

Data Policy Models for Funding Bodies

A report prepared for the Wellcome Trust, to evaluate models for policy and advice on access, governance and management of data generated by funded research, with reference to the Expert Advisory Group on Data Access (EAGDA).

Data Policy Models for Funding Bodies

1. Executive Summary	3
2. Recommendations	4
3. Introduction	5
4. Overview of the Expert Advisory Group on Data Access (EAGDA)	6
i. Establishment and Purpose	6
ii. Membership	6
iii. Reporting	6
iv. Activities	7
5. Research Findings	9
Q1: How are UK funders currently advised and aware of issues and topics relevant to health-related research?	9
Q2: How could UK funders of health-related research be advised and made aware?	10
Q3: What are the issues and topics that UK funders need to be aware of and advised on both now and in the future?	11
Q4: Is there any gap in advice or awareness currently drawn on by, or available to, UK funders?	13
Q5: How should the quality and success of advice and awareness be evaluated?	13
Q6: What was effective about the EAGDA model and what does evaluating it tell us about evaluation methods?	14
Q7: What new model for wider cross-funder advice and awareness is recommended and how might such a model be evaluated?	17
6. Conclusion	19
7. Appendices:	20
Appendix 1: Project Methodology	20
Appendix 2: Summary of Empirical Research	23
Appendix 3: Reflections on the future of EAGDA made during the final meeting	24
Appendix 4: Bodies in the UK landscape for provision of policy or direction of data sharing	26
Appendix 5: Direction from UK bodies on the sharing of human research data, 2012-2018	35
Appendix 6: Data related issues	45
Appendix 7: Cross-funder Policy Committees acting in the UK and elsewhere	47

Project Manager: Dr Sarah Coy (sarah.coy@dph.ox.ac.uk)
Centre for Health, Law and Emerging Technologies (HeLEX), University of Oxford

September 2018

1 Executive Summary

There was universal agreement that the work of EAGDA was of extremely high quality and that it tackled valuable questions, however how to use the advice generated within the expert committee was more challenging.

Challenges to implementation included difficulty engaging the funding organisations on the issues raised, which were deemed to be important, but not necessarily urgent. Part of this challenge was getting the right people in the room to action the recommendations – too senior and they were too busy to attend, too junior and they lacked the authority to make immediate decisions. In instances where funding bodies considered action to have been usefully taken, it was recognised that a champion, with access to senior decision-makers, was involved to push the agenda forward.

A further challenge was the mismatch in time-lines for conducting robust research to develop advice, compared with reactive policy-development in response to changes in the political climate. Both qualities, rigour and agility, were deemed valuable, but not easy to support using the expert panel model, that maintained buy-in from expert (and very busy) panel members while allowing for meaningful collaboration at short notice.

There was difference of opinion in whether the group should disseminate findings more widely, and it was finally agreed that a compromise could be to report headlines directly to the data users, but that the full details of recommendations were intended for funders only, to allow them the freedom to choose not to follow advice. With the caveat that lengthy reports were unnecessary, and considering alternative methods of [more succinct] reporting would be helpful.

Members and funders generally agreed that having some sort of EAGDA in the future would be useful, as the challenges it addressed would not be solved soon, and would continue to be of extreme importance to the wider community, and thus benefit from cross-funder discussion. Before determining exactly what that model might look like, it was clear that a number of questions needed to be raised and discussed amongst the funders first.

Once funders have agreed on the basic principles for the future of EAGDA, a new model can be developed to meet their needs, which allows big-picture appraisal of the environment for data sharing, with cross-funder views represented, and a clear route for implementation of advice, and review and evaluation of progress made as a result of the advice.

2 Recommendations

The following recommendations have arisen from the review of the policy and advice landscape, and the evaluation of EAGDA.

- 1) **We recommend that the funders involved in EAGDA consider a new model for its replacement as there is still considerable appetite for such a body to provide advice on pressing issues relating to data management and sharing**
- 2) **Before a new model for EAGDA can be devised, we recommend that the funders involved agree upon the following questions:**
 - What is the aim?
 - What is the scope and remit?
 - Who should be included?
 - How can buy-in from the funding organisations be achieved, and at what level within the organisation?
 - What is the process for accountability?
 - Should horizon scanning be incorporated, and if so how can it be usefully conducted?
 - What is the level of financial resource (and organisational support)?
- 3) **We recommend inclusion of several ‘vital’ elements. The new model must:**
 - Maintain a cross-funder approach
 - Include multi-disciplinary expert members
 - Avoid duplicating efforts (and therefore needs a method for keeping up-to-date with other work in this area)
 - Consider, and support, the processes required to enable advice to be implemented
 - Ensure there is a clear communication strategy between members and funders
 - Encourage funders to focus on their own working relationships when reporting back (with a clear process for disseminating findings internally)
 - Have sufficient resources to support a secretariat if one is seen to be needed
- 4) **A clear process for evaluation would help demonstrate the impact and reach of the model’s outputs. For this we recommend:**
 - a. A clear definition of success
 - b. A procedure for recording action and activity beyond the group, for sharing subsequent actions with the group, and for demonstrating the path from the new model through the funders and to other audiences
 - c. Funders consider whether the new model should be involved in setting data standards, and if so how this would be supported
- 5) **The new model would benefit from a clear strategy for communication and dissemination, including whether and how to engage with different audiences, particularly:**
 - Funding bodies
 - Data users
 - Data policy network
 - General public

3 Introduction

This research was commissioned by the Policy Unit at the Wellcome Trust (WT) to investigate potential models for providing strategic cross-funder advice on data access, use and management. It was conducted by the HeLEX Centre, University of Oxford (during August – September 2018).

The aim of the research was to inform policy decisions on how funders can work with the research community to ensure the best policy and operational approaches to access, governance, and management of data generated by funded research. It sought to provide an overview of the different organisations and groups operating in this area, with specific consideration of the role of the Expert Advisory Group on Data Access (EAGDA).

The research involved:

- a review of existing literature to map the UK landscape for provision of advice on data sharing;
- interviews with EAGDA committee members;
- interviews with key WT staff involved with EAGDA;
- focus groups with policy officers in WT, the Economic and Social Science Research Council (ESRC), the Medical Research Council (MRC) and Cancer Research UK (CRUK).

The full project methodology is outlined in Appendix 1, with a summary of empirical research in Appendix 2.

4 Overview of EAGDA

Establishment and Purpose

EAGDA was convened in 2012 by the ESRC, MRC, CRUK and the WT, in response to a perceived need to discuss challenges in data sharing, to provide advice and policy guidance that could be adopted more widely than by the individual institutions involved. The EAGDA Terms of Reference describe its remit as:

“The key functions of the Expert Advisory Group (the “Group”) will be to:

- a. advise UK funders¹ on emerging scientific, ethical and legal issues associated with the governance of data access for human genetics and cohorts studies.
 - i. EAGDA can bring relevant issues to the funders’ attention; and
 - ii. Examine and respond to particular issues as directed by the funders.
- b. support these studies - acting as a recognised UK authority on best practice in the governance of data access to which studies (through their funders) can refer new and problematic issues of principle they confront.
- c. maximise UK input into international policy discussions - strengthening the UK interface with relevant international groups and helping to ensure that funders and their representatives articulate a common set of effective messages internationally.”

The group promoted good working practices, consistent governance, and transparent decision making for managing and using data from cohort studies.

Focus Group 1: “Rather than just being advice to one funder, by joining up with a series of funders it was a much more collaborative exercise, and helped us all to be taking decisions together because it was the same unknown for everyone. It made much more sense to share the learning and to share the advice”

Membership

EAGDA comprised of 13 experts,² including leading academics and representatives from government bodies. Expertise spanned genetics, epidemiology, social sciences, statistics, IT, data management and security, law, and ethics. Additionally the group included representatives from each of the EAGDA funders, with the WT providing a secretariat.

Reporting

Primarily the EAGDA reported to the funding bodies that supported its activity. Funders would submit formal responses to the reports that were collected by the secretariat. Reports made recommendations directed towards the following key constituents:

- Funders
- Data Repositories
- Universities
- Research Institutions
- Consortia
- Learned Societies
- Data Controllers
- Publishers

¹ Current funders involved in this project are the Wellcome Trust, MRC, CRUK and the ESRC.

² <https://wellcome.ac.uk/sites/default/files/eagda-members-jan16.pdf>

Reports were disseminated to an EAGDA mailing list and through research networks. EAGDA activity was reported on subpages of the WT website, with summary information also provided on the funder’s individual websites.

Activities

EAGDA provided support to current and future studies and, where relevant, their Data Access Committees (DACs), across the fields of genetics, epidemiology and the social sciences – promoting best practice and encouraging consistency in governance and decision making. It also sought to enhance input into policy discussions on data access, both in the UK and internationally, including the following topics:

- Trustworthy Research environments
- Risks/harms of data breaches and sanction mechanisms
- Enabling linkage including guidelines, governance and advice on best practice.
- Considerations for commercial access
- Challenges for data infrastructures
- Legislation that concerns data access and reuse outside of the domain of scientific research
- Legal control and responsibility for data
- Sample access strategies and data management plans
- International links and reciprocity
- Clinical trial data and transparency
- Public trust in data sharing
- Recording metrics and affording recognition to data sharing activities

It achieved this by commissioning research and providing expert advice. Table 5.1 catalogues the group’s key outputs.

Data Terms	Report	Date
Sanctions and accountability	Report with recommendations	Apr-17
Data management plans	Report with recommendations	Aug-17
Best practices for infrastructure development and data curation	Minutes	Oct-16
Governance of data access	Report with recommendations	Jun-15
Risks of harm from data misuse	Review	Jun-14
Incentives to support data access	Report with recommendations	May-14
Protecting participants’ confidentiality	Statement with recommendations	Oct-13

Table 1: EAGDA outputs

Additionally, the EAGDA fed into funder discussions on:

- Health and Social Care Information Centre (HSCIC) and data sharing problems following the Partridge Review
- Establishing [clinicalstudydatarequest.com](#) to enable access to clinical trials data from multiple pharmaceutical companies
- Possible UK funder synergies with the National Institutes of Health (NIH)’s Big Data to Knowledge (BD2K) initiative
- Public trust and commercial access to data
- Funders’ response to researchers not fulfilling data sharing requirements
- Best practice in sample sharing
- Responding to Nuffield Council on Bioethics (NCoB)’s report on uses of biological and health data
- Principles for Trustworthy Research Environments
- The UK Anonymisation Network (UKAN) Anonymisation Framework

- Funders' responses to the Science and Technology Committee's inquiry into 'The Big Data Dilemma'
- WT's cohort strategy and ESRC's Longitudinal Studies Review
- Responses to the PACE (Pacing, graded Activity, and Cognitive behaviour therapy; a randomised Evaluation) tribunal decision
- Responses to the DeepMind/Royal Free data sharing agreement

EAGDA's term of funding ended in June 2017; this review, commissioned by the partner funders (co-ordinated by WT), evaluates the Group's role, activities and operations to ascertain whether it should continue for a further period, or be modified or terminated.

Appendix 3 summarises reflections on the future of the EAGDA made during the group's last meeting, 18th April 2017.

5 Research Findings

The research was split into 3 stages (fully outlined in Appendix 1):

Stage 1: information gathering

Stage 2: analysis and benchmarking

Stage 3: check and challenge

Seven questions were addressed within the research, the findings for which are outlined below, drawn from a mixture of desk-based and empirical research.

Q1 How are UK funders currently advised and aware of issues and topics relevant to health-related research?

The UK landscape for policy and advice relating to data sharing consists of a variety of organisations or initiatives with specific interests in this area, including:

- i. Funding bodies
- ii. Projects and initiatives to drive, direct or study the UK landscape
- iii. Regulatory bodies
- iv. Think tanks

There are a number of different organisations and groups recognised to be active in this area (see Appendix 4), considering issues (such as data availability, access to expert advice on information governance and data protection, etc.) and cross-cutting topics (such as artificial intelligence, cellular computing, etc.) relating to data sharing and access. As the conversation has progressed from whether data should be shared, to how best to manage sharing, this space has become increasingly crowded. The different needs of different types of data, the different technologies implicated and the different stakeholders involved means that these different players have overlapping interests, but maintain particular view points, or represent specific constituencies, which lends a specific focus to their work. Crucially, whilst the need for effective and robust governance of human participant data is recognised by funders and by the research community, conversations that relate to emerging requirements and future directions continue to occur largely in siloes.^{3,4}

EAGDA was therefore established to provide strategic advice across the funding bodies at a time when data sharing was on the rise and the number of difficult questions for funders was increasing. It was intended to provide governance advice at the data access committee level and inform better structuring of data access mechanisms, as well as providing a more strategic and overarching outlook through horizon scanning and forward looking activities. The significance of tackling these challenges at a funder, rather than research institution level, is the recognised need to apply standards or (where possible) universal approaches. Because of the procedure for consideration of research grant proposals and programmes of work, allowing detailed consideration of data management practices alongside an overview of the data to be collected, used, and potentially shared, the funders are well-placed to guide their communities in this issue (compared with, for example government agencies that will not have such a well-established direct link with the data users, to know the sorts of questions arising, or the practical implications). However, there is considerable overlap between communities, with researchers (or 'data-users') seeking funding from different

³ Appendix 4 provides a breakdown of the individual bodies represented in each category.

⁴ Appendix 5.i catalogues key directives published by UK funding bodies and other relevant organisations between April 2012 and July 2018 that are relevant to the governance of human participant data. In addition to the organisations listed in Appendix 4, this includes guidance published by the EAGDA. A number of oversight and advisory bodies have additionally been established whose remit relates to the sharing of human participant research data; Appendix 5.ii provides examples.

funding bodies, therefore finding common ground and a shared direction across funders is hugely beneficial to all involved.

EAGDA was perceived by interviewees and focus group participants as unique in its function and remit as a cross-funder policy committee which convened 'leading researchers on issues of data access and sharing who advise the funders on technical and legal issues in data sharing'. Nonetheless, there are other bodies with cross-funder participation that tackle questions relating to data management and governance, operating in the same space. Such bodies, in the UK and elsewhere, are summarised in Appendix 7 together with the issues they are addressing, and features of their organisational structure.

Findings from empirical research

Focus group participants and interviewees mentioned many of the organisations listed in the appendices as potential sources of information on these issues. However there was a distinct perception about the level at which different bodies pitched their advice, or the specific questions that would be of interest to them, or their constituents, and how that differed from EAGDA. There was universal recognition for the unique position that EAGDA held, with several participants asserting that one of the key differences with EAGDA was perhaps that the agenda was set both by the funders involved, but also by the expert members, which provided a different emphasis than in other bodies that were attending to very specific community needs or answering immediate questions. The scope to follow leads and explore questions in more depth, over a longer period of time, often collecting robust evidence to guide discussion, to hear from experts and to draw upon a variety of experiences and viewpoints, perhaps allowed a more 'academic' consideration of challenges than other bodies might be able to support.

Q2 How could UK funders of health-related research be advised and made aware?

Some of the questions tackled by EAGDA have been considered by funders through the use of informal or ad hoc committees. The focus group with MRC representatives revealed that there are 'data policy groups' where members of the MRC and ESRC meet to share knowledge. Individuals within WT and CRUK reported communicating regularly with the members of these groups and WT has its own internal group to discuss data governance topics. Many interviewees and focus group participants also stated they made use of contacts within their own networks when faced with a data governance query, particularly if relating to practical or procedural approaches to take, and in fact often would call upon these contacts more readily than considering EAGDA as a possible source of advice.

One example of an ad hoc committee, was a group convened by the MRC⁵ to craft their 'policy and guidance on research data'. Calling on colleagues from a group of funders that had 'similar, well-developed research data policies' that included WT, the ESRC, CRUK and the National Cancer Research Institute (NCRI). While the group was specifically intended to assist with the MRC's agenda, its discussions regarding research data sharing, access, and management advice would have been helpful to all funders.

The 'Concordat on Open Research Data,' which aims to 'provide clear and practical principles for working with research data that cover the many roles needed to support the research process' was the product of another example of multi-stakeholder group working.⁶ The Higher Education Funding Council for England (HEFCE), WT, Universities UK (UUK), and Research Councils UK (RCUK) were the four original signatories and it has subsequently been endorsed by other funders including CRUK and the Scottish Funding

⁵ <https://mrc.ukri.org/documents/pdf/data-sharing-from-population-and-patient-studies/>

⁶ <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>

Council (SFC). These informal groups or multi-stakeholder collaborations are narrower than EAGDA in scope but cover very similar subject matter. A preference of funders for ad-hoc groups or collaborations may be partly driven by the scepticism some of the empirical research revealed about the utility of a standing committee like EAGDA, because of the resource required to support it. However others challenged this viewpoint, citing the difficulty convening a group of experts at short notice, given how busy and time-limited these individuals can be. Having a group already identified, with common goals, shared experiences, and having already made a commitment to contribute, had significant benefit (not least as a framework to organise regular meetings in advance).

There are other, more 'stable', bodies and spaces, which are not 'policy committees' but within which cross-funder deliberation on health data sharing policy matters takes place (see Appendix 7 for further detail), that are relevant to this area, including bodies tackling similar issues on an international level, often with significant involvement of UK funders. These organisations operate at a different level again from the work of EAGDA, with additional responsibilities, constituents and objectives.

Q3: What are the issues and topics that UK funders need to be aware of and advised on both now and in the future?

Since the establishment of EAGDA in June 2012 the data 'playing field' has evolved;⁷ evaluating the requirements of, and need for, a future model demands understanding of the nature of issues upon which such a body would advise. While this question will ultimately depend on the funders' intentions with regard to the remit of the body (a question that should be broached by the funding bodies in response to this review), a review of central issues relevant to data access in the academic literature reveals that they can be meaningfully distinguished as (a) broad, strategic issues or (b) more specific, ethico-legal, technical or scientific issues.

a) Strategic issues

Given the broad support for promotion of good practices in data management and sharing, there is increasing discussion of the underlying principles and criteria to govern these practices. Two proposals that have been getting traction and receiving policy support in this area are the 'FAIR Principles' (that data should be Findable, Accessible, Interoperable and Reusable)⁸ and the Global Alliance for Genomics and Health (GA4GH) 'International Framework for Data Sharing' (which outlines the Global Alliance's use of data governance policies).⁹ There are also relevant initiatives by research consortia, aiming to address this demand in more particular contexts, such as the principles that should govern the operation of data safe havens.¹⁰ A cross-funder body could consider adopting a principled position as to whether there is scope for reaching a consensus on a set of relevant criteria, allowing the funder community to play a significant role in their creation or amendment in the future, specifically considering their communities' interests in, and reactions to the policies recommended.

Another important strategic question relates to the range of stakeholders that should be engaged in designing and coordinating data governance principles and policies, as well as determining ways of engaging with them. EAGDA had noted¹¹ that:

⁷ See Appendix 3 for reflection of future issues from the EAGDA chair

⁸ Wilkinson MD, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016. 3

⁹ Rahimzadeh V, Dyke S, Knoppers BM. An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health. *Biopreservation and Biobanking*. 2016. 14

¹⁰ Burton P, *et al.* Data Safe Havens in Health Research and Healthcare. *Bioinformatics*. 2015. 31. 20

¹¹ Minutes from EAGDA 4/17 meeting.

There is currently little coordination between the multiple groups working on issues relating to data use and governance in research and conversations continue to be siloed. There's inertia and people have good reasons to maintain their siloes so funders need to take these challenges seriously and work out what problems they want to address.

While specific funders have expressed commitment to 'urge strategic thinking across research communities to debate and establish appropriate data governance principles and standards for different settings, with discussions among scientific stakeholders being necessary',¹² there is still much-needed progress to be made. The work of some of the funders,¹³ as well as other voices in the literature,¹⁴ provide directions for the adoption of a governance framework that is more inclusive for the people who provide the data, including low or middle-income researchers, patients, research participants, groups or communities. A cross-funder body should raise awareness of stakeholder inclusion issues, making sure that funders are attentive to the social justice potential of open research and are working towards realising it.

b) Ethico-legal, technical, or scientific issues

Beyond the strategic concerns, there is a set of emerging 'ethico-legal, technical or scientific' problems around data sharing for health research. Ensuring effective governance requires that all processors in the data stream understand their roles and responsibilities regarding the data they handle; and movement of data between contexts (such as across disciplinary divides, over national boundaries, and beyond expected usage) introduces governance challenges that arise from a disparity of cultural expectations with regards to data ownership, alongside governance challenges resulting from local regulation and the need for technical compatibility. While a list of such problems can hardly be definitive, some examples that have already been faced and (to an extent) tackled by large research projects are outlined in Appendix 6.

Findings from empirical research

Funding body representatives describe the role of EAGDA as being potentially two fold, providing advice around data access mechanisms at the data access committee level, and providing a horizon scanning function, offering something more forward looking to co-develop strategy. Concerns were raised that horizon scanning can never be an accurate science, but the sentiment amongst EAGDA members and funder representatives was that the cross disciplinary expertise that EAGDA provided offered a unique forum for engaging discussions around emerging issues, before they would otherwise come to the attention of those responsible for formulating policy. One example given was that of blockchain, a secure method for data storage and distribution that has since been proposed as a potential method for storing records of consent, as well as for annotating genome records. The committee admitted that blockchain was raised as a potential technology to be discussed, but without clear and immediate relevance to the funders, other issues were prioritised instead, which demonstrates the friction between solving existing problems and identifying future topics or issues. Across different members and funders there was lively discussion about how much focus should be placed on horizon scanning, and whether this was the strict purpose for the group, however the overarching conclusion was that it was an important element, and should remain on the agenda.

¹² Burton PR, Banner N, Elliot M, Knoppers M, Banks J, Policies and Strategies to Facilitate Secondary Use of Research Data in the Health Science. *Int. J Epidemiology*. 2017. 46. 6.

¹³ Bull S, Ensuring Global Equity in Open Research (WT-commissioned report) 2016.

¹⁴ Kaye J, *et al*, Including all voices in international data-sharing governance. *Human Genomics*. 2018. 12. 13

Q4: Is there any gap in advice or awareness currently drawn on by, or available to, UK funders?

While there are several organisations and groups interested in this area, it was widely recognised by both interviewees and focus group participants that such bodies tended to be answering specific questions or interested in very particular issues, while EAGDA operated at a different level that allowed a more comprehensive appraisal of the wider issues. There is overlap with other groups but not direct duplication of effort, and on this basis it was highly valued.

One interviewee reported: "Rather than just being advice to one funder, by joining up with a series of funders it's was a much more collaborative exercise, and helped us all to be taking decisions together because it was the same unknown for everyone. It made much more sense to share the learning and to share the advice"

The issues and topics identified in the focus groups and interviews as being of relevance to UK funders when considering data access and sharing mirrored those identified in the literature (Appendix 6). One of the questions discussed more widely was whether it was possible to develop cross-funder consensus.

Participants were unanimous in the understanding that it would not be possible or necessarily appropriate, to try to reach consensus, given the different types of data, and different interests and priorities of the communities represented by the funders involved in EAGDA. The challenges described and examples of contemporary solutions developed through complex projects demonstrate that a cross-funder body could not itself aim to develop solutions for these issues. This would not be possible even if such a body had much more 'real resources' and 'convening powers' than EAGDA. However, by bringing knowledge gleaned through their professional networks an expert advisory body could serve to inform funders of developments in the data landscape and emerging approaches to governance, and thus help to develop a common direction of travel, along which funders could align their internal policies. It is crucial, therefore, that a successor to EAGDA communicates with other nodes of the data policy network and remains aware of the nature of the challenges, so as to allow funders to have a sound high-level understanding and develop appropriate policies to support the research community.

In response to the specific question of whether there is a gap, it is difficult to definitively identify holes in this crowded landscape, but it is fair to say that there are questions that have not yet been solved, there will be new challenges in the future, and if EAGDA was not continued there is not another organisation, body or group that would obviously take on its function.

Q5: How should the quality and success of advice and awareness be evaluated?

Several interviewees raised the question of evaluation of advice, and measuring subsequent impact. A challenge with the current model for EAGDA, as outlined further in question 6, is the conversion of advice into practical actions. Implementing advice required specific buy-in from the different funders, dependent on their internal structures for policy development and strategy. Funders provided an immediate response to the reports produced within the committee, including recommendations on specific topics and issues, however without procedures in place for further follow up, the committee rarely heard of subsequent actions that had been taken. There was no in-built process for evaluating the success or quality of the advice, or its impact. The metrics available were largely limited to the number of reports produced, as a measure of the productivity of the committee, the number of issues tackled,

and if any publications arose, the usual metrics relating to citations and impact factors may be relevant, although these have not been formally collected or recorded.

How this could be improved upon in the future is challenging, given the advisory nature of the committee. The committee did not have authority to require action, or to enforce inaction, and the funders were not bound to report action or excuse inaction. While within organisations there may be anecdotal appreciation for where EAGDA advice specifically had benefit or prompted change, there is no formal record.

One question for future consideration between the funders should therefore relate to evaluation of success, and thus first to consideration of what is meant by success. If in the future, outputs should continue to be advisory, funders must retain the freedom not to take the advice, but perhaps mechanisms to explain their decisions in the short and longer term would provide a useful demonstration that the advice had been adequately considered, alongside other priorities. Where advice had led to specific activity, a record of this could also be beneficial to share with other funders, and more widely, as a demonstration of best practice. This could be published or disseminated through other means, to amplify the impact of the committee findings.

A separate area relating to standards and evaluation of impact is the issues raised within EAGDA itself, surrounding data management and sharing. During focus groups and interviews, several participants questioned whether EAGDA should have focused on creating and setting standards relating to data quality, to directly address challenges to data sharing and management, although it wasn't absolutely clear on what authority it would implement such recommendations. If EAGDA determined that this was an area to consider in the future, one demonstration of the impact of the committee would be to measure awareness of, and adoption of, the 'EAGDA standards', as a concrete output. Subsequent review of the standards, including their impact on the field of data sharing, could lead to a further (measurable) method for evaluating impact of the committee itself.

Q6: What was effective about the EAGDA model and what does evaluating it tell us about evaluation methods?

What did EAGDA do well?

Findings from the interviews

In general, all of those interviewed agreed that EAGDA had been a successful initiative. The right disciplines had been represented within the committee, and the involvement of different disciplines enabled the group to have lively discussions on wide-ranging topics. There was a feeling the members truly enjoyed the experience. Even when not specific to the topic of data sharing, participants felt that they benefited from the free-wheeling nature of the discussions.

The cross-funder approach allowed individual funders to expand their networks and provided access to experts in different fields from their own. This allowed funders to learn from a diverse group of experts, and allowed the experts to advise the significant players involved in sharing biomedical data for research purposes. The funders agreed that while they could not implement the advice uniformly, it was helpful to know what was happening at other funding agencies.

Interviewees expressed pride in the reports and papers produced, and the funders acknowledged their benefit. Several peer reviewed papers resulted from the work which benefited the authors and raised the visibility of the issues and of EAGDA members, even if they were not usually badged as outputs from EAGDA, and thus did not lead to greater awareness of the committee itself. The high quality of the Secretariat was acknowledged,

and the excellent work they did in supporting EAGDA, which was vital to the smooth running of the committee and to ensuring precious time was used most effectively.

EAGDA was also able to provide, when needed, a rapid response to breaking issues. This was beneficial for the funders; they were able to show that they were aware of these issues, were convening experts to consider them, and were able to respond to their authorities, user communities, and the public. A cross-funder response was also seen to give greater weight to the issue and assisted those funders that may not have had other avenues through which to address questions, especially if it was outside their immediate area of expertise.

Findings from the focus groups

The focus group (FG) discussions with the funders tended to balance their approval of EAGDA with their concerns. FG1 confirmed the view of the interviewees that EAGDA was beneficial because it meant that no funder “went out on a limb” and imposed different requirements on their grantees than that of other funders. However, they also noted that the effectiveness of EAGDA should not be judged on the uniform implementation of advice but through the provision of principles that could be meaningfully taken forward by all the funders in their own context. They felt that no funder would support a model that devolved any of its decision-making authority.

FG2 noted that EAGDA had the ability to thoroughly respond to a variety of issues as opposed to a group set up for each issue addressed. Similarly, FG4 recognised the strength of the committee in the breadth of topics it covered, which distinguished it from other bodies that focused only on specific topics.

FG2 saw the reports as having value due to their thoroughness which made the outputs of EAGDA as something authoritative and comprehensive. There were questions however as to whether such outputs required a standing committee. A participant in FG3 highlighted the value of the academic papers that were delivered quickly as they drove the system in the right direction, though they acknowledged that it was difficult to assess their direct impact.

Participants in FG4 discussed the response to papers that highlighted how the linking of anonymised genomic datasets with other datasets could enable individuals to be identified. This impacted upon research participants’ privacy and was recognised as a growing issue due to the increasing accessibility and complexity of data sets, the wider use of demographic and administrative data in combination with genomic data, and the increasingly sophisticated methods of data analysis and interpretation. The implications of this development led EAGDA to provide funders with interim advice whilst a proper consideration of the issues could be undertaken.

This prompt response to an unexpected issue was highlighted as one of the key strengths of EAGDA. The ability to draw upon the expertise of EAGDA was felt to provide valuable guidance in an area where participants felt ‘no one was sure what to do with the new information.’ Its recommendations allowed for a quicker, more joined-up response from the funders that could then respond on the basis of EAGDA’s advice.

Box 1: The response to the Homer case

What did not work so well?

Findings from the interviews

Members reported that they felt that there was a lack of implementation of their advice, or at least a lack of information from the funders on this point. This contrasted with the funders who said that they had had a process in place through which the EAGDA members were informed at each meeting as to the actions taken in response to their deliberations. This ‘disconnect’ points to several things. There may have been a lack of understanding by the members as to their role being truly advisory with no firm expectation that the advice would be taken. Members seemed to assume that their advice would be taken on board and changes would be made. The funders, on the other hand, considered the advice but in some cases were able to move it forward, had chosen not to, or were still undecided whether changes could or would be made. The funders moved at a slower pace than was recognised

by the members and the funders were not as clear as they could be with the members on this point. This also points to the inherent differences between the funders and the impossibility of complete harmonisation across them. But, it could be that there was simply a lack of communication that could be addressed in future, so that the funders are clearer with the members as to their plans and actions.

Another point raised was the members felt there could have been better communication of their reports to the community. Again, this points to a lack of agreement as to the role of the members. If it is only advisory, then reports will be only for the funders and therefore no further communication is needed. However, the members seem to believe that they were speaking not only to the funders, but also to the wider scientific community. Whether this was because, as recognised expert academics, they are accustomed to disseminating and highlighting their work, or because they were unsure whether their messages were being passed on by the funders is unclear. Any future model will need to clarify the audience that the members are addressing and whether it is only the funders or is a broader network. It was mentioned in both interviews and focus groups that, while it had done an excellent job, the Secretariat was seen as overworked, which may have contributed to a lack of dissemination.

There was concern expressed by the members that, as there were many groups now doing work in the areas of data management and data sharing, duplication would occur. When EAGDA began, they were alone in addressing their issues, but this has now changed. The members strongly felt that any new model for EAGDA needed to ensure that it was not duplicating the efforts of other groups.

Findings from the focus groups

FG2 noted that towards the end of EAGDA, as issues around data sharing had changed, that the membership no longer represented the expertise needed, resulting in a lack of members who could answer the questions that that funder had. This may be because it was set up originally with cohort studies in mind. A standing committee, although able to address multiple issues, may not have the flexibility to respond quickly to breaking issues. An ad-hoc group could identify issues and the commission a group to study it, although the cost of this could be problematic.

The speed at which EAGDA worked was also questioned, as at times it seemed that by the time EAGDA responded, questions had already been answered elsewhere by other bodies. This linked to comments from FG4 that the topic of data sharing was a 'very crowded space'. While there was still room for a body like EAGDA, like the Interviewees, it was acknowledged that there was a need to assess the wider landscape and map the activities of other groups to avoid overlapping efforts.

The structure of EAGDA preparing formal recommendations and the funders' formal responses was seen by FG4 as overly bureaucratic, driven by the process rather than by the advice. This contrasted with FG2 stressing the need for specific implementable advice which one assumes would drive them to make changes. FG1 suggested a possible compromise – reports could include 'digestible' outputs and recommendations that could then be considered for implementation.

The differences between the way in which the funders work was again highlighted in FG4 where it was felt that there were different levels of buy-in across the funders and some benefited more from the outputs of EAGDA than others, as a funder may have already been working on a specific issue and thus not need EAGDA's input. This point was also raised in the FG3 where some felt EAGDA's recommendations were more pertinent to other funders, as that funder was already covering these issues. Such disconnect could make it difficult to maintain commitments from multiple funders in a future EAGDA.

What does evaluating EAGDA tell us about evaluation methods?

The tools drawn upon in this study to evaluate EAGDA and its impact, particularly for the funders involved, have largely focused on qualitative methods. Discussions with expert members and funders have provided experiential detail on the benefits, and challenges, of the approach to expert advising. As Q5 identifies, there is a lack of quantitative evidence surrounding the work of the committee, which could be useful for future evaluation, to more clearly demonstrate the reach and impact of the advice, and other outputs, provided by the committee.

Q7: What new model for wider cross-funder advice and awareness is recommended, and how might such a model be evaluated?

New models were considered at the workshop of EAGDA members and funders. Based on discussions during the interviews, the following were formulated:

1. EAGDA-like – This model would be essentially the same as EAGDA with an increased focus on feeding back responses, opinions and updates to the EAGDA members, a facet of EAGDA that they felt was lacking. Staying with this model does not, however, fully address the varying needs of the funders.
2. EAGDA-fast – This model would also follow the EAGDA model but would increase the membership and decrease the time spent on any one issue. These changes were suggested in response to a suggestion that, where an issue arose that needed a very fast response, a broader pool of members would be useful so that if members were not available at any one time, a replacement could be secured. Committee responses could be prepared based on expertise and at short notice. This model leaves horizon-scanning activities to others which might be a disadvantage. It was noted at the workshop that with this model, the committee could lose a sense of coherence due to the revolving membership.
3. EAGDA-layered – This model takes EAGDA and adds additional layers of interaction, in response to the comments that EAGDA was very good at providing specific advice especially to funders who were less than knowledgeable about an issue. The committee would horizon-scan like EAGDA and when a specific issue was identified, funders could decide individually if this was something they needed to take forward. If so, that funder could return to the EAGDA membership to find individual experts to work with them. This allows a building of relationships between members and funders and the opportunity for funders to receive help on specific issues as needed. On the other hand, it adds layers of bureaucracy and additional expense to the EAGDA process.
4. Funder-driven workshops – This was suggested as interviewees stated that this model worked well for other funders. Funders would choose an issue that arisen and organise a multi-day workshop focusing on that. Invitees would be experts such as those who served on EAGDA in addition to others as decided. During the workshop, attendees would discuss the issue and other related topics, with time set aside for in-depth sessions where a single or small group of funders could seek help on creating policy on the issue in their specific funding context. Such workshops can bring a broad group of experts from the UK and abroad and the sharing of ideas could produce policies, statements and collaborations. Workshops do need considerable planning and follow-up and would be resource-intensive. There is also no surety that the funders would receive the specific answers that they might need. The need for such a workshop would need to come from other sources, such as a group like EAGDA or one of the many other groups doing horizon-scanning in the biomedical sciences.

These models and variations of them were discussed amongst the workshop attendees. It became clear that because of the differing characteristics of the funding bodies represented, no one of these models, or variations thereof, was preferred or would be an ideal fit. Instead,

the impression gained was that the funders would like to have a suite of models at their disposal to be used when appropriate. For example, if there was an issue that was of interest to only two funders, then they together could plan a workshop to focus on that issue in the context of those funders. For general horizon-scanning involving multiple funders, a body like EAGDA would be more appropriate, as long as it was aware of what activities were being examined by other similar committees or bodies. Limiting the funders to one model was not seen as useful or wanted. But no model could be simply a 'talking shop' as it would be difficult to keep senior figures engaged when there were no concrete outcomes.

FG4 expressed a desire for a nimbler, less bureaucratic model and felt that the creation of the process of formal reports and responses meant it lost its key strength of being a resource for practical advice on topical issues. FG3 agreed – they felt the EAGDA model had issues around its agility and timeliness and was rarely a first port of call for advice. There was a need to decide whether a future model should provide advice or respond to specific questions, although there was no consensus on this.

More than one focus group noted that any future model needs to produce implementable advice, although no process for this was agreed. But it was clear from the workshop discussions that no funder would be comfortable with being required by an outside committee such as EAGDA to implement any recommendations. The differing contexts of the funders makes this impossible. For example, interviewees and workshop attendees mentioned monitoring funders' outputs and compliance, yet FG3 felt that such a recommendation was a good example of something that may not seem hard to implement but in practice would require a radical shift in how funders are set up and operate. Complete harmonisation across funders is not possible and any recommendations made by any future committee must allow for flexibility across funders.

As noted, there was much discussion on dissemination of EAGDA work. Any model will need to ensure there is a clear communication strategy between members and funders, including what will be disseminated, to whom and via what channels. For example, funders will need to decide how 'deep' the dissemination will go. It was noted in FG3 that perhaps ways to involve the research community to help them embed best practices could be considered; FG4 suggested that an external 'profile' would have been useful for EAGDA and could be considered in the future. Once any decisions about the reach of outputs is decided, all of the proposed models will require resources for a secretariat which will need to be properly resourced and staffed so it can fulfil the agreed responsibilities.

In addition to considering future models they might want to use, it is recommended that the funders also focus on their own working relationship. It was said many times that a cross-funder model was useful and wanted. But little time and effort has been spent on how the funders, in practice, can work together. It was noted at the workshop that EAGDA brought together funders in a different way than had been done before. Normally, funders that are funding a project sit in an advisory capacity overseeing that project. But in the EAGDA example, the funders were not funding parts of an 'EAGDA project' (except for the WT funding the secretariat), they were learning from each other and the EAGDA expert members and then making decisions regarding policy or funding. It is recommended that the funders take time to consider how this 'funders' relationship' might be best used in the future. There are signs that there will be more cross-funder activities in the future; the bringing together of the various funders under the umbrella of UK Research and Innovation (UKRI) shows this as a strategic move. As more and more funders grapple with similar issues, such as data sharing, in their own specific contexts, the greater the opportunities for them to learn from each other. A committee such as EAGDA could provide continued opportunities to work together. This could also lead to some harmonisation, where possible, across funders with clearer policy directions for researchers and the public.

6 Conclusion

The combination of desk-based and empirical research included in this review process has allowed EAGDA to be situated within the wider landscape of research policy and advice, relating to issues and topics in data sharing and access. It is clear that this is an area that still needs particular consideration, and that there is benefit from enabling a cross-disciplinary panel of experts to spend time working through the challenges in meticulous detail. One difficulty for the funding bodies for which the advice was generated, is how to manage the importance of the issues against the relative lack of urgency to tackle them. In the long-term it is widely acknowledged that addressing the topics and issues relating to data sharing will lead to considerable benefit for research, and society more generally, including through more efficient use of data, leading to higher quality research, and improved health outcomes. In the short-term, however, it continues to be a nice-to-have rather than a must-have.

The value the EAGDA brought to the conversation included high quality discussion drawing upon a range of expertise. The relative objectivity of the experts involved, underpinned by the freedom of the funders not to act on the advice, allowed far-reaching consideration of the appropriate direction of travel in this area, allowing the funders to gain access to big-picture perspectives in which to situate their own data priorities. This freedom brought with it consequences relating to the lack of accountability, without any pressing need to demonstrate use of or adherence to advice, or even to show that it was being appropriately fed into the internal structures of the funding bodies for closer scrutiny. Without tools to evaluate the impact of the advice given, it is little wonder that a nagging concern persisted at the value of the lengthy [and academic] reports, and the lack of practical outputs to show for the time and resource involved in the committee.

However these fears are allayed by the few concrete examples of value added, for example the EAGDA response to the Homer paper, which demonstrated that with the right channels to vocalise the work of the committee, it could be galvanised to inspire progress.

The starting point for considering the future role of such a group could therefore be to examine the feedback loop, and reach consensus amongst the funders about how to show that the advice is being picked up and considered, even in the absence of clear evidence of action.

Clearer agreement on the exact remit and scope of the group, and what it should cover, will provide a starting point for these discussions, with the onus being on individual funders to consider their commitment to the group, and its findings.

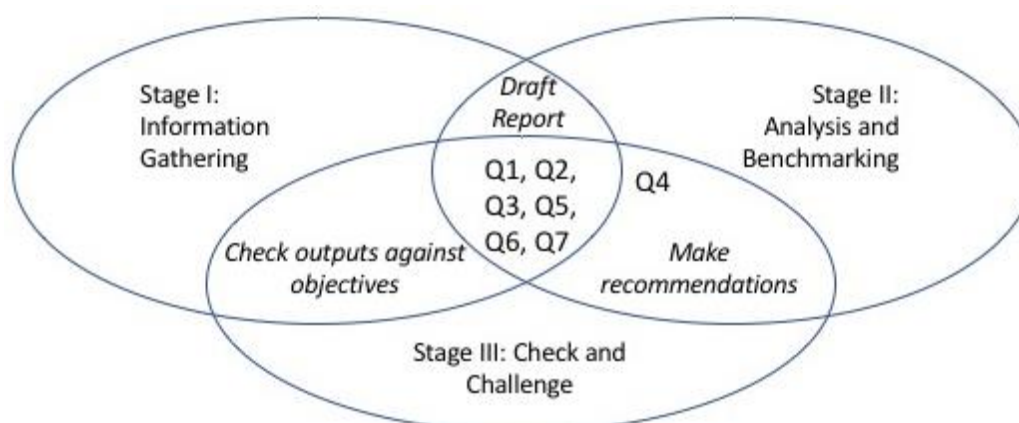
Perhaps, then, the real outcome from this research is not the identification of a new model for a body such as EAGDA, but the realisation that working together, as they did with EAGDA, is the future for the funders. The topic, such as data sharing for EAGDA, is not the important point. Instead, the funders need to focus on how they can and should work together when issues such as data sharing arise. From these discussions and through choosing a process, with a specific scope and remit, models will be found that will aid them in their work.

7 Appendices

Appendix 1: Methodology

The research will proceed through three stages. A distinction is drawn throughout between operational issues (such as data availability, access to expert advice e.g. on information governance and data protection, institutional support for data management, etc.), topics of strategic significance (such as artificial intelligence, incl. machine learning, Internet of Things (IoT), and cellular computing) likely to cross disciplinary boundaries and the methods by which advice may be made available to funders. **Stage One** (information gathering) will a) identify existing resource available to research funders to identify current and emerging issues and topics of significance, b) seek funder, researcher, and expert others' (e.g. industry leaders, access committee chairs and policy leads) views on relevant issues, topics, and methods, and c) seek their views on how the quality and success of an advisory model ought to be evaluated. **Stage Two** (analysis and benchmarking) will a) provide a gap-analysis of advice and awareness between the contemporary map of relevant resource available to UK funders and the desiderata of different constituencies, b) offer an evaluation method for an advisory model based on criteria of quality and success identified through stage one, and c) benchmark this evaluation method against the past EAGDA model. Stage two will culminate in production of a draft report. **Stage Three** will enrich the draft report through peer and stakeholder check and challenge. Draft recommendations will be tested and developed through a workshop with funders, researchers, and expert others. This will provide a robust evidence base to inform funders' discussion on how best to work with the research community to ensure the best policy and operational approaches to access, governance, and management of data generated by funded research.

The final report will address seven research questions. Each question will be considered through every stage of work, except question 4 which will be addressed only in stage two and three:



KEY QUESTIONS:

Q1: How are UK funders currently advised and aware of issues and topics relevant to health-related research?

Q2: How *could* UK funders of health-related research be advised and made aware?

Q3: What are the issues and topics that UK funders need to be aware of and advised on both now and in the future?

Q4: Is there any gap in advice or awareness currently drawn on by, or available to, UK funders?

Q5: How should the quality and success of advice and awareness be evaluated?

Q6: What was effective about the EAGDA model and what does evaluating it tell us about evaluation methods?

Q7: What new model for wider cross-funder advice and awareness is recommended and how might such a model be evaluated?

APPROACH: The following approach will be taken to each of the three stages (information gathering; analysis and benchmarking; check and challenge) in order that the final report might robustly answer each of Q1-Q7. **Approach to Stage One: Information gathering**

1. First, a list of key UK funders will be drawn up and agreed with Wellcome. They will be identified via existing contacts held by the research team, desk-based research and snowball techniques, sampling purposively for individuals with responsibility for research funding strategic policy and funding allocation for projects involving human research data and metadata. Key members of the established funding bodies (including Wellcome, ESRC, MRC, Cancer Research UK and NIHR) will be emailed and invited to semi structured interviews or expert focus groups, depending on which is deemed most appropriate for the respondents. Data collection will have five aims: To establish which are the key groups, activities and initiatives *currently* drawn on by UK funders for advice and awareness and the method of engagement (Q1);
2. To identify which key groups, activities, and initiatives UK funders *know* to be a potential source of advice or awareness *but do not currently* draw upon, with reasons where applicable (Q2);
3. To seek views on current issues and topics of strategic importance (Q3);
4. To seek views on the appropriate criteria to apply to a model of advice and raising awareness (Q5);
5. To seek views on the extent to which the EAGDA model of advice satisfied the criteria of quality and success associated with evaluation criteria (Q6).

Parallel to the above, members of the following groups, activities and initiatives will be invited to semi-structured interviews or expert focus groups (whether identified by funders or not):

- Members of UK research community engaged in studies requiring data access, use and management relating to human participants.
- Chairs or designates of relevant data access committees, including CPRD, UK Biobank, PBPP in Scotland, NHS Digital (IGARD), Public Health England (ODR) and CAG in England and Wales, and ADRN approvals panel.
- Members of national bodies placed to provide advice or identify current issues for research relating to data access, use and management. Including select members of the Data Steering Group (esp. Office of National Statistics, National Data Guardian for Health and Care, techUK, Government Data Service, Open Data Institute, Government Office for Science, Industry, e.g. Mastodon C, Microsoft UK Research, Yoti, Google DeepMind, Experian, Omidyar Foundation), NHS Digital Centre's Research Advisory Group (esp. MRC, Representative of Academia, Office for Strategic Coordination of Health Research, Health Informatics Sub-group, Academic Health Sciences Network, Health Research Authority, Wellcome, NHS Digital Data Director), the Information Commissioner's Office (and the associated UK Anonymisation Network), the Farr Institute, The Royal Society, The Royal Statistical Society, Genomics England, the NHS Digital Academy, British Library and the UK Digital Catapult Centre.

Members of other organisations or initiatives interested in health data use including, for example, Understanding Patient Data, Richmond Group, BMA, MedConfidential, Independent Cancer Patients' Voice, UsemyData; There are four aims associated with these interviews/expert focus groups:

1. to identify any *additional* key groups, activities, and initiatives perceived by respondents as potential sources of advice or awareness to UK funders and to seek views on the (non) use of these sources relative to criteria of quality and success identified (Q2);
2. To seek views on current issues and topics of strategic importance and method of engagement (Q3);

3. to seek views on the appropriate criteria to apply to a model of advice and raising awareness (Q5);
4. To seek views on the extent to which the EAGDA model of advice satisfied the criteria of quality and success associated with proposed evaluation criteria (Q6).

Relevant persons will be identified via existing contacts held by the research team and Wellcome, snowball technique and supplemented with desk-based research. The team is extremely well placed to identify and reach out to potential respondents, many of whom are personally known to team members through previous and current roles and projects. All persons invited to participate will be given the choice of meeting in person or telephone or SKYPE interview at a time to suit them. The research team has extensive experience designing and conducting semi-structured interviews and focus groups. Previous projects adopting similar methods include grants 1-3, 8, 9 (Appendix 1) and commissioned reports for Wellcome.

Approach to Stage Two: Analysis and benchmarking

Interviews and focus groups will be recorded, transcribed and thematic analysis undertaken using a framework approach. Relevant topics identified by respondents will be supplemented by additional desk-based research. Topics, including those identified by funders, researchers and others, will be clustered according to themes derived iteratively, through inductive and deductive analysis and according to immediacy (e.g. current issue or horizon scanning).

Criteria to underpin an evaluation method for an advisory model will be identified and clustered from the empirical data collected. Identification of appropriate evaluation methods will be supplemented via desk-based research. Desk based analysis of original empirical data will allow the EAGDA model to be benchmarked against the criteria of quality and success described by respondents.

Report will be drafted to address seven research questions (identified above as Q1-Q7). It will include recommendations of appropriate alternative evaluation methods with regards to current, potential future, models of advice, communication and input to strategic and operational matters concerning data for health-related research. Draft conclusions on (i) current and future structures, groups and initiatives that are, or could, enable research funders to identify emerging issues for data use and access and management and to formulate strategic policy decisions in response to these, (ii) gap analysis of available advice against current issues and topics identified as of strategic importance, and (iii) draft recommendations on current, and potential future, models of advice and awareness. Draft recommendations will be presented for check and challenge through stage three.

Approach to Stage Three: Check and challenge

Key respondents (e.g. funders, EAGDA members, researchers) will be invited to a workshop to check and challenge draft conclusions and recommendations. Opportunity will be given in the workshop to comment and constructively engage with recommendations contained in draft report. Final Report will then be drafted and submitted to funder.

TIMELINE (WEEKS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Stage I																	
Stage II																	
Stage III																	

Appendix 2: Empirical Research Summary

Interviews	Focus Groups
19 Invitations sent out (Aug 18) 13 Interviews conducted (Aug/Sep 18): - 8 Interviews with EAGDA committee members (independent of the funders) - 5 Interviews with policy officers sitting on EAGDA representing the funders	4 Focus Groups conducted (Aug 18): - WT (3 participants) - ESRC (4 Participants) - MRC (2 participants) - CRUK (6 participants)

Key Theme	Sub-theme	Sub-sub-theme
EAGDA Effective	Cross-funder model Output Output Criteria for success USP Impact Composition Approach Engagement	Deliberation vs practical implementation
EAGDA Role/Function	USP Criteria for success Unclear Mandate/remit/scope Cross-funder model Relationship with funders	
EAGDA Challenges	Engagement Engagement Engagement Engagement Cross-funder model Buy-in Duplication of effort Implementation Approach Composition Output	w/ right people w/ output w/ funders
New Model	Role USP Cross-funder model UKRI Composition Impact Output Criteria for success Needs Approach	
New Model Challenges	Cross-funder impact Engagement Engagement w/funders Output Logistics Buy-in/being a priority Approach duplication of effort	
Landscape	Data governance challenges public perceptions/perspectives new challenges cross-funder model other bodies/alternatives to EAGDA	

Table 2: Qualitative research coding framework

Appendix 3: Reflections on the future of the EAGDA made during the final meeting, 18 April 2017

The Chair's paper on the future of the EAGDA identified that changes in the external environment since the EAGDAs inception can be broadly categorised as follows:

- *Institutional developments*, including new bodies set up to do/support data use such
- *Knowledge base developments*: increasing understanding of the potential for better use of data across disciplines, including admin data and other big data applications; advances in statistical understanding of disclosure and anonymisation issues.
- *Legal developments*: e.g. GDPR; National Data Guardian; Digital Economy Bill.
- *Technological developments*: in data science; linkage; machine learning; data security and encryption; statistical disclosure control techniques; the platform economy.
- *Policy developments*: increasing awareness of data issues (both in terms of potential benefits, and potential risks) as a major area for policy development, by funders, learned societies and government.

SCOPE

- This is a timely point to consider EAGDA's role and remit, which has been restrictive as it has focused on cohorts and longitudinal studies when many of the issues it addresses have broader reach and significance; does it remain valid to limit the groups scope given the emergence of complex questions about archiving, data discovery, encryption technologies and data infrastructures, etc. that require similar cross-disciplinary thinking and approaches to which the group could also contribute?
- The boundaries of what constitutes health or health-related data are becoming more blurred as different data types can increasingly be linked across sources.

The current EAGDA model

- It is unclear to what extent funders' policies and practice are being informed by EAGDAs discussions and advice. Members judged that EAGDA had become disconnected from delivery and implementation and from strategic funding priorities and felt that buy-in was required at an executive board level to offer the strategic power for EAGDAs work to be useful.
- EAGDA raises issues that funders wouldn't necessarily see or be able to discuss themselves; it has not managed to instigate the resources and convening power required to take issues forward, act on them in a coherent way and build consensus
- EAGDA is not equipped to map the landscape of data use and management at the moment or fulfil a watching brief on law, technology and ethical issues. This would require both resource and operational and strategic commitment from a wider group of funders to cover off the breadth of the space.

RECOMMENDATIONS

- There are many UK bodies working in this space; the focus of an advisory group should be to provide big picture thinking and advice across disciplines.
- Clear Terms of Reference are required to guide the group's operation: What does it mean to be 'advisory'; what is the body's scope; who is going to act and how as a result of advice it gives? This will necessitate a deeper connection to funders and government, especially if the constituency EAGDA serves is widened beyond cohort and longitudinal studies.
- An ability to respond to issues identified in the right way will require substantial strategic resource and convening power; which needs to be recognised through willingness to organise and 'buy-in' to the initiative at an executive board level; feedback should also be provided to the advisory group on the impact of their findings or recommendations.

- A future EAGDA could work on a two-part model:
 - Raising issues funders wouldn't necessarily identify
 - Consensus building about how to take issues forward, bringing a wider audience on board to work out how to do things about it, e.g. convening, consensus building, connecting with other players in the landscape.
- Allowing the group to have some authoritative oversight of other initiatives in this space, convening and bringing together different perspectives, including different disciplines and data types, might require the backing of a larger group of funders.
- If EAGDA funders wish to continue the group with its present group of funders, the group will only really be relevant if it has a more public-facing role and can push its deliberations and recommendations out better and further, both to stimulate real progress and hold the funders to account.

There are two feasible directions for future activities:

- Being outward facing, reaching out into the research community and beyond (e.g. into policy making) and making messages more powerful and coherent. This would necessitate a larger group of funders and stakeholders around the table to bring the right convening power across disciplines.
- Given the complex landscape, focus on helping funders get their priorities and strategies right and coherent with inwardly-directed advice for them, involving practical guidance.

OTHER CONSIDERATIONS

- Reconciling the competing demands of producing outputs that are specific enough to be directly useful but broad enough to be widely relevant, while also understanding that the issues are constantly broadening, is a challenge
- The research field is not isolated from broader issues e.g. in the private sector over the responsibilities of data platforms over content hosting; the 'right to be forgotten' etc., and funders should have a grasp of the implications of these sorts of issues on research data use.
- There is currently little coordination between the multiple groups working on issues relating to data use and governance in research and conversations continue to be siloed.
- Identifying evidence gaps and issues on the horizon is challenging given these siloes: there is no single list of key issues in data ethics, for example, as most interested parties are concerned with specific issues e.g. data protection/privacy, data security, standards etc. depending on their own priorities and perspectives. A future group should look to connect between siloes, connect better with the law and with new technologies.
- There is now an 'industry' in data use and management: many more people are interested in this space than in 2012. This is positive but has the disadvantage that multiple conversations take place in the absence of clear and agreed principles, which means there is duplication of discussions and no substantive normative moves forwards. A group should seek to alter the 'ground state' so that progress can be made beyond these early discussions repeated in different contexts.

Appendix 4: Bodies in the UK landscape for provision of policy or direction of data sharing

i. Key UK funders of research involving human participant data

Government Funding Councils and NHS

Arts and Humanities Research Council
Biotechnology and Biological Sciences Research Council
Engineering and Physical Sciences Research Council
Economic and Social Research Council
Medical Research Council
Natural Environment Research Council
Science and Technology Facilities Council
Innovate UK
National Institute for Health Research

Academies

British Academy for the humanities and social sciences
Royal Academy of Engineering
The Royal Society
The Academy of Medical Sciences

Major Medical Research Charities (research spend > £75 million per annum) ¹⁵

British Heart Foundation
Cancer Research UK
Wellcome Trust

Other Research Charities

Leverhulme Trust
The Nuffield Foundation
The Health Foundation

Miscellaneous

JISC - Joint Information Systems Committee
National Endowment for Science, Technology and the Arts (Nesta)

¹⁵ Association of Medical Research, <https://www.amrc.org.uk/Pages/Category/member-directory?Take=3>)

ii. Projects and initiatives to drive, direct or study the UK landscape for data derived from human subjects

Body	Creation	Purpose	Category
The Farr Institute of Health Informatics Research	£20M in 2013 by the MRC	To accelerate the development of a UK-wide platform for health informatics research, including digital infrastructure to securely share data, support linkage at scale and develop standards. Created as an overarching brand to represent and accelerate research led by the four e-Health Informatics Research Centres	Infrastructure and delivery
UK Health Informatics Research Network (UK HIRN)	Brought together 10 government and charity research funders (2.18M MRC)	To promote best practice, build partnerships (particularly with academia, industry and the NHS) and co-ordinate activities in outward-facing priority areas	Infrastructure and delivery
Four e-Health Informatics Research Centres	Brought together 10 government and charity research funders; £19M (£5.8M MRC)	To create infrastructure for cross-sector collaboration and to develop a broader research community and harnessing expertise from across government, public sector, academia and industry.	Infrastructure and delivery
UKSeRP	Developed by the Health Informatics Group at Swansea University, with support from the Farr Institute of Health Informatics Research funded by MRC	Provides ISO27001 certified data curation, management, sharing and analysis facilities to the UK research communities through capabilities drawing from operation of SAIL	Infrastructure and delivery
MRC Medical Bioinformatics initiative	£39M	Aimed to improve informatics and computerised approaches to understanding mechanisms of human disease, facilitate research linkage, and build analytical capabilities tailored to medical research and integration and interpretation of data	Primary research into data science, including its sharing and reuse
MRC Capital Investment in Genomics England Data Infrastructure	2014 MRC invested £24M	Aligned to the Medical Bioinformatics initiative and to support the Clinical Research Infrastructure initiative and UK 100000 Genomes Project to establish main data centre and make resources securely available to researchers and clinicians	Infrastructure and delivery

NIHR Biomedical Research Centres	The first five BRCs were established in 2007	The NIHR BRCs are partnerships between: an 'Academic Partner', usually a research university or similar specialised institution; and an 'NHS Host', usually in the form of an NHS Foundation Trust. The aims of the BRCs are to: drive innovation in the prevention, diagnosis and treatment of ill-health through early translational (experimental medicine) research; translate advances in biomedical research into benefits for patients, the health system and for broader economic gain; and provide a key component of the NHS contribution to UK's international competitiveness.	Translation of research findings into healthcare practice
NIHR Health Informatics Collaborative programme	The HIC investment was £11M total from NIHR, which was matched by the partner NHS Trusts and the BRCs in the form of funds, existing resources and other contributions.	It was conceived in response to a 'grand challenge' from England's Chief Medical Officer: "to show that data collected in the course of routine patient care across the five largest biomedical research centres can be re-used for collaborative, translational research	Primary research into data science, including its sharing and reuse
NIHR Collaboration for Leadership in Applied Health Research and Care	The nine CLAHRC pilots were established in October 2008, and have completed their five years of funding. There are now 13 CLAHRCs around the UK	They are collaborative partnerships between a university and the surrounding NHS organisations, focused on improving patient outcomes through the conduct and application of applied health research.	Translation of research findings into healthcare practice
Academic Health Science Centres	Department of Health and Social Care in 2009 made a £60M investment to establish the first 5, these were renewed for a further 5 years and an additional one created	Focus on discoveries arising from basic and applied research; developing innovation and collaboration with industry; and on excellence as a provider of multi-professional education	Translation of research findings into healthcare practice
Academic Health Science Networks	There are 15 AHSNs across England, established by NHS England in 2013	To deliver a step-change in the way the NHS identifies, evaluates and adopts new technologies and are predicated on partnership working and collaboration between the NHS, academia, the private sector and other external partners, including industry	Translation of research findings into healthcare practice
NHS Genomic Medicine Centres	Thirteen Genomic Medicine Centres (GMCs) have been established by NHS England	Deliver an end-to-end and country-wide genomic medicine pathway, establishing the infrastructure to make genomic medicine a routine part of NHS care.	Translation of research findings into healthcare practice

NHS Global Digital Exemplars	Funding of the GDEs is through the £4.2B 'Paperless 2020' programme announced by Jeremy Hunt in February 2016, overseen by the National Information Board.	internationally recognised NHS providers delivering exceptional care, efficiently, through the use of world-class digital technology and information	Translation of research findings into healthcare practice
Administrative Data Research Network	ESRC	aimed to facilitate researcher access to administrative data routinely collected by Government Departments;	Infrastructure and delivery
Business and Local Government Data Research Centres	ESRC	These Data Research Centres make data, routinely collected by business and local government organisations, accessible for academics in order to undertake research in the social sciences of mutual benefit to data owners and in ways that safeguard individuals' identities	Infrastructure and delivery
RCUK Digital Economy Hubs	ESRC - £1.4 million	To build on and/or enhance work of The Big Data Network, as well as to conduct complementary activities which will contribute to the development of infrastructure for research using social media data	Infrastructure and delivery
Civil Society Data partnership Projects	ESRC	Building on relationships between academic researchers and civil society organisations to demonstrate the value of improved data infrastructure, enabling collection and analysis of data which is of interest to civil society organisations and through enabling the sector to better utilise its own data.	Translation of research findings for use by civil society organisations
Centre for Doctoral Training (CDT) in new forms of data	ESRC	Real-time data analysis, data linkage and interoperability of data from different sources. The substantive focus is on 'new forms of data' in a broad sense and on the opportunities provided by these data to study social processes as they unfold.	Primary research into data science, including its sharing and reuse
New and Emerging Forms of Data Policy Demonstrator Projects	ESRC as part of Big Data Network Phase 3	To demonstrate the potential value of New and Emerging Forms of Data to policymaking, either in their own right or when brought together with data from other sources. Seeks to inform the development of a larger investment; a new international, interdisciplinary 'Knowledge Hub' – drawing on existing and planned infrastructures and other investments made by the Research Councils and others to build a critical mass in capability, to access and analyse these new and emerging forms of data for policy-driven, impactful research in the social sciences	Primary research into data science, including its sharing and reuse

The Alan Turing Institute	EPSRC and five founding universities (Cambridge, Edinburgh, Oxford, UCL and Warwick). The Institute was established in 2015, with a five year £42M government investment, together with a matched £25M investment from the partner universities	To undertake data science research at the intersection of computer science, mathematics, statistics and systems engineering; provide technically informed advice to policy makers on the wider implications of algorithms; enable researchers from industry and academia to work together to undertake research with practical applications; and act as a magnet for leaders in academia and industry from around the world to engage with the UK in data science and its applications.	Primary research into data science, including its sharing and reuse
UK Health Data Analytics Network (UK-HDAN)	EPSRC	To develop an understanding of the UK health data analytics landscape; has drafted a research roadmap, as working document that aims to provide a clear statement of the research challenges and opportunities within health data analytics.	Infrastructure and delivery
UK Data Service	Funded by the ESRC to meet the data needs of researchers, students and teachers from all sectors	Provide access to high-quality local, regional, national and international social and economic data, including major government sponsored surveys, cross-national surveys, longitudinal studies, UK census data, international aggregate, business data and qualitative data. Additionally provide: support for policy-relevant research in the higher education, public and commercial sectors; guidance and training for the development of skills in data use; the development of best practice data preservation and sharing standards; the sharing of expertise with international data providers to remove barriers to accessing data	Infrastructure and delivery
Digital Curation Centre	Created in response to the JISC Continuing Access and Digital Preservation strategy, as a national centre for solving challenges in digital curation that could not be tackled by any single institution or discipline	An internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert advice and practical help to research organisations wanting to store, manage, protect and share digital research data; including access to a range of resources including how-to guides, case studies, online services and training programmes. Also offer consultancy and support with issues such as policy development and data management planning.	Infrastructure and delivery

Genomics England	Established and funded by UK Department of Health and Social Care	Genomics England was set up to deliver the 100,000 Genomes Project. It aims to bring benefit to patients, create an ethical and transparent programme based on consent, enable new scientific discovery and medical insights, and kickstart a UK genomics industry.	Infrastructure and delivery
UK Biobank	UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council , Department of Health , Scottish Government and the Northwest Regional Development Agency. It has also had funding from the Welsh Government , British Heart Foundation , Cancer Research UK and Diabetes UK . UK Biobank is supported by the National Health Service (NHS).	UK Biobank is a major national and international health resource, and a registered charity in its own right, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses.	Infrastructure and delivery
CLOSER	Funded by the ERC and the MRC.	Brings together 8 world leading longitudinal studies with participants born throughout the 20 th and 21 st centuries, with the aim to maximise the use, value and impact of these and other longitudinal studies to help improve understanding of the key social and biomedical challenges. Supports researchers in finding data from Longitudinal studies and produces longitudinal resources for the academic community. Works closely with the government, policymakers, think tanks and the third sector to influence policy.	Infrastructure and delivery
Health Data Research UK	Joint investment led by the MRC, together with the NIHR (England), the Chief Scientist Office (Scotland), Health and Care Research Wales, Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), the EPSRC, ESRC, the British Heart Foundation and the Wellcome Trust.	HDR UK supports world-leading research to develop cutting-edge analytical tools and methodologies to address the most pressing health research challenges. It will work with biological, clinical, social and environmental data, as well as emerging forms of data like that from wearable technology. As a national informatics research programme, HDR UK can capitalise on the UK's unique research strengths and data assets.	Infrastructure and delivery

Open Government Partnership	More than 70 nations participating – co-chaired by government and civil society representatives.	A multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. OGP has a strategy that has a heavy emphasis on policy mechanisms to enable access to government data.	Infrastructure and delivery
Health Data Finder for Research	NIHR	The Health Data Finder for Research enables researchers to find information about the UK healthcare data sets that are available for research and to direct them to the relevant data custodian experts to request access to these data sets. The aim of the Health Data Finder for Research is to help researchers navigate the UK health data landscape.	Infrastructure and delivery
The Research Advisory Group	MRC	The Research Advisory Group is the ‘intelligent customer’ to NHS Digital representing the research community. This strategic partnership will enable researchers to deliver better research quicker through the use of high quality NHS Digital data, and NHS Digital to improve its data and data services whilst demonstrating trustworthiness to the public.	Infrastructure and delivery
High-Level Group	EU	The HLG is a tripartite group of innovation experts from the EU institutions (European Council and Commission), Member State governments, leading international businesses, and academia. It is tasked with “prepare a series of recommendations ...on how to redesign, develop and manage an encompassing innovation policy in the EU.” The group has specific working groups on digitalisation/data, and the modernisation of governance systems.	Infrastructure and delivery
ELIXIR	EU	Unites Europe’s leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research	Infrastructure and delivery
Centre for Data Ethics and Innovation	UK Government Department for Digital, Culture, Media and Sport	To advise the government on the measures which are needed to ensure safe and ethical innovation in data and AI	Policy Advice

iii. Regulatory bodies in the UK that govern human participant data and the individuals who process it

- Care Quality Commission (CQC)
 - Established to regulate and inspect health and social care services in England.
- Health and Care Professions Council (HCPC)
 - Regulates 16 categories of health and care professionals including Biomedical scientists and Clinical scientists
- General Medical Council
 - Chief responsibility is 'to protect, promote and maintain the health and safety of the public' by controlling entry to the register of medical practitioners within the United Kingdom
- Human Fertilisation and Embryology Authority
 - a statutory body that regulates and inspects all clinics in the UK providing *in vitro* fertilisation (IVF), artificial insemination and the storage and use of donated human gametes and embryos. Also regulates human embryo research.
- Medicines and Healthcare products Regulatory Agency (MHRA)
 - Divided into three main centres:
 - MHRA Regulatory (the regulator for the pharmaceutical and medical devices industries)
 - Clinical Practice Research Datalink (CPRD)
 - National Institute for Biological Standards and Control
- Nursing and Midwifery Council (NMC)
 - Regulator for nursing and midwifery professions in the UK. Sets and reviews standards for their education, training, conduct and performance.
- Professional Standards Authority for Health and Social Care
 - Promotes the health and wellbeing of the public and oversees the nine UK healthcare regulators
- Health Research Authority (HRA)
 - Protect and promote the interests of patients and the public in health and social care research, by promoting transparency, making sure research is ethically reviewed and approved, provide independent recommendations on the processing of identifiable patient information where it is not always practical to obtain consent, for research and non-research projects, and oversee committees and services, including NHS Research Ethics Committees (RECs)
- Human Tissue Authority (HTA)
 - Regulator for human tissue and organs. Aim to make sure that laws laid down in The Human Tissue Act 2004, The EU Tissues and Cells Directives, and The EU Organ Donation Directives are followed by setting standards that are clear and reasonable, and in which the public and professionals can have confidence.
 - Under the Human Tissue Act 2005, the HTA is required to license the storage of relevant material for 'research in connection with disorders, or the functioning, of the human body', and as such it is responsible for the licensing of biobanks within the UK.

iv. Think tanks publishing on human participant data governance and use in the UK ¹⁶

- The King's Fund <https://www.kingsfund.org.uk/leadership>
- The Health Foundation <http://www.health.org.uk/>
- Nuffield Trust <https://www.nuffieldtrust.org.uk/>
- Policy Exchange <https://policyexchange.org.uk/>
- Reform – NHS – <http://www.reform.uk/our-work/area/nhsreform/>
- Institute for Public Policy Research <http://www.ippr.org/big-issues/health-and-social-care>
- Demos <https://www.demos.co.uk/research-area/health-social-care/>
- Westminster Health Forum
http://www.westminsterforumprojects.co.uk/publications/westminster_health_forum
- Chatham House (global health)
<https://www.chathamhouse.org/research/topics/global-health>
- Adam Smith Institute (no search function) <https://www.adamsmith.org/>
- RAND Europe <https://www.rand.org/randeurope/research.html>

¹⁶ <https://blogs.bodleian.ox.ac.uk/tw-knowledge-hub/think-tanks-on-healthcare/>

Appendix 5: Direction from UK bodies on the sharing of human research data, 2012-2018

- i. Key directives published by UK funding bodies and other relevant organisation relevant to the governance of human participant research data, April 2012 – July 2018

Body	Body Type	Created by	Publication Title	Publication Type	Date
UK Research and Innovation (UKRI)	Public Funding Body		Research and Innovation Infrastructure Roadmap	Strategic Plan	In progress
			Strategic Prospectus	Strategic Plan	May-18
			Guidance on best practice in the management of research data	Guidance	Jul-15
Wellcome Trust	Research Charity		Developing an outputs management plan	Guidance	Jul-17
			Wellcome Sanger Institute Data Sharing Policy	Policy	Jun-17
			Consultation on the future EU-UK relationship on research and innovation	Consultation	Jan-18
			Longitudinal Population Studies Strategy	Strategic Plan	Jul-17
			Transforming UK Translation	Strategic Plan	Jul-17
			Regulation and governance of health research: five years on	Report with recommendations	Nov-16
			Building and Sustaining Data Infrastructures: Putting Policy into Practice	Report with recommendations	Oct-16
Embedding Cultures and Incentives to Support Open Research	Report with recommendations	Oct-16			
EAGDA	Advisory Group		DataTerms		In progress

			Sanctions and accountability	Report with recommendations	Apr-17
			Data management plans	Report with recommendations	Aug-17
			Best practices for infrastructure development and data curation	Minutes	Oct-16
			Governance of data access	Report with recommendations	Jun-15
			Risks of harm from data misuse	Review	Jun-14
			Incentives to support data access	Report with recommendations	May-14
			Protecting participants' confidentiality	Statement with recommendations	Oct-13
Medical Research Council	Research Council		Research Changes Lives 2014 - 2019	Strategic Plan	Dec-13
			Mapping the Landscape of UK Health Data Research & Innovation	Report with recommendations	Oct-17
			Data Sharing Policy	Policy	Sep-16
			Data Management Plan - Guidance for applicant	Guidance	Current
			Data Management Plan – Guidance to reviewers	Guidance	Current
Economic and Social Research Council (ESRC)	Research Council		ESRC Research Data Policy 2018	Policy	Mar-15
			Data management plan: guidance for peer reviewers	Guidance	Current
			Multidisciplinarity on ESRC Grants	Report with analysis	Jan-18
			Funding Guide	Guidance	Oct-17
			Impact, innovation and interdisciplinarity expectations	Guidance	current

			Strategic Plan 2015	Strategic Plan	Jan-15
			Delivery Plan 2016-2020	Strategic Plan	May-16
			Delivery Plan 2015-2016	Strategic Plan	Jun-14
			Delivery Plan 2011-2015	Strategic Plan	Dec-10
Alan Turing Institute	National Institute (flagship?)	Bulk of funding from EPSRC - joint venture between them and the University of Cambridge, the University of Edinburgh, the University of Oxford, University College London (UCL) and the University of Warwick with : Queen Mary University of London, University of Leeds, University of Manchester and University of Newcastle joining in 2018	Annual Report	Report	Aug-17
UK Health Data Analytics Network	Network	Supported by EPSRC - Collaborates with Farr Institute and Alan Turing Institute - Coordinated by Manchester Informatics	Research Roadmap	Report	Jan-17
UK Data Service	Repository		Annual report 2016-2017	Report	Oct-17

		Funded by ESRC and phase one of its Big Data Network	Strategies for Research Centres	Guidance	Apr-12
			Data Management	Guidance	Current
The Farr Institute of Health Informatics and Research	Research Collaboration	Consortium of 10 organisations - led by MRC with Arthritis Research UK, British Heart Foundation, CRUK, Chief Scientist Office, ESRC, EPSRC, NIHR, National Institute for Social Care and Health Research (Welsh Body), and Wellcome Trust	Joint British Academy and Royal Society review: Data management and use: Governance in the 21st century (contribution)	Report with recommendations	Jun-17
			Research Ethics Framework	Guidance	Current
Nuffield Council on Bioethics	Independent charitable body	Nuffield Foundation, Wellcome Trust, and MRC	The collection, linking and use of data in biomedical research and health care: ethical issues	Report with recommendations	Feb-15
Digital Curation Centre	Centre of Expertise	Consortium of Universities of Edinburgh and Glasgow, UKOLN at the University of Bath, and STFC	Overview of funders' data policies	Guidance	
			Data Management Planning Tool	Guidance	
British Academy	National academy/think tank	Receives an annual grant from Department for Business, Innovation, and skills	Joint British Academy and Royal Society review: Data management and use: Governance in the 21st century (contribution)	Report with recommendations	Jun-17
			Data management and use: Governance in the 21st century: Landscape Review	Review	Jun-17

			Data management and use: Governance in the 21st century: Seminar Report	Report	Jul-17
			Data management and use: Priorities for Data Governance	Workshop discussion	Oct-17
The Royal Society	Independent academy	Receives an annual grant from Department for Business, Innovation, and skills alongside gifts from foundations, corporations and individual donors, as well as legacies in the form of bequests and contributions	Joint British Academy and Royal Society review: Data management and use: Governance in the 21st century (contribution)	Report with recommendations	Jun-17
			Data management and use: Governance in the 21st century: Landscape Review	Review	Jun-17
			Data management and use: Governance in the 21st century: Seminar Report	Report	Jul-17
			Data management and use: Priorities for Data Governance	Workshop discussion	Oct-17
			Science as an open enterprise	Report with recommendations	Jun-12
Cancer Research UK	Charity		CRUK Research Strategy: Progress Report	Report	Jun-17
			Policy on Data Sharing and Preservation	Policy	Jun-17
			Data Sharing Guidelines	Guidance	Current
National Institute for Health Research	entity representing publicly-funded research in the NHS	Funded by the Department of Health and Social Care	NIHR open access policy	Policy	Current
			Policy on patient data	Policy	Current
The Health Foundation	Charity		Sharing to improve: four case studies of data sharing in general practice	Review	May-18
Leverhulme Trust	Research Funder		Academic fellowships in 2018.		
			List of research fellowships in 2018.		

			List of research project grants in 2018.		
			List of visiting Professorships		
Government Statistical Service	Network across government and other public bodies		Data for the Public Good	Strategic Plan	Nov-13
UK Statistics Authority	Independent body		Code of Practice for Statistics	Guidance	Feb-18
ADRN	Initiative	Funded by the ESRC			
Genomics England	Initiative	Established by UK Department of Health	The 100,000 Genomes Project Protocol	Guidance	Nov-17
			Data Working Group Report	Report with recommendations	Mar-13
European Genome-phenome Archive	Repository	Part of the ELIXIR Core Data Resources - ELIXIR is a European intergovernmental organisation that is made up of life scientists, computer scientists and support staff			
UK Biobank	Initiative	It was established by the Wellcome Trust, the MRC, the Department of Health, Scottish Government, and the Northwest Regional Development Agency. It has also had funding from the Welsh Government, British Heart Foundation, Cancer Research UK and Diabetes UK.	Data Management & Sharing Plan	Guidance	Dec-12
NHS Digital	Executive non-departmental public body	Created by statute and sponsored by the Department of	GDPR Strategy Document	Strategic Plan	Oct-17

		Health and Social Care			
CLOSER	Initiative/Repository	Funded by ESRC and MRC	Data harmonisation	Research	Ongoing
			Data linkage	Research	Ongoing
			How to access the data	Guidance	Current
Health Data Research UK	Initiative	HDR UK is a joint investment led by the Medical Research Council, together with the National Institute for Health Research (England), the Chief Scientist Office (Scotland), Health and Care Research Wales, Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), the Engineering and Physical Sciences Research Council, the Economic and Social Research Council, the British Heart Foundation and Wellcome	Information Commons for Research	Strategic Plan	
METADAC	Initiative	MRC, Wellcome, and ESRC	METADAC Terms of Reference	Guidance	Mar-17

			Better governance, better access: practising responsible data sharing in the METADAC governance infrastructure	Journal Article	Apr-18
Global Alliance for Genomics & Health	policy-framing and technical standards-setting organization	Core funders are the Canadian International Data Sharing Initiative, National Institutes of Health, and Wellcome	Framework for Responsible Sharing of Genomic and Health-Related Data	Framework/Guidance	Sep-14
Independent Expert Advisory Group on the Data Revolution for Sustainable Development	Advisory body	United Nations Secretary General.	A World That Counts: Mobilising The Data Revolution for Sustainable Development.	Report with recommendations.	Nov-14
AMRC Data Advisory Group	Advisory Body	Association of Medical Research Charities.	AMRC Data Policy	Policy	Nov-13
Health And Social Care Transparency Panel	Advisory Body	UK Govt., Department of Health	HSCTP Terms of Reference	Guidance	Feb-12

i. UK oversight and advisory bodies whose remit relates to the sharing of human participant research data

Body	Creation	Purpose
<p>National Data Guardian</p>	<p>In 2014 by the UK Government, Department of Health. Chair: Dame Fiona Caldicott</p>	<p>The NDG's role is to help make sure the public can trust their confidential information is securely safeguarded and make sure that it is used to support citizens' care and to achieve better outcomes from health and care services. The NDG is guided by three main principles: (a) encouraging clinicians and other members of care teams to share information to enable joined-up care, better diagnosis and treatment, (b) ensuring there are no surprises to the citizen about how their health and care data is being used and that they are given a choice about this and (c) building a dialogue with the public about how we all wish information to be used, to include a range of voices including commercial companies providing drugs and services to the NHS, researchers discovering new connections that transform treatments, and those managing the services.</p>
<p>Independent Group Advising on the Release of Data</p>	<p>Succeeded the Data Access Advisory Group (DAAG) on 1/02/2017 – NHS Digital. Chair: Chris Carrigan</p>	<p>One of the roles of NHS Digital is to ensure a systematic and coherent approach to the scrutiny of requests for data releases. IGARD was established to replace the Data Access Advisory Group of NHS Digital and its primary purpose is to have oversight of all requests for dissemination of confidential information as defined in Section 263 of the Health and Social Care Act 2012 including personal, confidential, anonymous or sensitive data as defined by relevant legislation. Its secondary purpose is to make general recommendation or observations to NHS Digital via approved routes (e.g. Caldicott Guardian) about processes, policies and procedures including transparency measures such as registers and produce an annual report.</p>
<p>Confidentiality Advisory Group</p>	<p>NHS Health Research Authority – Chair: Dr Tony Calland MBE</p>	<p>CAG provides independent expert advice to the Health Research Authority and Secretary of State for Health on whether an application to process identifiable patient information without consent should or should not be approved, under s.251 of the NHS Act (2006).</p>
<p>Richard Doll Centenary Archive Data Access</p>	<p>Nuffield Department of Population Health – Chair: Professor Dame Anne Mills</p>	<p>The Committee provides advice and governance for the Archive on data access and sharing procedures. The Committee monitors data sharing requests, decisions made for active studies and monitors the lifecycle of any data agreements. It also</p>

Oversight Committee		decides on requests for data access and sharing for inactive studies that no longer have an active Custodian.
ESRC Data Infrastructure Strategic Advisory Committee	UK Data Service – Chair Gordon Marshall	DISAC is responsible for strategic oversight and the provision to the ESRC of independent assurance for the UK Data Service and investments made under the Big Data Network.
Information Governance Alliance		Core members of the Information Governance Alliance are the Department of Health, NHS England, NHS Digital and Public Health England. Representatives from the Information Commissioner's Office and the National Data Guardian's Office also sit on the Board. They offer advice and support, develop networks, publish guidance, endorse guidance produced by others, and work with local and national organisations to improve knowledge and practice of information governance across the health and care system. All NHS Digital guidance, training and resources on information management in health and care are now delivered through the IGA.
Institutional Ethics committees and NHS Research Ethics Service	Individual institutions and Health Research Authority	These committees review research involving human participants or personal data and advise researchers on how to ensure research involving human participants is conducted in a way that respects the dignity, rights and welfare of participants, and minimises risk to participants.

Appendix 6: Data related issues around which a future advisory model may be required to offer advice

Further details of initiatives and studies that have implemented solutions to some of these challenges, and provide a framework for future studies.

a. Managing the expectations of stakeholders in an inclusive manner

One key challenge is that of reconciling open science and unfettered access to data with the legal and ethical responsibilities of stewards of research datasets and researchers.¹⁷ Research participants are one of the key stakeholders with high expectations in this context, and an increasing body of evidence demonstrates their preference to be involved in decisions around its future use, have the ability to opt in or out of downstream research, or be represented through independent oversight.

METADAC provides a governance infrastructure for managing the reasonable expectations of research participants who provided data and samples to five longitudinal studies in the UK. DataSHIELD has developed methods for distributed computing and parallelized analysis to enable anonymised data to be drawn from multiple longitudinal studies and accessed remotely to enable data reuse. The Secure Unified Research Environment (SURE) is an online computing system developed to manage anonymization for outputs of population health and health services research conducted via an online data centre that allows researchers to access and analyse linked health-related data for approved studies.¹⁸ Data is minimally processed by its custodians before upload, and scored against a checklist at the point of download to determine the privacy risk and assign an anonymization treatment.

b. Harmonizing linked datasets and managing relevant risks

While the expectations of stakeholders need to be managed in any occasion, particular problems arise in the case of pooling information between studies, many times on an international level. One such challenge is the potential conflict of purposes between these sources and their implications for data management. For instance, the intersection of biological research and clinical practice, governance systems are required that on one hand provide for the avoidance of prohibited re-identification, and on the other permit re-identification where clinically relevant findings may need to be fed back to participants. With regards to the potential inclusion of both genomic and clinical data in data sets that may span multiple legal jurisdictions, Privacy-Preserving Record Linkage (PPRL) has been hailed as the current gold standard on this challenge by a joint task force of the IRDiRC and the GA4GH.¹⁹

The DataSHaPER platform offers an example of how linked datasets can be harmonised, allowing data to be integrated across bio-clinical studies;²⁰ its model has been applied to large

¹⁷ Laurie G, *et al.* Charting Regulatory Stewardship in Health Research: Making the Invisible Visible. *Cambridge Quarterly of Healthcare Ethics*. **2018**. 27. 2. 333-347.

¹⁸ O'Keefe C, *et al.* Anonymization for outputs of population health and health services research conducted via an online data centre. *J Am Med Inform Assoc*. **2017**. 25. 3. 315

¹⁹ Baker D, *et al.* Privacy-Preserving Linkage of Genomic and Clinical Data Sets. *IEEE Transactions on Computational Biology and Bioinformatics*. **2018**.

²⁰ Fortier I, *et al.* Quality, quantity and harmony: the DataSHaPER approach to integrating data across bio-clinical studies. *Int. J Epidemiology*. **2010**. 39. 5. 1383.

international data sharing projects, including BioSHARE.²¹ From a governance perspective it is noteworthy that it was created by a multidisciplinary consortium of experts that was pulled together and coordinated by three international organizations (P3G, PHOEBE and CPT).

c. Overcoming data access barriers and delays in the case of large datasets

Another set of issues comes from the application of some of the above challenges in the context of routinely-collected (i.e. administrative) data of patients and service users as they pass through health service and other public sector organisations for health research. The Secure Anonymised Information Linkage (SAIL) databank in Wales offers an example of an initiative exploring the potential reuse of electronically-held, person-based, routinely-collected data, whilst the Registered Access policy model by the GA4GH aims to streamline access to data that requires agreement to basic terms and conditions such as the use of DNA sequence and health data in research. Registered Access aims to address the open / controlled access binary whereby qualified researchers have to apply for access on a project-by-project basis, each requiring independent review, and permit research data to be more freely available online under a model whereby registered users agree to use the data appropriately. Examples where this has been applied so-far include the Wellcome Trust Case-Control Consortium, the DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources (DECIPHER) and the Critical Assessment of Genome Interpretation (CAGI).

²¹ Doiron D, *et al.* Data harmonization and federated analysis of population-based studies: the BioSHaRE project. *Emerging Themes in Epidemiology*. **2013**. 10. 12

Appendix 7: Cross-funder Policy Committees acting in the UK and elsewhere

Body	Body Type	Funders who Participate	What issues are they dealing with?	What are the organisational features of this body?
Health Data Research Institute UK	Research Institute	MRC-led, British Heart Foundation, Chief Scientist Office, Scotland, EPSRC, ESRC, Wellcome and others – list here	Development of capacity and methods to accelerate pace and scale of health and biomedical data science, including: partnerships to facilitate rapid application of the institutes research; secure and interoperable research environments; cross cutting analytical methods, tools and standards	Geographically distributed institute across a number of centres of excellence in the UK, with an appointed Director responsible overseeing scientific direction and expenditure. Aimed to build connectivity and enhance communication between institutes and with the wider community
Dementias Platform UK	Public-private research partnership	MRC-led, NIHR, ARUK, Stroke Association et al – list here	Enhancing research into dementia by accelerating discovery of new ways to understand, diagnose and treat dementia. Provides integrated access to data from 50 cohort studies involving over 2 million individuals.	Academic director and executive leadership team comprising academic and industry representatives, each reporting to funder appointed Oversight Board (comprising UK and international experts in Dementia); expert scientific steering group (20 members); Company Partner Forum comprising industry representatives; 4 international expert advisors
The Farr Institute (UK)	Research collaboration	MRC-led, consortium of ten funders – list here	Analyses clinical, biological, population and environmental data from multiple sources and collaborates with the government, public sector, academia and industry to unleash its value for public benefit.	21 academic institutions and health partners; International Advisory Board tasked with identifying strategic research and development opportunities and initiatives for the development of expertise in the UK and internationally (12 expert members representing scientific research, clinical practice and industry)
Understanding Patient Data (UK)	Research Partnership	Wellcome, MRC, ESRC, DHSC, PHE – see here	Working with patients, charities and healthcare professionals to communicate the uses of health information, explaining how and why data can be used for care and research, what is permissible, and how personal information is protected	Steering Committee comprising funder representatives and other organisations. Advisory group provide broader perspective through stakeholder representation, including academics, public involvement and

				engagement champions, communication experts and policy professionals
Research Data Alliance	Community-Driven Organization	A list of important research funders across borders – e.g. the European Commission, the Japanese Science and Technology Agency – list here	Enabling data sharing across domain, research, national, geographical and generational boundaries. Working Groups and Interest Groups draw from academia, industry and government to develop and adopt infrastructure that promotes data sharing and data-driven research, and accelerate growth of a cohesive data community.	Nominated Council responsible for oversight, strategy and sustainability; Secretariat, and establishment of Council Subgroups, which may include members outside the council; Technical Advisory Board comprising academics with technical expertise and IT specialists; Organisational Advisory Board representing members
Force 11	Community-driven organization	Such funders as the Uni of Oxford, Elsevier and others – see here	Investigating the capacity of digital publishing and information technology to enrich scholarly communication, such as through publication of software tools and research communication by means of social media channels.	Cross- disciplinary membership drawing from a broad range of sectors including commercial and non-profit publishers, libraries, scholarly societies, universities, other private and public sector organizations
Open Research Funders Group (ORFG)	Membership-based partnership	A list of important funders including Wellcome, the Bill & Melinda Gates Foundation and the American Heart Association – list here	Developing solutions for monitoring open access and open data policy compliance and tracking impact, including compilation of resources that evaluate impact, identifying best practices for policy development, and streamlining workflows	Membership is open to non-profit organizations around the globe that maintain or aid charitable, educational, or other activities serving the public good.
Crossref Funder Advisory Group	Not-for-profit membership organization	Individuals representing various funders across borders – including MRC, Wellcome, RCUK list here .	Reviewing Crossref infrastructure and services to better serve funders, looking at models for funder engagement with the database and development of a registry of grant identifiers	Invitation only advisory group of funder representatives, chaired and facilitated by an appointed Crossref representative. Quarterly meetings convened by conference call

<p>Science Europe</p>	<p>Membership-based partnership</p>	<p>43 funders across Europe including UK research councils – list here</p>	<p>Enhancing opportunities for collaboration between funding organisations and contributing to the development of research policy in Europe.; also contributing to global debate on science policy through its roles in the Global Research Council</p>	<p>Governing Board of 10 elected members who make, guide and monitor strategic direction, meeting at least twice per year; General Assembly comprising representatives of each member organisation. Cross- Disciplinary Scientific Advisory Committee (30 members) advising on key European research policy topics. Annual Symposium gathering independent scientific experts with high level representatives from Science Europe Member Organisations and external stakeholders. Secretariat teams comprise management, policy and research; communication and administration; 11 staff in total.</p>
<p>Global Research Council</p>	<p>Virtual organisation comprising heads of science and funding agencies from around the world</p>	<p>Science and engineering funding agencies from around the world</p>	<p>Promoting the sharing of data and best practices for high-quality collaboration among funding agencies worldwide through improved communication and identifying areas of common concern</p>	<p>Governing Board consisting of up to 12 heads of national research councils, distributed by continent. Executive support group of experienced senior level officials from participating research councils. Programme Committee tasked with considering suggestions for topics from participants and providing recommendations to the Governing Board. Executive Secretary provided by one of the participating organisations. Annual Global and Regional Meetings.</p>
<p>Global Alliance for Genomics and Health</p>	<p>Not-for-profit membership organisation</p>	<p>Wellcome, NIH (US) and CanShare (Canada) core funders – see here</p>	<p>Working together to create frameworks, standards and tools overcome technical and regulatory hurdles and enable the responsible, voluntary, and secure sharing of genomic and health-related data.</p>	<p>Executive Leadership Team; Steering Committee; Strategic Advisory Board and Secretariat. Foundational Workstreams and Technical Work Streams each informed by Driver Projects, comprising real-world genomic data initiatives sourced from around the globe. All draw on a Partner</p>

				Engagement team, comprising clinical expertise in 4 theme areas and representing diverse geographic areas
CODATA	Committee set up by the International Science Council, an NGO	The International Science Council currently has 141 Member Organizations, 39 Member Unions and Associations and 29 Affiliated Members. ²²	Promoting the policy, technological and cultural changes that are essential to promote Open Science, under the guiding principles that research data should be findable, accessible, interoperable and reusable, and that data susceptible for reuse should be as open as possible and as closed as necessary.	Standing Committees and strategic executive led initiatives; Task Groups and Working Groups; biannual General Assembly of member delegates

²² For a full list of members see <<https://council.science/about-us/our-members>> accessed 18 Sep. 18