

# GBC Working Group on Open Data Strategies

## Consultation Paper

### **IMPORTANT NOTICE**

This consultation paper was prepared by the GBC Secretariat based on the discussions and inputs of a Working Group comprised of experts drawn from GBC member and observer organisations. This paper does not present recommendations or policy proposals: rather, it presents a series of ideas for discussion and feedback by GBC's stakeholders and communities. These ideas reflect the key themes that emerged from the Working Group's discussions, but will not necessarily always fully reflect the views of individual Working Group members. None of the ideas, viewpoints or any other content in this paper should be taken to represent the policy or positions of the organisations to which Working Group members are affiliated.

## Executive Summary

Research funders around the world have set in place policies requiring that the researchers they fund seek to maximise the availability of research data and related research outputs with as few restrictions as possible. Whilst there is a growing recognition of the benefits of open research data and of open research practices more broadly, significant cultural and technical challenges remain in ensuring the successful implementation of these policies. For life science and biomedical research, this includes a key need to ensure biodata resources, which play a pivotal role in the delivery of these policies, are developed and sustained.

The Global Biodata Coalition (GBC) is a coalition of research funders that aims to stabilise and ensure sustainable funding for global biodata infrastructure. The GBC Working Group on Open Data Strategies brought together representatives from GBC Member and Observer organisations to consider ways that funders could cooperate to enhance implementation of their policies to support open research data, and consider the implications of these policies for biodata resources. It had two key objectives:

- To explore ways that funders of biological, biomedical and life science research could better share knowledge and collaborate in developing, implementing and evaluating open data policies and mandates; and in doing so,
- To maximise the value of biodata globally, and support and strengthen biodata resources.

Based on its discussions, the Working Group has identified a series of options as set out below. As GBC is a coalition of funders of research, particular emphasis is placed on those aspects of open data policy that funders can work together to improve.

These options are intended as a starting point for further discussion with GBC's stakeholders and communities - including research funders, biodata resource managers, publishers and the wider life sciences research community. We invite comments, input and ideas on this consultation paper from all those with an interest in maximising the value of open research data. GBC will use the outcomes of this consultation to develop a White Paper describing how GBC and its member organisations will work together to enhance the implementation of open data policies, as part of GBC's core mission to strengthen and sustain the global biodata resource infrastructure.

## Proposed Options

### 1. Aligning policies and data management plan (DMP) requirements

Funders should explore opportunities to work together to:

- a. Develop a common framework to encourage aligned funder policies**—this would set out high-level consensus elements of funder open data policies based on the experience of funders who have developed and implemented such policies. The goal is not to harmonise existing funder policies, but to identify a common framework of shared beliefs and best practices. This will assist funders who have yet to establish policies and aid international research collaborations.
- b. Identify and share good practices in the implementation of DMP requirements**—there is considerable scope for funders to share and learn from experiences of rolling out requirements for DMPs. The GBC is a valuable forum for these discussions, which could explore the potential to develop consensus cross-funder guidance — for example on how DMPs can best be structured, costed, assessed and monitored.

- c. **Innovate and test new approaches to maximise the value of DMPs**—there is potential for funders to collaborate on pilot approaches for making DMPs machine-actionable and accessible, and assess the value for researchers, funders and biodata resources.
- d. **Develop data sustainability planning as an element of DMP policy**—in many cases, specific tools and data resources - such as data portals and coordination centres - are developed in order directly to support projects and initiatives; DMPs should address the sustainability of data beyond the term of such projects/initiatives, including the use of biodata resources with long-term sustainability as underlying infrastructure or the safe migration of data at termination of project or initiative.

## 2. Working with biodata resources to advance policy implementation

Funders should explore opportunities to:

- a. **Align support for biodata resources to the requirements of open data policies**—ensuring their role in implementing open data policies is adequately supported (either directly, or through working with other organisations where direct funding or policy support for data infrastructure is outside the funder’s legal mandate). This might include:
  - i. assessing the current and future demand for biodata resources in light of open data policies;
  - ii. considering carefully the existing global resource landscape when making decisions over creating new resources or supporting major bespoke project databases;
  - iii. targeting support for biodata resources of importance to researchers to fully implement relevant community standards, including the FAIR principles.
- b. **Explore the potential to develop shared criteria and resources to assist researchers in finding and selecting suitable repositories**, including:
  - i. specifically recommending Global Core Biodata Resource (GCBRs);
  - ii. bringing together existing guidance and resources to create a common set of criteria that could be used across GBC funders;
  - iii. exploring the scope to develop a ‘Yellow Pages’ or “wizard” style resource to signpost repositories, potentially through leveraging an existing tool or service and drawing on GBC’s Global Inventory as a key source.
- c. **Develop common cross-funder principles on data preservation, sharing and retention across the data lifecycle** to set clear expectations for biodata resources and researchers, and common metrics for tracking usage, utility and impact across biodata resources.

## 3. Embedding good practice: training, incentives and culture change

Funders should work with research institutions and the broader research community to:

- a. **Explore opportunities to collaborate in promoting and funding training in data skills** for research professionals, including potentially building a resource of available opportunities.
- b. **Recognise the key roles of specialist data managers, data stewards and data scientists in research institutions and biodata resources** — potentially through developing a common statement or principles to recognise and support these roles.
- c. **Accelerate collaborative efforts to incentivise and reward open data practices, which may include:**
  - i. identifying good practice in embedding the principles of the San Francisco Declaration on Research Assessment (DORA) and Coalition for Advancing Research Assessment (CoARA) to recognise a full and diverse range of research outputs and contributions, in particular efforts to generate, curate and share high-quality data sets and data resources.

- ii. agreeing common approaches to implement persistent identifier schemes and emerging schemes to attribute contributions, and encouraging GCBRs and other resources to actively implement attribution and contribution tracking.
- iii. exploring the potential to develop funding tools—including prizes and targeted funding calls — to support innovation in open data practices and celebrate researchers and data managers who have pioneered open approaches.

#### **4. Embedding global equity in open data policies**

Funders should actively work together to ensure equity is at the heart of their approaches to promote open data. Specifically they should:

- a. **Adopt a common commitment to develop open data policies in an equitable, ethical and efficient manner** — GBC could serve to promote cross-funder dialogue, sharing of good practice and engagement with global stakeholders to ensure the benefits of open data are realised equitably, and in line with the CARE principles.
- b. **Monitor the implementation of open data policies to ensure global equity** — develop collaborative mechanisms to monitor the implementation of funder policies in line with the CARE principles and related frameworks, including access and usage of data by as yet under-served communities.

# Table of Contents

<b>Executive Summary</b>	<b>2</b>
<b>1. Introduction</b>	<b>6</b>
1.1 The challenge	6
1.2 The opportunity—a role for the Global Biodata Coalition	8
1.3 The GBC Board Working Group on Open Data Strategies	8
1.4 Scope	9
<b>2. Advancing Open Data Strategies—Areas for Funder Cooperation</b>	<b>11</b>
2.1 Overview	11
2.2 Aligning funder policies and their implementation	11
2.3 Working with biodata resources to advance policy implementation	14
2.4 Embedding good practice: training, incentives and culture change	16
2.5 Ensuring global equity in implementation of open data policies	18
<b>Annexes</b>	<b>19</b>
Annex A: Membership of the GBC Board Working Group on Open Data Strategies	19
<b>References</b>	<b>20</b>

# 1. Introduction

## 1.1 The challenge

1. In recent years, research funders around the world have introduced policies requiring that researchers manage and share research data and related research outputs (including research software and materials) in a manner that maximises the societal benefit that flows from the funded research, including making these data widely available to other researchers in a timely and responsible manner (see Neylon 2017, Davidson et al. 2019).
2. The development of these policies has been driven by a growing belief among research funders in the benefits of opening up access to research data (see Gaba et al. 2020). In particular:
  - **Amplifying benefit** — making research data available to researchers can enable it to be accessed, combined and re-used in new and innovative ways to advance discovery and its application to benefit society, including for research avenues that might never have been envisaged by those generating the data.
  - **Enabling reproducibility** — providing access to underlying research data and associated outputs is vital to enable published research findings to be scrutinised and replicated, with increasing attention focused over recent years on the ‘reproducibility crisis’.
  - **Increasing efficiency** — facilitating the openness, discoverability and re-use of research and its outputs could potentially reduce unnecessary duplication and waste, and increase the broader efficiency of the research enterprise.
3. For public and charitable research funders, there is a more fundamental belief that research outputs that have been funded from the public purse or from charitable funds should be publicly accessible (see OECD 2022). Policies to encourage open data have developed alongside, and sometimes flowed from, the related drive for open access to published research findings. Open access policies aim to ensure that published research findings can be freely accessed and re-used by researchers and the wider public. In addition, there has been a related policy drive from governments to unlock the economic potential of ‘big data’, and rapid developments in data science, artificial intelligence and machine-learning tools (see European Commission 2020).
4. The development of funder policies for open data is also occurring in the context of a broader global movement to embed open research practices. This movement has rapidly gained momentum and political support over recent years, and the landscape is rapidly evolving. Examples of key recent developments have included:
  - Key international declarations—including perhaps most notably the UNESCO Recommendation on Open Science in November 2021 (see UNESCO 2021).
  - Major national policy statements—including, for example, the US Office of Science and Technology Policy’s memorandum in August 2022 on “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research” (the ‘Nelson memo’, see OSTP 2022)
  - Key regional level policies and platforms—for example:
    - The European Commission’s focus on open science—including establishment of a Open Science Policy Platform and work to shape and implement a European Open Science Cloud (see Burgelman et al. 2019);
    - Nationwide data initiatives and platforms in Australia—including the Australian Research Data Commons and Australian BioCommons (see Barker et al. 2019);

- The African Open Science Framework (See Smith and Veldsman 2018).
  - Development of community standards—including the FAIR (Findable, Accessible, Interoperable and Reusable) principles for stewardship of scientific data; and CARE principles for indigenous data governance (see Wilkinson et al. 2016, Carroll et al. 2020).
  - The growth of community-led organisations to address challenges in data management and open science — for example, the Research Data Alliance and FORCE11.
  - The development of policies on open data from other key actors, in parallel with funder requirements — most notably, scientific journals and research institutions (See Hrynaszkiewicz et al. 2020).
5. In addition, the COVID-19 pandemic sparked an unprecedented focus on opening up access to research findings and data to aid the global effort to combat the pandemic, with initiatives such as the rapid establishment of the COVID-19 Data Portal in Europe and the COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University in the USA (see Harrison et al 2021, Gardner 2022).
  6. Despite gathering recognition for the importance of research data sharing and open science, barriers remain in embedding open data practices within the research community and ensuring the successful implementation of funder policies (See Anger et al. 2022, Hughes et al. 2023). Different disciplines are also at very different stages in terms of developing resources, practices and cultures for research data management and sharing, and different data types raise distinct challenges. However, a few high-level challenges are common across the board:
    - **Building and sustaining resources, tools and standards** — data sharing requires community databases, tools and standards to ensure long-term preservation and accessibility of research data. For some research fields and data types this infrastructure has yet to be developed. For others, high quality resources and tools are in place, but their long-term funding and sustainability is a critical concern.
    - **Developing data skills** — there is a need to develop skills across the research community and to develop and support the careers of specialist data managers and stewards to support research teams and curate data.
    - **Establishing incentives** — sharing research data carries a cost in terms of time and resource, so there is a need to ensure that researchers are incentivised to manage and share research data and receive appropriate credit for doing so.
    - **Addressing ethical, legal and social issues** — where data relate to human subjects or otherwise raise ethical, sovereignty or security concerns, a careful balance needs to be struck between the benefits of making these data open and the need to limit access where safeguards are required.
  7. As noted above, biodata resources play a pivotal role in the delivery of funder policies. Their long-term funding and sustainability are crucial to enable researchers to manage and share their data effectively and ensure the long-term preservation of these data. As funder policies around open data become more stringent, biodata resources will face increasing demand from researchers, and there is a need for funders to work with them and support them in implementing their policy requirements. These pressures on resources come alongside the challenges they already face in light of the broader trends highlighted above—for example, in dealing with the rising data volumes and complexity; in embedding the FAIR principles and other community standards; and in meeting the needs of users adopting advanced AI tools.
  8. Overcoming the many and significant challenges to the implementation of open data policies will require funders to work together, and to work with other key stakeholders including

biodata resource managers, research institutions, journals, learned societies and the broader research community.

## 1.2 The opportunity—a role for the Global Biodata Coalition

9. The Global Biodata Coalition (GBC) is a forum for research funders to better coordinate and share approaches for the efficient management and growth of biodata resources worldwide. The GBC aims to stabilise and ensure sustainable financial support for the global biodata infrastructure.
10. As a key part of its work, the GBC will identify and maintain a list of Global Core Biodata Resources (GCBRs) that are crucial for sustaining the broader biodata infrastructure. The GBC concluded the first GCBR selection process in December 2022 publishing an initial list of 37 resources (see Global Biodata Coalition 2022). In early 2023, it convened the GCBR Forum to bring together senior leaders from these resources to discuss issues relating to sustainability and other opportunities to work together to address common challenges. GBC is running a second round of selection for GCBRs in 2023 to give further resources a rapid opportunity to apply and ensure a more representative and inclusive GCBR list.
11. While GCBRs will be a key initial focus for the GBC's work to develop and implement approaches to advance sustainability, there is a vital need to characterise and maintain an overview of the entire global infrastructure of biodata resources. In 2023, the GBC published a global inventory of biodata resources, providing a fuller picture of the global landscape of biodata resources than has been available previously (see Imker et al 2023). Using a machine learning-enabled natural language processing approach to mine the literature, the work identified a corpus of 3,112 unique resources based on articles published between 2011 and 2021.
12. Open data is at the very heart of GBC's mission and values. Its member and observer organisations are passionately committed to embedding open data practices and to unlocking the huge potential benefits to society from making research data widely available. In particular, its member funders recognise the key need to support and sustain biodata resources in order to unlock the potential of open data.
13. The GBC is well placed to facilitate cross-funder dialogue and coordination in the implementation of open data policies. First, and most importantly, it brings together leading global funders in the life sciences and biomedicine — including many of the funders who have been amongst the earliest pioneers in the development of data sharing and open access policies. Second, with the announcement of the initial GCBR list and establishment of the GCBR Forum, it is well positioned to facilitate a dialogue between funders and biodata resources to explore the implications and needs for biodata resources relating to open data policies.

## 1.3 The GBC Board Working Group on Open Data Strategies

14. In early 2022, the GBC's Board of Funders established two cross-funder Working Groups: one to explore the topic of Open Data Strategies, and the other focusing on Sustainability. Both Working Groups bring together senior representatives from GBC Member and Observer organisations and are chaired by the Chair of the GBC Board. Each Working Group was tasked with developing a Consultation Paper to explore the challenges, and set out ways forward for



consideration by the GBC Board and consultation with the GBC's wider stakeholder community.

15. The Working Group on Open Data Strategies (referred to as “the Working Group” in this paper) had representatives from several GBC partner funders, namely: the Australian National Health and Medical Research Council (NHMRC), the Canadian Institutes of Health Research (CIHR), the European Commission, the Swiss National Science Foundation (SNSF), UKRI, the US National Science Foundation, the US National Institutes of Health and Wellcome. The full membership is listed at **Annex A**.
16. The Working Group was tasked with the following key aims:
  - To explore ways that funders of biological, biomedical and life science research could better share knowledge and collaborate in developing, implementing and evaluating open data policies and mandates; and in doing so,
  - To maximise the value of biodata globally, and support and strengthen biodata resources.
17. This Consultation Paper was developed by the GBC Board Working Group on Open Data Strategies with the support of the GBC Secretariat. The Working Group met four times to discuss the issues and formulate the ideas presented here for wider discussion and consultation. To help inform the discussions, the Secretariat undertook a web-based survey to collect information on existing policies, guidance and approaches from the organisations represented on the Working Group.
18. The Consultation Paper was discussed by the GBC Board of Funders at its in-person meeting in March 2023. It was then developed further by the Working Group based on the Board's feedback and inputs. GBC is keen to receive comments and feedback on the Paper from its key partners and stakeholders - including funding bodies (including those already engaged in GBC and those who are not), biodata resources (including the GCBR Forum) and the broader open data and research communities, including resource users and publishers.
19. Based on the feedback received, the GBC Secretariat will develop a White Paper setting out plans to enhance sharing of good practice for individual funders and explore collaborative approaches amongst funders to advance biodata resource sustainability for approval by the GBC Board.

## 1.4 Scope

20. In progressing its discussions, the Working Group agreed the following broad parameters for the scope of this Consultation Paper:
  - **The definition and scope of ‘open data’ incorporates digital assets resulting from research** — for the purposes of this Consultation Paper, open data encompasses digital assets that are the products of research supported by funders, recognizing research data and associated code are often inseparable and that many biodata resources are necessarily also software providers.
  - **Recognition that not all data can be fully open** — as discussed further below, all funders recognise that access to some data needs to be controlled to protect the privacy of individuals or safeguard against other risks of harm. The concept of “as open as possible,

as closed as necessary” has entered into widespread use among funders. This encapsulates the philosophy that data should be made openly available wherever this is appropriate, with proportionate controls and limits introduced where required that serve to maximise access and reuse of data in a safe, secure and ethical manner. Our discussion of open data policies recognises this concept is a vital and core element.

- **Build on previous and ongoing work around open data, and focus specifically on issues of collaboration amongst funders** — research data management and open data have been the subject of much previous and ongoing work from international and national organisations. The scope of this Consultation Paper is limited to open data issues of specific relevance to funders and focuses on promoting knowledge exchange and collaboration among funders on these issues. The Working Group emphasised the need to build upon the work of others and seeks to acknowledge these activities where appropriate. It also emphasised the importance of seeking broad feedback from a broad range of stakeholders — including researchers, resource managers and journals — in helping develop and refine the options presented.
- **The options presented in this Consultation Paper are relevant to all research funders, with an initial emphasis on those engaged in GBC** — the options presented in this Consultation Paper for cross-funder collaboration are focused initially on those engaged with GBC (Members and Observers) with the hope that collaborative activities proposed could extend to include other research funders in all parts of the world.
- **The GBC and Working Group fully endorse recognised community principles** — the GBC and the members of the Working Group all fully support and endorse the FAIR, CARE and TRUST principles. The options presented here are targeted at accelerating the uptake of these principles for the data outputs resulting from the research that funders support, ensuring that data and metadata are managed and shared in a manner that maximises their value.

## 2. Advancing Open Data Strategies—Areas for Funder Cooperation

### 2.1 Overview

21. The discussions of the Working Group focused on four broad topics where it felt there was a strong potential opportunity and need for greater cross-funder cooperation in order to advance the implementation of open data policies, and achieve the shared goal to maximise the value of research data. These topics are set out in the Sections that follow, namely:
  - Aligning funder policies and their implementation (Section 2.2)
  - Working with biodata resources to advance policy implementation (Section 2.3)
  - Embedding good practice – training, incentives and culture change (Section 2.4)
  - Ensuring global enquiry (Section 2.5)
22. In each section, a brief summary of the issues considered by the Working Group is provided to introduce the options proposed by the Working Group. As noted above, these options are presented as a starting point for discussion and feedback from GBC Board members and our wider stakeholder communities.

### 2.2 Aligning funder policies and their implementation

23. The funding organisations represented on the Working Group all have policies setting out expectations for the researchers that they fund in managing and sharing research data. Some—such as the US National Institutes of Health (NIH), the UK Medical Research Council and Wellcome—have policy requirements dating back almost two decades, which have been reviewed and updated periodically over time. For example, a new NIH data sharing policy came into force in 2023 following a major review built on a previous policy that had been in place since 2003. Other funders — such as the Swiss National Science Foundation (SNSF) and Australian NHMRC — have introduced policy requirements much more recently. Links to the funder policies discussed in this section can be found in the **References** section below.
24. Funder policies diverge significantly in terms of their format and style and in the details of their implementation, as is discussed further below. Some are standalone policy statements, and others are subsections of broader documents setting out granting policies and procedures. They also vary somewhat in scope—whilst all cover research data, some also cover software and/or materials. For example, Wellcome updated its previous Data Management and Sharing Policy in 2017, expanding its scope to cover research data, software and materials.
25. At their heart, however, the core outcomes the policies are seeking, and the high-level expectations they set for researchers are very consistent, in particular:
  - **Maximising benefit.** Policies require researchers to manage and share research data in a timely and responsible manner and with as few restrictions as possible, with the ultimate goal of maximising its value to society.

- **Recognising limits on openness.** Policies recognise that not all data can be shared openly, and that proportionate limits and restrictions on access may be required to safeguard the privacy of research participants or guard against other risks of harm. Many also allow limits on sharing whilst researchers seek appropriate intellectual property protection, where this is in line with maximising the societal benefit. As noted above, the concept of making data “as open as possible, as closed as necessary” has gained traction as an underpinning tenet in this regard.
- **Requiring Data management plans.** Policies increasingly require researchers to have in place a Data Management Plan (DMP), or equivalent plan, which documents how they will manage and share the data resulting from their research. Many funders require this to be included as part of funding applications; whereas others require research institutions to ensure these plans are in place as a condition for receipt of the funds.
- **Provisioning costs.** Policies recognise that data sharing has a cost and allow costs associated with data management and sharing to be included in funding applications.
- **Encouraging data deposition.** Policies recommend the deposition of data in stable and recognised community repositories wherever these exist for a particular data type (this issue is discussed further in Section 2.3 below)

26. In considering current funder policies regarding research data, the Working Group highlighted some emerging questions:

- **Specification of which data to share** — many funder policies have explicit requirements that data and associated code underlying published research findings should be preserved and made available by the time of publication wherever it is possible to do so. Beyond this, there is often a lack of specificity over which data should be retained and shared, and it is left to researchers to define this as part of DMPs. It is recognised that practices in this regard will vary considerably between research disciplines and data types.
- **Adoption of FAIR and other emerging community standards** — funder policies increasingly make reference to the FAIR principles, and encourage the adoption of key elements —such as the use of persistent identifiers, community metadata standards and specific data licences. However, policies currently stop short of an explicit requirement for FAIR or other standards. This likely reflects that the FAIR principles are still aspirational at this stage, and there are no mature or well-established mechanisms to assess compliance in binary fashion.

27. Whilst the requirement for a DMP is increasingly common across funder policies, funders currently vary significantly in the nature of guidance they provide to researchers and their processes for implementing this requirement. For example:

- **Format of DMPs:** Some funders provide specific templates for DMPs, whereas others do not specify a set format and provide more general guidance on the elements that a DMP should include. Some core elements required — including specification of the timing and mechanism for sharing, where the data will be deposited and the standards that will be adopted — are consistent, although the specific detail requested may vary considerably.
- **Costing of DMPs:** There are also varying levels of guidance for researchers on how to cost DMPs, in terms of the categories of costs that should be considered, which are eligible for funding, and how these should be included in funding applications.
- **Assessment of DMPs:** There are also a range of processes through which DMPs are requested and assessed at the point of funding. For some funders, DMP are required as part of grant applications and considered by reviewers and funding panels as part of normal peer review and/or by their own scientific program officers. Other funders do not

require the completion of a DMP until funding is awarded. For some, such as the European Commission, it may be required as an early project deliverable to be assessed by program officers. For others, formal submission may not be required and responsibility left to the research institution to ensure an adequate DMP is in place.

- **Monitoring of DMPs:** Most funder policies indicate that DMPs will be monitored and that there are potential consequences if researchers fail to manage and share data in line with policy requirements. The extent to which monitoring happens in practice is, however, highly variable and many funders have struggled to track this in any meaningful way.
28. Whilst DMPs have been a valuable tool in implementing funder policies, their full potential has yet to be realised for either funders or researchers. It is recognised that ideally research groups would use DMPs as living documents which they develop as their research progresses. However, there is little sign at present that DMPs are being used in this way at large. Similarly, there is significant and as yet untapped potential for funders to use DMPs to more systematically plan and target support for data resources and tools required for their delivery, and to link these plans to data outputs. Through the Research Data Alliance, there have been discussions over the potential benefits that could be gained for both funders and researchers from making DMPs machine-readable and encouraging their publication. However, there has been little tangible progress as yet to implement these steps in practice.
29. In light of its discussions, the Working Group felt there was considerable scope for continued dialogue and coordination between funders in implementing funder policies and DMP requirements. The options identified are summarised in **Box 1** below.

**Box 1—Alignment of policies and DMP requirements**

It is proposed that funders consider the following options for cross-funder cooperation:

- a. **Develop a common framework to encourage aligned funder policies**—this would set out high-level consensus elements of funder open data policies, based on the experience of funders who have developed and implemented such policies. The goal is not to harmonise existing funder policies, but to identify a common framework of shared beliefs and best practices. This will assist funders who have yet to establish policies and aid international research collaborations.
- b. **Identify and share good practices in the implementation of Data Management Plan (DMP) requirements**—there is considerable scope for funders to share and learn from experiences of rolling out DMPs. The GBC is a valuable forum for these discussions, which could explore the potential to develop consensus cross-funder guidance — for example on how DMPs can best be structured, costed, assessed and monitored.
- c. **Innovate and test new approaches to maximise the value of DMPs**—there is potential for funders to collaborate on pilot approaches for making DMPs machine-actionable and accessible, and assess the value this brings to researchers, funders and data resources.
- d. **Develop data sustainability planning as an element of DMP policy**—in many cases, specific tools and data resources - such as data portals and coordination centres - are developed in order to directly support projects and initiatives; DMPs should address the sustainability of data beyond the term of such projects/initiatives, including the use of biodata resources with long-term sustainability as underlying infrastructure or the safe migration of data at termination of project/initiative.

## 2.3 Working with biodata resources to advance policy implementation

30. A key common element of funder policies on open data for the life and biomedical sciences is an expectation that data be deposited in biodata resources that serve as recognised community repositories, where these exist for a particular data type. Most funders state a clear preference that researchers utilise curated subject-area resources to deposit data where these are available. Where they are not available, funders may encourage researchers to utilise generalist repositories (such as Dryad, Zenodo, FigShare, DataVerse, Mendeley Data or the Open Science Framework) or an institutional repository.
31. The development and long-term sustainability of biodata resources to meet the needs of the research community is therefore vital to the successful implementation of funder open data policies. In addition to enabling the long-term storage and preservation of the data, they also play a key role in assisting researchers to ensure their data and metadata meet community standards and help to reduce the burden of compliance with open data mandates on individual researchers and their institutions. As both funders and journals increasingly develop more stringent policies, this creates new demands on biodata resources and increases the need for funders to provide stable, long-term support.
32. Many funders, including those represented on the Working Group, actively support and fund biodata resources—including deposition databases and knowledge bases. However, decisions over which resources are prioritised for support may not always be informed or driven by a detailed knowledge of their relative importance to that funder's researchers to fulfil its policy requirements. And, in many cases, supported researchers are relying on resources supported by other funders, sometimes in other parts of the world, to which they are not themselves contributing funding.
33. The Working Group recognised that the legal mandate of some funders does not extend to direct policy or financial support of data infrastructure in their jurisdiction. For these funders, there may be opportunities for them to coordinate and work with other bodies at national and international levels to help ensure the resources on which their funded researchers depend are adequately supported and equipped to implement these funders' open data policies.
34. The scope for funders to develop approaches to better coordinate internationally to more equitably and sustainably fund globally important biodata resources is addressed in the Consultation Paper of the GBC Working Group on Sustainability. These issues were not considered in detail by the Working Group on Open Data Strategies. The Working Group did however discuss some key related issues relating to the implementation of open data policies, in particular:
  - The need to actively consider how best to identify and address gaps in resource provision for particular data types or research fields which come to light as a result of implementing open data requirements, including decisions over when it is appropriate to work with the community to proactively establish a new resource.
  - The need to consider risks of creating data silos and fragmenting the existing landscape when funding major data-intensive projects which might propose creating bespoke project databases and portals (as highlighted above). For such projects, DMPs should be carefully assessed and data should be housed in established and stable biodata resources where possible.

- The scope to utilise DMPs to develop a fuller picture of current and likely future demand amongst researchers for specific data resources — this could both inform decisions over which resources to prioritise for funding and aid resources in planning for future needs.
35. The Working Group also considered the additional demands that had been placed on biodata resources as a result of the uptake of key open data standards, such as the FAIR principles and TRUST principles for digital repositories (See Wilkinson et al 2016, Lin et al. 2020). Recent years have also seen the related roll out of new models for accreditation, including CoreTrustSeal. Given the key role of resources in implementing open data policies in line with these emerging standards and best practice principles, there is a need for funders to ensure that they are adequately resourced. The NIH has run a funding initiative since 2021 to “support existing data repositories to align with FAIR and TRUST principles and evaluate usage, utility, and impact” (see US National Institutes of Health 2023), alongside its schemes to support data repositories and knowledge bases. This provides one model for how funders could support resources to fully align with these emerging standards .
  36. Whilst funders actively encourage and sometimes require researchers to deposit data in recognised community repositories, they do not normally mandate which repositories should be used. For some data types and research fields, the use of particular repositories has become standard practice and may be required as a condition of publication. In other cases, however, researchers may not be clear which resources are available to them and what their funder’s preference would be. The Working Group therefore considered the role that funders could play in helping researchers identify repositories that meet their open data requirements.
  37. The funders represented on the Working Group currently adopt a range of different approaches to help “signpost” in this way. For example:
    - The NIH and SNSF set out criteria for the desired features that repositories should have. Accreditation schemes, such as CoreTrustSeal, also help to fulfil this role but awareness of these schemes is as yet limited in the research community as a whole.
    - Several funders provide lists of resources that meet their criteria for particular data types, many also refer to curated catalogues such as FAIRSharing and Re3Data.
    - The UK MRC has supported development of an ‘Innovation Gateway’ by Health Data Research UK which signposts repositories and tools.
  38. The options proposed by the Working Group for funders to cooperate in supporting biodata resources to help deliver funder open data policies are summarised in **Box 2** overleaf. The Working Group felt the GBC was well placed to lead and coordinate this work, with ongoing work to select Global Core Biodata Resources and develop the Global Inventory of biodata resources constituting key activities on which to build.

**Box 2—Working with biodata resources to advance policy implementation**

It is proposed that funders consider the following options for cross-funder cooperation:

- a. Align support for biodata resources to the requirements of open data policies**—ensuring their role in implementing open data policies is adequately supported (either directly, or through working with other organisations where direct funding or policy support for data infrastructure is outside the funder’s legal mandate). This might include:
  - i. assessing current and future demand for biodata resources in light of open data policies;
  - ii. considering carefully the existing global resource landscape when making decisions over creating new resources or supporting major bespoke project databases;
  - iii. targeting support for biodata resources of importance to researchers to fully implement relevant community standards including the FAIR principles.
- b. Explore the potential to develop shared criteria and resources to assist researchers in finding and selecting suitable repositories**, including:
  - i. specifically recommending Global Core Biodata Resource (GCBRs);
  - ii. bringing together existing guidance and resources to create a common set of criteria that could be used across GBC funders;
  - iii. exploring the scope to develop a ‘Yellow Pages’ or “wizard” style resource to signpost repositories, potentially through leveraging an existing tool or service and drawing on GBC’s Global Inventory as a key source.
- c. Develop common cross-funder principles on data preservation, sharing and retention across the data lifecycle** to set clear expectations for biodata resources and researchers, and common metrics for tracking usage, utility and impact across biodata resources.

## 2.4 Embedding good practice: training, incentives and culture change

39. The successful implementation of open data policies requires that the research community is adequately skilled and incentivised to manage and share data. In short:
  - There is a need to ensure that researchers across the board have the skills necessary to manage their data in line with community best practices, and that key specialist roles are nurtured, including data managers and data stewards based at research institutions (sometimes in research libraries or sometimes embedded in research teams) and biodata resources who serve as key members of research teams and provide essential support to the wider research community.
  - There is a need to recognise the time and effort that is required to manage and share high quality data that can be accessed and used by others, and ensure researchers and data specialists are appropriately supported, recognised and rewarded.
40. The Working Group considered the potential for funders to adopt a greater leadership role in training researchers and data professionals. It was emphasised that the training of researchers and the development of associated curricula is primarily the role of research institutions. Nevertheless, funders could potentially make a positive contribution to advocating for the inclusion of data management skills in researcher training at all levels. They could also highlight and proactively support research professionals to take up training modules and providers, such



as the Data and Software Carpentry programmes or the RDA-CODATA Summer Schools in Data Science (see Bezuidenhout et al. 2019, Teal et al. 2015).

41. Several funders on the Working Group had already taken proactive steps to recognise the key roles of data managers and data stewards as members of research teams, and provision support within research grants. The Working Group recognised however that there was a continuing need to support, recognise and develop career paths for these skills.
42. The challenges around incentives and recognition for data management and sharing have been explored in depth in other reports, including for example the work of the US National Academies' roundtable on aligning incentives for open science, and the European Commission's Open Science Policy Platform (see National Academies 2022).
43. The Working Group emphasised the continued need for funders to build on this work and actively demonstrate the value they place on efforts to advance open research data. Although culture change is necessarily a slow process, there are encouraging recent developments on which to build, such as:
  - the gathering momentum and support for international efforts to reform research assessment systems to recognise a full and diverse range of research outputs and contributions
  - The rapid development and increasing adoption of persistent identifier systems such as ORCID and DataCite
  - The increasing number of exemplars of the benefits and impacts of open data and open research activities, including in the context of the COVID pandemic.
44. The options identified by the Working Group in relation to training, incentives and recognition are outlined in **Box 3** below.

**Box 3—Embedding good practice: training, incentives and culture change**

It is proposed that funders should work with research institutions to:

- a. **Explore opportunities to collaborate in promoting and funding training in data skills** for research professionals, including building a resource of available opportunities.
- b. **Recognise the key roles of specialist data managers, data stewards and data scientists in research institutions and biodata resources**—potentially through developing a common statement or principles to recognise and support these roles.
- c. **Accelerate collaborative efforts to incentivise and reward open data practices—which may include:**
  - i. identifying good practice in embedding the principles of the San Francisco Declaration on Research Assessment (DORA) and Coalition for Advancing Research Assessment (CoARA) to recognise a full and diverse range of research outputs and contributions, in particular data sets, contributed curation and data resources.
  - ii. agreeing common approaches to implement persistent identifier schemes and emerging schemes to attribute contributions, and encouraging GCBRs and other resources to actively implement attribution and contribution tracking.
  - iii. exploring the potential to develop funding tools—including prizes and targeted funding calls — to support innovation in open data practices and celebrate researchers and data managers who have pioneered open approaches

## 2.5 Ensuring global equity in implementation of open data policies

45. Throughout its discussions, the Working Group emphasised the key need for funders to actively consider the implications of open data policies for all stakeholders globally, and ensure that they are implemented in an equitable manner for global benefit.
46. Particular consideration in this regard needs to be given to the differing challenges facing researchers in the low and middle-income countries, where many may not have access to the same level of funding, technologies or institutional support available to researchers in better resourced settings. Whilst open data has the potential to benefit all and empower researchers in low and middle income settings and enable researchers all over the world to collaborate to address shared global challenges, there is also a risk it could exacerbate existing inequities—particularly if researchers are forced to make research data rapidly available, without access to the tools or skills to utilise these data fully.
47. There may also be particular ethical or cultural issues impacting on how data may be managed and shared—these include, for example, heightened risks of harm or stigma to individuals or communities and concerns regarding sovereignty and benefit sharing. These issues are also being actively addressed by funders around the world in the context of research involving indigenous communities—with the development of the CARE principles providing one framework for addressing the challenges involved.
48. These challenges might necessitate tailored approaches to implementation that respect and address these concerns. They also require funders to adopt a collaborative and inclusive approach which actively engages stakeholders and ensures their concerns are recognised and addressed adequately, as summarised in **Box 4** below.

### **Box 4—Promoting global equity in open data policies**

It is proposed that funders consider the following options for cross-funder cooperation:

- a. **Adopt a common commitment to develop open data policies in an equitable, ethical and efficient manner** — GBC could serve to promote cross-funder dialogue, sharing of good practice and engagement with global stakeholders to ensure the benefits of open data are realised equitably, and in line with the CARE principles.
- b. **Monitor the implementation of open data policies to ensure global equity** — develop collaborative mechanisms to monitor the implementation of funder policies in line with the CARE principles and related frameworks, including access and usage of data by as yet under-served communities.

## Annexes

### Annex A: Membership of the GBC Board Working Group on Open Data Strategies

Name	Affiliation
Warwick Anderson (Chair)	Chair of GBC Board of Funders
Michael Ball	UK Research and Innovation (UKRI) - MRC
Ishwar Chandramousliwaran	US National Institutes of Health (NIH)
Christine Choirat	Swiss National Science Foundation (SNSF)
Michael Dunn	Wellcome
Davina Gherzi	National Health and Medical Research Council (NHMRC), Australia
Chris McMaster	Canadian Institutes of Health Research (CIHR)
Kostas Repanas	European Commission
Jennifer Weller	US National Science Foundation (NSF)

#### **GBC Secretariat**

Guy Cochrane  
Chuck Cook  
Rachel Drysdale  
David Carr

## References

- Anger M, Wendelborn C, Winkler EC, Schickhardt C (2022). Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies - a qualitative expert interview study. *Plos one*. 2022 ;17(9):e0273259. <https://doi.org/10.1371/journal.pone.0273259>
- Barker M, Wilkinson R and Treloar A (2019). The Australian Research Data Commons. *Data Science Journal*, 18(1), p.44. <http://doi.org/10.5334/dsj-2019-044>
- Bezuidenhout L, Drummond-Curtis S, Walker, B, et al. (2021). A School and a Network: CODATA-RDA Data Science Summer Schools Alumni Survey. *Data Science Journal*, 20: 10, pp. 1–12. <https://doi.org/10.5334/dsj-2021-010>
- Burgelman JC, Pascu C, Szkuta K, et al. (2019) Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*. 2019 ;2:43. <https://doi.org/10.3389/fdata.2019.00043>
- Carroll SR, Garba I, Figueroa-Rodríguez OL, et al. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), p.43. DOI: <http://doi.org/10.5334/dsj-2020-043>
- Cole NL, Reichmann S, Ross-Hellauer T (2023). Toward equitable open research: stakeholder co-created recommendations for research institutions, funders and researchers. *R. Soc. Open Sci.*10: 221460. <https://doi.org/10.1098/rsos.221460>
- Davidson J, Engelhardt C, Proudman V, Stoy L, & Whyte A (2019). D3.1 FAIR Policy Landscape Analysis (v1.0). FAIRsFAIR. <https://doi.org/10.5281/zenodo.5537032>
- European Commission (2020). The Economic Impact of Open Data: Opportunities for value creation in Europe. <https://data.europa.eu/en/node/7773> (accessed on 9 March 2023)
- Gaba JF, Siebert M, Dupuy A, Moher D, Naudet F (2020). Funders’ data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PLoS ONE* 15(8): e0237464. <https://doi.org/10.1371/journal.pone.0237464>
- Gardner L. (2022). The COVID-19 Dashboard for Real-time Tracking of the Pandemic: The Lasker-Bloomberg Public Service Award. *JAMA*. 2022 Oct;328(13):1295-1296. <https://doi.org/10.1001/jama.2022.15590>
- Global Biodata Coalition (2022). Global Core Biodata Resources: Concept and Selection Process. Zenodo. <https://doi.org/10.5281/zenodo.5845115>
- Gomes DGE, Pottier P, Crystal-Ornelas R, et al. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings. Biological Sciences*. 2022. <https://doi.org/10.1098/rspb.2022.1113>
- Harrison PW, Lopez R, Rahman N, et al. (2021). The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. *Nucleic Acids Research*. 2021 Jul;49(W1):W619-W623. <https://doi.org/10.1093/nar/gkab417>

Hrynaskiewicz, I, Simons, N, Hussain, A et al. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19(1), p.5. DOI: <http://doi.org/10.5334/dsj-2020-005>

Hughes LD, Tsueng G, DiGiovanna J, et al. Addressing barriers in FAIR data practices for biomedical data. *Scientific Data*. 2023 Feb;10(1):98. <https://doi.org/10.1038/s41597-023-01969-8>

Imker, HJ, Schackart, KE, Istrate, A and Cook, CE (2023). A Machine Learning-Enabled Open Biodata Resource Inventory from the Scientific Literature. Zenodo. <https://doi.org/10.5281/zenodo.7768416> (preprint)

Lin D, Crabtree J, Dillo I, et al. The TRUST Principles for digital repositories. *Scientific Data*. 2020 May;7(1):144. <https://doi.org/10.1038/s41597-020-0486-7>

National Academies of Sciences, Engineering, and Medicine (2022). Open Scholarship Priorities and Next Steps: Proceedings of a Workshop in Brief. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26557>

Neylon C (2017). Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes* 3: e14673. <https://doi.org/10.3897/rio.3.e14673>

OECD (2021). Recommendation of the Council concerning Access to Research Data from Public Funding, OECD/LEGAL/0347. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347> (accessed 9 March 2023)

OSTP (2022). Ensuring Free, Immediate, and Equitable Access to Federally Funded Research (memorandum for the heads of executive departments and agencies from Dr. Alondra Nelson) <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

Smith I and Veldsman S (2018). Data Driving Sustainability—the African Open Science Platform Project. *ELPUB* 2018, Jun 2018, Toronto, Canada. <https://doi.org/10.4000/proceedings.elpub.2018.25>.

Teal, TK, et al (2015). Data Carpentry: Workshops to Increase Data Literacy for Researchers. *International Journal of Digital Curation*, 10(1): 135–143 <https://doi.org/10.2218/ijdc.v10i1.351>

UNESCO (2021). UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949> (accessed on 9 April 2023)

US National Institutes of Health (2023). Support for existing data repositories to align with FAIR and TRUST principles and evaluate usage, utility, and impact (NOT-OD-23-044) <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-044.html> (accessed 9 March 2023)

Van den Eynden V and Bishop L (2014). Incentives and motivations for sharing research data, a researcher's perspective. [https://repository.jisc.ac.uk/5662/1/KE\\_report-incentives-for-sharing-researchdata.pdf](https://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf)

van Rossum J (2021). Analysis of international funder data policies (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5643352>

Wilkinson M, Dumontier M, Aalbersberg I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

### **Funder Policies referenced in this Consultation Paper**

US National Institutes of Health - Data Management and Sharing Policy (2023)  
<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>

MRC - Data Sharing Policy (2016)  
<https://www.ukri.org/publications/mrc-data-sharing-policy/>

NHMRC - Open Access Policy (2022)  
<https://www.nhmrc.gov.au/about-us/resources/nhmrc-open-access-policy>

US National Science Foundation - Dissemination and Sharing of Research Results  
<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>  
[https://www.nsf.gov/pubs/policydocs/pappg22\\_1/pappg\\_11.jsp#XID4](https://www.nsf.gov/pubs/policydocs/pappg22_1/pappg_11.jsp#XID4)

Swiss National Science Foundation - Policy on Open Research Data (2017)  
<https://snf.ch/en/dMILj9t4LNk8NwyR/topic/open-research-data>

Wellcome - Data, software and materials management and sharing policy (2017)  
<https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy>