Concordat on Open Research Data

The Concordat on Open Research Data has been developed by a UK multi-stakeholder group. This concordat will help to ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical, disciplinary and regulatory frameworks and norms, and with due regard to the costs involved.

Published 28th July 2016

Foreword

The UK is a world leader in terms of open research data: EPSRC's data framework has been adopted by many research funders, and the UK Data Archive has retained social science and humanities data for almost 50 years. This *Concordat* is a testament to the research community's ability to build on that expertise and steer this fast developing policy.

The UK is on course to make all taxpayer-funded research publications available in an open access format. Open research data is the next step in achieving the UK's open science ambitions. I see open access to research data as a fundamental good: combining research publications with their data will help drive transparency, improve co-operation and strengthen the UK's position as a global science leader.

The *Concordat*, for the first time, proposes a series of clear and practical principles for working with research data that cover the many roles needed to support the research process. It is not a rulebook, but a set of expectations of best practice developed by the research community itself.

This is not a Government owned document, nor should it be. The research community has worked hard to arrive at the consensus delivered in this report and I would like to thank the members of the UK Open Research Data Forum for their valuable contributions. I would also like to thank Professors Nick Wright, Rick Rylance and Duncan Wingham for their leadership.

Rt. Hon Jo Johnson MP

Minister of State for Universities and Science

Definitions

In this concordat, the following definitions have been adopted:

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). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as 'relational' or 'functional' components of research, thus signalling that their identification and value lies in whether and how researchers use them as evidence for claims.

They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.

The primary purpose of research data is to provide the information necessary to support or validate a research project's observations, findings or outputs.

Open research data are those research data that can be freely accessed, used, modified, and shared, provided that there is appropriate acknowledgement if required;

Not all research data can be open and the concordat recognises 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.

Introduction

This concordat will help to ensure research data gathered and generated by members of the UK research community is, wherever possible, made openly available for use by others in a manner consistent with relevant legal, ethical and regulatory frameworks and disciplinary norms, and with due regard to the costs involved.

The benefits from opening up research data for scrutiny and reuse are potentially very significant; including economic growth, increased resource efficiency, securing public support for research funding and increasing public trust in research. However, the concordat recognises that access may need to be managed in order to maintain confidentiality, protect individuals' privacy, respect consent terms, as well as managing security or other risks.

Openness implies more than disclosure of data. All those engaged with research have a responsibility to ensure the data they gather and generate is properly managed, and made accessible, intelligible, assessable and usable by others unless there are legitimate reasons to the contrary. Access to research data therefore carries implications for cost and there will need to be trade-offs that reflect value for money and use.

Commitment to the principles set out in the concordat will help demonstrate to government, business, international partners, other researchers and the wider public that, where appropriate, they can expect to see research data made open for the benefit of all. Such commitment will also ensure the results of research are properly open to scrutiny, with the data that underlies the concepts and arguments set out in published papers made accessible for testing and validation by other researchers, reinforcing the vital principle of self-correction.

The intention of this Concordat is to establish sound principles which respect the needs of all parties. It is not the intention to mandate, codify or require specific activities, but to establish a set of expectations of good practice with the intention of increasing access to research data as the desired position for research for the public benefit. It is recognised that in some fields opening up research data is still in the early stages of development and adoption and that the widespread adoption of open research data is best viewed as a journey in which the research community will participate over the coming years. Sharing research data in a manner that is useful and understandable requires putting research data management systems in place and having research data experts available from the beginning of the research process. It is recognised this Concordat describes processes and principles that may take time to establish within institutions, given there is currently a deficit of knowledge and skills in the area of research data management across the research sector in the UK.

This concordat sets out ten principles with which all those engaged with research should be able to work. By committing to the principles outlined in this concordat, the research community can demonstrate that they:

- are acting in an appropriate manner concerning research data;
- conform to all ethical, legal and professional obligations relevant to their work;
- nurture a research environment that makes data open wherever practical and affordable;
- use transparent, robust and fair processes to make decisions concerning data openness;
- have appropriate mechanisms in place to provide assurances as to the integrity of their research data; and
- recognise the importance of data citation and credit acknowledgement.

Following a similar process to that outlined in other UK concordats, this concordat recognises the different responsibilities of researchers, their employers, and funders of research. It also recognises the vital role others play in this, including professional, statutory and regulatory bodies; journals and publishers; academies and learned societies. By outlining these responsibilities, the concordat helps stakeholders to understand clearly the roles they play in producing the economic and social benefits of increased access to research data, delivering meaningful efficiency gains through the open sharing of data between researchers, developing the next generation of researchers and building public trust in the integrity of published research.

It is recognised that research is often carried out with partners from other countries and this has implications for open research data in terms of collaboration and the potential effects of the Concordat on international collaborators. It is not the intention to create barriers to international collaboration, but rather to assist the UK in playing a leadership role internationally.

The concordat adopts an approach that is supportive and developmental; recognising that open access to research data will be an ongoing process which:

Applies to all fields of research for the public benefit - The principles outlined in the concordat are relevant to all disciplines in which research data is gathered and analysed.

Emphasises responsibilities and accountabilities - The concordat implies the need for cooperation between different stakeholders and identifies the different roles they play in supporting open research data. The best way to ensure open research data becomes a reality for research in the UK is for all those involved to acknowledge and discharge their specific responsibilities and to work together towards developing a sustainable open research data environment.

Recognises the autonomy of researchers - researchers are a diverse group of people operating in many different cultures and contexts. They must have the freedom to strengthen policies and procedures relating to research and research data as appropriate to their circumstances: there can be no 'one size fits all' approach. The concordat provides a flexible framework to help researchers ensure they are able to fully discharge their responsibilities and to help employers ensure they have the mechanisms in place to meet the highest standards.

Complements existing frameworks - Extensive statutory and regulatory standards already exist to govern research practice and data access where it is deemed necessary. Similarly, conditions of grant from funding bodies will often be accompanied by specific guidelines that themselves create obligations. The concordat does not supersede or replace these, but addresses directly the issues related to open research data.

Principle #1

Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice.

In many fields, data is already widely shared and there are a number of excellent examples of open data in fields such as crystallography, genetics, archaeology and linguistics. These disciplines have benefitted both in terms of progressing research but also in enhancing resource efficiency and therefore securing funder support for their efforts. In addition to establishing practical arrangements for making research open, these fields have developed a culture of transparency and sharing; and this is a powerful asset in protecting against research fraud or innocent mistakes. These actions also enhance the reputation of the institutions in which the research is being undertaken.

Access to data across many fields is also stimulating new types of thinking as researchers develop new understandings by bringing together data from a variety of sources. This is enabling new perspectives on multi-disciplinary problems across a wide variety of fields. In many instances, it is the linking of data from a range of public and commercial bodies alongside the data generated by academic researchers that is enabling the most exciting insights in, for example, the application of technology to complex sustainability related issues such as transport.

Open data can underpin innovation, for example when researchers with fresh perspectives use data in unexpected ways or when companies use data to help them develop new products. This can lead to substantial economic benefits and help growth.

It is not always appropriate to make research data openly accessible, and there are a variety of legitimate reasons to restrict access, however, the concordat takes as its starting axiom that, where possible, making research data openly available for inspection and use by others is an inherent good with many benefits. Within this new paradigm, the following expectations will be established:

Researchers will, wherever possible, make their research data open and usable within a short and well-defined period, which may vary by subject and disciplinary area and reflect the resources available to them to do so. Data supporting publications should be accessible by the publication date and should be in a citeable form. Where it is not possible to make data open for legitimate reasons, there should be no negative consequences for those researchers concerned.

Employers of Researchers will foster a research environment which recognises the value of open data and will seek to provide appropriate access to infrastructure systems and services to enable their researchers to make research data open and usable, having due regard to value for money. They will also recognise good data management as an important aspect of researchers' duties (see Principle #9).

Funders of Research will support open research data by appropriately acknowledging and supporting its costs, and by supporting the wider agenda with appropriate policy and investment activities.

Principle #2

There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable.

It is not always appropriate to make research data openly accessible and there are reasons why access must be restricted, including *inter alia*, to maintain confidentiality, guard against unreasonable costs, protect individuals' privacy, respect consent terms, as well as managing security or other risks.

Governance arrangements must be in place to establish if and how data that relates to or derives from individuals can or should be made available, while safeguarding privacy and confidentiality. These should draw upon well-established models and good practices for managed access to data and always be proportionate to the level of risk associated with the particular data holding. Studies may adopt a graded approach where less sensitive data types are made more readily available, and access to more sensitive data is more stringently controlled. Governance arrangements need to take full account of legal, regulatory and ethical requirements – including applicable data protection laws and relevant codes on research ethics and research integrity.

The research community values highly the involvement of companies in collaborative research which brings substantial societal benefits through innovation leading to economic growth. It is important that open research data does not deter companies from collaborating with universities and other research organisations. There is therefore a need to develop protocols on whether, when and how data that may be commercially sensitive should be made openly accessible, taking account of the weight and nature of contributions to the funding of collaborative research projects, and providing an appropriate balance between openness and commercial incentives. Many research projects rely on collaboration with voluntary or public sector organisations, and it is likewise important that open research data does not disincentives such collaborations. Research organisations are also under a public obligation to maximise the economic benefits of their research and the exploration of these issues is a legitimate reason to delay making research data open for an appropriate period.

The role of third-party data providers in the wider research environment is also important and it is recognised that such providers may impose legitimate restrictions on making data more widely available. Data licensing agreements can make it complex to make research data open, creating legitimate and genuine difficulties for researchers and research organisations.

There may be other valid reasons to restrict access to data, including the need to protect sensitive environmental or cultural sites, or cases where the costs of preserving or supplying the data are disproportionate. In addition, data should not be shared if it would infringe intellectual property rights, confidentiality requirements or any other legal restrictions.

Decisions on which data to preserve and make open should generally be made by individual researchers under the auspices of a verifiable and transparent process of oversight at an appropriate institutional level. Specific plans for sharing of data should be considered from the earliest stages of project planning and set out in the Data Management Plan. It is important, however, that constraints on openness must not be applied on a blanket basis but should be justified and justifiable case by case. Research organisations or individual researchers withholding data must therefore consider carefully the grounds on which they are acting and be prepared to justify their actions.

Principle #3

Open access to research data carries a significant cost, which should be respected by all parties.

Whilst the benefits of open research data are real and achievable, the necessary costs - for IT infrastructure and services, administrative and specialist support staff, training and for researchers' time - are significant. It is therefore vital that consideration of costs (both capital and recurrent) forms an important part of any obligation arising from the move to open research data recognising that such costs may fall outside of the defined time period of a particular project. Such costs should be proportionate to real benefits. It is recognised that the benefits and costs of open research data must be tensioned with those of the research portfolio as a whole.

It is UK policy that research organisations undertaking publically funded research are able to access resources for all legitimate costs through the so-called dual support system. It is therefore reasonable that appropriate costs of making research data open are met through those mechanisms whilst recognising the obligation to reduce costs through efficiency and sensible design of both obligations and infrastructure. All research funding organisations that impose a requirement for open research data must do so in a manner that is consistent with available cost recovery mechanisms.

For research organisations such as Universities or Research Institutes, these costs are likely to be a prime consideration in the early stages of the move to making research data open – particularly where the required cost recovery mechanism is not yet in place. Both IT infrastructure costs and the on-going costs of training for researchers and for specialist staff, such as data curation experts, are expected to be significant over time. Significant costs will also arise from Principle #10 regarding the undertaking of regular reviews of progress towards open access to research data. All of these costs must be balanced with the benefits to the research portfolio as a whole.

Principle #4

The right of the creators of research data to reasonable first use is recognised.

The creation of original research data may often require significant expertise and hard work over many years. It is obvious that any undermining of the incentive to undertake such work would have a significantly negative impact on the advancement of global research and knowledge. Therefore it is vital the transition to open data must not reduce the willingness of researchers to undertake the journey to gather and generate original research data.

In some disciplines, such as astronomy and genomics, immediate sharing of research data is expected and provides significant benefits. However, this approach is not appropriate for all disciplines. If researchers across all disciplines were to be required to make newly-generated data or analyses of that data available immediately, many may conclude there is little advantage in pursuing original data-gathering, measurements or analyses. Rather it would be easier to simply wait for others to undertake the work and then to take advantage of their data. Such a situation would clearly be undesirable.

To prevent such negative outcomes, researchers who generate original data must have reasonable right of exclusive first use for an appropriate and well-defined period, which may vary by subject and disciplinary area. Such periods should be established as disciplinary norms through consultation led by learned societies. This should include an understanding that researchers first need to verify newly-obtained data (generally by repeating measurements) before they themselves can use the data for publications or other outcomes.

It should be noted, however, that even in disciplines where immediate sharing is not the norm, there may be circumstances in which research data should be made immediately open in the public interest, for example when it may be of significance and value in dealing with a public health emergency.

In some circumstances, this right of first use could include the withholding of initial datasets until later related datasets have been developed. This could be justifiable if such an action were to advance conceptual understanding of central concepts carrying implications for the research field, or in studies that address long-term changes or developments. This justification should not be used without serious consideration undertaken prior to the commencement of the long-term study; and in all cases data supporting and underlying publications should be accessible by as close to the publication date as possible and in citeable form (see Principle 8).

Any period of exclusive use should be considered from the earliest stages of project planning and set out in the Data Management Plan, and this should be balanced against the public interest in release and may be tested if someone makes a formal request for the data.

Principle #5

Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.

When users gain access to and use open research data - as indeed any data generated by others - it is vital they do so in a manner that respects the contexts and norms under which it was gathered and generated. It is thus essential that those who subsequently use the data respect and adhere to the same frameworks and observe any restrictions that may have been imposed during data collection or generation. This is widely recognised already in fields of research that rely on data of a highly personal nature from research participants (for example, patient data – see Principle #2); but it can apply equally in many other research fields.

All users of research data must formally cite the data they use. This is important both in those cases where the data has been generated as an inherent part of research, and where the primary aim of the research has been to create datasets that can be used by others. The obligation to recognise through citation and acknowledgement the original creators of the data must be respected in both cases. Publishers should enable the formal citation of data in articles to support these practices.

As stated in the existing Concordat on Research Integrity "*Individual researchers are responsible for compliance with ethical, legal and professional frameworks whilst it is the role of employers to support researchers in this through clear policies, awareness raising and providing clear advice and guidance*". Research organisations should therefore be proactive in revising such guidance and advice to reflect the issues of open research data. Learned societies should also play a strong role in establishing relevant ethical guidelines and promoting best practice across the disciplines that they nurture.

Production of open research data should be acknowledged formally as a legitimate output of the research process and should be recognised as such by employers, research funders and others in contributing to an individual's professional profile in relation to promotion, research assessment and research funding decisions. Such formal recognition should be accompanied by the development and use of responsible metrics that allow the collection and tracking of data use and impact. In general, data citations should be accorded appropriate importance in the scholarly record relative to citations of other research objects, such as publications.

Principle #6

Good data management is fundamental to all stages of the research process and should be established at the outset.

The careful management of data throughout the research process is crucial if the data arising from research projects is to be rendered openly discoverable accessible, intelligible, assessable and usable. It is essential therefore that the management of research data is considered from the beginning of the research process and due consideration is given to how research data are to be managed.

It is expected that research organisations should provide access to the necessary infrastructure to enable researchers to manage their data effectively, and provide guidance to individual researchers on the correct and relevant data management and storage methodologies for that research field. It is recognised that there is an existing complex network of institution and funder-derived discipline repositories already in existence and that the UK research community must debate further how this data ecology is developed and resourced. Infrastructure should be seen as a shared responsibility across the research community, rather than falling just on research organisations.

Individual researchers should consider how they will manage the data they collect and generate at an early stage of conceptualising their research and take advice from relevant experts on best practice in their field. It is recognised though that there is also a need for more specific guidance in many disciplines to guide researchers and that learned societies may play a key role in developing relevant discipline specific guidance.

A properly considered and appropriate research data management plan should be in place before a specific research project begins so that no data is lost or stored inappropriately. Wherever possible, project plans should specify whether, when and how data will be made openly available. It is recognised that good data management explicitly implies that not all data is worth preserving but that researchers must exercise judgement under appropriate guidance.

The importance of training in research data management cannot be overstated as an enabler of open research data, and all researchers should receive such training at an early stage in their careers, along with subsequent updating as appropriate (see Principle 9 below).

Principle #7

Data curation is vital to make data useful for others and for long-term preservation of data

Data curation is the process of preparing data for use by others and long-term preservation. This can be achieved in a number of ways, such as through peer review, adherence to community-specific data formats and standards, deposition in specific repositories and through appropriate descriptions, or dedicated data articles in journal publications. As methodologies vary according to subject and disciplinary fields, data type and the circumstances of individual projects, the choice of methodology should not be mandated.

In most cases, research data can be made accessible via data repositories and web interfaces, provided these repositories are able to guarantee persistence of the datasets for a reasonable time period (see Principle 8). In many cases an appropriate accessible data summary or description – a landing page or dedicated data article – with sufficient metadata could be the gateway to access or facilitate a request for a specific data set.

It is clear that there must be reasonable bounds on the resources consumed in providing such metadata and indeed the degree of curation that do not place unreasonable demands on researchers, or their employers. In addition, appropriate policies governing the curation of physical samples, non-digital data and artefacts are not well developed at present. The broad role of learned societies in establishing discipline specific norms is seen as crucial.

It is envisaged that tools to discover data (e.g. specialised search tools and perhaps subject catalogues) and to integrate data with the peer-reviewed literature will develop further to help potential users locate relevant data. It is essential therefore that research data is made open with appropriate metadata, using open standards, in a manner that is consistent with the use of such tools.

Open research data should also be prepared in such a manner that it is as widely useable as is reasonably possible, at least for specialists in the same or linked fields. This also applies to the supporting metadata, which, where possible, should provide details of how the data were collected or generated and information on, for example, the processing and quality control applied.

Researchers are encouraged to store research data in non-proprietary formats, wherever possible. If this is not possible (or not cost-efficient), researchers should indicate what proprietary software is needed to process research data. Those requesting access to data are responsible for re-formatting it to suit their own research needs and for obtaining access to proprietary third party software that may be necessary to process the data.

Principle #8

Data supporting publications should be accessible by the publication date and should be in a citeable form.

One of the most important principles of research is that all published results should be assessable by others. Such assessments - which may be undertaken by reviewers and editors before publication, and by others post-publication – constitute a fundamental underpinning of the advancement of knowledge, as well as helping to guard against fraud. It has therefore long been the expectation that publications should include sufficient details for research to be tested and validated wherever possible. Ensuring that the findings reported in publications can be replicated and/or reproduced is often difficult to achieve in practice. But the aim of replicability is critically important since it facilitates the process whereby each researcher builds on the achievements of prior work and thus advances the whole research field.

In this spirit (and recognising the issues raised under Principle #2), it is vital that the data supporting and underlying published research findings should, as far as possible, be made open by the time the findings are published and be preserved for an appropriate period. This could be achieved by depositing and providing access to relevant data and associated software (where possible) via a repository owned or operated by a discipline-specific research community and its funding bodies, a publisher, a research institution, a subject association, a learned society, national deposit libraries or a commercial organisation; or via other mechanisms that provide appropriate and sustainable services. The dataset should be citable in itself, for example through the use of persistent identifiers, such as Digital Object Identifiers (DOIs) to ensure clarity of which exact dataset is under discussion or examination.

It is recognised however that in some disciplines there are very well established legitimate disciplinary norms (under the guidance of learned societies) that permit limited time delays in releasing data relating to initial publications. These arrangements might be reviewed in time as the culture of open research data becomes more established.

It is neither practical nor cost-effective to make all data open for an unlimited amount of time. Nevertheless, data underlying publications should be retained for 10 years from the date of any publication which fundamentally relies on the data, unless specified otherwise by the funder of the research.

It is important that open data is, in general, freely available, without for example payment or subscription requirements. It is recognised that in circumstances involving high costs of data preparation or transfer, for example with exceptionally large data sets, reasonable costs of data supply may be passed onto those requesting access.

Principle #9

Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders.

The development of open research data depends on the ability of all involved to understand their responsibilities and to optimise their own opportunities. It is clearly of little use making research data open if researchers in general lack appropriate data skills to make use of the opportunity. Underpinning this is recognition that curating, archiving, manipulating and analysing data requires a set of skills distinct from those utilised to collect, generate, or measure the data in the first place. In some cases, an individual researcher may well be capable of acquiring the necessary skills through self-directed learning, but for most, specialist tuition will be essential.

All stakeholders therefore have responsibilities to facilitate the development of appropriate data skills amongst the wider research community. For research institutions this should include the provision of researcher training opportunities provided in an organised and professional manner. It is imperative also that funding organisations, alongside research institutions, support the provision of such training through appropriate funding routes. Individual researchers must also ensure their own data skills are at a level sufficient to meet their own obligations whilst understanding the benefits to themselves of a higher level of understanding.

The specialised skills of data scientists are crucial in supporting the data management needs of researchers and institutions. Research institutions and funders should work together to help build under-pinning capacity and capability in this area, and to attract and retain such specialists by developing well-designed and sustainable career paths for them.

Principle #10

Regular reviews of progress towards open research data should be undertaken.

The journey towards open research data will require considerable efforts over the medium term. The importance of open research data is widely accepted but implementation is not straightforward. Progress will require the coordinated efforts by a number of actors and across a number of areas. The difficulties involved should not be underestimated and new issues will emerge as progress is made. There will also be developments internationally which will have an impact on UK policy and practice.

It is vital therefore that researchers, research organisations and funders remain committed to the development of open research data. This should be manifested in the undertaking of regular reviews that monitor progress and register issues to be addressed. Such reviews should not be over-burdensome but rather flexible and recognise that developments will take time. Their essence should be one of identifying and sharing best practice. This would be best achieved through engagement with community activities, such as the UK Open Data Forum, that bring together the full range of stakeholders.

Long-term commitment from all stakeholders will ensure the benefits of open research data are realized in practice through sensitive implementation and will help to secure the UK's position as an international research leader. This will be to the mutual advantage of all involved; providing a strong incentive to support open research data.

Annex 1: The Concordat Working Group

Rick Rylance – AHRC and RCUK

Duncan Wingham, NERC and RCUK

Nick Wright – Newcastle University

Rachel Bruce – Jisc

William Hammonds – Universities UK

Jamie Arrowsmith – Universities UK

Ben Johnson – HEFCE

Mark Thorley – NERC

Tim Jones – Warwick University

Michael Jubb – Research Information Network

Iain Hrynaszkiewicz – Springer Nature

Maja Maricevic – British Library

David Carr – Wellcome Trust

Matthew Woollard – Essex University

Tim Bradshaw – Russell Group

Annex 2: Useful References

Costs and benefits of sharing research data

Principle #1 - Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice.

Principle #3 - Open access to research data carries a significant cost, which should be respected by all parties.

The economic and scientific case for sharing research data is well made in a number of detailed reports. The Research Data Alliance's 2014 "The Data Harvest Report" is subtitled "Sharing data for knowledge, jobs and growth".

The introduction recommends that "We believe the storing, sharing and re-use of scientific data on a massive scale will stimulate great new sources of wealth. It turns data into a type of infrastructure, transforming the enterprise of science so anyone, anywhere, anytime can use and re-use data. It will mean new products and services, new companies and jobs. New trade flows will develop, and the competitiveness of nations will again be in play."
(https://rd-alliance.org/data-harvest-report-sharing-data-knowledge-jobs-and-growth.html)

In 2012, The Royal Society published "Science as an Open Enterprise", which focuses specifically on the scientific benefits of data sharing – in terms of the need for open enquiry as a cornerstone of scientific practice. (https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/)

The Knowledge Exchange (a cross-European partnership made up of national education technology agencies) produced a report in 2014 concerning the motivations of researchers and research groups in sharing data, covering multiple subjects across five countries. (http://www.knowledge-exchange.info/projects/project/research-data/sowing-the-seed)

The ERAC (European Research Area and Innovation Committee) report (2016) focuses on the key opportunities and challenges in sharing research data, alongside examining terminology and practice (https://era.gv.at/object/document/2402)

Further evidence from other perspectives is widely available, a 2012 post from the Open Knowledge Foundations provides a useful range of links detailing economic perspectives on open data (http://openeconomics.net/2012/10/03/the-benefits-of-open-data-evidence-from-economic-research/) and this video presentation from Stephen Gray of the Jisc funded CAIRO project details the benefits of sharing data in the live and performing arts (http://find.jorum.ac.uk/resources/10949/18273)

Specific work has been done on estimating the costs of curation – Neil Beagrie's work for Jisc adapts an established approach from the museums sector (http://beagrie.com/krds-i2s2.php) and the EC funded Collaboration to Clarify the Costs of Curation (http://www.4cproject.eu/) provides tools and guidance to support such analysis.

In some subject areas disciplinary-level data sharing collaborations have emerged – studies of the impact and value of a number of disciplinary data centres have taken place in recent years in the UK, and a summary and synthesis is available: Beagrie, N. and Houghton J.W. (2014) The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres,

Jisc.http://repository.jisc.ac.uk/5568/1/iDF308__Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

This recent post on the Efficiency Exchange details the savings realised by the ESRC supported UK Data Service (http://www.efficiencyexchange.ac.uk/7426/the-uk-data-service-best-practice-in-efficiency-effectiveness-and-value-for-money/)

Most major research funders are now mandating research data sharing as a condition of project funding, and providing detailed guidance alongside this. The UK research councils released expanded guidance in 2015 (http://www.rcuk.ac.uk/documents/documents/rcukcommonprinciplesondatapolicy-pdf/),

Jisc have developed guidance, and detailed case studies, in support of the EPSRC policy (https://www.jisc.ac.uk/guides/meeting-the-requirements-of-the-EPSRC-research-data-policy) which sits alongside EPSRCs own guidance (https://www.epsrc.ac.uk/about/standards/researchdata/), and there is guidance from the Wellcome Trust (https://wellcome.ac.uk/funding/managing-grant/developing-data-management-and-sharing-plan) and the European Union (Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

The Jisc-supported Digital Curation Centre offers guidance on making the case for research data management at an institutional level (http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm) alongside a wealth of advice, guidance and support for the practice of RDM (http://www.dcc.ac.uk)

Jisc's Research Data Network (http://researchdata.network) supports discussion and debate around new developments in the field of research data. In particular it has had a focus on the planned Jisc Research Data Shared Service, but aims to capture wider emerging practice. working on a toolkit that will cover these aspects as we develop latest practice through the next 24 months.

Data sharing, ethics and the law

Principle #2 - There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable.

Principle #4 - The right of the creators of research data to reasonable first use is recognised.

Principle #5 - Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.

Sharing research data poses ethical and legal questions, particularly around consent and data protection, and around the exploitation and ownership of intellectual property.

Jisc offers guidance on data protection for research data (https://www.jisc.ac.uk/guides/data-protection-and-research-data), with the ESRC's "Ethics Guidebook" situating the issue within institutional ethical approval processes. (http://www.data-archive.ac.uk/create-manage/consent-ethics). Medical data presents specific issues, detailed within a Digital Curation Centre Briefing paper (http://www.dcc.ac.uk/resources/briefing-papers/legal-watch-papers/sharing-medical-data) and the Medical Research Council offers guidance specifically around sharing data from patient and population studies (http://www.mrc.ac.uk/research/policies-and-resources-for-mrc-researchers/data-sharing/data-sharing-population-and-patient-studies/)

Guidance on "first use" and other sharing restrictions based around intellectual property is usually subject specific, and can be found within funders guidance, normally an embargo period of up to three years can be set.

Good practice in sharing research data

Principle #6 - Good data management is fundamental to all stages of the research process and should be established at the outset.

Principle #9 - Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders.

Principle #10 - Regular reviews of progress towards open research data should be undertaken.

There are a number of useful sets of training and educational material for research data management available under an open licence for reuse. These include:

- The CARDIO materials, allowing institutions to assess their data management needs and gaps in provision: http://cardio.dcc.ac.uk

- The MANTRA materials, aimed at researchers: http://datalib.edina.ac.uk/mantra/

- The RDMROSE materials, aimed at librarians and information professionals: http://rdmrose.group.shef.ac.uk

- The DMPONLINE tool, specifically supporting the development of project data management plans: https://dmponline.dcc.ac.uk

A Jisc "short guide" provides an overview for researchers: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data

Training is often provided at an institutional level – reports from the Knowledge Exchange (http://www.knowledge-exchange.info/event/rdm-training) and Universities UK (http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/data-skills-training-in-english-universities.aspx) examine ways in which this is delivered.

The "Directions for Research Data Management in UK Universities" report (published in 2015 by ARMA, RLU, RUGIT, SCONUL, UCISA and Jisc) (http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf) contains detailed recommendations around skills and systems required for institutional RDM implementation.

The Re3data supports researchers and research managers by offering a list of more than 1,500 research data repositories (http://www.re3data.org).

Metadata and curation

Principle #7 - Data curation is vital to make data useful for others and for long-term preservation of data

Principle #8 - Data supporting publications should be accessible by the publication date and should be in a citeable form.

The FAIR principles – research data should be findable, accessible, interoperable and reuseable – are enumerated and expanded in this Scientific Data comment by Barend Mons (et al) (http://www.nature.com/articles/sdata201618). These principles outline the technical affordances required to ensure that the maximum benefit from data reuse can be realised.

The DCC's "Data Asset Framework" supports institutions in making curation decisions around research data http://www.data-audit.eu.

It is widely agreed that research data should be uniquely identified with a persistent identifier. The British Library offers guidance on digital object identifiers (DOIs) for research data, and runs the UK "datacite" service. (http://www.bl.uk/aboutus/stratpolprog/digi/datasets/WorkingWithDataCite_2013.pdf)

Similarly, the reliable identification of individual researchers is facilitated via a UK-wide ORCID consortium supported by Jisc (https://www.jisc.ac.uk/orcid)

Jisc is working on a common metadata profile with UK higher education institutions to aid in the discovery of research data. (https://rdds.jiscinvolve.org/wp/2016/03/11/core_metadata_profile/ https://rdds.jiscinvolve.org/wp/2016/03/18/how-much-metadata-is-enough/)
This builds on best practice and common metadata profiles but extends them to deal with administrative aspects.

Publishers are also focusing on developing policies and specifications for research data published alongside articles. The Data Citation Implementation Pilot (DCIP) publisher early adopters group is developing a roadmap to inform policy development (https://www.force11.org/group/dcip/eg3publisherearlyadopters), Springer Nature have standardised journal policies using four templates (http://www.springernature.com/gp/group/data-policy)
http://www.esrc.ac.uk/funding/guidance-for-grant-holders/research-data-policy/
http://www.nerc.ac.uk/research/sites/data/

HIGHER EDUCATION **hefce**
FUNDING COUNCIL FOR ENGLAND

RESEARCH
COUNCILS UK

Universities UK

wellcome