Realising Open Data Principles in UK Research Institutions

Key insights from the STAR (**S**ustainable & **TrA**nsparent **R**esearch data) project, led by the UK Reproducibility Network (UKRN)

Pen-Yuan Hsing
University of Bristol,
United Kingdom

Jessica Wheeler
University of Bristol,
United Kingdom

Lorna Duncan
University of Bristol,
United Kingdom

Rosalind Strang
University of Bristol,
United Kingdom

Neil Jacobs
UK Reproducibility Network

Abstract

We report on the state of open research data (ORD) policy and practice across UK research institutions through the STAR (**S**ustainable & **TrA**nsparent **R**esearch data) project. Through qualitative interviews, focus groups, and workshops involving 52 university staff across 21 UK institutions, we investigated the progress and challenges in ORD practices since the 2016 publication of the Concordat on Open Research Data.

We observed that while institutions have made progress establishing ORD specialist roles, developing policies, and creating repository infrastructures, systematic monitoring processes and widespread adoption remain stalled. Key challenges include capacity constraints in institutional repositories, limited workload recognition, insufficient funding for long-term archiving, and varying disciplinary interpretations of ORD relevance.

Based on workshops with participants, we recommend recognition of ORD in academic career frameworks, development of disciplinary-relevant data sharing practices, improved infrastructure for monitoring ORD practices, and enhanced support for external disciplinary repositories. The study emphasises the need for a values-driven rather than compliance-driven approach to ORD implementation, calling for deeper engagement with diverse academic communities to ensure ORD requirements remain meaningful and relevant across disciplines. These findings provide insights for research institutions and funding bodies in developing more effective and inclusive ORD policies.

Introduction

Open research is increasingly considered a key component of research practice on a global level.[1] This is underpinned by transparency and access concerns and values, including around equity and oversight; systemic power imbalances; and accumulative quality, progress, and cost efficiency in research learning (e.g., Gwinn, 2006; Peters, 2014). Open research applies across the research lifecycle, from design to final outcomes, making it possible for wide audiences of researchers and stakeholders, now and in the future, to engage with and utilise diverse research learnings, far beyond the relatively small extractions filtered into traditional peer-reviewed publications. It is also commonly held that open research should be FAIR ('findable', 'accessible', 'interoperable', 'reusable').[2]

Recognising its national and international importance, several multi-stakeholder working groups have reported in recent years on the implementation of open research in the United Kingdom (UK). One such initiative is the Concordat on Open Research Data (the Concordat).[3] The Concordat, published in 2016 and signed by Universities UK, sets out ten open research data (ORD) principles to support the UK research community to make the research data they generate openly available for use by others, wherever possible. The principles aim to be comprehensive and actionable, relating to ethical, legal and professional obligations, addressing areas such as practicality, affordability, transparency, robustness and fairness, mechanisms and infrastructure, data integrity, citation and credit acknowledgment. Recognising concerns in the use and storage of human-subject data, and those relating to commercially driven data, the Concordat set out a now well-established approach to open research practice, aiming for all research data to be 'as open as possible, as closed as necessary'.

The STAR Project

The STAR (**S**ustainable & T**rA**nsparent **R**esearch data) project is led by the UK Reproducibility Network (UKRN) and supported by several UK research institutions, with contributions from the Data Curation Centre and other experts.[4] It uses qualitative methods to evaluate the implementation of the Concordat's principles in UK research institutions since 2016. It set out to reveal progress and barriers to ORD; and explores ways in which institutions are and could better be supported in the curation, publication, and reuse of ORD.

Methods

Throughout 2024, we held interviews, focus groups, and workshops with 52 university staff developing and implementing ORD policy. They include those employed in research institution professional services; directors, managers and service providers in research and ethics services, library services, and information technology (IT) support services;

---

[1] Such as the UNESCO Recommendation on Open Science (https://www.unesco.org/en/open-science) or the European Union Horizon 2020 programme (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

[2] E.g., as described here: https://www5.open.ac.uk/library-research-support/research-data-management/fair-principles

[3] Published here: https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf

[4] Project webpage: https://www.ukrn.org/activities/star/

dedicated open data support services; and research institution leaders, such as Pro-Vice Chancellors for Research and UKRN research partners.

The interviewees represented 21 universities, selected to be inclusive of diverse UK locations, research funding breadths and scales, research community size, and disciplinary focus, including traditional academic disciplines, vocational/professional and arts orientations, and institutions with more focused, as well as those with broader-ranging and multiple, disciplinary remits.

To provide structure for thinking about ORD practices during the interviews, our topic guide and guiding questions were built on 24 concrete tasks derived by the Digital Curation Centre from the ten principles of open data in the Concordat (UK Reproducibility Network, 2024).

Two workshops in late 2024 focused on outputs and dissemination, and co-produced (with project partners and participants) recommendations for tackling challenges as reported below.

This paper offers an overview of current findings after the conclusion of the project in December 2024, following an assessment of Concordat-related institutional tasks, analysis of qualitative interviews, site visits, institutional policy reviews and workshop discussions.

Progress in the Open Research Data Journey

Areas of Significant Progress

Despite the diversity of institutions, within which the Concordat had been explicitly adopted in some but not all cases, participants reported similar areas of substantial ORD progress. ORD specialists were typically in post (or in partial ORD post) and providing expert professional ORD service support. ORD workers were supporting the development of ORD policies, which institutions had approved or were in the process of approving, sometimes informed by the Concordat. They were supporting the creation, selection and development of bespoke repository infrastructures, ensuring institutional research community access to online data repositories (e.g., Figshare) or bespoke institutional data stores. Some had undertaken substantial work transferring institutionally held ORD to new, better recognised and more easily creditable repository formats. ORD workers routinely provide qualified technical support to curate and upload ORD, supporting data management plans and ethics applications to include considered ORD elements. ORD workers often deliver training to diverse researcher communities in ORD practices. ORD practices were regularly described as standard elements of postgraduate training programmes, with interviewees noting that postgraduate students and early career researchers more often attend training events.

Typically, there were recognised pockets of excellence, researchers and groups who could be characterised as ORD active (having actively sought to publish ORD) or as pioneers (being involved in early ORD initiatives).

There were also, occasionally, established disciplinary strongholds. In these cases, there is advanced discipline-based normalisation of ORD practices, often via discipline-based repositories and with less reliance on institutional ORD support.

Areas of Stalled Progress

In contrast, participants reported that ORD practices were typically more stalled in relation to the implementation of systematic ORD monitoring and recording processes.

A lack of sufficiently detailed routine evidence and monitoring of ORD progress was consistently recognised by participants, both in relation to the extent to which institutional research communities were making and using ORD. Similarly, institutions

were only beginning to recognise and promote ORD successes and did not yet consistently recognise researchers' ORD practices in ways that would support career progression.

While high-quality ORD infrastructure and expert professional service support was offered across a diverse range of sampled institutions, and with more researchers publishing open data via institutional repositories, looking across the whole research community in any single institution, there was a sense that ORD practices remained exceptional, rather than normalised.

Institutional ORD 'early embedded, relatively small-scale uptake'

In terms of the current scale of ORD outputs and disciplinary breadth, institutional ORD practices could be characterised as 'early embedded with relatively small-scale uptake', with institutional buy-in, support service provision, policy and infrastructure development, but with limited uptake. Within the diverse sampled UK research institutions, ORD practice was taking hold, but not normative across disciplines.

Drivers and Blocks

ORD practices were most often driven by journal publication requirements and discussed as part of data management plans where these were required in funding applications. Tangible benefits (such as ORD reuse benefits) were not yet widely reported or experienced by researchers (outside of distinct disciplines where ORD practices are already normalised, with active ORD reuse programmes and dedicated repositories). There was concern that ORD commitments, made as part of funded research applications, were not yet consistently met, and that lack of institutional or funder monitoring reduced the impetus among research teams to dedicate time and resources to ORD, deprioritising it in a list of already challenging project output requirements. Meanwhile, institutional reputations, research cultures, resources and workloads continue to galvanise around established REF criteria (the Research Excellence Framework, a UK-wide assessment of university research which informs block grant funding), funding attainments and high-impact peer-review journal publications.

Key Issues for the Future Progress of ORD within UK Research Institutions

Scaling (or Capacity) Issues

While STAR project participants actively engage in training and increasing ORD knowledge and practice within their institutions, they recognised that wider ORD uptake could create capacity issues. Current institutional infrastructures, while impressive in terms of expert support, accessibility and functionality of ORD repositories, tended to have considerable capacity restrictions.

Sometimes, there were already examples of research teams publishing ORD, and finding they could not do so via institutional repositories or needed to dramatically reconsider and reduce the amount of data to publish, because of limits on storage capacity and inadequate within-project funding to support ORD plans. In such cases, the storage volume of institutional repositories and the costs of long-term storage were acute issues.

Bottlenecks: Future Data Extent and Quality Considerations

In the longer-term, both institutional repository capacities and numbers of expert staff supporting access may be bottlenecks to broad ORD adoption. This gives the impression that current infrastructures can serve only a limited scale or ambition of ORD practice. Future institutional ORD guidance will require more thorough considerations of the extent, quality, curation and costs of ORD as part of research project planning and in relation to desired outputs.

The Need for Within-discipline 'Data Stewards'

Establishing peer-led, within-discipline ORD 'data stewards', guiding and supporting ORD practices, including appropriate curation, based within diverse research communities, and separate from professional service support, was seen as a next step for universities. Devolving the ORD data knowledge- and skill-sharing role beyond professional services was deemed important, partly to prevent overwhelming them, but also to support disciplinary engagement to ensure that ORD practices and sensibilities are shaped to be relevant to diverse research communities.

The Need for External Repositories

The need for discipline-centric, funded and supported national and international repositories, as already seen in some areas of ORD excellence (e.g., astrophysics, biology), was recognised as an important component relating to the future scaling of ORD practice. Questions arose about the potential role and need for additional national ORD infrastructures, including externally funded repositories able to provide long-term ORD capacity and support. This is particularly important for institutions with smaller research support budgets, where their ability to pay for hosting, or subscribing to, an institutional data repository is limited with uncertain sustainability.

Financial Climate

Future institutional ORD repository and support service resourcing concerns were heightened by financial challenges, with pressures on institutions perceived to impact the workloads and capacity of those in teaching, research and support service roles. There was sentiment that the financial pressures ahead were likely to be even more demanding and posed a potential threat to ORD ambitions. Interviewees suggested that existing financial support (internal and externally-sourced) for institutional data infrastructure and support should be inventoried, and if appropriate, ringfenced for long-term data archival and preservation purposes.

Workload Recognition and Resourcing Issues

There were challenges involving unclear delineation and delegation of data curation responsibilities. Some data specialists are already sometimes overwhelmed performing ORD data preparation work for and with academics. This support is provided in recognition that many researchers currently lack appropriate skills, need additional support and may struggle to dedicate time to perform time-consuming ORD curation. Personalised support is sometimes offered by our interviewees to academics, especially to those needing to demonstrate appropriately publishing data in recognised repositories, to meet publication or grant requirements. It is possible that much more of this work is being performed by support service workers than is likely sustainable and scalable.

Despite this workload, some interviewees felt a sense of isolation within their institution's wider administrative hierarchy. In other words, they perceive that ORD

concerns often do not have a 'seat at the table' in senior-level decision making. In a few cases, data support workers must stress the risks of non-compliance with data management or safeguarding regulations to insert broader ORD considerations into the university research agenda.

In addition to concerns shared by professional staff who are ORD support workers, our interviews also revealed that ORD practices are rarely recognised in the hiring and promotion of academic staff. Due to this, and the perception that funding bodies do not have strong data-sharing expectations for funded projects, our interviewees described a lack of incentive for their institutional academic communities to deploy already-stretched time and labour to the curation, sharing, and reuse of ORD.

Funding of ORD Resources

There was recognition that funding bodies do not yet explicitly require, and funding applications do not confidently request, sufficient funds within application budget projections to support ORD practices and long-term data archiving.

Overall, the skills development and time/effort required to effectively curate and store open data is not sufficiently recognised or costed, despite data management plans. It is unclear that institutional finance services are yet able to accurately estimate projected ORD costs.

Monitoring Activities

Lack of monitoring (by funders as well as institutions) of ORD implementation following research project completion reduces the weight placed on ORD project commitments, especially where these have also been insufficiently costed into project proposals and so may require 'hidden' and unrecognised professional research labour to complete.

Those with ORD support roles reported some technical monitoring difficulties. Commonly reported methods, such as tracking peer-reviewed publications by university researchers, typically involve labour-intensive manual checks for associated data availability statements to find open research datasets, and for compliance with project data management plans and funder requirements.

Furthermore, publications are often produced at the end of a project, by this stage there is less likely to be project capacity to address outstanding ORD issues, again relegating ORD work to 'hidden' workloads.

ORD Reuse and the Intrinsic Perceived Value of ORD

Lack of widespread reuse of ORD in many fields of research reduces the intrinsic perceived value of ORD work. Those we interviewed cite two major points of interaction with academic researchers: during proposal development when a data management plan is required by funders, and at the end of a project when preparing datasets for publication to meet that requirement. During these interactions, ORD data management practices are focused on compliance, rather than the potential impacts of data reuse. As such the inherent value of ORD is rarely the focus. Recognition and celebration of the successful reuse of ORD, or impacts of ORD reuse, is currently limited, whether institutionally or by major funders, such as UK Research and Innovation (UKRI, the primary public research funding body in the UK).

Making ORD Meaningful

The meaning and relevance of 'data' and following from this, of 'open data', varies greatly across fields of research. To make ORD practices inclusive currently requires some

challenging 'translation work', which can be intellectually stimulating (e.g., exploring interpretations of the meaning and value of 'data' with a sculptor or composer), but also a time-consuming and frustrating conundrum (e.g., when the question of meaning is framed more as 'what does this have to do with me and my work?'), as shared by those performing ORD support service roles.

The current language of ORD is not felt to be inclusive. ORD guidance lacks an appropriate diversity of case studies and examples. The case was repeatedly made by participants for a better range of examples of effective and diverse ORD practices, across different research traditions and fields. There is a need to engage with different conceptualisations of 'repositories' and of 'data' (for instance, those not limited to digital formats, such as 'archives', 'galleries', and 'artefacts'). ORD guidance could also helpfully use case studies to demonstrate a range of meaningful impacts of ORD reuse.

There is a need within institutions for engagement among those disciplinary communities who are less likely to self-evidently view ORD as relevant to their work. This is important because future funding may well be shaped by more tightly specified future ORD requirements and regulations. Broader-based engagement now could allow for more values-driven and discipline-relevant shaping of the specification of ORD practices and requirements.

Compliance-driven vs Value-driven ORD?

There is anticipation that ORD practices will gain prominence in future research assessments, which could be during peer review, grant applications, or the next REF in 2029. Our participants welcome the potential for recognition of their institutions ORD practices. They also welcomed the push a stronger REF steer would bring, within their institution and beyond, to progress ORD practices.

However, this welcome was cautious and was accompanied by apprehension that the adopted metrics for assessment of 'ORD practice excellence', without prior and more inclusive disciplinary involvement, will not capture what might be meaningful or useful ORD practice in different disciplines with differing epistemologies. They fear the creation of burdensome, bureaucratic and for some researchers meaningless, compliance-driven, rather than values-driven, approaches to ORD practice.

Key Recommendations

The challenges we discussed were the focus of two online workshops with project participants in July and October 2024. During these events, participants made the following recommendations to place ORD on the agenda for future research policy—for institutions and funding bodies—in the UK:

- The curation, sharing, and reuse of ORD should be formally recognised in university hiring and promotion frameworks.

- Methods for formally costing data labour throughout the research lifecycle should be developed and be explicitly expected by funders in grant applications.

- Methods for deciding what data to share and how should be developed within disciplinary communities.

- The value of reusing ORD should be stressed in research training and considered in research assessments.

- Deeper engagement with academic communities representing diverse disciplines—including, but not limited to, arts and humanities subject areas—must

be conducted so that any future expectations and rules on ORD do not benefit some fields of research while becoming burdensome and irrelevant for others.

- Interoperable infrastructure is needed for easier monitoring of ORD practices on an institutional scale. For example, this could include machine-readable, structured data availabilities statements for research outputs that could be aggregated, replacing the manual work required now.

- Funding bodies should follow up with data management plans in grant applications through reviews or audits with the goal of encouraging ORD practice rather than being punitive.

- In addition to institutional repositories, external disciplinary data repositories should be supported and recognised as another venue for publishing ORD. This is especially important for institutions with limited resources which cannot maintain their own repositories.

- Disciplinary data stewards should be defined and designated among an institution's faculty and departments.

- Data support workers among the professional staff in institutions should be given a voice in research policy decision making and connected with groups with relevant remit, such as research ethics and integrity teams.

- Existing funding for data support and infrastructure should be inventoried and ringfenced as appropriate.

- ORD requirements should account for smaller institutions with limited support capacity, and the rollout of such requirements must be gradual so that institutional support capacity has ample time to adjust and scale.

Implementing open research data practices in UK research institutions entails complex transformations of research culture and practice. While foundational infrastructure and support services are now largely in place, widespread adoption requires overcoming considerable challenges around capacity, resources, and disciplinary relevance. Moving forward, success will depend on shifting from compliance-driven approaches toward value-driven engagement that recognises disciplinary diversity, provides adequate resources for data stewardship, and creates meaningful incentives for researchers and data support workers.

Acknowledgements

References

Gwinn, M. (2006). The Access Principle: The Case for Open Access to Research and Scholarship. *Emerging Infectious Diseases* 12(9), 1473. https://doi.org/10.3201/eid1209.060643

Peters, M. A. (2014). Open Science, Philosophy and Peer Review. *Educational Philosophy and Theory* 46(3), 215–219. https://doi.org/10.1080/00131857.2013.781296

UK Reproducibility Network, Jacobs, N., McCutcheon, V., Horton, L., Lister, A., Davidson, J., … Donaldson, M. (2024, February 21). UKRN Institutional tasks implied by the UK Open Research Data Concordat. https://doi.org/10.31219/osf.io/7ew63