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.

Foreword

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, tr anscripts 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

- 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 p2on p2.6(ned6(ea)10.5(r)-5.9(l)2.6(y)8.9( )on p2)2d ftirati

Complements existing frameworks - Extensive statutory and regulatory standards already exist to

Principle #1

Open access to research data is an enabler of high qualityresearch, 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 bcanc.8(t)76.6(A

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

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

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.2(ed)0.a evuo0ow3.5(i)2J 0.00t re w l-11.2(a)10.5(t)-6.6(euo55l)2.6o(r)6.9(m)-3.9(s)ou.2( )]TJ

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

Principle #10

Regular review 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)

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

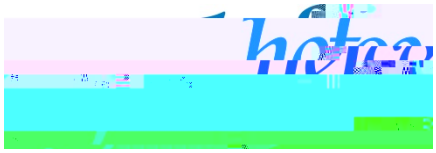
Principle #6 - Good data management is fundamental to all stages of the research process and should be established at the outset.

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: htt10. cs10.5(m)-5.1EMC /P <</MCID53csapsd0.5(mi)95(mo /P .d /P c.5(5(c)158.5./MCID51362(sc.5(4(.



Universities UK