This is a repository copy of *Navigating data governance associated with real-world data for public benefit: an overview in the UK and future considerations*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/203983/

Version: Published Version

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

BMJ Open

Navigating data governance associated with real-world data for public benefit: an overview in the UK and future considerations

Monica Catherine Jones ![ORCID],[1] Tony Stone,[2] Suzanne M Mason,[3] Andy Eames,[4] Matthew Franklin ![ORCID] [3]

[1]Faculty of Medicine and Health, University of Leeds, Leeds, UK
[2]School of Health and Related Research, The University of Sheffield Faculty of Medicine Dentistry and Health, Sheffield, UK
[3]School of Health and Related Research, The University of Sheffield, Sheffield, UK
[4]Health Informatics, NHS Sheffield CCG, Sheffield, UK

Correspondence to
Monica Catherine Jones;
m.c.m.jones@leeds.ac.uk

SUMMARY

Real-world data encompass data primarily captured for the provision or operation of services, for example, electronic health records for direct care purposes, but which may have secondary uses for informing research or commissioning. Public benefit is potentially forfeited by the underutilisation of real-world data for secondary uses, in part due to risk aversion when faced with the prospect of navigating necessary and important data governance processes. Such processes can be perceived as complex, daunting, time-consuming and exposing organisations to risk. By providing an overview description and discussion around the role of six key legal and information governance frameworks and their role regarding responsible data access, linkage and sharing, our intention is to make data governance a less daunting prospect and reduce the perception that it is a barrier to secondary uses, thus enabling public benefit.

INTRODUCTION

The use of the terms 'real-world data' (RWD) and 'real-world evidence' (RWE) in the context of health decision-making has grown substantially in the last 20 years, although unified and consistent definitions of these terms remain elusive.[1 2] Often referred to as 'administrative', 'observational', 'routine', 'large' or even 'big' data sources; over the last decade, they have become of increasing interest to those conducting health technology assessments, to provide policy-makers with evidence to inform decision-making and develop guidance on the reimbursement and administration of new health technologies within a care system.[2–4] For example, the National Institute for Health and Care Excellence (NICE) for England and Wales, within their RWE framework, suggests RWD is: 'Data relating to patient health or experience or care delivery collected outside the context of a highly controlled clinical trial';[5] relatedly, NICE's definition of RWE is simply: 'evidence generated from the analysis of

RWD'.[5] As such, RWD includes administrative (eg, Hospital Episode Statistics), registry (eg, National Cancer Registration and Analysis Service) or survey data on populations (eg, Health Survey for England) among other routine data sources, which can comprise stand-alone or linked datasets.[6–8]

RWD have a range of potential uses that could lead to public health and non-health benefits (eg, well-being, independence and empowerment).[9–12] For example, as a complement or substitute to primary data collection within RCTs, for example, to reduce or avoid responder burden or loss to follow-up. Alternatively, RWD can be analysed in its own right to assess associations between factors involving medical and/or public health (eg, frailty and mortality), or for the purpose of causal inference (eg, estimating average treatment effects).[13–15] Such data could be used to inform the development of new health technologies, for example, artificial intelligence (AI) and other digital health technologies that use routine data within risk prediction and/or machine learning processes.[16 17] Secondary uses of data can also inform more than just academic research, for example, when it is employed by local and national commissioning agencies and governing bodies to inform performance monitoring, policy decision-making, guide public spending decisions (eg, commissioning new care services across health and social care sectors), and Population Health Management. However and related to all these example contexts which could lead to public benefit[10 11]: the use and sharing of RWD presents particular challenges, especially around data protection and associated data governance. Public benefit is potentially forfeited by the underutilisation of RWD for secondary uses, in part due to risk aversion

when faced with the prospect of navigating necessary and important data governance processes, although there are other barriers to sharing and using RWD which can include organisational features and process factors.[9–11 18]

The aim of this article is to explore how we can make data governance a less daunting prospect and reduce the perception that it is a barrier to secondary uses, relative to processes and gateways for responsible use of data with public benefit as a focal point. As such, within this article, we focus on the legal and information governance (IG) structures associated with secondary uses of data relating to individuals primarily captured for the provision or operation of health and care services; in this article, we particularly focus on the UK's Data Protection Act 2018, General Data Protection Regulation (GDPR) 2018, common law duty of confidentiality, National Health Service (NHS) Act 2006 section 251, Control of Patient Information (COPI) Regulations 2002 and the Health and Care Act 2022.[19–23 23] We also make comparisons between the UK's and other countries' (ie, US) frameworks for data protection and security, and associated initiatives such as the Five Safes framework and the increasing interest and use of Trusted Research Environments (TREs). A TRE is a highly secure computing environment that provides remote access to health data for approved researchers to use in research.[24] They are usually owned and managed by public funded bodies such as the NHS in the UK.

WHAT IS IG?

IG refers to a framework and set of practices that are designed to ensure the effective and secure management of information within an organisation. It encompasses the policies, procedures, systems and controls that govern the collection, storage, use, sharing and disposal of information.[20] In the context of healthcare, IG is of utmost importance due to the sensitive and confidential nature of patient data. It aims to protect patient information, ensure its accuracy and accessibility, and promote the responsible and lawful use of data. Effective IG involves the development and implementation of policies, staff training, appropriate technology infrastructure, risk assessment and management, regular audits, and compliance monitoring. By adhering to these principles and practices, organisations can ensure the confidentiality, integrity and availability of patient information while promoting trust and confidence in health and social care systems.

INTRODUCING THE BEDROCK: GDPR, DATA PROTECTION ACT 2018 AND THE COMMON LAW DUTY OF CONFIDENTIALITY

Since 2018, the UK has had new rules around how to handle personal data: the GDPR, as enacted in the UK by the Data Protection Act 2018.[19 20] A guide exists for understanding UK GDPR which explains the general data protection regime that applies to most UK businesses and organisations processing data relating to identifiable,

living persons.[20] In summary, this guide defines data protection principles, rights and obligations, summarises key points, answers frequently asked questions, and contains practical checklists to help people comply with GDPR. Additionally, the NHS Health Research Authority (HRA) provides GDPR operational guidance for research,[25] and the Information Commissioner's Office (ICO) issues data protection guidance.[26] An important aspect for consideration is that the UK GDPR defines data 'controllers' and 'processors':

▶ A controller determines the purposes and means of processing personal data.
▶ A processor is responsible for processing personal data under the direction of a controller.

As such, key aspects of the GDPR are to help define and operationalise responsibilities of data controllers, while also improving the rights and notification of the data subjects. GDPR covers the whole of Europe, but confidence, privacy legislation and tort law are country specific. Overall, this helps people to know what to do and how to do it when dealing with data relating to living persons.

The common law duty of confidentiality (also referred to as the 'common law duty of confidence'), unlike GDPR and the Data Protection Act, is not defined by a written document but is instead based on previous court decisions and principles established by the Courts and Tribunals Judiciary. Broadly, there is a legal precedent that information given in confidence, or under an expectation of confidence, must not be disclosed without the information provider's agreement unless there is another valid lawful basis.[27] In contrast to GDPR, the duty of confidence applies to information provided by individuals even after their death. Health and care information relating to a person is generally considered to be owed a duty of confidence. This type of information is often termed 'confidential patient information'.

COMMON LAW DUTY OF CONFIDENTIALITY AND THE NHS ACT 2006 SECTION 251 (ENGLAND AND WALES)

Obtaining agreement from each patient to access their health and social care information for large-scale or population-level analyses is rarely practical and important historical records of patients who have since died would be unavailable. Where it can be shown that there are no alternative practical means and that there is demonstrable potential public benefit, the common law duty of confidentiality can be set aside. This enables confidential patient information to be shared for specific purposes without the explicit agreement of the individual patients, without the controller being in breach of their duty of confidence when circumstances are justified. The NHS HRA decides whether to approve an application but their decision is strongly guided by advice from the independent Confidentiality Advisory Group (CAG) with due regard to guidance by the National Data Guardian.[28] The CAG expects applicants to provide a mechanism

for patients to opt-out from their data being used for secondary purposes on a project-by-project basis. The final decision of whether to share data or not rests with the data controller, but all parties must comply with UK Data Protection Legislation and NHS (or other appropriate) data security standards at all times.

A MOMENT OF REALISATION: COVID-19 PANDEMIC AND THE COPI NOTICES

When the COVID-19 pandemic started, there was a realisation of the requirement for greater and more timely access to a wider variety of data to manage the public health response to the pandemic. The Secretary of State for Health and Social Care issued Notices under Regulation 3 (4) of the Health Service COPI Regulations 2002[29]: these notices directed health and care providers and, particularly, NHS Digital (then national collator of information about health and social care in England) to share confidential patient information with authorised organisations for the purposes of managing the COVID-19 pandemic. This has been very useful for implementing public health interventions, but also for monitoring the pandemic, informing key decisions affecting population health and enabling research focused on delivering public benefit. It is important to note that the COPI Notices did not remove the requirements for organisations to comply with UK Data Protection Legislation and NHS (or other appropriate) data security standards.

The COPI Notice(s) issued over the course of the COVID-19 pandemic were limited in scope to permit the processing (including sharing) of data solely for the purpose of supporting the government's response to COVID-19 . The majority of these COPI notices expired on 30 June 2022.[30–33] Most COVID-19-related projects requiring ongoing processing of confidential patient information (in England and Wales) have now transitioned to use another legal basis for the processing of this data, mainly section 251 of the NHS Act 2006. The public benefit achieved due to the timely sharing of patient data enabled by the COPI notices has highlighted that further public benefits could be achieved through the appropriate sharing and use of such RWD for purpose beyond purely the response to the COVID-19 pandemic.[9–12]

INSTIGATING FURTHER CHANGE: THE HEALTH AND CARE ACT 2022

The Health and Care Act 2022 dismantles many of the structures established by the Health and Social Care Act 2012[34] and is intended to reinforce the ambitions of the NHS Long Term Plan.[23 34 35] The Act puts Integrated Care Systems (ICSs) on a statutory footing and provides for the Care Quality Commission to assess how local authorities deliver their adult social care functions. It has established an Integrated Care Board and an Integrated Care Partnership in every part of England. The Act enables NHS England (among others) to publish data specifications

detailing information which providers would be obliged to submit to NHS Digital, and makes it a criminal offence to share that data inappropriately. The Act has pushed forward the publication of mandatory information standards for the processing of information (including collection and storage), which requires health and social care providers to comply with such standards rather than, as previously, have regard to them; this requirement also extends to private providers.[35] The Act requires NHS Digital to promote the effective and efficient planning, development and provision of health services and of adult social care in England.[35] The Act also makes it clear that NHS Digital may share information for purposes connected with the provision of healthcare or adult social care or the promotion of health including for research purposes.[35] Alongside other considerations within the Act's 'Health and Adult Social Care: Information' section, the Act outlines requirements for providers to share 'anonymous' information—the nature and sharing of anonymous data aligns with the Data Protection Act and GDPR 2018,[19] but also common law duty of confidentiality; as such, the Act aligns with current data governance legislation rather than extending it.

The main challenge of the Health and Care Act 2022[23] in the UK is to implement significant changes to the organisation and operation of the NHS. The Act's creation of ICSs brings together NHS providers, local authorities and other health and care organisations to plan and deliver care for their populations. The challenge lies in ensuring that these ICSs operate effectively, with clear lines of accountability and governance, and that they provide high-quality, patient-centred care, while also addressing health inequalities and managing costs effectively, which could be achieved through appropriate sharing and use of RWD.[12]

WHAT ELSE IS NEEDED: A CHANGE IN THE DATA LANDSCAPE

Bringing data together on a project-by-project basis (eg, through one time data extracts) is inefficient. NHS HRA support is required for each project wishing to link together patient-level health or care data held by more than one organisation and can only be used for that project. The Digital Economy Act (DEA) 2017 created a framework for enabling public authorities to share information with accredited researchers for approved projects seeking to deliver public benefit.[36] Before data are made available to researchers, it is linked and depersonalised by independent, accredited data processors to ensure the data is no longer reasonably likely to reidentify persons or businesses. Researchers and their projects must be accredited before gaining access and the data can only be accessed within an accredited safe environment. Finally, all research outputs are independently checked to ensure they meet statistical disclosure control guidelines; this is known as the Five Safes framework, which we have described in table 1 when comparing alignment of the framework to the UK's healthcare IG regulations.[37]

Table 1   Comparing alignment of the UK healthcare IG regulations against the Five Safes dimensions

Five Safes dimensions	Dimension description	Alignment with UK healthcare IG regulations
Safe project	Security measures in place to ensure that projects involving healthcare information adhere to privacy and data protection regulations	The Data Protection Act 2018, which incorporates the General Data Protection Regulation principles, requires organisations to have clear purposes for processing data, obtain consent from individuals and implement appropriate security measures to protect the data.
Safe people	Ensuring that only authorised individuals have access to healthcare information and that their access is appropriate and secure	UK healthcare IG regulations require organisations to implement strict access controls and authentication mechanisms to verify the identity of individuals accessing the data. The National Health Service (NHS) also provides guidelines and training programmes to ensure that healthcare professionals understand their responsibilities when handling patient information.
Safe data	Protection and management of healthcare data to prevent unauthorised access, loss or corruption	UK healthcare IG regulations mandate the implementation of robust security measures, including encryption, firewalls and secure storage, to protect healthcare data from breaches and cyberattacks. The NHS Digital Security and IG Standards provide guidelines on data protection and secure handling of healthcare information.
Safe settings	The physical and virtual environments where healthcare data is stored, processed and accessed	UK healthcare IG regulations require organisations to ensure that the settings where data are processed or stored, such as hospitals, clinics and healthcare systems, meet certain security standards. This includes measures to prevent unauthorised physical access to data storage facilities and the use of secure networks and infrastructure to protect data during transmission.
Safe outputs	Procedures in place to ensure that outputs generated from healthcare information analysis and research are appropriately managed and reported	UK healthcare IG regulations require organisations to deidentify data for research purposes to protect patient privacy. The NHS Digital Code of Practice for Data Release provides guidance on the safe and ethical use of data, ensuring that outputs are properly anonymised and comply with legal and ethical requirements.

IG, information governance.

Unfortunately, health and care data are explicitly excluded from the DEA 2017, despite the public benefits that could be achieved by including health and care data within the Act. This exclusion means that it is currently necessary to seek dual approvals for projects seeking to combine and link health and care data with data from other domains (eg, economic and crime data) made accessible under the DEA 2017. This dual approval process is not documented and requires coordination between approval bodies.

Discussion of the differences between IG regulations
Navigating healthcare IG regulations can be complex, especially when it comes to understanding the differences between regulations in different countries. In the UK, healthcare IG is primarily governed by several key regulations and guidelines. As such, in table 2, we discuss some of the key differences between the healthcare IG regulations in the UK to aid researchers in navigating this field.

Researchers navigating the field of healthcare IG in the UK should familiarise themselves with these regulations, guidelines and organisations. Seeking advice from research ethics committees, NHS Digital, the HRA and the NIHR can provide valuable support and ensure compliance with the relevant regulations.[25] Staying up to date with evolving regulations is also essential as healthcare IG is an evolving field that responds to technological advancements and changing societal expectations. The UK's IG jurisdiction and other countries, such as the USA with Health Insurance Portability and Accountability Act, also have several key differences in their approach to data protection and privacy, with some examples provided in table 3 using the UK and USA as case studies.[38]

PRACTICAL EXAMPLES OF APPROACHES TO DATA GOVERNANCE
Health Data Research UK (HDR UK) is a national institute in the UK that aims to harness the power of health data to improve patient care, enable ground-breaking research and drive innovation in healthcare.[24] It brings together expertise from universities, research organisations, NHS trusts and industry partners to facilitate the secure and responsible use of health data. Another regional initiative is the Yorkshire Health and Care Record (YHCR), which is a collaboration between the NHS and local authorities in the Yorkshire and Humber region.[39] The YHCR aims to create a comprehensive digital health and care record for the region's population, integrating data from multiple

Table 2 Summarising some key differences between IG regulations, and associated organisations and principles

IG regulations, organisation and principles	Key differences
Data Protection Act 2018 (DPA) versus General Data Protection Regulation (GDPR)	The DPA 2018 and GDPR are two essential regulations that govern data protection and privacy in the UK. The DPA 2018 is the UK's implementation of the GDPR, and it provides the legal framework for processing personal data. Researchers must adhere to the principles of data protection outlined in these regulations, such as obtaining valid consent, ensuring data security and providing individuals with rights over their data.
NHS Digital versus Caldicott Principles	NHS Digital, the national information and technology partner to the health and care system in England, has specific guidelines for handling patient data. They provide detailed guidance on data standards, IG and security management for researchers working with health data. The Caldicott Principles, established in 1997 and revised in 2013, define the guidelines for protecting patient information and ensuring its confidentiality.
Confidentiality versus Security	Confidentiality and security are crucial aspects of healthcare IG in the UK. Researchers must comply with regulations that aim to protect patient information from unauthorised access, disclosure or breaches. They must adopt appropriate security measures to safeguard data, including encryption, access controls and secure storage.
Research Ethics versus Research Governance	Research involving human subjects must comply with ethical standards. Independent research ethics committees review and approve research proposals to ensure that they meet ethical guidelines. Additionally, the Research Governance Framework provides a framework for researchers to navigate ethical considerations, governance and best practices in research.
National Institute for Health Research (NIHR) versus Health Research Authority (HRA)	The NIHR is a UK government-funded organisation that supports and funds health research. It provides a wealth of resources, funding opportunities and guidance for researchers conducting clinical trials, studies and health-related research. Researchers can access the NIHR's infrastructure, expertise and support services to navigate the regulatory landscape effectively. The HRA plays a crucial role in regulating health research in the UK. It provides guidance and resources for researchers, including the Integrated Research Application System, which streamlines the process of gaining approvals and permissions for health research projects. The HRA ensures that research studies adhere to ethical standards and safeguards the rights, safety and well-being of research participants.
Local variations versus Devolved administrations	It is important to note that while the UK has a unified healthcare IG framework, there may be variations in specific regulations or guidelines across different regions. Devolved administrations in Scotland, Wales and Northern Ireland have some autonomy in healthcare governance, which may lead to nuanced differences in regulations and guidelines.

IG, information governance; NHS, National Health Service.

sources to provide a holistic view of an individual's health and care information.

These initiatives demonstrate an innovative approach to data governance and sharing. Both HDR UK and YHCR have implemented robust data governance practices, with examples including:

► Consent and patient engagement: HDR UK and YHCR prioritise patient engagement and consent. They involve patients and the public in decision-making processes related to data sharing, ensuring transparency and accountability. Patients are provided with clear information about how their data will be used and given the choice to opt out if they wish.

► Secure data infrastructure: Both initiatives invest in secure data infrastructure to protect patient information. This includes employing encryption techniques, access controls and anonymisation methods to minimise the risk of reidentification.

► Data access frameworks: HDR UK and YHCR have established data access frameworks that govern how researchers and organisations can access and use health data. These frameworks outline the criteria for accessing data, including obtaining appropriate approvals, demonstrating research merit, and adhering to strict data security and privacy standards.

► Ethical approvals and governance bodies: HDR UK and YHCR have established ethical approvals processes and governance bodies to ensure that research projects using health data adhere to ethical guidelines. These bodies review research proposals and monitor ongoing projects to ensure compliance with data protection regulations and ethical principles.

HDR UK and YHCR are provided as examples for descriptive purposes only, noting regionally and nationally, within and across countries, such institutions and initiatives are developing and evolving.

Table 3 Some differences in approaches to data protection and privacy between countries—a UK versus US case study

Consideration	UK	USA
Legal framework	The primary legislation governing data protection is the Data Protection Act 2018[19] supplemented by the General Data Protection Regulation (GDPR). The GDPR sets out the data protection standards across the European Union and the UK	The USA operates under a sector-specific approach, with Health Insurance Portability and Accountability Act (HIPAA) specifically addressing the privacy and security of health information.
Scope	The GDPR and UK data protection laws apply to all personal data, regardless of the sector	HIPAA focuses solely on protected health information (PHI) held by covered entities, including healthcare providers, health plans and healthcare clearinghouses
Consent requirements	Under the GDPR, organisations must obtain freely given, specific, informed and unambiguous consent from individuals to process their personal data. Consent is one of several lawful bases for processing data	HIPAA allows the use and disclosure of PHI without individual consent for treatment, payment and healthcare operations, among other permitted purposes.
Individual rights	The GDPR provides individuals with robust rights, including the right to access their data, rectify inaccuracies, request erasure and restrict processing. Individuals also have the right to data portability	HIPAA grants individuals similar rights to those granted in the UK, but they are more limited and specific to healthcare-related data.
Data transfers	The GDPR places restrictions on the transfer of personal data to countries outside the European Economic Area unless certain safeguards are in place	HIPAA does not specifically address international data transfers, but other regulations, such as the Privacy Shield and Standard Contractual Clauses, may apply
Enforcement and penalties	GDPR violations can result in significant fines, with penalties of up to €20 million or 4% of global annual turnover, whichever is higher	HIPAA violations can lead to civil and criminal penalties, including fines up to US$1.5 million per violation category, but the enforcement may vary across different states

HOW THE NHS CONNECTS DATASETS AS AN INVESTMENT FOR PUBLIC BENEFIT

There are initiatives focused on integrating various datasets within the NHS to create a comprehensive and interoperable health information system, with the goal being to improve patient care, enhance research capabilities and drive efficiencies within the healthcare system. By connecting datasets from different sources, such as electronic health records, hospital data, primary care data and social care data, the NHS aims to gain a holistic view of patient health, enabling healthcare providers to make more informed decisions and improve health outcomes. This initiative represents a significant investment in data infrastructure and technology to enable seamless data sharing and analysis for public benefit.[9–11]

While the specific details and progress of this initiative may vary over time, it underscores the importance of data governance and privacy considerations. The NHS is committed to ensuring that patient data is handled securely and in compliance with legal and ethical standards. Robust data governance frameworks, including consent mechanisms, anonymisation techniques and strict access controls, are put in place to safeguard patient privacy and maintain public trust. Investing in data connectivity and interoperability within the NHS holds the potential to unlock valuable insights, enable research collaborations, and enhance the delivery of healthcare services. However, it is crucial to ensure that these efforts are accompanied by strong governance mechanisms to protect patient privacy and maintain data security.

OUR RECOMMENDATIONS AND FUTURE CONSIDERATIONS

Understanding this regulatory landscape can be challenging for health and care providers, commissioners, research institutions and the general public—here we have provided an overview of six key legal and IG frameworks of which people should be aware to make the responsible use, linkage and sharing of data a less daunting prospect. We have provided references to further information and useful guidance, such as ICO's guide to UK GDPR[20]; although such guidance is not infallible, it is useful to guide secondary uses of data.

Through reflecting on good practice exemplars both nationally, for example, HDR UK, and regionally, for example, the YHCR, it is possible to work within these necessary legal and IG frameworks to enable analysis to answer a range of health and social care questions.[24 39 40] As another example, NHS Digital's national TRE in England uses the COPI Notice for its legal gateway for access to healthcare data.[41] The direction of travel in the UK is strongly towards TRE/Secure Data Environments and is government led, although there is still some merit in pairwise collaboration while this is put in place.[42] Pairwise collaboration in the context of TREs can have several benefits—table 4 gives an overview of some

Table 4 An overview of merits related to pairwise collaboration in the context of Trusted Research Environments (TREs)

TRE merit	Rationale
Enhanced expertise	Pairwise collaboration allows researchers from different institutions or organisations to bring together their diverse expertise and knowledge. This can lead to a more comprehensive understanding of complex healthcare issues and foster innovative approaches to research.
Increased data sharing	TREs aim to create secure environments for data sharing and analysis. Pairwise collaboration encourages researchers to share their data within these trusted environments, facilitating access to larger and more diverse datasets. This can lead to more robust and reliable research findings, as well as the potential for new discoveries.
Reduced redundancy	Collaboration between research teams can help prevent duplication of efforts. By sharing research goals, methodologies and preliminary findings, researchers can identify areas where their work overlaps and streamline their efforts. This can save time, resources and funding, allowing for more efficient progress in healthcare research.
Accelerated innovation	Collaborative research efforts often lead to accelerated innovation. Pairwise collaboration enables researchers to pool their resources, technologies and intellectual capital, creating synergistic effects. This collaborative environment can promote the development of new treatments, therapies and technologies that can have a positive impact on healthcare outcomes.
Improved data privacy and governance	TREs prioritise data privacy and governance, providing secure environments for data access and analysis. Pairwise collaboration within these environments ensures that researchers adhere to strict data protection protocols, maintaining patient confidentiality and upholding ethical standards. This helps build trust among researchers, institutions and the public, fostering a robust research ecosystem.
Access to specialised resources	Collaboration can provide researchers with access to specialised resources that may not be available at their own institutions. By partnering with other researchers, they can tap into unique datasets, equipment, facilities or expertise, enabling them to conduct more comprehensive and advanced research studies.
Cross-pollination of ideas	Pairwise collaboration facilitates the exchange of ideas, perspectives and methodologies. Researchers from different disciplines or institutions can bring fresh insights and alternative approaches to problem-solving. This cross-pollination of ideas can spark creativity and foster interdisciplinary research, leading to novel discoveries and advancements in healthcare.

of these benefits. It is important to note that while pairwise collaboration has its merits, it should be conducted within the framework of the TREs' guidelines and governance mechanisms to ensure data security, privacy and ethical considerations.

Enabling access to data for secondary uses through TREs complying with the Five Safes Framework reassures data suppliers and the public, compared with provisioning data extracts for individual projects—see also tables 1 and 4. The Five Safes model also improves efficiency by making it possible to make the same data available for many different, suitably approved, projects without the need for the data provider to disseminate multiple data extracts to different organisations. This efficiency was demonstrated by the use of NHS England's NHS COVID-19 Data Store TRE and the ONS Secure Research Service (TRE) hosted Public Health Research Database to generate evidence to inform the response to the COVID-19 pandemic. As of March 2022, NHS Digital's TRE is supported by a government announcement that up to £200 million is being invested to enable more secure and efficient access to NHS data through TREs and digital clinical trial services, with an initial focus on NHS Digital data to help researchers understand the impact of COVID-19 on cardiovascular disease and cancer.[43] This national TRE has been extremely useful as an indicator as to the power of unlocking data, hopefully paving the way for extended uses, with clear responsibilities and

checks/balances, for public benefit as recommended in the Goldacre Review.[12] Another example is Clinical Practice Research Datalink. This is one of the most extensive linked databases of primary, secondary and other care data. It has given many research opportunities in the past. There has been an effort to combine this data at the national level, but still, there are legal challenges, as described in this paper.

When enabling and using data for secondary uses, it is also important to have shared learning to avoid making mistakes in the future. A key partner in the responsible use of data is the public who both have a vested interest in the use and protection of data for which they are the subject, but also the beneficiaries and even funders (eg, via charities and public funded agencies) of care programmes informed by the associated generated evidence base. The public should be fully informed in a transparent and useful manner as to the secondary uses, users and public benefits of utilising RWD, including how data governance can support the whole process.[9 44]

CONCLUSION

Our article is a necessarily shortened overview of key legal and IG aspects to consider when accessing, storing or using RWD. It is geared towards those who want to use data for secondary uses, for public benefit, but feel they do not have sufficient understanding of data governance

to get started, or members of the general public who want to grasp a better understanding of data governance. We hope this article informs and encourages responsible access to health and care data for the realisation of public benefit.

ORCID iDs
Monica Catherine Jones http://orcid.org/0000-0001-7929-2808
Matthew Franklin http://orcid.org/0000-0002-2774-9439

REFERENCES

1 U.S. Food and Drug Administration (FDA). Real-world evidence: U.S. food and Drug Administration (FDA). 2022. Available: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

2 Makady A, de Boer A, Hillege H, *et al*. What is real-world data? A review of definitions based on literature and Stakeholder interviews. *Value Health* 2017;20:858–65.

3 Joore M, Grimm S, Boonen A, *et al*. Health technology assessment: a framework. *RMD Open* 2020;6:e001289.

4 Thiese MS. Observational and Interventional study design types; an overview. *Biochem Med (Zagreb)* 2014;24:199–210.

5 National Institute for Health and Care Excellence (NICE). NICE real-world evidence framework: National Institute for health and care excellence. NICE. 2022. Available: https://www.nice.org.uk/corporate/ecd9/chapter/overview

6 NHS Digital. hospital episode Statistics (HES): NHS Digital. 2022. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

7 NHS. *Health Survey for England: NHS Digital*. 2022. Available: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england

8 GOV.UK. *National Cancer Registration and Analysis Service*. 2020. Available: https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras

9 Baxter S, Franklin M, Haywood A, *et al*. Sharing real-world data for public benefit: a qualitative exploration of Stakeholder views and perceptions. *BMC Public Health* 2023;23::133.

10 Byrne N. *What is public benefit?: NHS Digital*. 2023. Available: https://digital.nhs.uk/blog/data-points-blog/2023/what-is-public-benefit

11 National Data. What do we mean by public benefit? evaluating public benefit when health and adult social care data is used for purposes beyond individual care: National data Guardian. 2022. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1124013/NDG_public_benefit_guidance_v1.0_-_14.12.22.pdf

12 Department of Health & Social Care. Better, broader, safer: using health data for research and analysis: Department of health & social care. 2022. Available: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis/better-broader-safer-using-health-data-for-research-and-analysis#executive-summary

13 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.

14 Frieden TR. Evidence for health decision making—beyond randomized, controlled trials. *N Engl J Med* 2017;377:465–75.

15 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887–92.

16 Jiang F, Jiang Y, Zhi H, *et al*. Artificial intelligence in Healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43.

17 Agrawal R, Prabakaran S. Big data in Digital Healthcare: lessons learnt and recommendations for general practice. *Heredity (Edinb)* 2020;124:525–34.

18 van Panhuis WG, Paul P, Emerson C, *et al*. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;14::1144.

19 legislation.gov.uk. *Data Protection Act 2018: legislation.gov.uk*. 2018. Available: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [accessed Feb 2022].

20 Information Commissioner's Office. *Guide to the General Data Protection Regulation (GDPR): ICO*. 2021. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/ [accessed 20 Feb 2022].

21 legislation.gov.uk. *The Health Service (Control of Patient Information) Regulations 2002: legislation.gov.uk*. 2002. Available: https://www.legislation.gov.uk/uksi/2002/1438/contents/made [accessed 21 Feb 2022].

22 legislation.gov.uk. *National Health Service Act 2006: legislation.gov.uk*. 2006. Available: https://www.legislation.gov.uk/ukpga/2006/41/contents

23 Legislation.gov.uk. *Health and Care Act 2022*. Available: https://www.legislation.gov.uk/ukpga/2022/31/contents/enacted [accessed 24 May 2023].

24 Health Data Research UK (HDR UK). Health data research UK (HDR UK): health data research UK (HDR UK). 2022. Available: https://www.hdruk.ac.uk/

25 Health Research Authority. GDPR guidance for researchers and study Coordinators: HRA. 2018. Available: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/

26 Information Commissioner's Office. *Information Commissioner's Office*. 2022. Available: https://ico.org.uk/

27 NHS Digital. *Data sharing standard 7b – Duty of Confidentiality: NHS Digital*. 2022. Available: https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance/data-sharing-standard-7b---duty-of-confidentiality

28 Health Research Authority. Guidance for CAG applicants: HRA. 2021. Available: https://www.hra.nhs.uk/about-us/committees-and-

services/confidentiality-advisory-group/guidance-confidentiality-advisory-group-applicants/

29 NHS Digital. *Control of patient information (COPI) notice: NHS Digital*. 2022. Available: https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/control-of-patient-information-copi-notice [accessed 20 Feb 2022].

30 Department of Health & Social Care. *Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002– general [withdrawn on 1 July 2022]: Department of Health & Social Care*. 2022. Available: https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information/coronavirus-covid-19-notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002-general--2

31 Department of Health & Social Care. *Coronavirus (COVID-19): notification to organisations to share information [expired on 31 October 2022]: Department of Health & Social Care*. 2022. Available: https://www.gov.uk/government/publications/coronavirus-covid-19-notification-to-organisations-to-share-information [accessed 24 May 2023].

32 Department of Health & Social Care. *COVID-19: notification to GPs and NHS England to share information [expired on 30 April 2023]: Department of Health & Social Care*. 2023. Available: https://www.gov.uk/government/publications/covid-19-notification-to-gps-and-nhs-england-to-share-information [accessed 24 May 2023].

33 Department of Health & Social Care. COVID-19: notification to Gps and NHS England to share information, may 2023 [set to expire 1 July 2023]: Department of health & social care. 2023. Available: https://www.gov.uk/government/publications/covid-19-notification-to-gps-and-nhs-england-to-share-information-may-2023 [accessed 24 May 2023].

34 Department of Health & Social Care. *Health and Social Care Act*. London: DHSC, 2012.

35 The Kings Fund. *The Health and Care Act 2022: our work to inform and make sense of the legislation: The Kings Fund*. 2022. Available: https://www.kingsfund.org.uk/projects/health-and-care-act-2022-make-sense-legislation

36 legislation.gov.uk. *Digital Economy Act 2017: legislation.gov.uk*. 2017. Available: https://www.legislation.gov.uk/ukpga/2017/30/contents/enacted

37 UK Data. *What is the Five Safes framework?: UK Data Service*. 2023. Available: https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

38 US Department of Health and Human Services. Health information privacy: US Department of health and human services. 2023. Available: https://www.hhs.gov/hipaa/index.html [accessed 24 May 2023].

39 Yorkshire and Humber Care Record (YHCR. Yorkshire and Humber care record (YHCR): Yorkshire and Humber care record (YHCR). 2022. Available: https://yhcr.org/

40 Health Data Research (HDR) UK. Health data research innovation gateway: HDR UK. 2022. Available: https://www.healthdatagateway.org/ [accessed 21 Feb 2022].

41 NHS Digital. *Trusted Research Environment service for England*. 2021. Available: https://digital.nhs.uk/coronavirus/coronavirus-data-services-updates/trusted-research-environment-service-for-england

42 Department of Health & Social Care. *NHS England's protection of patient data: Department of Health & Social Care*. 2023. Available: https://www.gov.uk/government/publications/nhs-englands-protection-of-patient-data/nhs-englands-protection-of-patient-data#technical-measures-and-controls [accessed May 2023].

43 NHS. *NHS Digital data to help researchers understand the impact of COVID-19 on cancer*. 2022. Available: https://digital.nhs.uk/news/latest-news/nhs-digital-data-to-help-researchers-understand-the-impact-of-covid-19-on-cancer [accessed May 2022].

44 Franklin M, Stone T, Baxter S, *et al*. *Unlocking Data to Inform Public Health Policy and Practice Sheffield: University of Sheffield*. 2022.