with two-factor authentication.

5 Work in the DSH+ is monitored (including researcher queries and access locations). Access logs, and other information generated by the researcher (including applications and project documentation) are kept and the DSRB can audit them.

6 The RIGM provides quarterly reports on new projects, data access, movement of data, and data deletion to the DSH+ Operations Group, which reports to the University's senior Research Compliance Committee.

7 Cybersecurity controls. The University's standard cybersecurity controls apply to the DSH+. The standard controls include roles-based access control with two-factor authentication, with access restricted to the University network through a VPN. The DSH+ processes are based on the University's policies on information management and security classification. Encryption keys are managed and unique to a project.

8 Physical access within the University is managed and only available in fixed physical locations with identity-based access. Data in the DSH+ is tagged so that it can be tracked and activity relating to it logged.

5.6.6 Output checking for statistical disclosure control

The DSH+ provides a way for researchers to extract data from the system (for example, if the data is necessary for publication). Output checking is a two-stage process. The researcher reviews their outputs according to statistical disclosure control principles, then requests a review by the Output Checker. The review is based on guidelines for checking output based on microdata research and the Handbook on Statistical Disclosure Control (SDC) for Outputs.

Roles and responsibilities

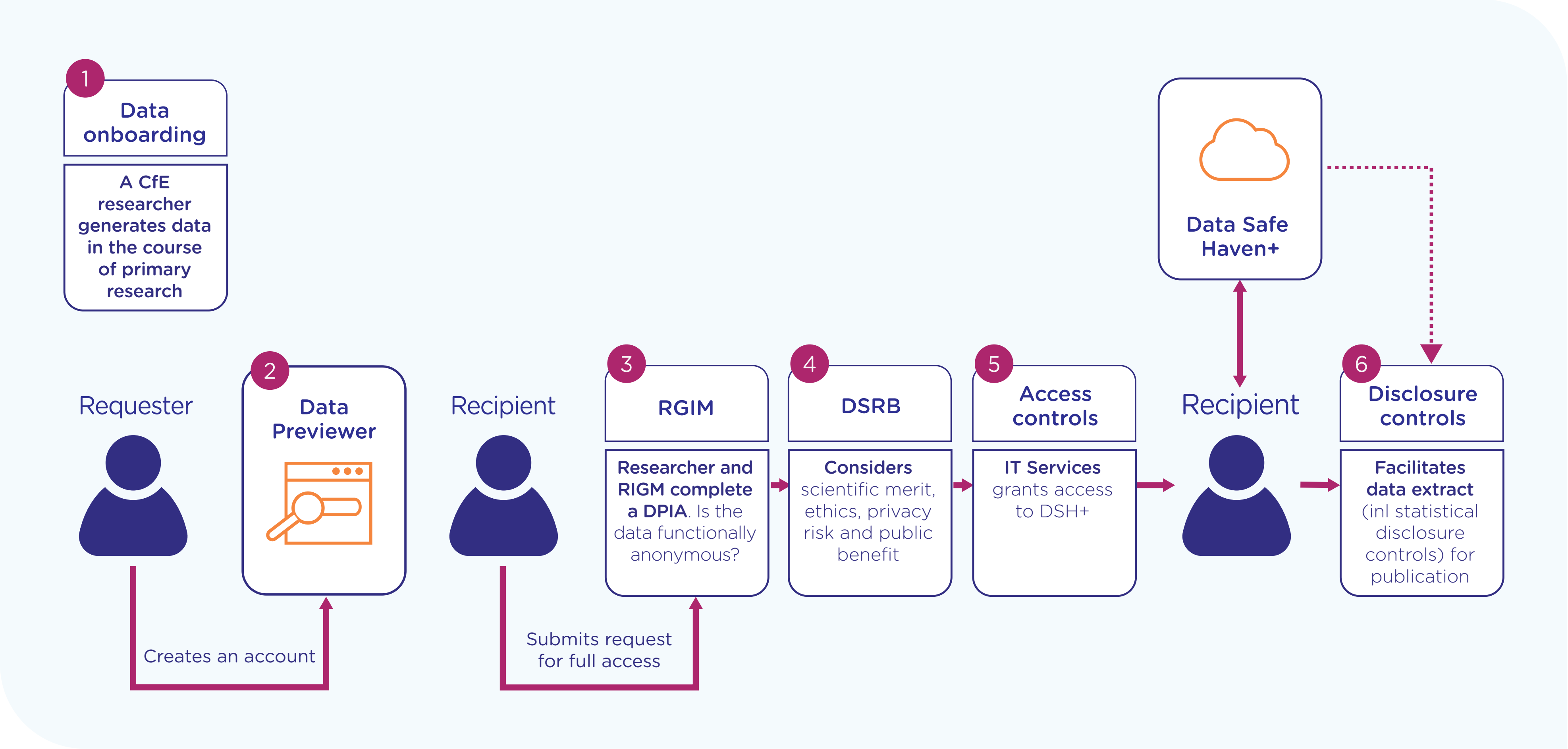
1 The CfE's Output Checker reviews outputs in line with statistical disclosure control principles.

2 Research IT takes the technical steps necessary to allow the researcher to extract data.

What controls are applied?

1 Manual review to make sure no information is released that could identify an individual or organisation. Output checking is a two-person process, the output checkers will work with the researchers, however, they have the final say on whether data is released.

5.7 CfE - DSH+ simplified process diagram



1 Data onboarding

A CfE researcher generates data in the course of primary research

Requester

Creates an account

2 Data Previewer

Recipient

Submits request for full access

3 RGIM

Researcher and RIGM complete a DPIA. Is the data functionally anonymous?

4 DSRB

Considers scientific merit, ethics, privacy risk and public benefit

5 Access controls

IT Services grants access to DSH+

Data Safe Haven+

Recipient

6 Disclosure controls

Facilitates data extract (inl statistical disclosure controls) for publication

CONTENTS

Copyright 2021, Privitar LTD

Where do we go from here?

The conversation around sharing health data is evolving. Since we started work on this report we've seen new proposed EU legislation on information governance, an updated version of the ADF and the UK government's National Data Strategy published, as well as innovative work on data sharing structures (for example, the Web Sciences Institute's white paper on a data foundation).

Data sharing in response to the COVID-19 pandemic led to the government issuing notices requiring health organisations to process confidential patient information, in line with UK GDPR. The government has also launched the Goldacre Review to learn lessons ahead of the Data Strategy for Health and Social Care.

As we see new ways of handling and sharing data, we encourage stakeholders to make sure the guidance continues to provide a useful reference point for practitioners. The best way to do this is to set up and maintain a shared, industry-led conversation on how practitioners translate guidance and principles into actionable processes. We hope that regulators and standard setters use elements of these case studies as examples for their guidance.

In producing this report, we were struck by the challenges of anonymous data. It is difficult for practitioners to know when they have done enough to turn personal data into anonymous data. It is also difficult to compare the effect of different types of controls. For example, if your risk assessment identifies a possible re-identification risk, you can either mitigate that risk using technical means (which will affect the usefulness of the data) or require data recipients to not exploit it (for example, by contract) and monitor them to check they comply with the contract.

We believe that regulators have a significant opportunity to encourage data sharing by issuing updated guidance on some of the most challenging aspects of regulatory requirements. For example, guidance on interpreting anonymous, on applying controls (including PETs) and on managing re-identification risk.

We've seen organisations rely on the concept of 'functional anonymity' where data can be considered anonymous in a given situation. So it can be anonymous for the end user, but identifiable for the organisation providing the information. To conclude that the data is functionally anonymous, organisations take account both of the data itself and the controls applied to the data. Regulators could provide clarity by explicitly endorsing this approach, or by setting out the types of controls that organisations can consider (for example, are contract controls better, worse or just different to working with data in a secure research environment?).

We also want to hear from professionals working on health data sharing. We welcome your reflections on whether these case studies are in line with or different from your experience. Views from readers sharing data in other sectors are also welcome. Do the experiences in healthcare offer useful takeaways for sharing data in your sector? You can reach us on policy@privitar.com.

CONTENTS

Copyright 2021, Privitar LTD

About the authors

Guy Cohen, Head of Policy

guy.cohen@privitar.com

Guy joined Privitar in 2016, before which he worked in the UK Civil Service, in the Department of Health, the Cabinet Office and HMRC. Guy has been a fellow at Cambridge University's Centre of Science and Policy, a member of the Royal Society Privacy Enhancing Technologies Working Group, and is the technical editor for the IEEE Data Privacy Process Standard.

Marcus Grazette, Europe Policy Lead

marcus.grazette@privitar.com

Marcus joined Privitar in 2019, after more than 10 years at the Foreign and Commonwealth Office. He has also worked as a consultant with EY. He specialises in European public policy. His academic background is in law and he has a MA in European public policy from l'université Paris 1 Panthéon-Sorbonne.

Dr. Hector Page, Research Scientist

hector.page@privitar.com

Hector is a research scientist at Privitar. In 2018 he was the lead author for a report commissioned by the UK Government Statistical Service on Differential Privacy. Before joining Privitar he worked on spatial cognition at the UCL Institute of Behavioural Neuroscience. He has a PhD in Computational Neuroscience from the University of Oxford.

CONTENTS

Copyright 2021, Privitar LTD

APPENDIX A - Glossary and legal terminology

Cell suppression: The hiding of particular values in a dataset. It is typically used when creating aggregate data from an underlying dataset. Selecting the cells to suppress based on which contain small counts is called small count suppression.

Data linkage: The process of bringing together data about the same individual from different sources. For example, data from a GP practice can be linked to data about hospital admissions relating to the same individual. Linkage usually requires a common 'record key' which exists in both data sources (for example, an NHS number).

Data provider: The party who controls access to the data and is able to determine with whom the data is shared.

Data sharing: The process of providing data or access to data. This includes providing an extract, which the recipient processes in their environment or allowing the recipient to access data in a hosted environment.

De-identification: The process of reducing the probability that data can be linked to a specific known individual (known as re-identification). This can be achieved by controls applied to the dataset itself (for example, generalisation) or to the context in which the data exists (for example, access control). De-identification can reduce the identifiability of data until it becomes anonymous.

Protocol: The document describing the client's proposed research.

Provisioning: the data engineering process to create the data extract, apply data transformations and share the extract with the client.

Precedent set review pathway: Developed by the Confidentiality Advisory Group (CAG) to speed up reviews for applications that are similar to previous applications reviewed by CAG. Precedent set categories identify situations as ones they commonly see, and any application falling into one of these categories will be processed under the precedent set review pathway. Each category has a CAG sub-committee which processes applications in 30 days. For more information, see this page from the Health Research Authority.

Trusted third party model: Proposed by the UK Administrative Data Research Network (ADRN) in their 2012 report "Improving Access for Research and Policy" found here. In this model, a trusted third party is responsible for creating a link between records for the same individual across multiple datasets. They receive direct identifiers (for example, name, address, NHS number) from data providers without any other information. They use this information to create linkage keys (referred to in the report as "study identifiers") to the data without carrying out this linkage itself. Data holders can then send their datasets containing linkage keys in place of direct identifiers to the recipient. Recipients are then able to link the data without access to direct identifiers and the trusted third party sees only the information necessary to set up linkage keys.

Synthetic data: Data generated by fitting a model to the data and then producing new data records from the model. These new records include statistical properties contained in the model. All or part of a dataset may be synthetic. In some contexts, this is referred to as overimputation. For a summary of the types of synthetic data, see the ONS's page on this topic.

Copyright 2021, Privitar LTD

Term	Definition
Personal data	Article 4, GDPR
Pseudonymous / pseudonymisation	Article 4, GDPR
Anonymous / anonymisation	Recital 26, GDPR, interpreted with reference to the ICO's 2012 code of practice, UK case law *(Information Commissioner v Millar)* and case law from the CJEU *(Breyer)*.
Legitimate interest	Article 6(1)(f), GDPR
Public interest	Article 6(1)(e), GDPR
Consent	Article 6(1)(a), GDPR
Patient information, confidential patient information	Section 251, NHS Act 2006
Privacy harms	Recital 75, GDPR
Processing	Article 4, GDPR
Transparency principle	Recital 39, GDPR
Confidential information	Case law. For example, in *Campbell v MGN Limited* [2004] UKHL 22, Lord Nicholls indicated that the right to private life is part of the duty of confidentiality [17] and that 'the touchstone of private life is whether in respect of the disclosed facts the person in question had a reasonable expectation of privacy' [21].

CONTENTS

Copyright 2021, Privitar LTD

APPENDIX B - Common data controls

1. Deletion (full redaction), either removing the records entirely (for example, removing an attribute or column), or replacing all values with a constant value such as “XXXXX” or “0”.

2. Clipping (partial redaction), where a value is partly deleted. For example, keeping only the first half of a postcode or the last four digits of a credit card number. Clipping makes some types of data more general (see generalisation).

3. Tokenisation, where a value is replaced with a randomly generated value, a token. Tokens may need to be in a specific format, for example, a specific length, to make sure tokenised data is compatible with other processing that may be applied to it. UK National Insurance numbers, for example, follow a specific format (two letters, six digits, one letter) and follow specific rules (for example, the second letter is never an ‘O’). Generating a token with the correct formatting allows tokenised data to pass a validation test.

4. Hashing, where a function is applied to the value to produce a fixed length output known as a “hash.” The function is one-way, so the hash cannot be converted back to the original value. This is a common technique, but has been shown to be vulnerable to attack. Hashing has many variations, including salted hashing, where a random string, a ”salt,” is added to the value before it is hashed.

5. Substitution, where a value is replaced by another value from a predefined list. This can provide a form of generalisation, where the substituted value is less precise than the original and many values map to it. For example, the values “Westminster” and “Lambeth” may both be substituted for the more general value “London.”

6. Field level encryption, where a value is encrypted to produce an output ciphertext derived from the input value and a cipher key.

7. Perturbation, where random noise is added to a value, for example, transactions might be perturbed by any full unit value in a range, so an input value of $182 may be perturbed by +/- $5 to generate an output value in the range $177-$187.

8. Generalisation, where a value is made less precise. The specific technique varies depending on the type of data.

   a. Binning, this most commonly involves transforming a specific value into a range, for example, £182 transformed to the range £180 - £190, or a midpoint £182 → £185.

   b. Rounding, for example, rounding to the nearest 100, so £182 → £200.

   c. Clipping some types of values. for example, clipping a postcode to just the first three digits, or clipping a time to give just the hour, not the minutes or seconds, so 09.06.55 → 09.

Copyright 2021, Privitar LTD

APPENDIX C - Common contract terms governing data use

1. Access to data is being provided for the research purpose described in your approved Data Access Application Form. Data provided shall not be used for any other purposes without the prior written consent of the Data Sharing Review Board (hereafter referred to as the Board).

2. The researcher is required to obtain researcher accreditation status to access data supplied by the Board within the DSH+.

3. The researcher shall not disclose the data pursuant to this Agreement to anyone.

4. The researcher will not attempt to identify any individual person or organisation through access and use of the data. In the unlikely event that a researcher inadvertently identifies a data subject via spontaneous recognition, the researcher will inform the Centre's Research Information Governance Manager as soon as possible.

5. The researcher will not attempt to link the data to any other external files, unless such data linkage exercise has been explicitly approved as part of their application, or approved subsequently as part of a special request to the Board.

6. Any incidents of unauthorised access to, processing of, or disclosing of data must be reported to the Board as soon as possible.

7. Any non-compliance with this Agreement will result in the immediate imposition of remediation, see Appendix A for a list of non-compliance behaviours. Also see the Secure Access Compliance Policy (add web link) for more information.

8. The Board reserves the right to monitor, record, and audit, or to request a written report from the researcher regarding the use and activities relating to the use of the data by the researcher during the lifetime of this Agreement.

9. The Board will retain all information submitted by the researcher (including queries, applications, appeals, project documentation) for the lifetime of the Board. The Board will retain and use this information for monitoring, management and improvement of the service and for the creation of a knowledge base. In the interest of transparency, the Board may publish the PIs name, Institution and project title of the individual projects approved by the Board on the University of Manchester website. In addition, the Board will publish at the end of the project a one page lay summary of the project as supplied by the researcher.

10. The Agreement is subject to review and without limitation whenever a change in the law, contracts for services with third parties, other procedures or other relevant circumstances takes place.

11. On termination of the Agreement for whatever reason, all access to the data related to the project shall cease immediately.

CONTENTS

Copyright 2021, Privitar LTD

Contact us:

e: info@privitar.com
t: UK +44 203 282 7136
w: www.privitar.com

PRIVITAR
Research & Policy

www.privitar.com

Copyright 2021 Privitar LTD