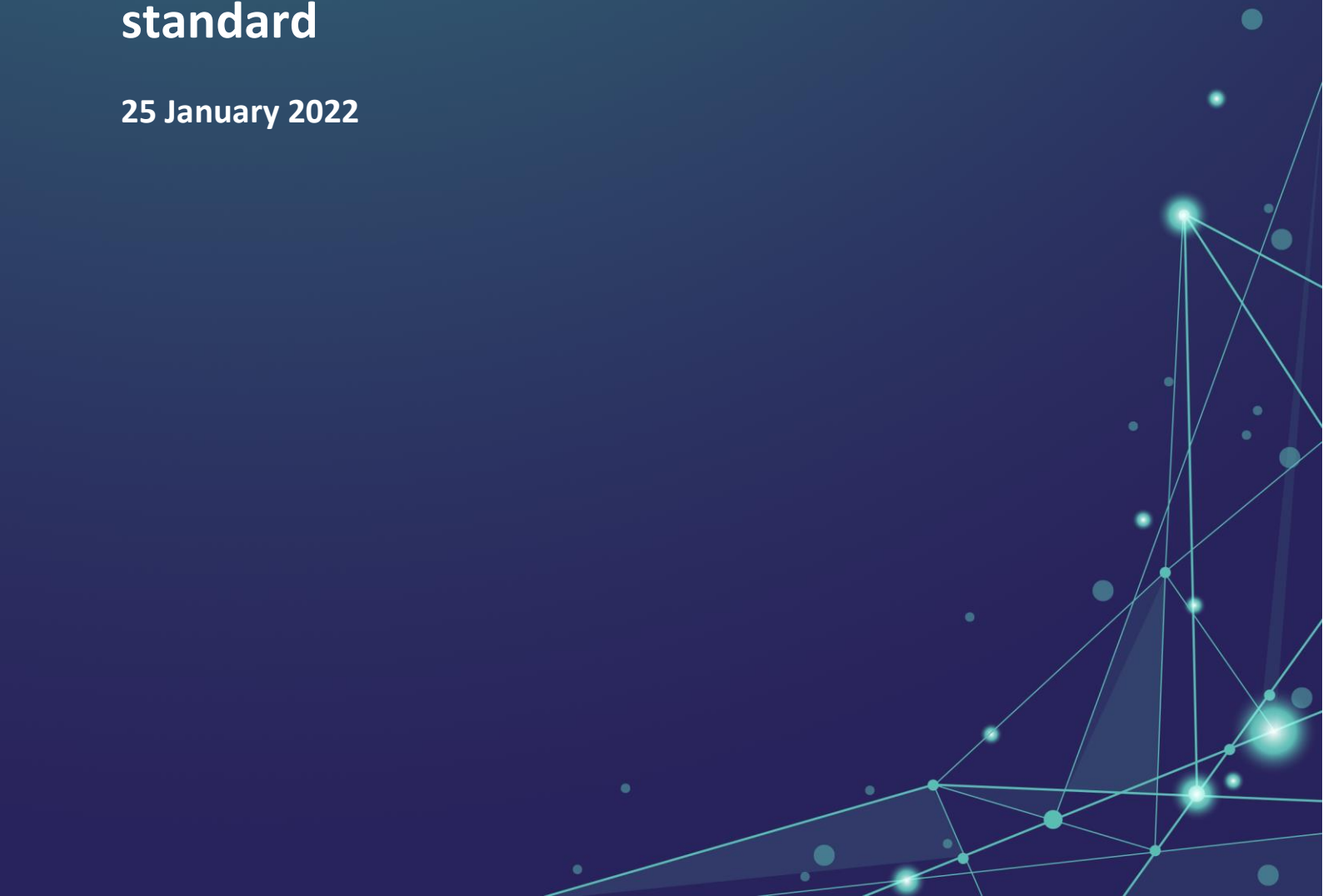
UK Health Data Research Alliance

# Improving transparency in the use of health data for research: Recommendations for a data use register standard

**25 January 2022**



UK Health Data Research Alliance

# Table of Content

UK Health Data Research Alliance

# Executive Summary

A data use register (also known as a data release register or list of approved projects) offers the public a clear record of how their data is being used, by whom and, most importantly, for what purpose. The April 2021 report 'Putting Good into Practice' by the National Data Guardian (NDG)[1] and Understanding Patient Data highlighted, *'transparency is required throughout the whole data life cycle, not just at the point of application'.*

There is also an expectation for the publication of *'clear statements of data users' credentials and sources of funding to protect against data manipulation, potentially non-altruistic motivations and hidden agendas.'* An appropriately designed and populated data user register that is openly available is a necessary step towards meeting the public's expectations and helping to demonstrate trustworthiness.

The UK Health Data Research Alliance brings together data custodians to establish best practice for the ethical use of UK health data for research and innovation at scale. It is therefore ideally placed to support the development and adoption of a data use register standard.

This paper builds on a Green Paper published in July 2021, which was initiated through the work of the Public Advisory Board of HDR UK. It incorporates valuable insights and feedback gathered through a period of consultation that included an open survey and several meetings and focus groups with various stakeholders. We are extremely grateful for the support received and the positive engagement of contributors from organisations representing health data research bodies, universities, data custodians and patient and public panels.

The response to the Green Paper has confirmed the need and value of data use registers in demonstrating trustworthiness through transparency in the use of health data for research. The recommendations remain largely unchanged given the strength of support received, particularly from the 75 members of the public who have provided feedback through surveys and focus groups. These recommendations are principally for data custodians but also require the active engagement of researchers and funders, in particular recommendation 5.

- Recommendation 1: All data custodians and controllers responsible for the collection, storage and sharing of data for the purpose of research, innovation and service evaluation should publish and actively promote a public record (data use register) of approved research studies, projects and other data uses.
- Recommendation 2: Data use registers should, as far as possible, be populated in near real time directly from information provided through the Data Access Request process to improve timeliness and accuracy of entries.
- Recommendation 3: Data use registers should be made available in both human readable and machine-readable formats to maximise their utility.

[1] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/977737/PGiP_Report_FINAL_1304.pdf

- Recommendation 4: Data use registers should have a consistency of format and content, based on the Five Safes framework, to enable ease of understanding and aggregation of registers.
  The recommended minimum data elements includes: organisation name, project title, lay summary, public benefit statement, latest approval date, dataset(s) name and access type (Table 1).
  A draft specification of all recommended data fields to be included in data use registers is included in Table 2.
- Recommendation 5: Researchers, data custodians and funders should use data use registers to close the loop on the impact of data use by including, where possible, links to research findings and other outputs as these become available. This is aligned with the #MakeItPublic strategy set out by Health Research Authority.

The analysis of the current state of data use registers across the UK highlights the different starting points for organisations and this will inevitably influence the pace of adoption and implementation. All organisations should have assessed their current state against the proposed standards within 3 months and a first wave of improvements should have been completed within 6 months. The Health Data Research Innovation Gateway provides a mechanism for organisations to meet the standard.

# Overview

## Purpose

Transparency in the use of health data for research and innovation to help demonstrate trustworthiness is the principal driver for this White Paper and so our public contributors have been critical to its development. We define a data use register (also referred to as data release registers or list of approved studies) as a public record of data an organisation has provided approved individuals and organisations secure access to for the purpose of research, innovation, and service evaluation. It typically contains information about the type of data being accessed, the purpose, date of approval and name of organisation using (or receiving) the data. Publication of data use registers is an opportunity for organisations to be as transparent as possible on how the data they hold is used.

National bodies have a legal obligation to make this information publicly available. Research databases are also required to share a minimum dataset, as outlined by the ethics committee in their conditions for ethical approval[2]. Publishing this information in an accessible and understandable format helps organisations demonstrate trustworthiness and, and potentially increase advocacy for data use.

By establishing a core set of standards on whom, how and why data is accessed, we hope to demonstrate the value and benefit of using health data, develop a culture of openness amongst data custodians, and generate better insight into health data usage.

A preliminary analysis of the current state of data use registers[3] has provided valuable insight into the practices of data controllers and custodians in publishing a register. Many of the data custodians and controllers regularly publish a public record of data uses, but almost half of them do not.

[2] QResearch. (2018, Dec 27). Conditions of Ethical Approval. https://www.qresearch.org/media/1155/257790-18em0400-qresearch-conditioins-of ethical-approval-27122018.pdf
[3] https://www.medrxiv.org/content/10.1101/2021.05.25.21257785v1

There is scope for considerable improvement and the development of minimum standards can be a first step to drive and influence change in practice.

## The role of the UK Health Data Research Alliance in creating a standard

The UK Health Data Research Alliance (the Alliance) brings together leading health, care and research organisations united to establish best practice to enable the ethical use of health data for research and innovation at scale. It comprises a variety of organisations including public bodies, NHS trusts, biobanks, cohorts, medical charities, health data research hubs and AI Imaging Centres of Excellence. All these organisations share the commitment to provide safe access to health data for research and innovation and drive best practice.

The Alliance Letter of Intent[4] highlights transparency of governance and operations as one of the founding principles. As such, publishing a register of projects, studies and/or organisations that have used data under a member's custodianship is a vital demonstration of this.

Data use registers are important for data custodians and the public, as well as to researchers and funders, but there has been a lack of standards or agreement of best practice. With this work, the Alliance has focused on aligning approaches to improving consistency and availability, and standardising the content, format, and frequency of publishing data use registers.

The standard is intended for both data controllers and data custodians that operate access arrangements on behalf of data controllers. Data custodians that are in the process of developing a public register will benefit from a framework and guidance on the recommended approach, whilst those with established registers are recommended to strive for alignment to the agreed standards. Alliance members can lead by example by adopting these standards and fostering best practice.

Development and adoption of these standards has not just been the preserve of data custodians. The perspective of patients and members of the public has been critical in identifying the minimum requirements of a data use register. Researchers and innovators are also impacted by the publication of these registers and how they influence the relationship between transparency and the competitive nature of research. They also have a responsibility and an interest to make the link to research outputs. Funders have a role to play in supporting adoption and influencing the behaviour of researchers and data custodians.

# Recommendations for a data use registers standard

The following recommendations for data custodians and controllers implementing data use registers have been shaped by the collective insight and experience of Alliance members and relevant stakeholders, the analysis of the current state of data use registers, and the shared learning on data transparency from other organisations. The recommendations have received widespread support with 93% of respondents endorsing

[4] UK Health Data Research Alliance (2020, Mar). Letter of Intent to Join the UK Health Data Research Alliance. https://ukhealthdata.org/wp content/uploads/2021/06/Alliance-Letter-of-Intent-Mar-20.pdf

them in the Community response to the Green Paper, with the majority of respondents (75 in total including focus groups) being members of the public (Appendix I)[5].

The high level of interest and input from lay representatives highlights the public interest in the need for transparency and the importance of involving the public in all stages of the data-driven research cycle. These recommendations have also been informed by recent insights from the HDR UK's Public Advisory Board[6].

However, feedback has highlighted a risk to accessibility, as the proposed recommendations are dependent on users with access to the internet and sufficient web competency. Addressing the digital divide is not limited to data use registers, it is common to all online communication models. Through implementation we will follow advice from relevant organisations to ensure we are as inclusive as possible in all communications around data use transparency.

It has also been noted that variation in infrastructure and capability amongst data custodians will impact the speed of adoption of 'high-level' or resource dependent standards. The minimum standard has been set so that it can be adopted by most data custodians.

Details of the responses can be seen in Appendix 1.

## Recommendation 1: Transparency

All data custodians and controllers responsible for the collection, storage and sharing of data for the purpose of research, innovation and service evaluation should publish and actively promote a public record (data use register) of approved research studies, projects and other data uses.

The consultation has indicated that clarification of responsibilities and scope is required.

Whilst all parties have a responsibility to be transparent about the use of individual-level data, the focus of this recommendation is the organisation or entity that provides access to data. In most cases this will be the data 'owner' or controller. We define data controller, as the individual or organisation that determines the purposes for which and the manner in which any personally identifiable data is or will be processed, as required by the Data Protection legislation. A data custodian is also responsible for the safe custody, transport, storage of, and access to data. In some instances, a data custodian may operate access arrangements on behalf of the data 'owner' or controller. An example of this is SAIL Databank, operated by the University of Swansea on behalf of a wide range of data controllers. Ultimately, it is the data controller that has the responsibility for ensuring that transparency of data use happens even where they delegate authority.

However, this recommendation should not prevent the research organisation(s) accessing the data also being transparent about what data they are accessing and for what purpose.

A data use register should be complementary to privacy information provided by organisations at the time of data collection.

[5] https://ukhealthdata.org/news/improving-transparency-in-the-use-of-health-data-for-research/
[6] https://www.hdruk.ac.uk/wp-content/uploads/2021/07/280621-PAB-Data-Access-procedures-paper-Building-trust-in-data-access-through-public-involvement-in-governance.pdf

The minimum scope of these standards in terms of data use is research. However, given that the distinction between research, innovation, service evaluation and audit can be blurred at times, we would encourage all uses of individual level data beyond direct care to be documented in a data use register.

Finally, the minimum scope in terms of who is accessing the data is external organisations. However, we would also encourage internal uses of individual level data beyond direct care to be documented in a data use register (including access enabled for the completion of audits and via honorary contracts).

Where multiple datasets are accessed from different data custodians in order to fulfil project requirements, this should be clearly stated, and information should be available in all relevant custodians' data use registers.

As stated in the Alliance Letter of Intent, signed by new members, all Alliance member organisations are expected to *'publish a register of active projects accessing the data under their custodianship and new data access requests received. The register should be easy to find (as defined by a simple, quick and intuitive user journey) on a public facing website, and in an accessible format that includes at least the minimum dataset defined by the Alliance Board.'*

The significance of accessibility along with the need for active promotion of data use registers was also highlighted by lay representatives in response to the Green Paper.

We would therefore expect Alliance members to lead by example for other UK organisations that provide access to health and social care data for research demonstrating the principle of transparency.

## Recommendation 2: Frequency

Data use registers should, as far as possible, be populated in near real time directly from information provided through the Data Access Request process to improve timeliness and accuracy of entries.

Regular updates and maintenance of data use registers are crucial to delivering public confidence and trust in why and by whom data is being used and demonstrating reliability. In the spirit of transparency, where possible, we recommend organisations update data use registers as soon as a data access request is approved and also specify when data is accessed. The consultation has highlighted two main reasons why a delay in making this information public might occur:

1. If the process for producing a data use register requires additional information that is not provided as part of the application process and this information is not requested until after approval.
2. To protect the privacy of the research team and/or nature of the project.

To address 1, the minimum data elements have been selected to include only information that would be reasonably expected to be collected as part of the data access request process. Therefore, whilst there may be enhancements to the data use register entry collected post approval, this should not be a reason to delay the initial publication of the data use. Of note, the Health Data Research Innovation Gateway (*the Gateway*) offers the opportunity to upgrade data access request processes without the need for each custodian to invest in new systems. We would therefore encourage organisations to explore using the Gateway as a route to improve both data access processes and the automation of publication of data use registers.

To address 2, whilst there may be legitimate cases where a data use register entry should be delayed or kept private, this should be the exception, not the rule. Access requests that have been agreed under the terms

of a non-disclosure agreement will make up some of these cases. However, in the spirit of transparency, we would encourage data custodians to disclose the number and type of organisations that fall within this category. Decisions to delay publication should, as far as possible, be reviewed by lay representatives to test for reasonableness.

Delays between approval and access to the data was also highlighted as a possible source of delay. To address this situation, the detailed specification (Table 2) includes both approval data and access data. Providing both of these dates also increases transparency of a custodian's 'fulfilment' process which may also help to reduce the time between approval and access.

During the transition to near real time updating, we also recommend that information in data use registers is no more than 3 months behind approvals.

## Recommendation 3: Format

Data use registers should be made available in both human readable and machine-readable formats.

The format of the register must be accessible for public and patients as our primary audiences. The format must also enable data controllers to meet legal obligations under principle 1 of the Data Protection Act[7] and provide insight for funders, researchers and other system stakeholders interested in the use of health data for research and innovation.

The public are more likely to find and understand information about data use that is embedded on an organisation's web page. However, not everyone will be able to find relevant websites or know where to look. In addition to making data use registers publicly available, custodians should also make an effort to promote their registers and include reference to them in their privacy information.

Whilst the power of the internet and the search functionality supports the discoverability and accessibility of data use registers to a global audience, there are people who either do not have access to or do not regularly use the internet. Consideration should be given to increase accessibility to all groups and increase equality in the accessibility to information about data use. However, this must not be an excuse to delay or limit the production of an online register which is currently the most accessible approach to sharing this information.

In addition to a searchable, well formatted data use register accessible on a web page or equivalent, researchers, funders, and other system stakeholders have identified the need for a machine-readable format that enables aggregation and analysis. This may have more detailed information included and can be seen as the canonical record at a point in time.

A machine-readable format is intended as data that can be processed by a computer without human intervention while ensuring no semantic meaning is lost (such as csv, JSON, XML) and must be structured data[8].

[7] UK Legislation. Data Protection Act 2018. Retrieved June 03, 2021, from: https://www.legislation.gov.uk/ukpga/2018/12/part/4/chapter/2/crossheading/the-data-protection-principles/enacted

[8] https://opendatahandbook.org/glossary/en/terms/machine-readable/

Both formats, human-readable and machine-readable, should be used by all custodians and controllers. However, with varying qualities and types of processes of data management across organisations, those two formats may be produced in different order by custodians.

## Recommendation 4: Content

Data use registers should have a consistency of format and content based on the Five Safes framework and an agreed specification to enable ease of understanding and aggregation of registers.

Use of the Five Safes framework[9] was endorsed as the best approach to drive standardisation of information about data use as it ensures that all aspects around who is accessing the data (Safe People), which data (Safe Data), for what purpose (Safe Projects), where data is analysed (Safe Settings) and what outputs have been generated (Safe Outputs) are considered.

Extensive research and consultation has informed the full specification[10] of recommended fields for a data use register (Table 2, GitHub). There is a residual risk of inconsistencies in the definitions of certain fields and different interpretations across organisations. Further work is likely to be required to ensure consistent definitions and standardised naming conventions are applied by publishers of data use registers.

Data custodians and controllers that have the resources and infrastructure are encouraged to strive for adoption of the full standard (or as much as relevant to their data assets). However, a tiered approach to the implementation of standards was identified as the most practical and sustainable method of adoption. Input from the public, researchers and custodians has been crucial to develop the data use register minimum data elements. These seven fields are considered the minimum for public and patient understanding of data use (Table 1).

As outlined above, this information would be expected to be captured through the data access approval process, so integration of data access approval processes with publication of data use registers may reduce the effort required to meet the minimum standard. All Alliance member organisations are expected to meet the proposed minimum set of data elements in their current data use registers and should publish as much, and as accurate, information as possible around data use, and strive to improve this over time.

### Recommended minimum standards for data use registers

**Table 1: Summary of the 7 minimum required fields for a data use register, grouped according to the Five Safes framework.**

| Five Safes | Field | Definition |
|---|---|---|
| **Safe People** | **Lead applicant organisation name** | The name of the legal entity that signs the contract to access the data. |

[9] https://www2.uwe.ac.uk/faculties/bbs/Documents/1601.pdf
[10] https://github.com/HDRUK/data-use-register

| Five Safes | Field | Definition |
|---|---|---|
| **Safe Projects** | **Project title** | The title of the project/research study/request that the applicant is investigating through the use of health data. |
| | **Lay summary** | A concise and clear description of the project, (e.g., as required by URKI in funding applications). It should outline the problem, objectives and expected outcomes in language that is understandable to the general public. Applicants should be encouraged to limit this description to a maximum number of words. |
| | **Public benefit statement** | A description in plain English of the anticipated outcomes, or impact of project on the general public. |
| | **Latest approval date** | The last date the data access request for this project was approved by a data custodian. |
| **Safe Data** | **Dataset(s) name** | The name of the dataset(s) being accessed. NB. Care will be required to ensure data uses accessing multiple datasets does not impact the machine-readable format. For example, if the data are provided in .csv form, then publishers should not delimit the list of datasets with the comma. |
| **Safe Setting** | **Access type** | Determines whether the data will be accessed within a Trusted Research Environment (TRE) or via the data release model. |

## Full specification of all recommended fields for data use registers

**Table 2: Summary of the recommended fields for a data use register, grouped according to the Five Safes framework (also accessible in GitHub).**

*Specifies the fields that have been proposed as the minimum standard for a data use register.

| Five Safes | Field | Definition |
|---|---|---|
| **Safe People** | **Lead applicant organisation name*** | The name of the legal entity that signs the contract to access the data. |
| | **Organisation ID** | A unique identifier for an organisation that is preferably an industry used standard such as Grid.ac. |
| | **Organisation sector** | The type of organisation that has signed a contract to access the data. |
| | **Applicant names** | The name of the Principal Investigator, as well as any other individuals that have been authorised to use the data. |
| | **Applicant ID** | ORCID identifier. This provides a persistent digital identifier that you own and control, and that distinguishes you from every other researcher. |
| | **Funders/ Sponsors** | The name of any funders or sponsors involved in the project. |

| Five Safes | Field | Definition |
|---|---|---|
| | **DEA accredited researcher?** | Depending on the type of data you are requesting, you might be required to become an accredited researcher. Most access to data in the Secure Research Service (SRS) will be by researchers accredited under the Digital Economy Act 2017 (DEA). This specifies the accreditation status of the principal applicant/researcher, as defined by the Office for National Statistics (ONS) Research Code of Practice and Accreditation criteria. |
| | **Sub-licence arrangements (if any)?** | Identifies whether there are any permissions for the applicant to share the data beyond the named parties. e.g., NHS Digital may approve a data release to the ONS, who then makes decisions about access to accredited researchers undertaking approved projects in their own trusted research environment. |
| **Safe Projects** | **Project ID** | A unique identifier for the project that is preferably an industry used standard, such as IRAS ID. However, for non-research projects, a unique reference number created by the data custodian on receipt of the application is sufficient. |
| | **Project title*** | The title of the project/research study/request that the applicant is investigating through the use of health data. |
| | **Lay summary*** | A concise and clear description of the project, (e.g., as required by URKI in funding applications). It should outline the problem, objectives and expected outcomes in language that is understandable to the general public. Applicants should be encouraged to limit this description. |
| | **Public benefit statement*** | A description in plain English of the anticipated outcomes, or impact of project on the general public. |
| | **Request category type** | This categorises the 'purpose of the share' (i.e., research, policy development, etc). |
| | **Technical summary** | A summary of the proposed research, in a manner that is suitable for a specialist reader. |
| | **Other approval committees** | Reference to other decision-making bodies that the project has already been authorised by. |
| | **Project start date** | The date the project is scheduled to start or actual start date. |
| | **Project end date** | The date the project is scheduled to finish or actual end date. |
| | **Latest approval date*** | The last date the data access request for this project was approved by a data custodian. |
| **Safe Data** | **Dataset(s) name*** | The name of the dataset(s) being accessed. |
| | **Data sensitivity level** | The level of identifiability of the data being accessed, as defined by Understanding Patient Data. |

| Five Safes | Field | Definition |
|---|---|---|
| | **Legal basis for provision of data under Article 6** | The lawful basis for processing is set out in Article 6 of the GDPR. At least one legal basis must apply whenever you process personal data. Processing shall be lawful only if and to the extent that at least one of the following applies. |
| | **Lawful conditions for provision of data under Article 9** | Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. This does not apply if one of the following applies. |
| | **Common Law Duty of Confidentiality** | If confidential information is being disclosed, the organisations holding this data (both the organisation disclosing the information and the recipient organisation) must also have a lawful basis to hold and use this information, and if applicable, have a condition to hold and use special categories of confidential information, and be fair and transparent about how they hold and use this data.<br>In England and Wales, if you are using section 251 of the NHS Act 2006 (s251) as a legal basis for identifiable data, you will need to ensure that you have the latest approval letter and application.<br>For Scotland this application will be reviewed by the Public Benefit and Privacy Panel.<br>In Northern Ireland it will be considered by the Privacy Advisory Committee. If you are using patient consent as the legal basis, you will need to provide all relevant consent forms and information leaflets. |
| | **National data opt-out applied?** | Specifies whether the preference for people to opt-out of their confidential patient information being used for secondary use has been applied to the data prior to release. |
| | **Request frequency** | Determines whether this is a 'one-off' request or a recurring dataset to be provided over a specific time period. |
| | **For linked datasets, specify how the linkage will take place** | Specifies whether applicant intends for the datasets to be linked with any additional datasets. Relevant information on the organisations undertaking linkages and how the linkage will take place must also be disclosed. As well as a summary of the risks/mitigations to be considered. |
| | **Approach to data minimisation** | Outline the measures taken to minimise the data in line with GDPR. |
| | **Description of the confidential data being used** | A description of the specific patient identifiable fields that have been included in the dataset(s) being accessed. |
| | **Release/Access date** | The date the data access was granted and active research started. |

| Five Safes | Field | Definition |
|---|---|---|
| **Safe Setting** | **Access type*** | Determines whether the data will be accessed within a Trusted Research Environment (TRE) or via a data release model. |
| | **How has data been processed to enhance privacy?** | Description of the tools or software used to reduce level of identifiable data being shared. |
| **Safe Outputs** | **Link to research outputs** | A URL link to any academic or non-academic research outputs, as they become available, including code used. |
| | **New data assets created** | Specifies the procedures and systems in place for the secure management, storage and archiving of data assets created as a result of linkage. |

## Recommendation 5: Link to outputs

Researchers, data custodians and funders should use data use registers to close the loop on the impact of data use by including, where possible, links to research findings and other outputs as these become available.

Data use registers have the potential to improve the efficiency of research by highlighting past and present research projects and data uses to researchers and funders, prioritising funding for high impact data assets, and identifying under-served areas of data collections or research that could be prioritised for future funding. These registers could also help close the loop on demonstrating public benefit by providing a link to the outputs and outcome of data use.

Demonstrating clear links between data use and impact (or lack of) will help to create a culture of transparency and openness was a key learning from the community. This is aligned with the #MakeItPublic strategy set out by Health Research Authority and needs to be part of a **system wide effort** involving researchers, data custodians, funders, and regulators. Research funding organisations widely highlighted the challenges around tracking outputs from the research they fund, so we have an opportunity to combine efforts to overcome some of these hurdles. We look forward to working with stakeholders, including research funders, to develop the levers to enable this to become routine practice.

This will not be easy, especially where there may be a significant gap between data access and output or impact. However, there are a range of developments and levers that can be used to support this recommendation. These include greater use of Digital Object Identifiers (DOIs) for data assets cited in research outputs, and making it a condition of future funding and/or access to data, and using stages. Having the Gateway as a common portal can help the process by centralising aspects of follow up with researchers and enabling automation where possible.

UK Health Data Research Alliance

# Towards implementation

Establishing a standard provides a consistent framework to support data custodians and controllers meet their moral, ethical and in some instances legal obligation to publish a register of approved research studies, projects and other data uses.

The readiness for change of each data custodian will vary and depends on the current depth and quality of their data use registers, and the infrastructure and resources available to support their transition to the recommended standard. This variability was evidenced through our research into the current state of data use registers, which highlighted a disparity in the content, functionality and timeliness of data currently published and more strikingly that almost half the organisations reviewed did not have a data use register publicly available.

It was highlighted during the consultation phase, that not all organisations have the resources to implement changes to their current practice. So, while we encourage all organisations to strive for adoption of the full specification of the recommended standard, the priority is for each data custodian to meet the first recommendation and at least have a public data use register, alongside the minimum data elements outlined in recommendation four. Furthermore, to ensure that the public are fully aware and understand how to access this information, it is essential that data use registers are actively promoted and are accessible to all communities.

The Gateway provides a 'ready-made' solution for data custodians without the infrastructure to develop their own data use register. The design and functionalities of the Gateway data use register have been heavily informed by the recommendations outlined in this paper. Adoption of the specification recommended in this paper across custodians will enable federation of data use registers between custodians and the gateway. As such the Gateway has the potential to provide a system-wide view on data uses. It provides an important step towards implementation of data use register standards and an exemplar model of how information of approved data uses is shared and managed.

Through the drive and leadership of the Alliance, there is great potential for this to become a priority for all data custodians moving forward.

UK Health Data Research Alliance

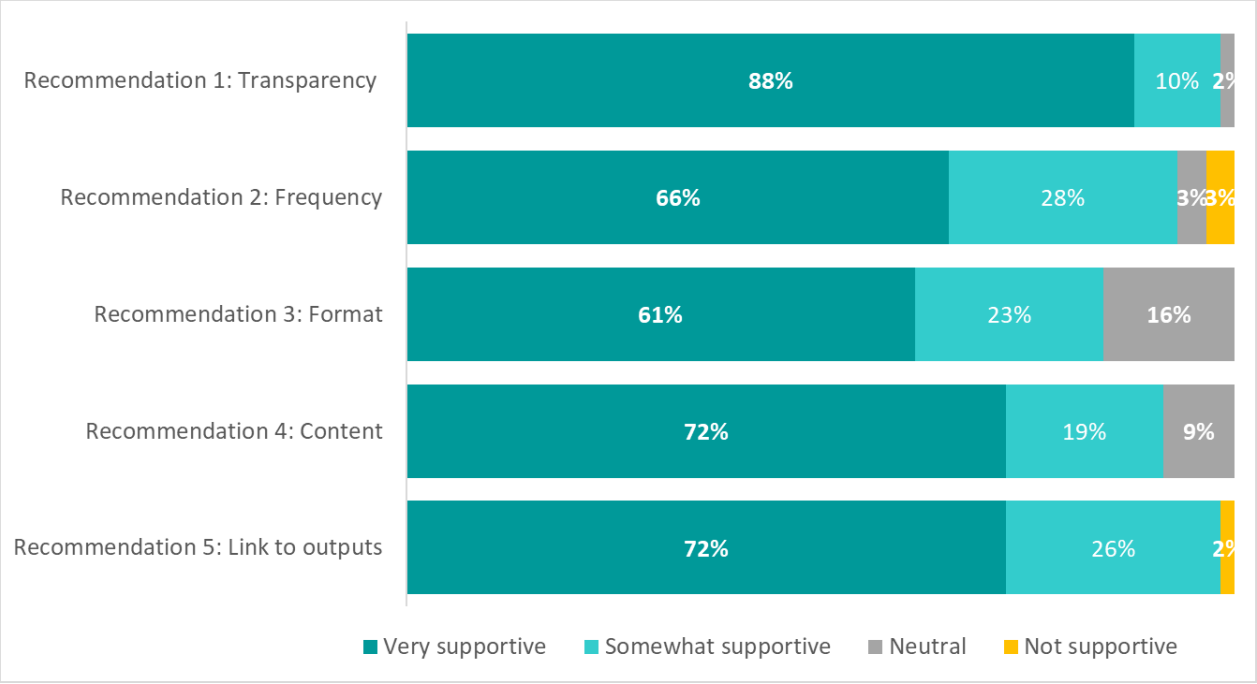
# Appendix I

## Community response to the Green Paper

In July 2021, the Alliance published a Green Paper on the emerging recommendations for a data use register standard, which went out for public consultation via an online survey. Sixty respondents completed the survey. In addition, focus groups were held with patient and public involvement panels, capturing a further 54 responses. A mix of researchers, custodians, innovators, and members of the public contributed to the consultation. The high level of interest and input from lay representatives emphasises the need for transparency and potential for increased public advocacy in data use.

The results from the survey, demonstrate the high level of community support for the recommended standard, with 93% of respondents either 'very or somewhat supportive' of all five recommendations (**Figure 1**).

**Figure 1: Sixty respondents communicate their level of support for the five recommendations on a data use resister standard**




The feedback received was also useful in capturing the community's perspective on the potential drivers and barriers to successful implementation of the standard, with the following themes emerging:

- **Accessibility through publication and promotion.** The considerable level of support for Recommendation 1 (Fig.1) highlights that the first key step to transparency is for data custodians to publish a register of any clinical or non-clinical research studies, projects and requests that have been approved for data use under their custodianship. However, it was also noted that the publication of registers must be supported by active promotion. To ensure accessibility, it's essential for data custodians and controllers to make the public aware of this information and how it can be accessed. Further to this, the proposed recommendations do not currently address the digital divide' or visual

impairments/disabilities that are likely to reduce accessibility. Whilst, this may not be addressed in the initial standard, it should be acknowledged as a risk to inclusivity and addressed in future iterations of the standard.

- **Sustainability by minimising the requirements.** Recommendation 4 proposes a consistency of format and content that is based on the Five Safes framework, ensuring that information around Safe People, Safe Projects, Safe Data, Safe Settings and Safe Outputs is captured. Whilst alignment to the Five Safes framework was well supported and praised as a "useful way of conceptualising information security", providing an attainable minimum standard by identifying the "key mandatory elements" was also noted as necessary for the long-term adoption and sustainability of data use registers.
  In view of this and supported by the input of various stakeholders we have proposed a minimum standard for publication (Table 1). The data elements recommended would be a first step to transparency and offer a foundation that can be developed over time.

- **Timeliness through automation.** The potential benefit of a more diversified research landscape through the provision of "real time" information (Recommendation 2) was recognised by the majority of respondents. By providing researchers with a timely and accurate view of past and present research projects and data uses, we could reduce duplication, whilst also encouraging funders to increase investment in underserved areas of research. However, the practical limitations of this recommendation were also noted. As a transition to real-time data is heavily dependent on the automation of data access management systems, it was acknowledged that not all organisations would be in a position to move on with real time updates of data use registers.

- **Closing the loop on impact through links to research outputs.** Demonstrating the impact of data used for research and innovation was recognised by all sections of the community as an essential final element in the data access 'life cycle', with 98% of respondents supportive of Recommendation 5. Patient endorsement and advocacy of health data usage through better communication of research outputs was a benefit obvious to most. However, respondents also acknowledged that the link to research outputs is not currently implemented in a consistent and effective manner.
  Linking data uses to both academic and non-academic research outputs was recognised as an ambitious target, requiring the joint effort of data custodians, funders and researchers. In addition, a better understanding of existing reporting platforms, data sharing repositories and a more effective use of acknowledgements and DOIs could help increase transparency of how data is used and what the impact of that data use is. Data use registers provide a great opportunity to establish and drive best practice in this important area.

UK Health Data Research Alliance

# Appendix II

## Organisations contributing to development of recommendations and consultation

**Table 3: List of organisations (including health data research organisations and universities, data custodians and patient and public groups) that responded to the public consultation and/or contributed through interviews and workshops.**

| | | |
|---|---|---|
| Anglia Ruskin University | Imperial College London | Public Health Scotland |
| Beat Kidney Stones | Imperial College London Neonatal Data Analysis Unit | QResearch |
| BHF Data Science Centre | Independent / HDR-UK / Genomics England | Queen Mary University |
| Bristol University | Institute of Cancer Research | Roche |
| Cancer Research UK | King's College London | SAIL Databank |
| Cardiff University | Lancaster University | SPNFT |
| City, University of London | London Hazards Centre | St. Bartholomew's Hospital |
| Clinical Practice Research Datalink (CPRD) | MedConfidential | The Brain Tumour Charity |
| DATA-Can Health Data Research Hub | Medical Research Council | The Renal Registry |
| Edinburgh University | Moorfields NHS FT | UCL and Ixico plc |
| Flat Iron | MRC CTU at UCL | University College London |
| Great Ormond Street Hospital NHS Trust | NHS Digital | University Hospitals Birmingham NHS Trust |
| Gut Reaction Health Data Research Hub | NHSX | University of Dundee |
| Guys and St Thomas NHS Foundation Trust | NIHR BioResource | University of Nottingham |
| HDR UK Cambridge Understanding Causes of Disease | NI Cancer Research Consumer Forum (NICRCF) | University of South Wales |
| Health and Social Care Northern Ireland | Nottingham University Hospitals NHS Trust | University of St Andrew |
| Health Research Authority | Nottinghamshire Healthcare NHS Foundation Trust | University of Warwick |
| HIC – The University of Dundee | Our Future Health | use MY data |
| HQIP | Oxford University Hospitals NHS Foundation Trust | Usher Institute, University of Edinburgh |
| Human Fertilisation & Embryology Authority | Patient & Public Involvement Panel | Wellcome Sanger Institute |
| Imperial College Healthcare NHS Trust | Public Health England | |