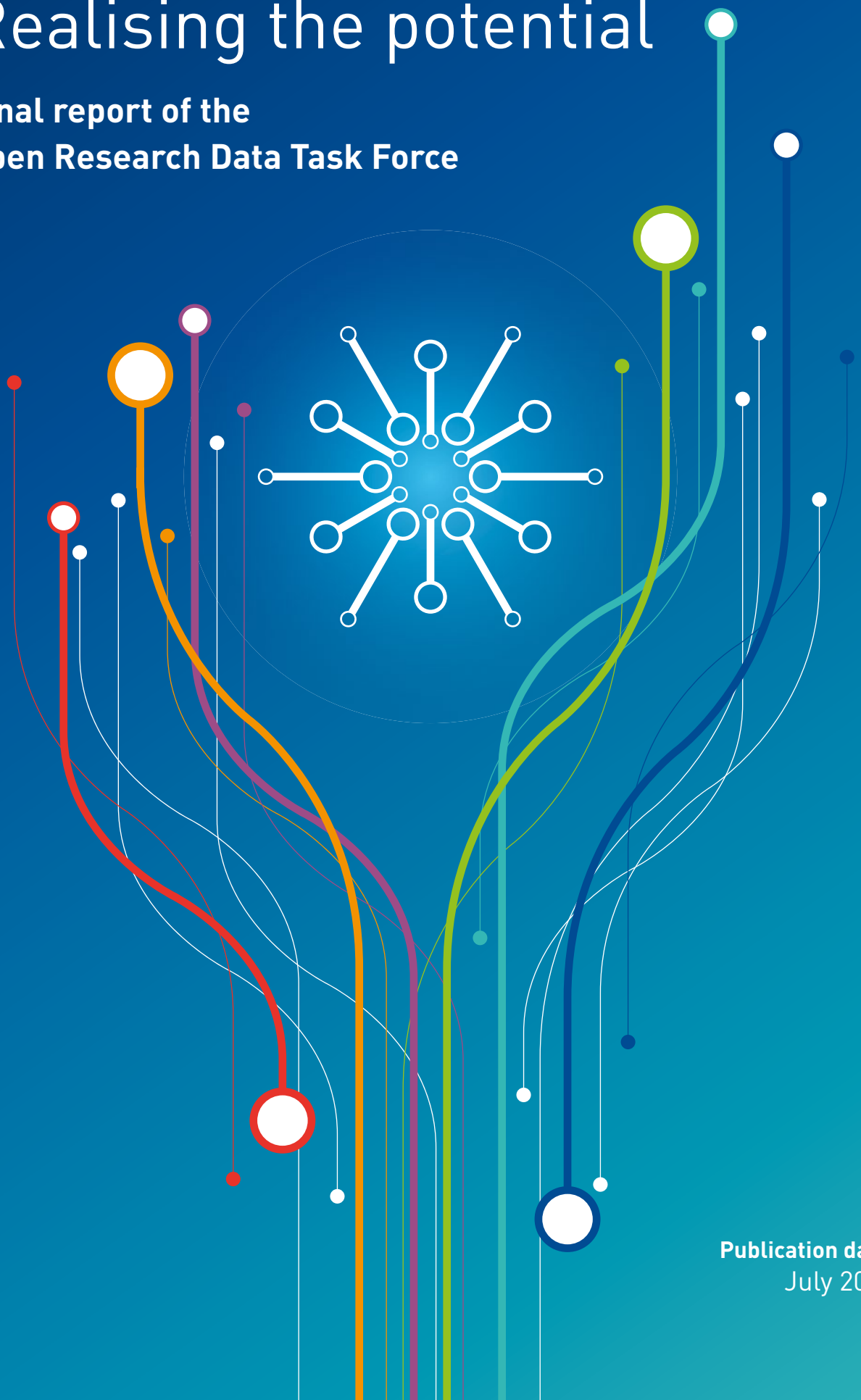
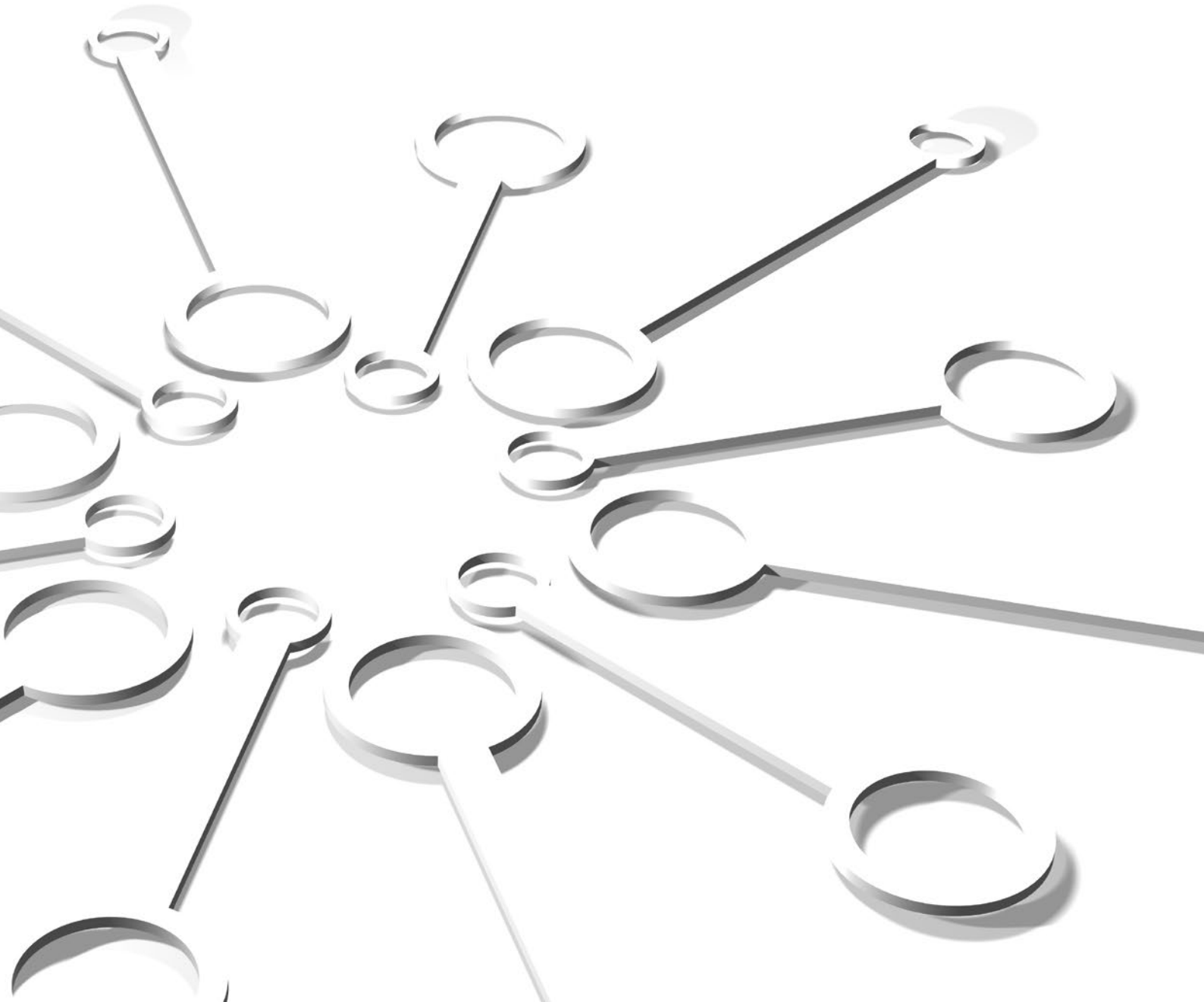
Realising the potential

Final report of the Open Research Data Task Force



Publication date:
July 2018



Executive summary

Introduction

Researchers are creating, gathering and using data in hitherto-unimagined volumes. These vast data resources dramatically increase the capacity of science to infer patterns in phenomena, whether physical, chemical, biological or human, or in the complex systems that are at the heart of most global challenges.

It is in this context that a new paradigm of *"open science"* has developed, that is more efficient, open to all, integrated across disciplines and societally engaged. Its necessary bedrock is:

- that published scientific results should be open access - digital, online, free of charge, and free of most copyright and licensing restrictions; and
- that the data acquired by individual scientists and scientific groups should be subject to a default position whereby it is made findable, accessible, interoperable and re-useable (FAIR);

The second of these aims forms the focus of this report and offers profound opportunities. Open research data (ORD) have the potential not only to deliver greater efficiencies in research, but to improve its rigour and reproducibility, to enhance its impact, and to increase public trust in its results.

The government has signaled its intention to position the UK at the forefront of the data revolution, and barriers to the effective use of data are falling steadily with increases in processing power and the wider use of artificial intelligence (AI). Yet national and international science systems have not fully adapted to the new opportunities presented. There is also increasing awareness of the potential for data access and use to impinge on personal privacy, social ethics, safety, security or intellectual property in novel and sometimes unpredictable ways. Meanwhile, varying and incompatible standards, inadequate services and infrastructure, and pressures on funding continue to inhibit the integration and use of open research data. Addressing these challenges and tensions will be critical to embedding ORD as an integral part of research practice and unlocking the benefits this entails.

The Open Research Data Task Force was formed to:

a | establish realistic ambitions for UK open research data (ORD) and the steps required to achieve them;

b | assess the current state of play across higher education institutions (HEIs), research organisations and disciplinary groups, and the international landscape;

c | identify options for building appropriate capacity and capability, and for enhancing the existing infrastructure; and

d | consider potential costs and benefits to government and the sector.

The Task Force and its remit

The Task Force has sought to build on the principles set out in *the Concordat on Open Research Data*, and to take account of wider moves towards ORD within the international landscape.

Our overarching aims in formulating our recommendations have been:

- **to enhance the quality and value of research by providing the incentives, the services, and the support researchers need as both creators and users of data;**
- **to facilitate the wider scrutiny and re-use of data;**
- **to increase public engagement and trust in research; and**
- **to promote the use of data to facilitate innovation and to enhance the benefit it brings to citizens in the UK and beyond.**

Moves towards open research data (ORD) are essential to achieving those aims, but progress to date has been at best uneven. It is therefore timely for UK government to consider whether and how it might intervene to enhance the capacity of the powerful UK science base in effectively exploiting the potential of ORD. Our report examines whether there are processes of incentivisation, coordination, policy development, infrastructure provision or stimulation in key foundational areas where intervention would benefit both the UK science system and society. Our findings are largely directed at research funders, research organisations and researchers themselves, but their efforts are most likely to bear fruit if underpinned by appropriate support and facilitation from government.

Better incentives, fewer barriers

Making ORD a norm requires change on the part of many researchers. At present, however, they must contend with fragmented policies and services, and for many the benefits of open research data remain unclear.

Adoption of ORD practices is highly variable across and within disciplines, and only rarely attracts tangible rewards. Improvements in data citation practice, the growth of data journals and endorsement of the **San Francisco Declaration on Research Assessment (DORA)** represent promising moves to improve matters. Rewards for ORD in future Research Excellence Frameworks and ensuring that ORD practice is assessed in applications for research funding will also help to stimulate change. Nevertheless, we recognise that cultural change will take time.

Alongside enhanced incentives, coordinated efforts are also needed to improve researchers' skills in handling and analysing data, to deliver increased capacity in data science, and to provide specialist support services. This will require new initiatives to recognise and reward data skills, develop training programmes and peer support networks, and implement sustainable career paths for growing numbers of specialist staff.

We therefore recommend that:

- **All stakeholders** act to strengthen the incentives for researchers to create and use open research data
- **Funders and research organisations** establish training programmes for researchers in data management, analysis and stewardship
- **Research organisations** strengthen the provision of specialist support services within research organisations, and increase capacity in data stewardship, research software and data science.

Active leadership

The range of organisations engaged in supporting the development of ORD is both broad and complex.

ORD services developed by the Research Councils and other funders operate alongside those from Jisc, individual research organisations, and a variety of UK and international providers. Learned societies and publishers also play important roles in developing services and establishing disciplinary norms, particularly in those communities showing high levels of engagement with ORD.

The fragmentation and variability that results is not unique to the UK, with similar challenges evident in other countries including Germany and the USA. Nevertheless, it is clear that some countries are moving more rapidly than the UK to establish national infrastructures and leadership arrangements for ORD. The UK should take note of and learn from these developments, while further strengthening its engagement with key international initiatives such as the European Open Science Cloud and the Research Data Alliance.

A careful balance must be struck between top down and bottom-up initiatives, and no individual organisation can or should take overall responsibility for the development of ORD policies and services in the UK. However, the current lack of coordination and consensus puts the UK's hitherto-leading role at risk. UK Research and Innovation has only recently come into being, and will need time to develop the stronger coordinating mechanisms that are now required, both domestically and internationally. However, it is uniquely placed to bring the key actors and stakeholders together to this end.

We therefore recommend that:

- **UK Research and Innovation** takes a co-ordinating role in overseeing the development of ORD policies, infrastructure and services, including:
  - Shared understandings of roles and responsibilities
  - Distribution of resources
  - Engagement with international and non-academic stakeholders
  - Development of guidelines and protocols on ORD
  - Establishing the evidence base and monitoring trends and progress towards ORD.

Clear expectations

ORD policy in the UK is framed by the principles set out in the Concordat on Open Research Data, but awareness of the Concordat remains low, and wide variations in policies and expectations of researchers persist.

The Concordat itself does not reflect the FAIR principles (Findable, Accessible, Interoperable, Re-usable), which have gained widespread support in recent years.

A single set of policy requirements for all researchers is neither feasible nor desirable. Nevertheless, embedding ORD across the whole UK research community will require increased consistency and clarity in the expectations placed on researchers. Adoption of the 'as open as possible, as closed as necessary' principle and a commitment to harmonisation across UKRI's nine councils must be accompanied by clear criteria surrounding necessary restrictions on access. Policy development efforts must both stimulate and reflect changes in practice and possibilities in different fields. There is scope to clarify requirements around data management plans (DMPs), strengthen mechanisms for quality assurance of data, and promote greater consistency in formats, metadata and licensing arrangements.

In the publishing community, efforts are already underway to define, through the **Research Data Alliance** and other initiatives, a common set of policy requirements for the data associated with journal articles. Further work is needed to promote widespread uptake of these and other requirements across the full range of publishers and learned societies with which UK authors publish.

Meanwhile, the contribution of ORD to enhancing the quality and value of research, and increasing public trust in its results, should be acknowledged via strengthened links with other research policy initiatives, including those for research integrity, knowledge exchange and impact.

We therefore recommend that:

- **Funders, research organisations and publishers** establish clear expectations on preservation of data and software, including the repositories to be used
- **UKRI** leads an effort to promote greater harmonisation of funder ORD policies, based on the FAIR principles and with a view to research data being 'as open as possible, as closed as necessary'
- **Publishers and learned societies** take steps to require data access statements, support open citation of data and facilitate data deposit and linking to other outputs.

User-friendly services

It remains difficult for many researchers to find appropriate services to handle the data they create, and to locate data they might wish to use.

Specialist services are well-established in a small number of fields, such as biosciences and astronomy, but more than half of UK researchers rely on their own resources to manage and preserve the data they create. Many remain unaware of relevant institutional and other support services.

The development of comprehensive, domain-specific services is a long-term endeavour that will require extensive international cooperation. The UK must continue to play an active part in these developments, but also ensure that well-resourced, generic services are made available to researchers in the meantime. There is a need for new guidance and exemplars to ensure that data meets appropriate quality standards; for tools to standardise and automate data management, documentation and curation processes; and for an increased focus on improving research software, and on recruiting and retaining software engineers. These efforts should be guided by the twin aims of ensuring that data meets the FAIR principles, and that it is effectively preserved in trusted, certified repositories.

There are growing concerns that ORD infrastructure and systems may fall under the control of providers whose long-term interests may not coincide with those of the wider research community. Work to improve transparency and openness in research data infrastructure must therefore be accompanied by efforts to promote a competitive marketplace for commercial ORD services alongside publicly-funded provision. Research organisations need greater support and guidance in negotiating with commercial providers, to ensure agreements are consonant with the FAIR principles and the aims of open science.

The UK is an active participant, and often a leading player, in a wide range of international ORD partnerships and infrastructures. This engagement must be sustained and, where possible, strengthened across the full spectrum of scientific disciplines. The UK should continue to work in close partnership with organisations such as the **Research Data Alliance** and **CODATA** to promote international alignment of approaches to ORD.

We therefore recommend that:

- **Research organisations and funders** take steps to ensure that all researchers have access to user-friendly services, both generic and domain-specific
- **Research organisations and funders** develop, with support from Jisc, a set of principles for negotiation with commercial providers of ORD infrastructure to maximise interoperability, retain data ownership and reduce the risk of 'lock-in'
- **UKRI** ensures the development of sustainable ORD infrastructure features prominently in its long-term research and innovation infrastructure roadmap, with relevant needs to be identified in close consultation with learned societies and subject communities
- **Research organisations and funders** take active steps to sustain and strengthen UK participation in international ORD services and initiatives.

Sustainable funding

ORD places additional demands on researchers and research organisations which are at present only rarely matched by new revenue streams.

Inevitable tensions arise in determining the balance of costs to be met from short-term project funds, institutional block grant funding, and long-term infrastructure funding. The sustainability of ORD services, and the development of underpinning infrastructures, are often overlooked, particularly in disciplines where large-scale project funding is rare. Evidence also suggests that researchers remain unwilling, or feel unable, to request funds for data management and sharing within grant applications. Meanwhile the infrastructure and services that research organisations themselves provide may fail to meet the full range of researcher requirements.

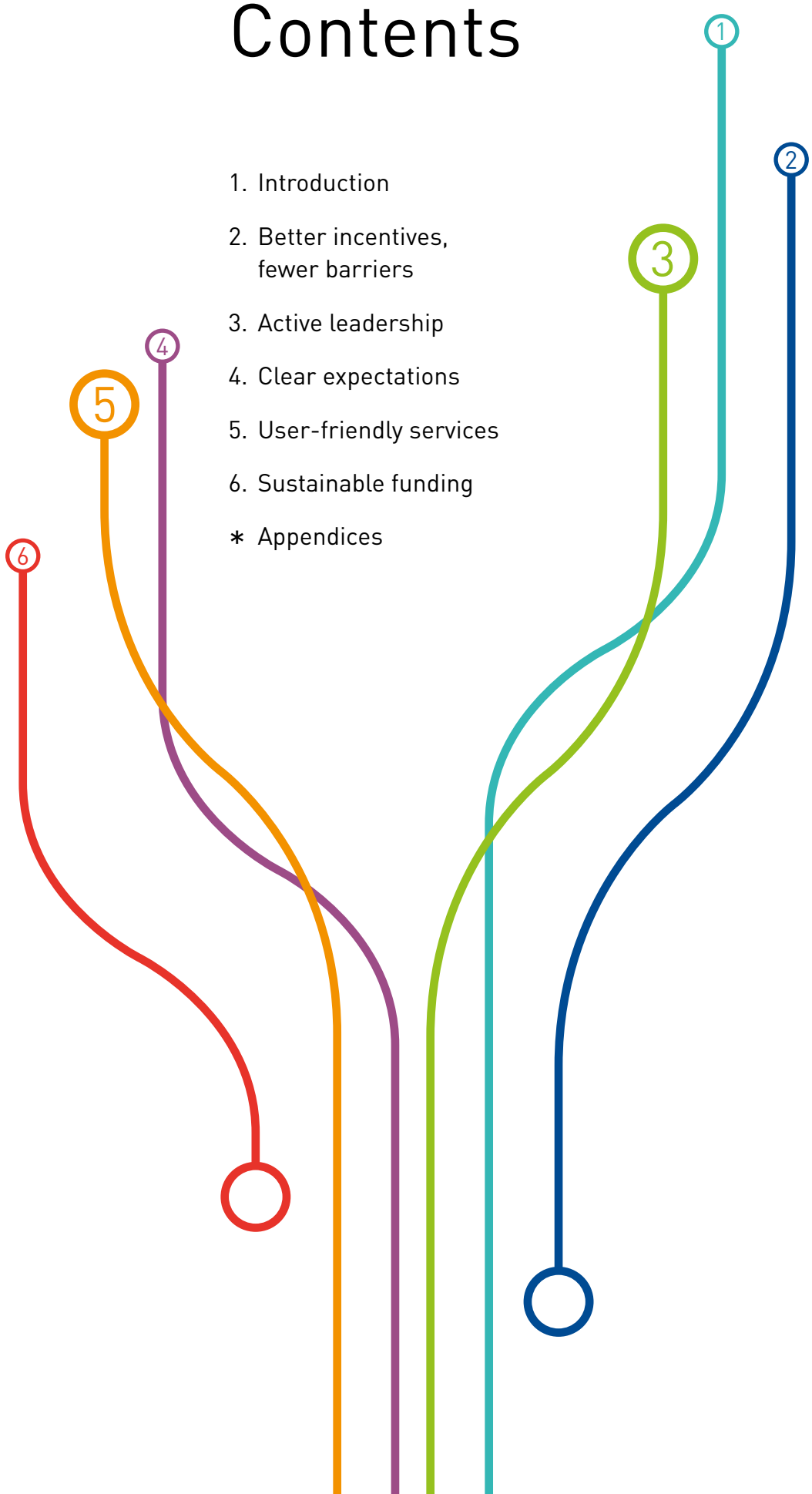
Addressing these challenges requires a greatly improved understanding of the costs, business and funding models for ORD throughout the research lifecycle. Efforts to date have tended to focus on the costs of preservation and archiving, with insufficient attention paid to 'pre-archive costs'. These include both the additional researcher effort needed to manage and share data, and the capital and revenue costs of ORD services and infrastructure. Allied to this, there is a need to develop an inventory of relevant infrastructures, and undertake a regular landscape analysis to identify strengths and gaps in provision. New funding streams may be required to ensure the medium to long-term sustainability of national research data infrastructure.

The costs involved can be tensioned against an increasing awareness of the productivity gains arising from ORD, and the potential for wider use of research data by industry and other non-academic users. Further, targeted work is needed to understand the opportunities for re-use of research data within specific industry sectors, and the potential for new mechanisms, such as 'data trusts' and dedicated funding schemes, to facilitate this.

We therefore recommend that:

- **UKRI and other stakeholders** work together to review the costs, business and funding models of current data services
- **All funders** require existing, funded data services to develop appropriate plans for ORD
- **UKRI, funders and research organisations** review levels of funding for ORD to ensure these remain appropriate to an increasingly data-rich research landscape
- **UKRI and other funders** consider the provision of funds to fill priority gaps, support data publishing, and make existing data readily usable in support of national research and innovation priorities.

Contents



1. Introduction

1.1 The Task Force and its remit

The Open Research Data Task Force was established by the then Minister of State for Universities and Science, Jo Johnson MP, in Autumn 2016.

It followed a **recommendation** from Professor Adam Tickell that a 'roadmap' for national open research data infrastructure be produced, with a set of recommendations. The Task Force was set up to lead and co-ordinate this work. The members comprised experts in the field, and representatives of research funders, Universities UK and Jisc (which also provided the secretariat). The Department of Business, Energy and Industrial Strategy was an observer.

The terms of reference for the Task Force covered:

a | establishing realistic ambitions for UK open research data (ORD) and the steps required to achieve them.

b | assessing the current state of play across HEIs, research organisations and disciplinary groups, and the international landscape.

c | identifying options for building appropriate capacity and capability, and for enhancing the existing infrastructure.

d | considering potential costs and benefits to government and the sector.

"Publicly funded scientists have a responsibility to contribute to the public good through the creation and communication of new knowledge, of which associated data are intrinsic parts. They should make such data openly available to others as soon as possible after their production in ways that permit them to be re-used and re-purposed."

International Accord on Open Data

1.2 Background and context

The work of the Task Force is set in the context of wider moves to make both the practice and the results of research more open; and support for greater public engagement with research, including citizen science.

In this context, there have been declarations from a number of leading bodies both in the UK and across the globe about the responsibilities of researchers as creators of new knowledge and understanding for public good.

It is crucial that the UK should play an active and leading role in international efforts to that end.

Embedding ORD as an integral part of good research practice would bring a wide range of benefits, including:

- increasing the efficiency of research, exploiting digital technologies and advanced methodologies to the full;
- improving the rigour, validity and reproducibility of research by ensuring that truth claims are accompanied by the evidence on which they are based, and making that evidence more readily accessible for scrutiny and interrogation;
- promoting more effective communication and collaboration between researchers, especially in interdisciplinary research and in addressing 'grand challenges';
- accelerating and increasing the impact of research both within the research community and beyond;
- making the processes and results of research accessible to wider audiences, including people and organisations in the public, commercial and voluntary sectors; and
- democratising research, by securing greater public engagement in the research process, and enhancing public trust in the results of research.

Evidence suggests that the benefits in terms of the cost-efficiency of research may **approach 50%**; and the use of research data from a single service such as the **European Bioinformatics Institute** contributed to research impacts conservatively estimated at £920m a year.

In the UK as in other countries, encouragement for ORD is closely related to efforts to promote access and use of some of the 2.5bn gigabytes of data created every day, especially the data generated by

the public sector. The **UK Government has accepted** the proposition that data, and the services that surround it, are essential parts of the national infrastructure. Open data has the potential to bring widespread benefits to the UK economy and society, contributing to the target of increasing investment in research and development to **2.4 per cent of gross domestic product** (GDP). The Government is thus seeking to improve data-related capabilities and capacity within Government, and to promote similar developments across society and the economy as a whole, making use of bodies including the **Open Data Institute**, the **Alan Turing Institute**, and the forthcoming Centre for Data Ethics and Innovation.

1.3 The Concordat on Open Research Data

For the academic community in particular, the **Concordat on Open Research Data**, published in July 2016 and endorsed by key funding agencies and Universities UK, set out ten key principles relating to ORD, covering:

1. ORD as an enabler of high-quality research and a facilitator for innovation
2. Legitimate restrictions on openness
3. The costs associated with ORD
4. The right of data creators to reasonable first use
5. Legal, ethical and regulatory frameworks
6. The importance of good data management practice
7. Data curation and long-term preservation
8. Access to data supporting publications
9. Developing capabilities and capacity in data skills
10. Regular reviews of progress towards ORD as a norm.

The intention of the Concordat was to establish expectations of good practice, recognising that policy, practice, infrastructure and services vary widely at present. The current report identifies a set of concrete actions to accelerate progress towards the adoption and use of ORD in the UK.

1.4 Definition of terms

We have adopted the definitions set out in the **Concordat on Open Research Data**, which can be briefly summarised as follows:

- Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical).
- Open research data are those research data that can be freely accessed, used, modified, and shared, provided that there is appropriate acknowledgement if required.

The Task Force recognises that not all research data can be open and that access may need to be managed in order to maintain confidentiality, guard against unreasonable cost, protect individuals' privacy, respect consent terms, as well as managing security or other risks. Accordingly, the Task Force endorses the principle that research data should be 'as open as possible, as closed as necessary'.

1.5 Opportunities and challenges for the research community

The continuing digital revolution offers profound opportunities for research.

New tools, and the vast data resources that are increasingly available, offer the potential for unprecedented insights into complex phenomena, dramatically increasing researchers' ability to discover and infer patterns and relationships, whether physical, chemical, biological or human, or in the complex systems that are at the heart of global research challenges.

But digital technologies pose disruptive challenges to established research systems and to the roles, responsibilities and relationships between the key stakeholders. Some of those challenges relate to the potential for greater public access to research and its results, and much wider public engagement with research (including active participation in the form of 'citizen science') as key characteristics of a democratic society. But it has proved difficult to establish the systemic and coordinated responses needed in a complex landscape.

Co-ordination is necessary in order to mitigate the risks associated with fragmentation, in a context where there is a need to develop greater awareness and understanding of:

- the ethical, legal and regulatory issues arising from the creation and use of research data;
- the challenges in improving research integrity and testing the reproducibility of results that involve complex data streams, algorithms and statistical manipulations;
- the need to sustain the core values, principles and processes that underlie high-quality research at the same time as securing greater public engagement;
- the challenge of developing systems that meet the needs and motivations of an increasing variety of research users and producers working within different contexts, disciplines and cultures;
- the varying standards and vocabularies – and variations in data quality – that make it difficult to create the integrated data resources needed for interdisciplinary research into global challenges such as sustainable development; and
- the costs associated with making data freely accessible, understandable and usable.

1.6 The UK in the global ORD ecology

UK institutions and individuals are prominent among those engaged with international initiatives such as the **Research Data Alliance**, **CODATA**, the **World Data System**, and working groups established by the **OECD** and bodies such as **Science Europe**, the **International Science Council** and **Science International**.

The importance of international co-operation in fostering moves towards open research in general and ORD in particular has been recognised by the OECD and in successive meetings of **G7 Science Ministers**. The G7 has established an **Expert Group** on Open Science which is considering ways to meet the infrastructure requirements in different areas of research, and to incentivise higher levels of adoption.

The European Union and its member states have also been active in promoting open research and ORD through a range of projects and initiatives. These include **OpenAIRE** and **EUDAT**; infrastructure and service developments such as the **European Bioinformatics Institute (EMBL-EBI)**, **ELIXIR** and **CLARIN**; setting open data as a default requirement for grants in the **H2020 programme**; and moves to establish a co-ordinating **European Open Science Cloud (EOSC)**.

The European Commission has recently issued a revised **recommendation** on access to preservation of scientific information, and **proposed a new directive** of the re-use of public sector information, including research data. Individual member states such as **Germany**, **France**, the **Netherlands** and **Finland**, as well as other countries such as **Australia** and **Canada**, have been active in a similar vein, seeking to establish co-ordinated approaches to open research, including the development of federated national research data infrastructures.

In the UK, a number of bodies – funding agencies, **Jisc**, individual research organisations, groups of researchers, the **Royal Society**, the **British Library**, and some publishers – have been active in promoting and supporting ORD, with progress stimulated by initiatives going back as far as the **e-science programme** established in 2001. But the challenges are considerable, and despite a number of reports and reviews over the past decade, the principles and aspirations set out in the Concordat are not yet reproduced in harmonised or co-ordinated sets of policies, services and infrastructure in the UK. This reflects in part the scale of the challenge the move to ORD represents, which should not be underestimated. Hence, as in other countries, while adoption of ORD practices is well-established in some disciplines, in others it remains at best a minority interest among researchers, with low levels of awareness and understanding, and varying cultural attitudes.

1.7 The Task Force and its work

The Task Force has met six times and has commissioned consultants to support its work. Full details of the Task Force membership can be found in Appendix 1.

A **review of the landscape** of policies, infrastructure and services to support ORD in the UK as compared with the rest of the world was published in June 2017. Case studies of developments in a range of subject areas and institutions are published as Annexes to this report. The Task Force was supported in its work by Jubb Consulting and **Research Consulting**. Their role was to assemble the evidence base, through a review of relevant literature and interviews with a number of specialist stakeholders in the UK and overseas. The consultants also drew on the experience and expertise of Task Force members, who have themselves consulted members of their respective communities. This report and its recommendations represent the considered views of the Task Force, taking full account of the evidence they have gathered and the views presented to them.

1.8 The report and its structure

The report is built around five key themes, with recommendations associated with each of them:

We make recommendations at the end of each section, and we include in Appendices 2 and 3 a tabulation of the recommendations and of the bodies with responsibility for their implementation. But we recognise that moves to embed ORD practice across the research community as a whole, and still more to promote its use in wider communities, will take time. One of our key messages is that there are no quick fixes: sustained efforts by all key stakeholders will be needed over several years if we are to realise the full potential of ORD.

Our overarching aims are:

- to enhance the quality and value of research by providing the incentives, the services, and the support researchers need as both creators and users of data;
- to facilitate the wider scrutiny and re-use of data;
- to increase public engagement and trust in research; and
- to promote the use of data to facilitate innovation and to enhance the benefit it brings to citizens in the UK and beyond.

The report refers at key points to the principles set out in the Concordat. It aims to provide a digest rather than a full account of the relevant evidence, and readers are directed to the **landscape report** and the case studies for further information, including references. This report focuses on ways to encourage researchers to make the data they create accessible and usable, much more than on ways of stimulating demand from potential research users. A priority for further work should be to develop greater understanding of potential research users' needs, and their capabilities and capacity to make use of ORD.

Further reading

(see also the bibliography in the **landscape report**)

Beagrie, N. and Houghton, J. (2016) The Value and Impact of the European Bioinformatics Institute **https://beagrie.com/static/resource/EBI-impact-report.pdf**

CESSDA SaW (2017) Benefits Factsheet **https://www.cessda.eu/Projects/All-projects/CESSDA-SaW/WP4/Cost-Benefit-Advocacy-Toolkit/Benefits-Factsheet**

European Commission (2018) Recommendation on access to and preservation of Scientific Information **https://ec.europa.eu/digital-single-market/en/news/recommendation-access-and-preservation-scientific-information**

G7 Expert Group on Open Science (2017) Annex 4 to G7 Science Communiqué **http://www.g7italy.it/sites/default/files/documents/ANNEX%204_WG%20Open%20Science_0.pdf**

Hall, W. and Pesenti, J. (2017) Growing the Artificial Intelligence Industry in the UK **https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf**

HM Government (2017) Industrial Strategy: Building a Britain Fit for the Future **https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf**

Knowledge Exchange (2017) Knowledge Exchange Approach Towards Open Scholarship **http://www.knowledge-exchange.info/event/ke-approach-open-scholarship**

OECD (2017) Open Research Agenda Setting, OECD Science, Technology and Innovation Policy Papers No.50 **http://dx.doi.org/10.1787/74edb6a8-en**

ICSU, IAP, ISSC and TWAS (2017) Open Data in a Big Data World: An international accord **https://www.icsu.org/publications/open-data-in-a-big-data-world**

RfII (2016) Enhancing Research Data Management: Performance Through Diversity: Recommendations regarding structures, processes and financing for research data management in Germany **http://www.rfii.de/en/dokuments/**

Royal Society (2012) Science as an Open Enterprise **https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf**

2. Better incentives

Concerted action from all stakeholders is needed to increase incentives and lower barriers if ORD is to become a common feature of the research landscape across the UK, and if already-existing data are themselves to become a key part of the research infrastructure. ORD must not only be made possible through training and support, but made rewarding through proper incentives for researchers and research organisations.

Researchers as both producers and consumers of data must be at the heart of any system to promote and support ORD. They operate within many different cultures and contexts, and as the Concordat stresses, they require high degrees of autonomy for success in their research. Currently they must interact with fragmented sets of policies and services: research group and departmental, institutional, disciplinary and community, national and international. Moreover, there is currently little evidence apparent to researchers as to the benefits they might receive – either in the short or the longer term - through making their data open (Concordat Principle 1).

Therefore there is a need to:

- reduce barriers, and increase incentives for ORD;
- ensure that researchers have the necessary skills, along with helpful technical services and support from specialist staff (Concordat Principle 9);
- encourage the re-use of data through secondary analyses and repeat studies; and
- to promote cultures conducive to the creation and use of open research data.

It is also imperative that further development of both policy and services should be founded on sustained dialogue with different research communities, and on a deep understanding of their practices, cultures and needs.

2.1 Progress towards ORD in different disciplines and subjects

ORD practices, data services, and community initiatives, vary hugely across different subjects and disciplines, and this is reflected in the case studies associated with this report, and in the disciplinary distribution of data repositories both nationally and globally (Figure 1).

In the UK, as in the rest of the world, high levels of adoption, supported both by large-scale services and a wide array of smaller community-led initiatives, tend to be concentrated in the medical and life sciences, along with some other areas, most notably geosciences. Such patterns of adoption are at best sparsely reflected in some other subjects, including many areas of engineering and the humanities. And even in areas with relatively high levels of adoption, researchers' practice in large-scale facilities may differ from that at individual or group level. These differences reflect in turn the significant variations between the cultures and practices – and in the nature of the data created or gathered – in different disciplines and sub-disciplines, and in different research environments.

Fine-grained differences are strongly influenced by:

- variances in the nature of research and of the data produced in different subjects, disciplines or areas of research, and in different research environments;
- constraints relating to data that is sensitive on a wide range of commercial, security or other grounds, or because it relates to identifiable individuals;
- general levels of research funding available and the scale of individual research projects;
- the availability, scale and accessibility of ORD infrastructures and services in different subject domains; and
- the relationships between data specialists and researchers in their specific subject domains.

The influence of these and other factors means that, as one recent **report** put it, 'data sharing practices depend on the field: there is no general approach'. These differences underscore the importance for all policy-makers and service providers of sustaining dialogue with all the different segments of the research community.

Figure 1 Disciplinary provision of research data repositories

(source: re3data.org – CC-BY, last accessed on 02/01/2018)

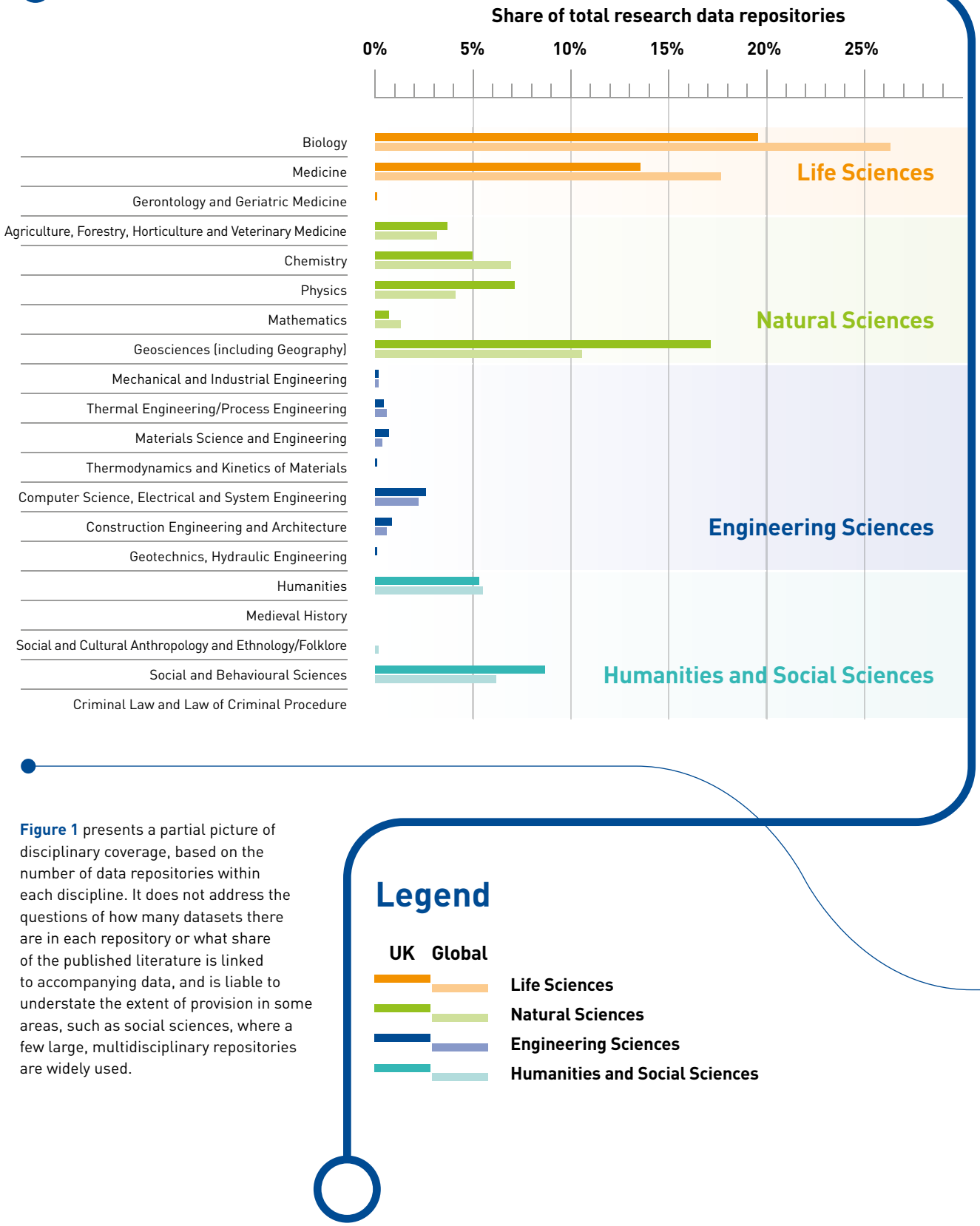


Figure 1 presents a partial picture of disciplinary coverage, based on the number of data repositories within each discipline. It does not address the questions of how many datasets there are in each repository or what share of the published literature is linked to accompanying data, and is liable to understate the extent of provision in some areas, such as social sciences, where a few large, multidisciplinary repositories are widely used.

2.2 Incentives and rewards for researchers

Research has long been a competitive endeavour, in which researchers have sought to enhance their reputations, and secure resultant career rewards, through success in their research.

Performance management regimes have become an increasingly-important feature of the research landscape over the past three decades, influenced in the UK by the requirements of the **Research Excellence Framework** (REF) and its predecessor the Research Assessment Exercise (RAE). In these assessment exercises, in recruitment and promotion processes, and in the reputation economy more generally, publications in scholarly journals and monographs continue to dominate. The RAE and REF rules have always allowed for data as one of the many different forms of output to be submitted for assessment, but such submissions represented less than 0.04% of **submissions to the 2014 REF**. When research performance is being evaluated, achievements in creating ORD are perceived - rightly or wrongly - as counting for little, if at all. In a few fields such as genomics and astronomy (see Annex: Case Studies), sharing of at least some kinds of data has come to be seen as an essential part of the research process, a cultural norm driven by the nature and needs of research. In other cases, researchers may be motivated by an altruistic desire to support the cause of open science; to ensure that their research and their findings are sound, and can stand up to external critique; or to extend their networks and potential for partnerships. But in many disciplines and subject areas, the evidence of benefits to ORD creators or users is at best anecdotal. There are moves, notably through the work of some publishers with **DataCite** and the **Force 11 Joint Declaration of Data Citation Principles**, to enable and encourage the citation of datasets alongside or instead of articles and other publications. The increasing adoption of **ORCID** as a unique identifier makes it easier to link datasets to their creators. The growth of data journals, the current interest in the UK in endorsing the recommendations of the **San Francisco Declaration on Research Assessment** (DORA), and associated moves to celebrate and develop indicators of good practice, may help to stimulate change.
But entrenched perceptions and cultures are unlikely to shift quickly.

Moreover, there is little evidence of researchers suffering negative consequences from failing to meet funders' and institutional requirements on data sharing or ORD. There is, for instance, considerable scepticism in many subject domains as to whether data management plans (DMPs) weigh significantly in the assessment of grant applications; or whether ORD practice – or its absence – is a significant factor in the peer review for scholarly publications.

The adoption and use of ORD were not assessed in REF 2014, while in REF 2021 units will be asked to provide information on 'open research' in the research environment section of their submissions. The funding bodies have also **stated** that panels will take account of 'activity to encourage the effective sharing and management of research data', but further guidance and consultation is awaited on how such activity should be recorded and presented, and how it will be assessed. The primary purpose of the REF is to drive and assess excellence in research, with the outcomes informing the allocation of funding, providing reputational yardsticks and accountability for public investment in research. Nevertheless, its power to influence the behaviours of researchers and research organisations means that it could provide a strong stimulus towards the adoption and use of ORD, especially if linked with further action from other funders. We note in this context that work is currently underway by the **Open Research Funders Group** to develop a blueprint to incentivise open research in funders' policies, assessment processes and communication and engagement activities. Improvements in these areas could in turn help to address concerns about the reproducibility of research results, and to stimulate greater public trust in and engagement with research and its results.

2.3 Barriers to ORD

Weak incentives are accompanied by barriers to more widespread adoption and use of ORD.

Variations in the requirements set by different funding bodies, research organisations and publishers make the policy landscape difficult to comprehend; and individual policies and services are not always well-attuned to disciplinary practices and cultures. Alongside researchers' concerns about misinterpretation or misuse of the data they have created, or about 'giving away' their hard-earned research capital, there can be practical barriers. The time and effort required to make research data open and accessible in accordance with the **FAIR principles** (Findable, Accessible, Interoperable, Re-usable) **can be considerable**; and those researchers who are keen to adopt ORD practices may find themselves stymied by a lack of practical guidance and specialist support.

Calls to remedy deficits in skills and capability in data management and curation have been a repeated refrain in reports for the past decade and more. Training materials and courses have been developed by a number of bodies including the **Digital Curation Centre** and the **Software Sustainability Institute**; and Jisc has provided support for specialists in libraries and other services in individual universities. At a more general level, the EU-funded FOSTER project has brought together training resources for open science in an online **handbook**.

But deficits remain all too evident:

- many researchers need to develop greater understanding and skills in managing and curating the data they gather or create;
- many researchers also need to develop higher-level skills in software, in data processing, in analysing large and complex data sets, and in the application of learning algorithms;
- there is a need for more capacity in the form of professional experts in their different subject domains in matters including data stewardship, software, data processing and analysis; and
- there is a similar need to improve the recognition and rewards for specialist staff in these areas, both in libraries and embedded in research groups, with clearer career paths less dependent on time-limited project funding.

Coordinated efforts are needed to address these issues, so that researchers are not left struggling with sub-optimal solutions, and to bridge what risks becoming a growing divide between data specialists and researchers in their different subject domains. More courses to develop capacity in data science – as recommended in a number of reports including the recent **Growing the Artificial Intelligence Industry in the UK** – are certainly needed. But so are much greater efforts to ensure that the research community as a whole is better equipped to deal with the challenges as well as the opportunities of digital research. Training to establish good scholarly practice in the digital age needs to start at undergraduate level; but there is also a need to tackle current skills deficits.

Steps to that end include encouraging funders and research organisations to:

- recognise data skills and practice in their reward systems;
- establish more training programmes for researchers in different subject domains;
- promote the development of peer support networks;
- ensure that specialist support is provided close to researchers in their institutions and subject domains, with clear responsibilities for the provision of that support; and
- provide sustainable career paths for professional data experts, building on the **existing commitments** to support professional technicians.

We therefore recommend that:

a All stakeholders should provide **further incentives for researchers**:

i. Funders and other stakeholders should take active and co-ordinated measures to enhance their **understanding of the needs of different research communities**.

ii. Research organisations should ensure that data skills and ORD practice - both in creating ORD and using existing data - are given due weight in **assessing researchers' performance and achievements**.

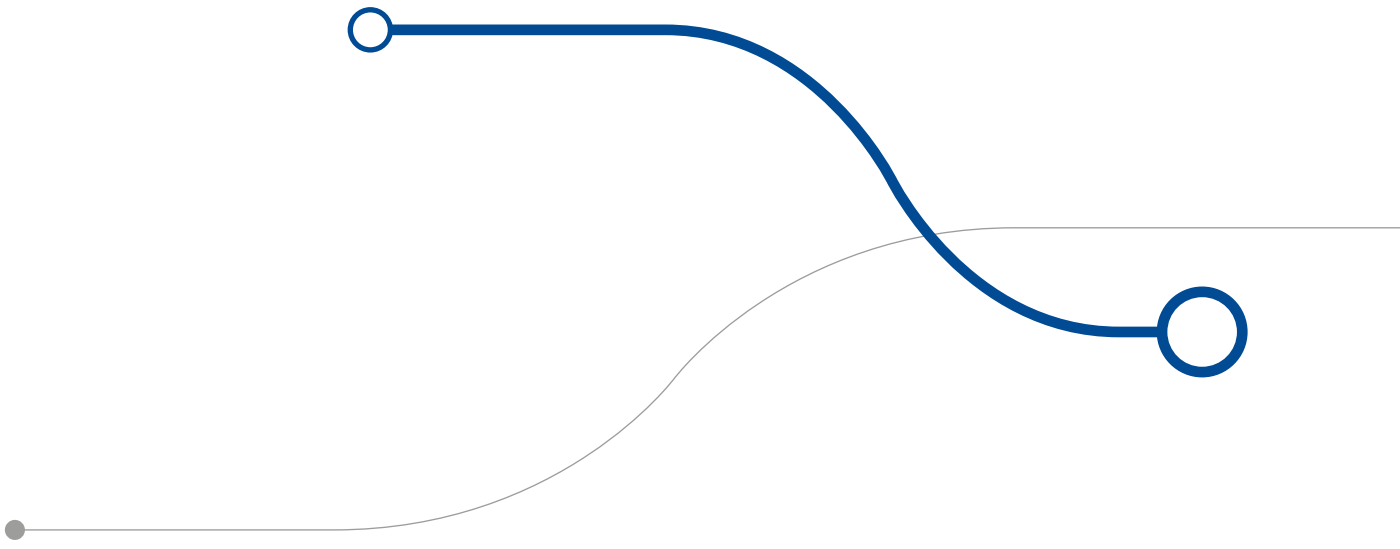
iii. Funders and publishers should establish **ORD training and guidance for peer reviewers in order to ensure greater consistency of approach**, and should monitor the impact of changes to peer review processes on researchers' uptake of ORD practices.

iv. Research England and other funding bodies should **consult and consider further with subject communities how best to assess and reward** ORD practice, with a view to strengthening expectations in this area in the post-2021 REF.

v. Publishers and other stakeholders should take further steps to **promote the re-use of data, and data citation** (Concordat Principle 8).

b Funders and research organisations should establish, in partnership with doctoral training centres and relevant specialist organisations, **training programmes for researchers at all levels** in data management and stewardship, in data processing, in analysing large and complex data sets, and in the application of learning algorithms (Concordat Principles 6, 7, 9).

c Research organisations should review their services to ensure that effective data management, ORD practice, and use of data are facilitated through **provision of specialist support services**, and help to expand the UK's specialist capacity in data stewardship, in research software, and in data science (Concordat Principle 9).



Further reading

(see also the bibliography in the **landscape report**)

CWTS Leiden and Elsevier (2017) Open Data: The Researcher Perspective. **https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf**

Data Citation Synthesis Group, Martone M. (ed.) (2014) Joint Declaration of Data Citation Principles: FORCE11. **https://doi.org/10.25490/a97f-egyk**

FOSTER (2018) Open Science Training Handbook **https://book.fosteropenscience.eu/**

Jubb, M (2016) Embedding cultures and incentives to support open research, Wellcome Trust **https://dx.doi.org/10.6084/m9.figshare.4055514**

Open Science Skills Working Group (2017) Providing Researchers with the Skills they Need to Practise Open Science, European Commission, **https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf**

Science Europe (2018) Guidance Document: Presenting a Framework for Discipline-specific Research Data Management **https://www.scienceeurope.org/wp-content/.../SE_Guidance_Document_RDMPs.pdf**

Stuart, D et al (2018) Research Data White Paper: Practical Challenges for Researchers in Data Sharing, Springer Nature. **https://doi.org/10.6084/m9.figshare.5975011.v1**

Van den Eynden, V et al (2016) Towards Open Research: practices, experiences, barriers and opportunities, Wellcome Trust **https://dx.doi.org/10.6084/m9.figshare.4055448**

Wilkinson, M D et al (2016) The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, **https://doi.org/10.1038/sdata.2016.18**

Working Group on Rewards under Open Science (2017) Evaluation of Research Careers Fully Acknowledging Open Science Practices: Rewards, incentives and/or recognition for researchers practising Open Science, European Commission **https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf**

3. Active leadership

Making ORD and its use part of good research practice depends on leadership from senior figures in the research community, and from funders, research organisations, and a wide range of specialist organisations. But leadership often comes in the form of community initiatives from below, as well as policy and service development from above. The complex landscape that results needs co-ordination, harmonisation of approaches and strategic direction. UKRI is in the best position to take a lead, working in partnership with research organisations and other key stakeholders.

The research ecology is complex, involving governments and funders from the public, voluntary and commercial sectors, universities and other research organisations, learned societies and disciplinary groups, researchers themselves, libraries and memory institutions, publishers, and a wide range of specialist services. Hence it is not surprising that the range of individuals and organisations with an interest in supporting the development of ORD is similarly broad and complex. Embedding ORD as a normal and integral part of research practice is an essential building block in enhancing public engagement and trust in research. It requires active and sustained measures from all the key stakeholders.

Figure 2 Key actors in Open Research Data in the UK

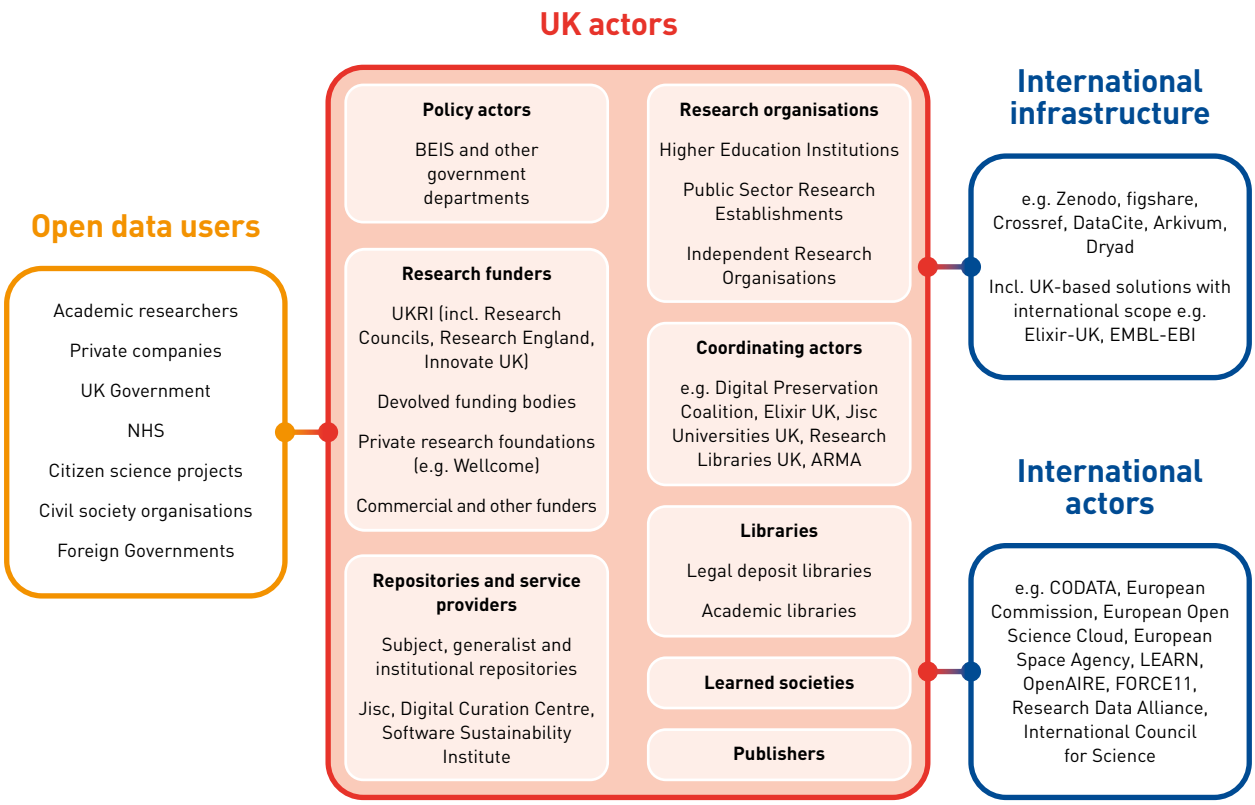


Figure 2 is an attempt to capture the major organisations and actors relevant to ORD in the UK research and innovation landscape. Due to the complexity of the landscape there will inevitably be information and entities that are missing.

3.1 Key actors and their roles

Over the past two decades in the UK, the development of policies on data management and ORD, and to some extent service provision, have tended to be driven by the Research Councils and major funders such as the **Wellcome Trust**.

But the nature and scope of their activities have varied significantly. Thus some Research Councils support well-established data centres, such as the **British Oceanographic Data Centre** (NERC) and the **UK Data Service** (ESRC), and the Wellcome Trust (along with MRC and BBSRC) provides crucial support for the **European Bioinformatics Institute**. But other funders do not support parallel services that operate across their full domains. Hence other organisations have played key roles too.

Jisc initiatives and services have stimulated and underpinned important developments across the university sector; and some individual universities and research organisations have adopted policies and established services to support ORD at a generic level within their organisations. Such developments have depended in large part on the sustained commitments of dedicated individuals in library and IT services, but also on the support of a champion at senior level within the institution. Universities have concerns about the costs associated with moves towards ORD, and policies in many cases remain aspirational rather than fully-implemented. There are many gaps, and well-resourced services are far from common across the HE sector.

Specialist services such as the **Digital Curation Centre** (DCC) and the **Software Sustainability Institute** (SSI) have also played critically-important roles in helping to develop and sustain communities of expertise and good practice in matters relating to ORD; and they have gained international reputations for their work. The development of effective communities of practice in the UK is reflected in the prominent role that UK representatives play in the activities of relevant international organisations and initiatives, including the **Research Data Alliance** (RDA) and **CODATA**.

Domain-specific services such as **Cambridge Crystallographic Data Centre (CCDC)** are also of crucial importance in their fields. Other important services stem from community-led initiatives, and some are supported by relevant learned societies, such as the **Chemical Database Service** (Royal Society of Chemistry) or at a smaller scale the **AlgaeBase** (British Phycological Society). And in some subject domains, such as environmental research, **funders** are working together at international level to co-ordinate their policies and services. In many subjects and disciplines, however, leadership and the development of ORD initiatives have been uneven at best.

Where members of research communities show high levels of engagement with ORD issues, some publishers, working typically through their journal editors, have helped to stimulate ORD practice. Policies on access to the data underlying the results presented in journal articles can have a powerful impact, particularly if they are accompanied by workflows and services that facilitate easy deposit and the provision of access to the data. At present, most publishers have been reluctant to institute policies that may be too far in advance of the research communities associated with their journals, and many have been reluctant to invest in workflows and services not seen as part of their core business. In some areas, including **earth and space science**, publishers are working together with their communities, and with funders, to coordinate the development of guidelines and approaches. Some major publishers have begun steps to harmonise their policies, and it is important that they should liaise with funders and other stakeholders to ensure common approaches. There remains a need for further work in areas including data deposition, the standardisation of data access statements and ensuring that data underpinning published articles is always citeable in its own right.

3.2 The need for co-ordination

It would be wrong to suggest that any individual organisation could or should take responsibility for the development of ORD policies and services across the whole UK research landscape.

Rather, there is a need to develop an ecology in which a diverse range of stakeholders work together to address and find solutions to researchers' various needs. But the current landscape is characterised by inconsistencies, gaps, overlaps, and lack of clarity – especially from the perspective of researchers – as to the roles and responsibilities of different organisations at local, national and international levels.

Fragmentation and variability have been key themes running through studies of research data and e-infrastructure policies and services for more than a decade. This is by no means unique to the UK. Recent reports and presentations have described the situation in the USA as characterised by 'radical scatter', and a recent report in Germany found over-dependence on individual initiatives and organisations, an absence of co-ordination, and lack of strategic direction. Such findings have led the German Rectors' Conference to approve the establishment of a federated National Research Data Infrastructure. Similar arrangements are being established in other leading research nations in Europe, including Denmark, Finland and the Netherlands, as well as Canada and Australia; and there are efforts to establish federal strategies in the USA too. Since collaboration with overseas partners is one of the key strengths of UK research ecology, it is important also that the UK should take note of these moves to create more

co-ordinated approaches to ORD. Senior representatives of UK funders and policy-makers, as well as practitioners, should strengthen their engagement with international initiatives including the EOSC and the RDA.

A great deal has been achieved by the various UK organisations involved in ORD. However, the lack of consensus on the appropriate roles and responsibilities of stakeholders within the overall research ecology (institutions, funders, learned societies, publishers, library and information services, research groups, community initiatives, national, international) is creating unresolved tensions, and puts the UK's hitherto leading role in the world at risk. There is now a need for coordinating mechanisms and incentives to promote cooperation across the research sector, with links also to international services and initiatives. In short, there is a need to develop a strategy, and organisational structures, to build on existing strengths, remedy weaknesses, and fill gaps. In doing so, a balance will be needed between top-down and bottom-up approaches: in some cases the right approach will be to promote and support community-led initiatives.

UKRI will need time to assume a leadership role, but it will have a central role to play. With the help of its e-infrastructure Advisory Board, it is uniquely placed to bring the key actors and stakeholders in the UK together, while taking proper account of:

- the range of policies and services (those currently in place, or which might be developed in the future) that operate most appropriately at local, national and/or international levels;
- the necessary degrees of autonomy for researchers and research organisations alongside the legitimate requirements of funders and policy-makers;
- the balance between generic policies and services that operate across all fields, or at an interdisciplinary level, and those that seek to address the needs of specific subjects and disciplines;
- the balance between small-scale community-driven or project-based initiatives and services on the one hand, and larger more formally-established services run by research institutions, funders and other organisations – including publishers and other commercial organisations - on the other;
- the opportunity to promote ORD when the Research Councils and other UK funders make agreements with funders overseas to support collaborative research; and
- the need for measures to facilitate and incentivise the re-use of ORD, not only in the research community, but by individuals and organisations far beyond that community, including those participating in citizen science.

We therefore recommend that:

a **UKRI should establish for itself a co-ordinating role** – while taking full account of the critical importance of active leadership from other stakeholders including research organisations, funders, specialist service providers, publishers, learned societies, and senior representatives of the research community – **in overseeing the development of ORD policies, infrastructure and services**, including:

i. establishing a **shared understanding of roles and responsibilities** across funders, research organisations, specialist organisations and different subject areas;

ii. considering the **most appropriate distribution of resources** for ORD between different actors within the UK research landscape;

iii. providing a focus for **strong engagement** with;

- international initiatives and organisations, and measures to promote collaboration in research
- organisations beyond the research community with interests as potential re-users of data, sources of data, or providers of data services;

iv. promoting in collaboration with relevant partners the development of **guidelines and protocols on ORD and its use** across the full range of subject areas; and

v. **establishing the evidence base and monitoring trends of relevance to ORD**, including developments in research methods and processes; trends in the availability and use of data; the balance between commercial service providers and more open systems; quality assurance mechanisms; and evidence as to usage and benefits (Concordat Principle 10).

Further reading

(see also the bibliography in the **landscape report**)

Association of American Universities and Association of Public and Land-Grant Universities (2017) Public Access Working Group Report and Recommendations **www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Intellectual-Property/Public-Open-Access/AAU-APLU-Public-Access-Working-Group-Report.pdf**

European Research Area and Innovation Committee, Working Group on Open Science and Innovation (2018) Assessment of the Amsterdam Call for Action **data.consilium.europa.eu/doc/document/ST-1202-2018-INIT/en/pdf**

RfII (2016) Enhancing Research Data Management: Performance Through Diversity: Recommendations regarding structures, processes and financing for research data management in Germany **www.rfii.de/?wpdmdl=2075**

4. Clear expectations

Policies on research data, and the expectations set for researchers, have developed piecemeal over time, and have tended to become more complex as the implications of ORD have become better understood. A single set of policy requirements for all researchers is neither feasible nor desirable. Nevertheless, embedding ORD across the whole UK research community demands greater harmonisation of existing policies, and increased consistency and clarity in the expectations placed on researchers.

Researchers face a confusing array of policies and conditions for ORD set by funders, universities and other research organisations; and some publishers and other agencies including data centres may set their own requirements. Policies and requirements vary in nature and scope, and the relationships between them are complex. At the most generic level, policy in the UK is now framed by the principles set out in the **Concordat on Open Research Data** endorsed by HEFCE, RCUK, Universities UK and the Wellcome Trust and published in 2016. Awareness of the Concordat remains low, however, and few policies make explicit reference to it. Nor do they – or the Concordat – refer to the readily-understandable **FAIR principles** (Findable, Accessible, Interoperable and Re-usable) that are being used increasingly as a basis for policy development in countries across the world.
It is unlikely that a single set of policies could cover all contexts, subjects and disciplines; but greater clarity in requirements and consistency in terminology will be necessary if the UK is to realise the full benefits of ORD.

4.1 The UK policy landscape

The Research Councils and other major funders have been developing their policies on research data for more than a decade. In response to comments about the differences between them, RCUK published in 2011 a set of seven **Common Principles on Data Policy** to provide an 'overarching framework' for Council policies.

That document was revised in 2015, and supplemented by lengthy **guidance** on the implementation of the principles; both documents remain extant. While there are some commonalities between the RCUK principles and those set out in the Concordat published in 2016, there are significant differences in language, tone and emphasis. Moreover, there are differences between the policy requirements and expectations of individual Councils, in addition to variations with regard to the allocation of costs; monitoring of ORD practices and compliance; skills and training; and the provision of data services. There are yet further differences between the Councils' requirements, the policies of other major funders such as the **Wellcome Trust** and the **European Commission**, and those of universities and other research organisations.

It is critically important that greater clarity in the expectations placed on researchers is founded on (but not constrained or determined by) a deep understanding of disciplinary norms. Some research communities are likely to remain in advance of others in their practical commitment to ORD, but there remains a need for much greater consistency in the policies, expectations and terminology adopted by funders, research organisations, publishers and others. This can draw on – and enhance the value of – the work of leading agencies in the UK but also of international organisations and initiatives, including the **RDA**, **CODATA**, **Science Europe**, the **Open Research Funders Group** and the **International Science Council**.

Policies should refer explicitly to the principles set out in the Concordat, and also to the FAIR principles. Adoption of the 'as open as possible, as closed as necessary' principle and a commitment to harmonisation, but not uniformity, of policies across UKRI's nine councils would help to clarify expectations. But these steps must be accompanied by clear articulation of the criteria surrounding necessary restrictions on access, particularly those which may arise with data provided by commercial partners, that is sensitive on security grounds, or that relates to identifiable individuals.

Table 1 The current and desired policy landscape for ORDs

Area	Current state	Desired future state
Policy coverage	Most policies lay stress on the data that underpins publications, although some policies – including those set by some publishers - refer more broadly to data of long term value, observational and experimental data, raw data as well as data refined through some standard procedure, or even more broadly to all the data acquired, assembled or created during the course of research.	Policies establish expectations that both stimulate and reflect changes in practice and possibilities in different fields. Where appropriate and feasible, policy coverage is extended to include the full range of data, along with software, workflows, algorithms, standard operating procedures and other materials required to validate truth claims.
Data management plans (DMPs)	Most policies refer to the need for DMPs to be prepared as a key part of any research project plan, and for them to be regularly reviewed and updated, but as the **Digital Curation Centre shows**, expectations as to issues to be addressed, assessment and monitoring procedures vary widely.	DMPs are required in all cases, with clear requirements on the issues to be addressed in them, domain-specific guidance on those issues, and clear assignment of individual and organisational responsibilities. DMPs are rigorously assessed and actively monitored.
Data formats	Many funders request in DMPs information about data formats and standards, but offer little guidance on the issue. Some funders', institutions' and publishers' policies refer to a preference for open or standard formats rather than proprietary ones, but many are silent on the issue.	Funders and other policymakers provide appropriate, discipline-specific guidance on data formats and standards, with a stated preference for open and standardised formats wherever possible.
Quality assurance	Some but not all policies point to the importance of quality assurance and standards. Few specify or provide guidance on possible quality procedures: standardised data capture, regular calibration, the use of templates, checking data manually and so on. Arrangements for the peer review of the data underlying publications – and guidance and training for peer reviewers – are still uneven.	Policies emphasise the need for quality standards and quality assurance, and provide or refer to guidelines on appropriate procedures to ensure that that data are accessible, understandable and usable. Strengthened but proportionate arrangements for peer review of data underlying publications are in place, underpinned by effective guidance and training for reviewers.
Metadata and documentation	Many policies refer to the importance of appropriate metadata and the documentation necessary to provide potential users with essential information about the origin and processing of research data. But while some prescribe the use of appropriate metadata standards, and/or documentation formats, many are silent on these issues, with important implications for discoverability and re-use.	Discipline-specific metadata standards and documentation formats are prescribed or suggested wherever possible. The Jisc **research data model** is further developed and more widely adopted.
Preservation	Policies make varying stipulations as to where, how, and for how long data should be preserved. A common requirement is that data should be preserved for a minimum of ten years, but some policies specify shorter or longer periods or suggest the need for judgements to be made as between costs and the value of the data. For published data associated with journal articles, or made accessible via links, it may be assumed – though not necessarily explicitly stated - that preservation should be in perpetuity; and some policies do make that stipulation for some kinds of observational data. There are yet further variations as to the repositories or data centres to be used for long-term preservation and curation.	Policies set clear expectations as to where, how and for how long different kinds of data should be preserved, with an appropriate balance between the value of the data, the risk of loss, and the costs involved. There is convergence of approaches to the preservation of similar types of data, and trusted repositories or data centres are widely used across all fields for curation and preservation.

Area	Current state	Desired future state
Access	Many policies require access to be provided via a recognised repository, though some allow for informal modes of access on request. Some policies prescribe the provision of links between data and publications, and a few journals and publishers require formal data access statements to be provided alongside all published articles, though there is as yet no common format for such statements. Many policies allow for embargo periods during which researchers have exclusive use of the data they have created, though the maximum periods allowed again vary.	Policies require access to be provided via a trusted and sustainable repository, and prescribe two-way links between data and publications, in all but exceptional cases. Journals and publishers routinely require data access statements, in standardised formats. Embargo periods are permitted, and vary by field, but there is increasing convergence.
Legal and ethical issues	Some policies refer to data protection and other legal and ethical issues that can surround the gathering of data, as well as data processing and subsequent access and use of that data. These are of particular prominence in areas of research involving human subjects, and biomedical and social science research funders therefore lay stress on the importance of proper ethical and regulatory procedures. But similar issues can arise in many other areas of research where policies may be less clear.	All policies lay explicit stress on the need for compliance with legal, ethical and regulatory frameworks, including the General Data Protection Regulation (GDPR), and provide links to relevant sources of guidance.
Ownership and licensing	Relatively few policies make explicit reference to the ownership of data and intellectual property rights, which may apply to many kinds of data created or collected by researchers. Ownership rights, and the contracts and licensing agreements associated with them, can be highly complex, but are often not adequately recognised in funders', research organisations' and publishers' ORD policies. Hence such policies may act as a disincentive to researchers collaborating with a range of partners in the commercial, voluntary and public sectors.	Policies clearly articulate (or acknowledge potential uncertainties relating to) ownership and intellectual property rights applicable to data created or collected by researchers. Appropriate licensing arrangements for different kinds of data from different sources are clearly set out. An approach of 'as open as possible, as closed as necessary' is married with robust safeguards to address the legitimate interests and concerns of research partners.
Data use	Relatively few policies place expectations on data creators to encourage and facilitate widespread use of their data.	Policies set clear expectations and provide incentives for researchers to promote and facilitate re-use of their data, and provide field-specific guidance on how this can be achieved.

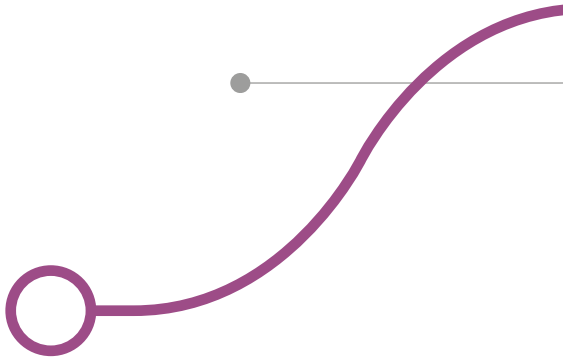
4.2 Publisher policies

As the prevalence of research data policies from research organisations and funders increases, publishers and editors are also paying more attention to standardisation and the wider adoption of data sharing policies.

Work to establish a requirement for data access statements to accompany journal articles has had an impact in some areas; but their use remains far from comprehensive. Meanwhile, major publishers such as **Elsevier** and **SpringerNature** have moved to adopt a standardised research data policy framework, and the Research Data Alliance (RDA) **Interest Group on Data Policy Standardisation and Implementation** is defining a common set of journal data policy requirements. Further work is needed to promote uptake of these and other requirements across a wider range of publishers and learned societies, whilst respecting disciplinary differences.

4.3 Related policies

The fundamental aims of enhancing the quality and value of research, and increasing public engagement and trust, are closely linked with many other aspects of research policy, including research integrity, knowledge exchange, research metrics and impact, and open access to publications.



The wide and growing array of policies relating to research can be daunting and confusing for researchers. It would be helpful if ORD policies were more closely and explicitly linked to policies and codes in relevant areas, as well as other features of research and innovation strategies, including infrastructure development, grand challenges, collaboration, information assurance, flows of knowledge between industry and academia, and international collaboration.

We therefore recommend that:

a | **The long-term aim** for funders, research organisations and publishers should be to establish **clear expectations as to**:

i. **what data, software and related material should be preserved and made accessible**, and in what formats;

ii. the **trusted, certified repositories** that should be used for this purpose; and

The initial focus should be on **data supporting published research findings** (Concordat Principle 8) before moving on to broader sets of data with long term value in different subjects and disciplines.

b | UKRI should lead an effort between sector organisations to **establish greater harmonisation of ORD policies**, using common terminology (as used in the Concordat), to include:

i. an explicit **commitment to both the Concordat and the FAIR principles**;

ii. a commitment to the principle of **'as open as possible, as closed as necessary'** as a basis for all policies, together **with a clear articulation of the criteria surrounding necessary restrictions on access** (Concordat Principles 2, 4) and the requirements of legal, ethical and regulatory frameworks (Concordat Principle 5);

iii. a clear set of expectations with regard to intellectual property rights in research data;

iv. a clear articulation of the relationships between ORD and other aspects of research policy, including **research integrity, knowledge exchange, research metrics and impact, and open access to publications** (Concordat Principle 8); and

v. greater consistency in **requirements for DMPs** and how they are handled and monitored, whilst preserving appropriate disciplinary differences.

c | Publishers and learned societies should: review their policies to require a **data access statement** with all articles and monographs; support open citation of data; and enhance workflows to facilitate data deposit and linking.

Further reading

(see also the bibliography in the **landscape report**)

Naughton, L., & Kernohan, D. (2016). Making sense of journal research data policies. Insights, 29(1), 84–89. DOI: **http://doi.org/10.1629/uksg.284**

Science Europe (2018) Guidance Document Presenting a Framework for Discipline-specific Research Data Management **https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf**

5. User-friendly services

The current provision of ORD services is uneven, with urgent needs for enhancements to existing services, and the creation of new ones. Additional specialist staff are needed to provide technical services and guidance, and support the development of communities of practice. Data creators require easily-discoverable exemplars of good practice, and user-friendly web-based tools. For data users, there needs to be a stronger focus not just on repositories, but on knowledge-bases and other services that add value to data by creating new interfaces that facilitate browsing, discovery and re-use.

Data services are provided, as we have noted, by a range of local, national and international organisations and initiatives, with relatively little co-ordination between them. In this complex and rapidly-changing landscape, it can often be difficult for researchers to find the services best suited to handling the data they have created, and also the location of data that they might wish to use (some of it created outside the research community). Moreover, when data creators and potential users find such services and sources, they are often not user-friendly.

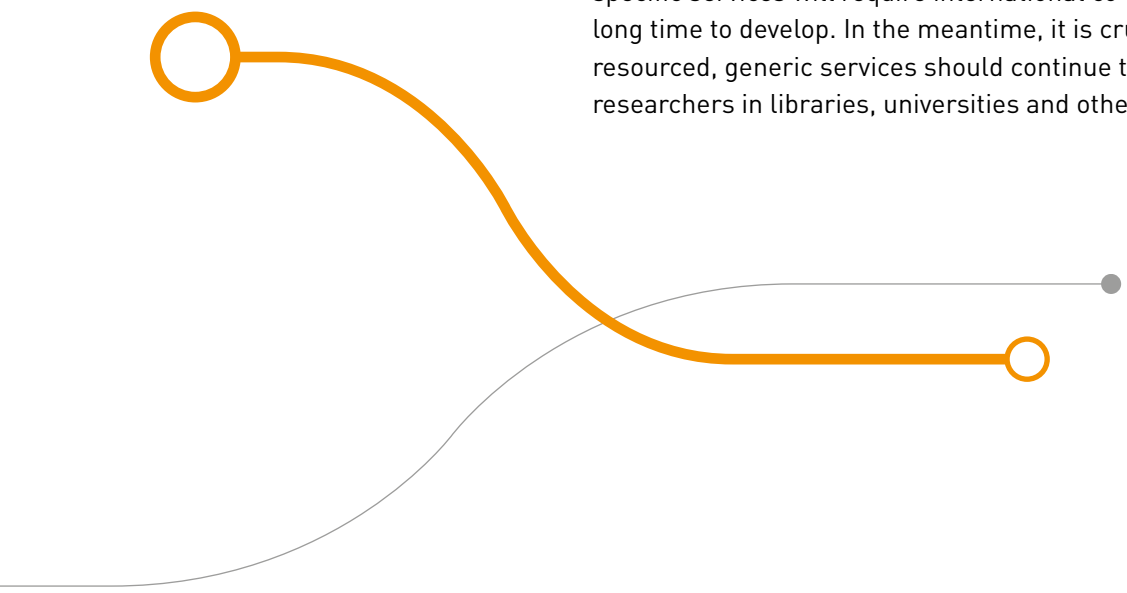
5.1 Services and their take-up

Given the complex array of services we have described in this report, it is not surprising that take-up of those services is itself patchy.

Surveys indicate that while in a relatively small number of areas – notably in the biosciences and astronomy – take-up of well-established specialist services is high, more than half of UK researchers rely entirely on their own resources to manage and preserve the data they create; and in such circumstances, the chances of its being made available to others in any form are minimal.

Since research almost invariably builds on what has come before, project-based, bottom-up approaches lead to a constant reinvention of some of the underpinning elements, such as tools developed for active research data management, alongside new innovation. Researchers typically look to what is known and/or available for their own discipline area, and may be ignorant of relevant institutional or other support services.

Evidence suggests that there is a positive feedback loop between the provision of well-resourced specialist domain-specific services – especially when they are developed and implemented in partnerships between researchers, funders and other agencies – and relatively high levels of take-up. But where domain-specific services are not available, and researchers must rely on generic services provided by research organisations or other suppliers, take-up tends to be significantly lower. The development of more user-friendly services closely targeted at the needs of specific research communities – as both creators and users of data – is essential. But a fully-comprehensive set of domain-specific services will require international co-operation and will take a long time to develop. In the meantime, it is crucial that more, and well-resourced, generic services should continue to be developed close to researchers in libraries, universities and other research organisations.



5.2 Meeting researchers' and other users' needs

User-friendly services are needed that provide researchers and data users with easily-understandable and ready means to address two key questions, sometimes characterised as the first and last mile: *"what should I do as I create my data and seek to make it open?"* and *"how do I actually use data that is already available?"*.

In a fragmented eco-system, researchers need help to find the answers, and to navigate their way around the various inadequately connected services relevant to their needs. Jisc's proposed **shared data service** is in part intended to provide the needed connectivity.

Issues need to be addressed from the very beginning of the research process if data are to meet the FAIR principles, and be appropriately curated and preserved. First, researchers need more easily-findable and readily-understandable guidance and exemplars on how to ensure that their data meets appropriate quality standards. They need to understand how to address issues including what is measured or recorded, the standards of measurement, and ways in which data are manipulated; and on how to choose between, and comply with, different standards. There is a need for similar guidance for data users too on how to ensure that the data they wish to use conforms to good standards.

Second, despite progress over the past two decades, the recording of research processes and results in many fields still requires significant manual effort; and this represents a significant barrier to widespread adoption of good data management practice. Tools to help standardise, and where possible automate, the processes involved in data management, documentation and curation would help to reduce, if not eliminate, those barriers.

Third, although relatively few researchers have received any formal relevant training, many of them use and modify software in the course of their research, and many create their own bespoke software. Guidance and clear exemplars on what software to use for different purposes, and on creating new software, are essential. The **Software Sustainability Institute** has helped to develop communities of practice in software carpentry. But much more needs to be done to improve research software, and thus to enable more researchers to work more effectively. Key issues remain the lack of support for researchers who are heavily engaged in coding, who get little recognition for such work; the difficulty in developing careers in this area; and conversely the problems research organisations have in recruiting and retaining high-quality software engineers.

In sum, researchers need more effective support in making sure that data and related material meets the **FAIR principles**, as outlined in Table 2. None of these issues is intractable, and there are many good resources and services to help researchers address them. But as **one expert advocate** of ORD put it, they are not always well co-ordinated or connected, and they are not always easily discoverable.

Table 2 Implementation of the FAIR principles

Principle	Current state
Findability	Making research data easily-discoverable by both machines and humans is inherently more complex than doing the same for scholarly articles and other publications. Findability depends critically on the quality of metadata, and the efficiency of harvesters and indexers. A number of specialist discovery services have been developed, by DataCite, the Data Citation Index and others. But detailed searching, whether via specialist or more general search engines, requires rich metadata that can be onerous to create, and many researchers are dissatisfied with existing tools and guidance for this purpose. Jisc has developed a canonical **research data model**, and some experts are seeking to develop metadata standards and schema in specific subject areas, such as that being developed by the **International Virtual Observatory Alliance (IVOA)** in astronomy. But more work is needed to develop the necessary schemas, ontologies and controlled vocabularies across all fields to make the kinds of searching that users are now accustomed to for text and images work effectively in searches for data. It is important that the UK is actively involved in international initiatives to that end, including those of the **World Wide Web Consortium (W3C), schema.org**, and the **European Open Science Cloud.** But there is also a need for more expert services to enable researchers to improve their metadata.
Accessibility	Researchers rely in large part on repositories and data services to ensure that their data – and associated metadata – are accessible, using open standardised communications protocols. But more guidance is needed for researchers in many subject areas on the repositories or data services available to them, and the criteria on which they should choose which to use. Data may also be held back pending clearance from the community being studied, or may simply be too big to be shared easily. Repositories and data centres need to do more to make the task of uploading data straightforward and user-friendly.
Interoperability	The heterogeneity of the data created or gathered by researchers means that conforming to the principle of interoperability is a huge challenge. While there is widespread agreement that the use of open, community-developed standards is crucial, lack of co-ordination has led to many overlapping and competing standards. There is tension between the drive for broad standards that facilitate interdisciplinary research on the one hand, and the needs of specific disciplinary communities on the other. Similar tensions can arise with domain-specific data centres and services that are responsive to the needs of their particular communities, but may have little incentive to interact with other domains. The UK must be actively involved in international efforts to resolve these tensions, and help researchers with readily-findable and user-friendly guidance and tools needed to achieve interoperability in practice.
Re-usability	A key objective should be to make data open-able and re-usable from the start, rather than trying to make it so at a later stage. At present, data are most re-usable where data types are simple and easy to describe, and when the community is organised and collaborative. Re-usability depends on comprehensive information about why, when, where and how data was created, and how it was manipulated. Documenting provenance in this way is essential to enable users to understand the data (names, labels and descriptions for variables, code and so on) and to verify that the research was conducted as described. For research with well-tried processes, such recording may be built into the processes themselves, and data provenance thereby verified as part of its creation. But such practice remains relatively rare; and bespoke processes may have to be recorded manually. Licencing arrangements for research data are often unclear, and poorly understood, while re-purposing data so that it can be re-used more generally by users from outside the research community poses even greater challenges. Much more work is needed in order to address these issues, and particularly create data commons and other services where both researchers and other potential users could gain access to data, and manipulate it in a variety of ways.

5.3 Digital preservation

Any guarantee of access to data for the long term rests with repositories, and it is crucial that they should conform to standards on key issues including the use of persistent identifiers; publicly-available metadata; data stability and persistence; access and use; machine-readability; and the sustainability of the repository itself.

In view of the complexities involved, individual research organisations and libraries will require continued support and the provision of cross-sector solutions, including those provided by Jisc. There are long-standing debates on all these issues, and it is important that the UK should be represented at senior level in forums, including those established by **Science Europe**, **CODATA** and the **Research Data Alliance**, where matters such as certification of repositories are being debated.

5.4 Openness of research data infrastructure

The services provided at local, institutional, national and international levels already involve commercially-owned services alongside those organised, financed and operated by different parts of the research sector.

Both open source and commercial services can legitimately serve the needs of ORD, but, as the **Metric Tide report** (2014) has recognised, there is a need for greater transparency and openness in research data infrastructure.

The **Forum for Responsible Research Metrics** is actively working to improve the research information management infrastructure that underpins metric use and which closely intersects with the data infrastructure. However, there are also concerns that underpinning systems are overly-controlled by providers whose long-term interests may not coincide with those of the wider research community. Discussions about moves towards open scholarly infrastructures are long-standing, and some progress has been made through organisations such as **NISO** and the development of open standards. The balance between commercial and publicly-funded provision will need to be kept under careful review, with close attention paid to issues such as responsiveness to changing needs, competition in the marketplace, interconnections and interoperability between different

elements in the eco-system, and the openness and adaptability of the system as a whole. Jisc has done important work in developing requirements to ensure maximum interoperability, and this should be developed further to build capabilities and capacity in research organisations, with support for them when it comes to negotiating with commercial providers. Universities and other research organisations must ensure that agreements are consonant with the **FAIR principles** and the aims of open science, that they and their researchers retain ownership of all relevant data, and that they have a viable exit strategy to avoid vendor 'lock-in'.

5.5 Local, national and international services

Many if not most of the domain-specific services used by UK researchers involve international partnerships.

There are many relatively small-scale community-led initiatives based in the UK or with strong UK participation, such as the **Cuneiform Digital Library** dealing with ancient texts, or the **Protein Circular Dichroism Data Bank**, dealing with the secondary structure and folding properties of proteins. At a much larger scale, the UK has been a major player in the development of international infrastructures such as the **International Virtual Observatory Alliance** in astronomy, and the **ELIXIR** organisation involving 21 countries across Europe in the life sciences. A wide range of UK organisations are also participating in global initiatives such as the **Global Alliance for Genomics and Health (GA4GH)** and **dataONE**. Such examples could be multiplied many times over. On the other hand, UK involvement in European Research Infrastructure initiatives in the humanities, such as **CLARIN** and **DARIAH**, or in the more generic **OpenAIRE** open access and repositories initiative, has been limited. It is crucial to the interests of the UK itself that it should sustain – and where necessary strengthen – its engagement and participation in these services at the most senior level, as well as in international organisations such as the **RDA** and **CODATA** with important roles in policy-development, community-building, training, and consolidation of standards and services.

We therefore recommend that:

a | Research organisations and funders should review their provision to ensure that, across all subjects and disciplines, researchers as both creators and users of data have **access to readily-understandable and user-friendly ORD services – both generic and domain-specific**.

b | Research organisations and funders, with the support of Jisc, should **develop a set of principles for negotiation with commercial providers of ORD infrastructure** to maximise interoperability, retain ownership of relevant data, and reduce the risk of vendor 'lock-in'.

c | The development of **sustainable ORD infrastructure should feature prominently as UKRI develops its long-term research and innovation infrastructure roadmap**, with relevant needs to be identified in close consultation with learned societies and subject communities.

d | Research organisations and funders should take active steps to ensure that the UK sustains and strengthens its **participation in the development of services that operate at international level, and engagement with initiatives on standards, software, tools and related matters that demand international co-operation**.

Further reading

(see also the bibliography in the **landscape report**)

Allen, R and Hartland, D (2018) FAIR in practice: Jisc report on the Findable Accessible Interoperable and Reuseable Data Principles **http://doi.org/10.5281/zenodo.1245568**

Bryant, R et al (2017), The Realities of Research Data Management, OCLC **https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html**

CWTS Leiden and Elsevier (2017) Open Data: The Researcher Perspective. **https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf**

European Commission (2018) Sustainable European Research Infrastructures: A Call for Action **http://ec.europa.eu/research/infrastructures/pdf/ri_policy_swd-infrastructures_2017.pdf**

Kaye, J et al (2017). Establishing a shared research data service for UK universities. Insights, 30(1), 59–70. DOI: **http://doi.org/10.1629/uksg.346**

OECD Science, Technology and Industry Papers (2017) Strengthening the Effectiveness and Sustainability of International Research Infrastructures **http://dx.doi.org/10.1787/fa11a0e0-en**

More...

OECD Science, Technology and Industry Papers (2017) Digital Platforms for Facilitating Access to Research Infrastructures **http://dx.doi.org/10.1787/8288d208-en**

Royal Society (2018) A snapshot of UK research infrastructures **https://royalsociety.org/~/media/policy/Publications/2018/snapshot-uk-research-infrastructures.pdf**

Stuart, D et al (2018) Research Data White Paper: Practical Challenges for Researchers in Data Sharing, Springer Nature. **https://doi.org/10.6084/m9.figshare.5975011.v1**

Tenopir, C et al (2016) Research Data Services in European and North American Libraries: Current Offerings and Plans for the Future **https://onlinelibrary.wiley.com/doi/full/10.1002/pra2.2016.14505301129**

Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management **https://doi.org/10.13140/RG.2.1.4929.1363**

Wolff, C et al (2016a) Ithaka S+R Jisc RLUK UK Survey of Academics 2015, Ithaka S+R, New York **https://doi.org/10.18665/sr.282736**

Wolff, C et al (2016b) Ithaka S+R US Faculty Survey 2015 Ithaka S+R, New York **https://doi.org/10.18665/sr.277685**

6. Sustainable funding

There is a need for an improved understanding of the costs, business and funding models for ORD throughout the research lifecycle, coupled with targeted funding schemes to fill priority gaps, incentivise innovation and promote data re-use. The additional costs associated with a more data-rich research landscape must be tensioned against an increasing awareness of the productivity gains arising from ORD, as well as the benefits accruing to researchers and research organisations, service providers, and consumers of ORD.

Developing and sustaining research data infrastructures depends on long-term funding. In a dual support system characterised by funding for short-term projects on the one hand and block grant distributed to institutions on the other, demands for long term funding – for national infrastructures in particular – bring inevitable tensions. Meanwhile, moves to embed ORD in practice place additional demands on individual researchers and research groups, as well as research organisations themselves, and these are rarely matched by new revenue streams.

In the competition for funds from the Research Councils, there are pressures to give precedence to research projects rather than long-term infrastructures. Larger projects may build their own infrastructure in-house, usually ad hoc and temporary, subject to the needs of the project. But project funding may only cover the set-up costs, with few incentives to build for wider reuse even where it is justified, and longer-term sustainability may therefore be left unaddressed. In disciplines and subject areas where researchers are less reliant on external funding, or where large-scale project funding is rare, infrastructure needs may go entirely unmet.

6.1 Project-based, grant-funded research

Most UK funders now require costed DMPs to be submitted as part of grant application processes.

Yet researchers are often unwilling or unable to request costs for data management and sharing, due to a perception that funds can be better deployed on other activities, and/or uncertainty over which costs should be met from grants, and which represent indirect costs of research. Partly as a consequence, most funding for data management services and activities remains indirect in nature. Notwithstanding the work of Jisc and others, only limited funds have been committed to foster experimentation in ORD, promote re-use of existing data, and support new developments such as data publishing.

Funding for current data infrastructures to support grant-funded research is usually provided by research organisations, Jisc, and/or funders of large facilities, and this has led to understandable tension, since approaches and expectations vary by discipline and funder. There are indications, for example, that infrastructure providers funded by research-performing organisations tend to focus on servicing their own organisation; and providers' different perspectives can inhibit collaboration across institutional boundaries. A gap often develops between the infrastructure and services that organisations must provide as a requirement of external funding and the much wider requirements of their researchers, who may not otherwise be supported in accessing or providing infrastructures in their own domains.

6.2 Understanding costs

The focus of most efforts to understand the costs of data management and ORD to date has been on long-term preservation and archiving at the end of the research life cycle.

This results in large part from the need for large-scale data archives and related services to apply for renewals of funding at regular intervals. At the archiving stage there has now developed, as we highlight elsewhere, a complex dynamic between researchers' practice and the expectations of - and the services offered by - funders, research organisations and publishers; and the focus has been on the relationships between scholarly articles and underlying data.

However, there is a need to broaden the focus to consider the costs of what may be termed the 'pre-archive' phase, including both:

- the additional researcher effort and directly-incurred costs - above standard planned research procedures and practices - that are needed to manage and preserve research data and make it shareable beyond the primary research team; and
- the capital and revenue costs of the full range of infrastructures and services supporting ORD practice throughout the research life cycle.

There have been few rigorous attempts to examine the full economic costs of those activities and services, and their sustainability, not least because they vary significantly between disciplines, and between individual projects producing different kinds of data at different scale. Nevertheless, there are concerns across the research sector that the costs are significant. Existing **work undertaken by Jisc** in this area provides a valuable starting point, but there is a need to develop a much better understanding of the scope and scale of costs involved, given the wide range of pre-archive ORD activities which it falls to researchers and research organisations to undertake (Table 3).

Table 3 Example 'pre-archive' activities associated with the move to ORD

Lifecycle Phase	Researcher / research team	Research organisation
Skills, Support and Outreach	• Maintain familiarity with institutional policies, procedures and support services • Attend relevant training	• Train researchers in good data management practice • Develop internal data management policy and guidelines • Liaise with data centres and other institutions • Advocacy, promotion and awareness-raising
Project Initiation and review	• Design, plan and cost research on open principles • Specify IT, software, licensing and technical support requirements • Draft and cost data management plan • Assign project roles and responsibilities for achieving openness • Monitor data management practice against the DMP	• Review and advise on DMPs and costings • Aid in experimental design and planning • Maintain registry/records of data management plans
Creation and management of data and related material	• Secure consent for data sharing • Transcription and digitisation • Metadata creation • Undertake data cleaning • Manage data transfer and access • Active data storage and backup • Ensure data security • Conversion to standard or open file formats • Anonymisation and/or pseudonymisation • Confirm copyright and licensing arrangements	• Provide data backup, risk/disaster management and contingency planning • Development of institutional standards and coordination of best practice • Enhance data skills development for researchers • Provide specialist analytical services and advice • Support and advise on metadata creation and documentation requirements • Advise and negotiate on copyright and IP

Archive phase

6.3 Sustainability of national and international research data infrastructures

The OECD defines sustainability for a research infrastructure in terms of its capacity to remain operative, effective and competitive over its expected lifetime.

Since research is so often an internationally-framed enterprise we must consider national infrastructures within this wider ecosystem. As data archives and services are long-term enterprises they will reflect evolving financial, governance and political contexts leading to great diversity across both disciplines and regions.

In Germany, for example, as in the UK, a complex range of actors results in no ready source of comprehensive information about the funding of research data infrastructures or projects, or of their costs and benefits. As a consequence, **Deutsche Forschungsgemeinschaft (DfG)** is looking to develop funding models to support general infrastructure covering the full range of disciplines, beyond individual research projects. And it acknowledges that over the next few years, data infrastructures will have to account for a larger proportion of research budgets. In a similar vein, the **High Level Expert Group on the European Open Science Cloud** has estimated that on average about 5% of total research expenditure should be spent on properly managing and stewarding data in an integrated fashion.

In considering the sustainability of national research data infrastructures, there is a need to develop an inventory of relevant infrastructures, and undertake regular reviews of the ecosystem to identify strengths and gaps. Existing services should be encouraged to develop plans to implement ORD, and common indicators should be developed to facilitate the initial evaluation, monitoring and follow-up of business plans. The difficulties of transitioning new infrastructure from the set-up to the operational phase should be explicitly acknowledged, together with a recognition that new funding streams may be needed to ensure medium to long-term sustainability.

6.4 Enabling uptake of ORD beyond the research community

Although our focus has been on the research community, encouraging use of ORD beyond that community will be essential if the wider benefits foreseen for the UK are to be realised.

This is likely to require a targeted approach by industry sector, and should include consideration of both the use of ORD by industry and other non-academic users, as well as the provision of data for research purposes by these organisations. In many areas, ORD has the potential to facilitate advances in machine learning and artificial intelligence. The development of 'data trusts', as recommended in the report **Growing the Artificial Intelligence Industry in the UK**, could help to facilitate the sharing of data between organisations holding data and organisations looking to use data to develop AI, as well as other applications. Consideration should also be given to the provision of funding for initiatives designed to make existing data more readily usable in support of national research and innovation priorities, as outlined in the government's **Industrial Strategy** and in UKRI's Strategic Prospectus.

We therefore recommend that:

a | UKRI and other funders should work together with universities and other research organisations to **review the costs, business and funding models adopted by the current range of data services** in the UK, their sustainability and potential for future development.

b | Funders should **establish a requirement for all funded data services to develop appropriate plans for ORD**, taking due account of commercial, ethical and other constraints on what data can be made open.

c | UKRI and other funders should work together with research organisations to review their current support for ORD, and **take measures to ensure that levels of funding appropriate to an increasingly data-rich research landscape are provided** – directly, via research project grants and other routes – to develop and sustain the portfolios of services needed to support ORD.

d | UKRI and other funders should give consideration to the provision of funds for new initiatives, including challenge funds to:

i. **fill priority gaps in specific areas**;

ii. **support data publishing**; and

iii. support exemplar projects to use already-existing data, and **pilots to make such data more readily usable in support of national research and innovation priorities**, particularly by users from beyond the research community.

Further reading

(see also the bibliography in the **landscape report**)

Beagrie, N. (2011) Keeping Research Data Safe 2 Activity Model **https://beagrie.com/krds.php**

European Commission (2016) Realising the European Open Science Cloud **http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf**

Hall, W. H, and Pesenti, J. (2017) Growing the Artificial Intelligence Industry in the UK **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf**

HM Government (2017) Industrial Strategy: Building a Britain fit for the future **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf**

Hodson, S et al (2017) Income Streams for Data Repositories RDA-WDS Interest Group on Income Streams for Data Repositories **http://dx.doi.org/10.5281/zenodo.46693**

Inroad (2018) Policy Brief #1 Insights on research infrastructure roadmap processes, funding and business models **http://inroad.eu/wp-content/uploads/2018/03/20180306_Policy_Brief_1.pdf**

Kejser B et al (2014), D3.1 Evaluation of Cost Models and Needs & Gaps Analysis Final Report (Revision 1) **http://www.4cproject.eu/d3-1/**

Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data (2016) "Funding research data management and related infrastructures" **http://www.knowledge-exchange.info/news/articles/24-05-2016**

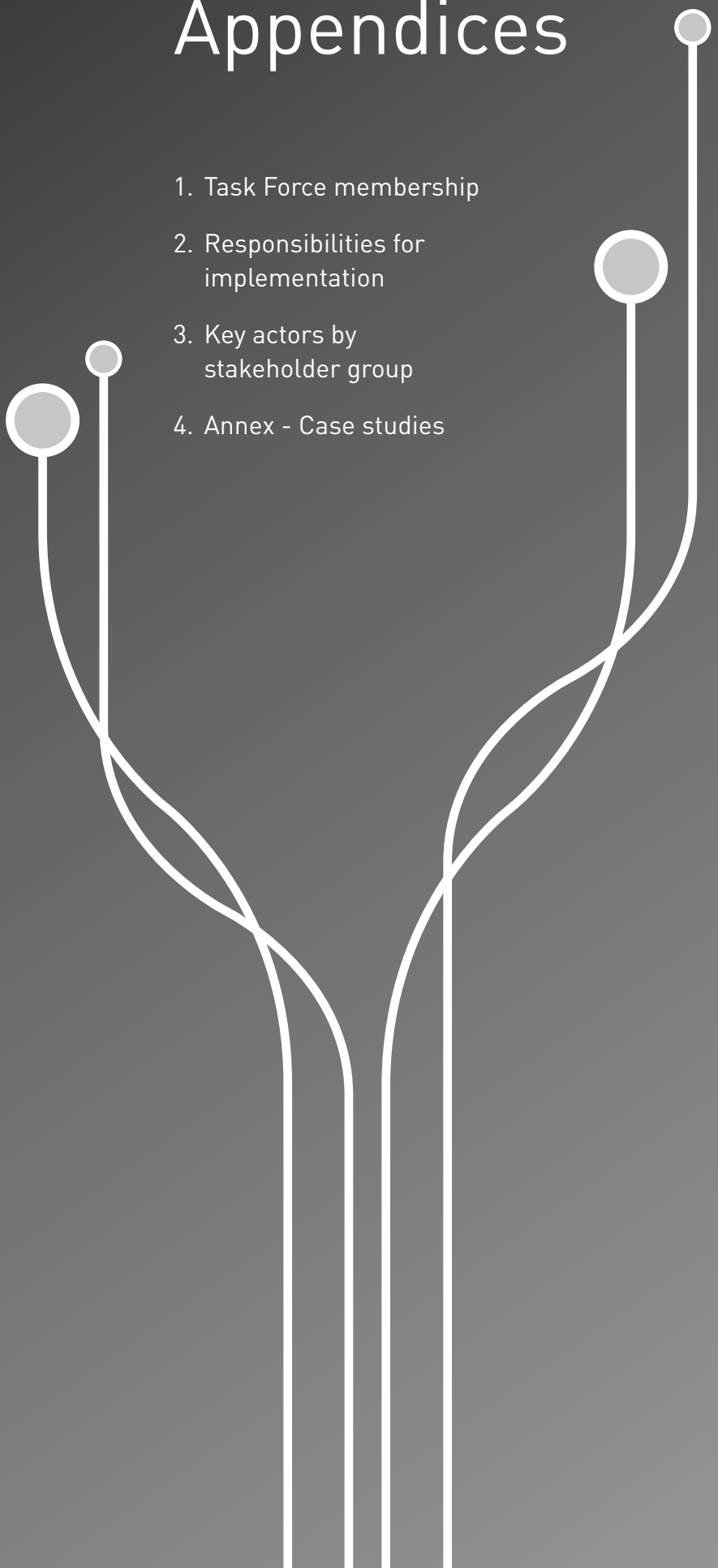
OECD (2017) Strengthening the effectiveness and sustainability of international research infrastructures **http://dx.doi.org/10.1787/fa11a0e0-en**

OECD (2017) Business models for Sustainable Research Data Repositories **http://dx.doi.org/10.1787/302b12bb-en**

UK Data Service (2013) Data Management and Costing Tool **http://www.data-archive.ac.uk/media/247429/costingtool.pdf**

Appendices

1. Task Force membership
2. Responsibilities for implementation
3. Key actors by stakeholder group
4. Annex - Case studies

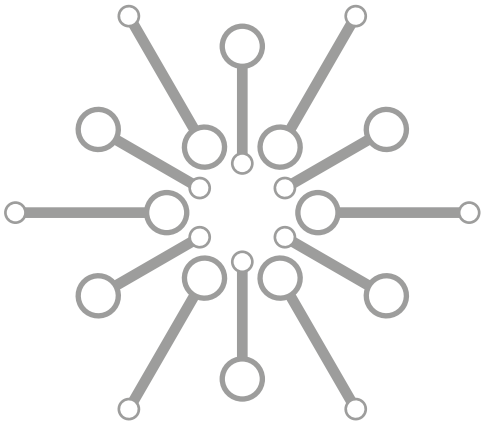


Appendix 1

Task Force membership

Professor Pam Thomas (PVC Research, University of Warwick) chaired the Task Force:

- **Professor Geoffrey Boulton** (University of Edinburgh, CODATA)
- **Professor Bill Spence** (Queen Mary University of London)
- **Professor David De Roure** (Oxford University)
- **Professor Andrew Prescott** (University of Glasgow)
- **Chris Awre** (University of Hull)
- **Professor Nick Wright** (Newcastle University)
- **Professor John Helliwell** (University of Manchester)
- **Stuart Taylor** (Royal Society)
- **Professor Carole Goble** (University of Manchester)



with representatives from HEFCE (RE/UKRI), The Wellcome Trust, Universities UK, BEIS and Jisc.

- **Claire Fraser** and **Jonathan Piotrowski** (now Research England) and **Ben Johnson** (HEFCE)
- **David Carr** and **Robert Kiley** (The Wellcome Trust)
- **Max Hastings** (Universities UK, now UKRI) and **Jamie Arrowsmith** (UUK)
- **Wayne Williams** (UKRI)
- **Alexandra Saxon** (RCUK)
- **Peter Ford** (BEIS) - observer
- **Rachel Bruce** and **Caroline Ingram** (Jisc)

Jisc provided the secretariat to the task force.

Appendix 2

Responsibilities for implementation

This Appendix provides further details to support the implementation of the Open Research Data Task Force's recommendations, including an indicative split of responsibilities across key stakeholder groups.

Appendix 3 provides further details on the key organisations within each group who are likely to have a role in supporting implementation. Further work will be required within and across these stakeholder groups to agree responsibilities and develop concrete work streams to support implementation of our recommendations.

For a more comprehensive analysis of relevant stakeholders readers should refer to the **Landscape Report**.

Indicative timeframes

We stress in the report that progress towards embedding ORD as a norm in research practices will require sustained efforts by all stakeholders over several years. In the table to the right, we indicate timeframes in which we recommend that significant activity should have taken place, and some progress should have been made, though we do not expect that all issues should be resolved. The periods are deliberately overlapping, since we recognise the close interconnections between many of the measures we recommend.

Short-term = within 1-3 years
Medium-term = within 4-6 years
Long-term = within 7-10 years

Responsibilities for implementation

L **= Lead Role** – these stakeholders should take responsibility for initiating and co-ordinating action against the recommendation.

S **= Supporting Role** – these stakeholders have a supporting role to play in the implementation of the recommendation.

Report reference	Recommendation	Concordat Principle(s)	Timeframe	Responsibilities for implementation						
				UKRI	Other funding bodies	Learned societies	Research organisations	Libraries	Publishers	Infrastructure and service providers
	Better incentives, fewer barriers									
2.a.i	Take active and co-ordinated measures to enhance understanding of the needs of different research communities	1, 2, 3, 4, 5, 6, 9	Medium-term	L	S	S	S	S	S	S
2.a.ii	Ensure that data skills and ORD practice - both in creating ORD and using existing data - are given due weight in assessing researchers' performance and achievements	1, 5	Long-term			S	L	S	S	
2.a.iii	Establish ORD training or guidance for peer reviewers of grant applications and of publications, and monitor the impact of changes to peer review processes on researchers' uptake of ORD practices	9	Medium-term	L	L	S			L	
2.a.iv	Research England and other funding bodies should **consult and consider further with subject communities how best to assess and reward** ORD practice, with a view to strengthening expectations in this area in the post-2021 REF	1	Short-term	L		S	S			
2.a.v	Take further steps to promote the re-use of data, and data citation.	8	Medium-term	L	L	S	S	S	L	S
2.b	Establish, in partnership with doctoral training centres and relevant specialist organisations, training programmes for researchers at all levels in data management and stewardship, in data processing, in analysing large and complex data sets, and in the application of learning algorithms	6, 7, 9	Medium-term	L	S		L	S		S
2.c	Review institutional services to ensure that effective data management, ORD practice, and use of data are facilitated through provision of specialist support services, and help to expand the UK's specialist capacity in data stewardship, in research software, and in data science	9	Medium-term				S	L		L

L = Lead Role **S = Supporting Role**

Report reference	Recommendation	Concordat Principle(s)	Timeframe	Responsibilities for implementation						
				UKRI	Other funding bodies	Learned societies	Research organisations	Libraries	Publishers	Infrastructure and service providers
	Active leadership									
3.a	UKRI should establish for itself a co-ordinating role in overseeing the development of ORD policies and services	-	Short-term	L						
3.a.i	Establish a shared understanding of roles and responsibilities across funders, research institutions, specialist organisations and different subject areas	3, 6, 7, 8	Medium-term	L	S	S	L	S	S	S
3.a.ii	Consider the most appropriate distribution of resources for ORD between different actors within the UK research landscape	3	Medium-term	L	S		S			S
3.a.iii(1)	Strengthen engagement with international initiatives and organisations, and measures to promote collaboration in research	10	Medium-term	L	S	S	L	S	S	L
3.a.iii(2)	Provide a focus for strong engagement with organisations beyond the research community with interests as potential re-users of data, sources of data, or providers of data services	2	Long-term	L	S	S	L	S	S	S
3.a.iv	Promote the development of guidelines and protocols on ORD and its use across the full range of subject areas	2, 4, 5	Medium-term	L	S	S	L	S	S	S
3.a.v	Establish mechanisms to build the evidence base and monitor trends of relevance to ORD, including: • developments in research methods and processes; • trends in the availability and use of data; • the balance between commercial service providers and more open systems; • quality assurance mechanisms; and • evidence as to usage and benefits	10	Short-term	L	S		S		S	S

L = Lead Role **S = Supporting Role**

Report reference	Recommendation	Concordat Principle(s)	Timeframe	Responsibilities for implementation						
				UKRI	Other funding bodies	Learned societies	Research organisations	Libraries	Publishers	Infrastructure and service providers
	Clear expectations									
4.a	Establish clear expectations as to: • what data, software and related material should be preserved and made accessible, and in what formats. • the trusted, certified repositories that should be used for this purpose. The initial focus should be on data supporting published research findings before moving on to broader sets of data with long term value in different subjects and disciplines	8	Short-term	L	S	S	S	S	S	S
4.b	Establish greater harmonisation of ORD policies, to include: i. an explicit commitment to both the Concordat and the FAIR principles ii. a commitment to the principle of 'as open as possible, as closed as necessary', together with a clear articulation of exceptions and the requirements of legal, ethical and regulatory frameworks iii. a clear set of expectations with regard to intellectual property rights in research data; iv. a clear articulation of the relationships between ORD and other aspects of research policy, including research integrity, knowledge exchange, research metrics and impact, and open access to publications v. consistency in requirements for DMPs and how they are handled and monitored, whilst preserving appropriate disciplinary differences	2, 4, 5	Short-term	L	L	S	S			
4.c	Require a data access statement with all articles and monographs; support open citation of data; and enhance workflows to facilitate data deposit and linking	8	Medium-term	S	S	S		S	L	

L = Lead Role **S = Supporting Role**

Report reference	Recommendation	Concordat Principle(s)	Timeframe	Responsibilities for implementation						
				UKRI	Other funding bodies	Learned societies	Research organisations	Libraries	Publishers	Infrastructure and service providers
	User-friendly services									
5.a	Review ORD provision to ensure that, across all subjects and disciplines, researchers as both creators and users of data have access to readily-understandable and user-friendly ORD services – both generic and domain-specific	1,3,6	Long-term	L	L		L	S		S
5.b	Develop, with the support of Jisc, a set of principles for negotiation with commercial providers of ORD infrastructure to maximise interoperability, retain ownership of relevant data, and reduce the risk of vendor 'lock-in'	6	Short-term	L	S		L	S		
5.c	The development of sustainable ORD infrastructure should feature prominently as UKRI develops its long-term research and innovation infrastructure roadmap, with relevant needs to be identified in close consultation with learned societies and subject communities.	3	Medium-term	L		S	S			S
5.d	Take active steps to ensure that the UK sustains and strengthens its participation in the development of services that operate at international level, and engagement with initiatives on standards, software, tools and related matters that demand international co-operation	3, 6	Long-term	L	L	S	L	S	S	S

L = Lead Role **S = Supporting Role**

Report reference	Recommendation	Concordat Principle(s)	Timeframe	Responsibilities for implementation						
				UKRI	Other funding bodies	Learned societies	Research organisations	Libraries	Publishers	Infrastructure and service providers
	Sustainable funding									
6.a	Review the costs, business and funding models adopted by the current range of data services in the UK, their sustainability and potential for future development	3, 6	Short-term	L	S		S	S		S
6.b	Establish a requirement for all funded data services to develop appropriate plans for ORD, taking due account of commercial, ethical and other constraints on what data can be made open	6, 7	Medium-term	L	L		S	S		S
6.c	Review current support for ORD, and take measures to ensure that levels of funding appropriate to an increasingly data-rich research landscape are provided – directly, via research project grants and other routes – to develop and sustain the portfolios of services needed to support ORD	3, 6	Medium-term	L	L		S			S
6.d.i	Give consideration to the provision of funds for new initiatives to fill priority gaps in specific disciplinary areas	3, 6	Short-term	L	L					S
6.d.ii	Give consideration to the provision of funds for data publishing	8	Medium-term	L	L	S	S	S	S	S
6.d.iii	Consider funding exemplar projects to use already-existing data, and pilots to make such data more readily usable in support of national research and innovation priorities, particularly by users from beyond the research community.	1, 3	Medium-term	L	L		S			

L = Lead Role **S = Supporting Role**

Appendix 3

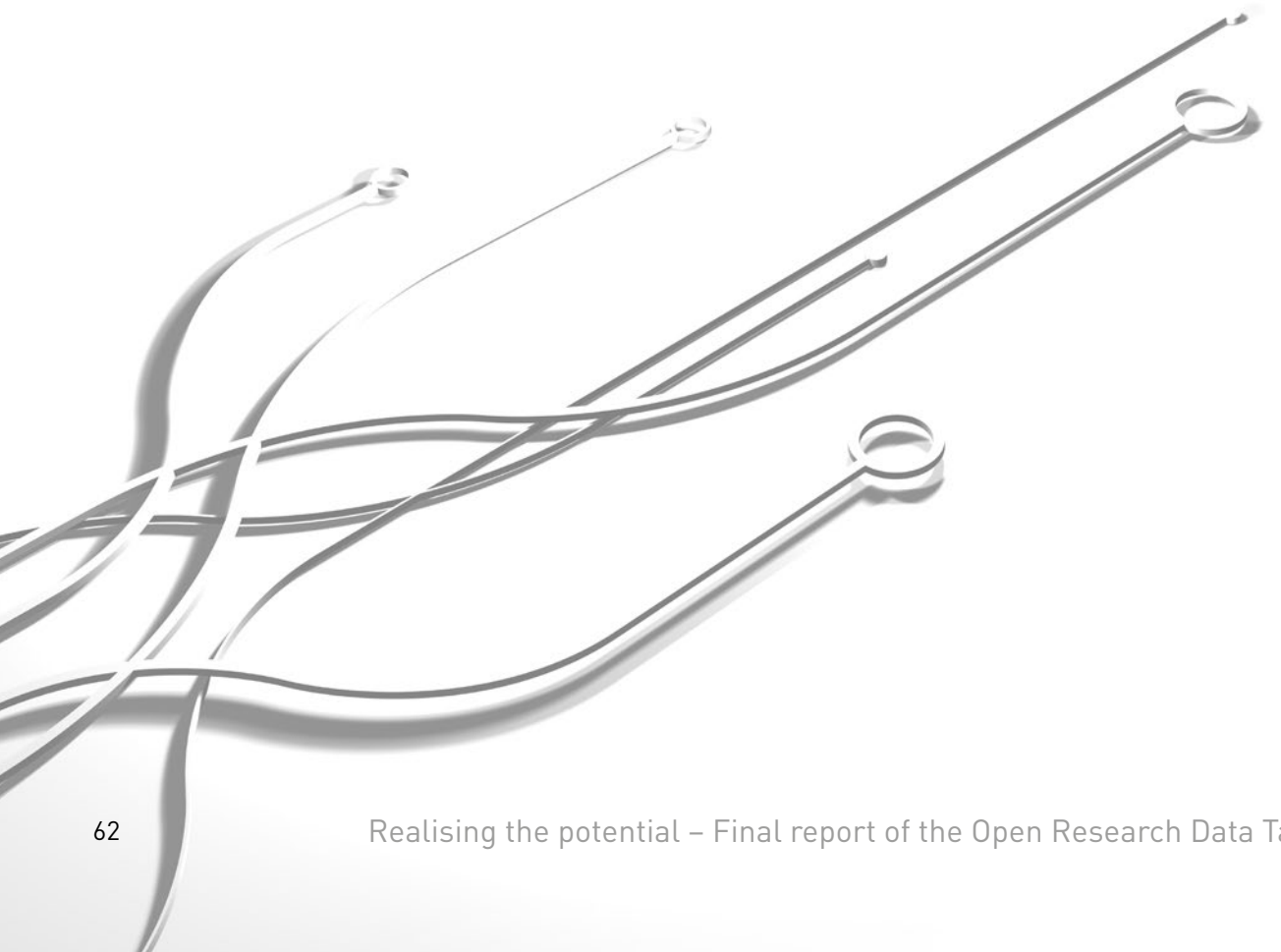
Key actors by stakeholder group

The table provides further detail on key actors within the stakeholder groups identified in Appendix 2.

Its focus is on those organisations who are active within the UK landscape, but in each case there will also be a need to take account of relevant international developments, actors and working groups. These include, among others, the European Commission - including relevant EC-projects and infrastructure (e.g. OpenAIRE, FOSTER Plus, Zenodo and the European Open Science Cloud) as well as a number of expert groups convened on its behalf - Science Europe, Knowledge Exchange, the Research Data Alliance, CODATA, the International Council for Science (ICSU), the European Universities Association, and a wide range of international service and infrastructure providers.

Stakeholder group	Description and key actors
UKRI	UK Research and Innovation includes the seven Research Councils, Innovate UK and Research England. In many cases UK Research and Innovation is identified as having a leading or co-ordinating role in taking our recommendations forward. The Task Force emphasises, however, that progress on ORD is reliant on collective action and partnership between multiple stakeholders, and a balance between bottom-up and top-down initiatives. The distinctive focus and remit of the nine councils within UKRI is acknowledged, and our recommendations will need to be implemented by each body in a way that reflects the different communities they serve.
Other funding bodies	Other funding bodies include the Higher Education Funding Council for Wales (HEFCW), the Scottish Funding Council and the Department for the Economy of Northern Ireland, particularly but not solely through their oversight of the Research Excellence Framework, managed on their behalf by Research England. This group also includes the National Institute for Health Research, government departments such as DfID, the Home Office and the Department of Work and Pensions, and private foundations and charities such as the Wellcome Trust and Cancer Research UK (many of which are members of the Association of Medical Research Charities (AMRC)).
Learned societies	Learned societies represent the interests of disciplinary communities and can act as intermediaries between researchers and stakeholders in industry, government and civil society. Many are also active publishers of scholarly journals and monographs, and may assume additional responsibilities for the promotion of open research data in this capacity. Priorities for initial engagement should include the Royal Society, the British Academy, the Academy of Social Sciences, the Royal Academy of Engineering, the Arts and Humanities Alliance and the Academy of Medical Sciences.

Stakeholder group	Description and key actors
Research organisations	Research organisations includes universities, public sector research establishments and independent research organisations. Universities UK and GuildHE act as formal representative bodies for higher education in the UK while other co-ordinating actors include the Independent Research Organisation Consortium (IROC) and the Association for Innovation, Research and Technology Organisations (AIRTO). Also included under this heading are higher education sector mission groups, including the Russell Group, University Alliance and Million+, and a range of professional and membership associations serving distinct groups of staff within research organisations. These include the Association for Research Managers and Administrators (ARMA), Vitae and the UK Research Software Engineer Association.
Libraries	Libraries have a distinct and important role to play in the promotion of Open Research Data. They host repositories, provide training and support to researchers, and can act as advocates for open practices both within their host organisations and externally. This category includes both academic libraries located with research organisations, the British Library and the other legal deposit libraries. Key co-ordinating actors are Research Libraries UK (RLUK), the Society of College, National and University Libraries (SCONUL) and Jisc. The latter is well-placed to support libraries and research organisations in sharing good practice, developing new capabilities and services and monitoring progress.
Publishers	Publishers are represented by trade bodies including the Publishers Association, the International Association of STM Publishers, the Association of Learned and Professional Society Publishers (ALPSP) and the Open Access Scholarly Publishers Association (OASPA). Jisc and Jisc Collections are actively engaged in representing the higher education sector's interests to the publishing community, and involved in defining best practices for data journals in partnership with some publishers.
Infrastructure and service providers	Jisc is the UK education and research network, and links to international networks for research data. It provides a wide range of services to the higher education sector, including those relating to data storage, repositories and archiving; cyber security; discovery and safe sharing; data standards and protocols; and specialist advice and guidance. Other infrastructure and services for ORD are provided by a wide range of international initiatives, funders, sector agencies, individual universities and research institutes and commercial providers. Key actors within the UK landscape include domain-based provision such as the NERC data centres, the UK Data Service, the Administrative Data Research Network (ADRN), the European Bioinformatics Institute (EBI) and the Research Data Facility, plus advisory and support services such as the Digital Curation Centre (DCC) and the Software Sustainability Institute (SSI). Many of these are funded in whole or in part by the Research Councils.

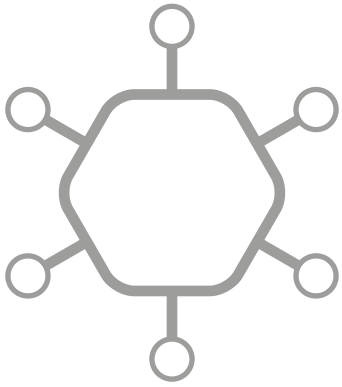


Annex

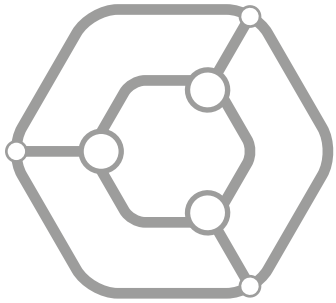
Case studies

1. Astronomy

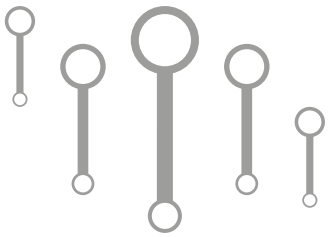
2. Biosciences



3. Crystallography



4. Digital Humanities



5. University of Bristol

University of BRISTOL

6. University of Salford

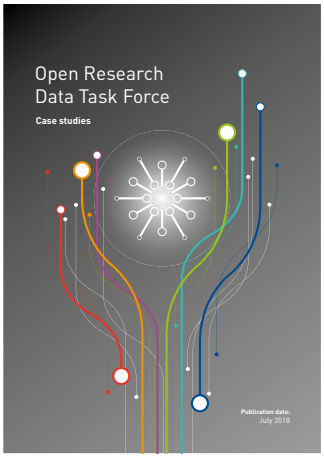
University of Salford MANCHESTER

7. Natural History Museum

NATURAL HISTORY MUSEUM

8. Germany

Download PDF



The Open Research Data Task Force was created in Autumn 2016 to establish realistic ambitions for UK open research data (ORD), to assess the current landscape, and to identify options to improve capacity, capability and infrastructure.

This report summarises the Task Force's findings, and identifies a set of concrete actions to accelerate progress towards the adoption and use of ORD in the UK, under five main headings:

- Better incentives, fewer barriers
- Active leadership
- Clear expectations
- User-friendly services
- Sustainable funding

The recommendations made herein are intended to deliver greater efficiency in UK research, to improve its rigour and reproducibility, to enhance its impact, and to increase public trust in its results.

Realising the potential

Final report of the Open Research Data Task Force

Report Design: www.crayfishdesign.com