IJDC | *Conference Paper*

# Making Reproducibility a Reality by 2035? Enabling Publisher Collaboration for Enhanced Data Policy Enforcement

Rebecca Taylor-Grant
Taylor & Francis

Matthew Cannon
Taylor & Francis

Allyson Lister
University of Oxford

Susanna-Assunta Sansone
University of Oxford

## Abstract

This paper describes a project which identified practical and pragmatic ways to increase the FAIRness and reproducibility of published research. Academic journals have supported Open Science through the implementation of data sharing policies for over ten years; some evidence has since emerged on the additional time, resources and expertise that policy enforcement requires as part of an editorial workflow. A series of publisher workshops facilitated by the EC-funded TIER2 project aimed to identify the key checks needed to enforce strengthened journal data sharing policies and to understand which editorial roles have the capacity to undertake such enforcement. The intended outcome of this work was to establish the workflows and resourcing which can support academic journals to enforce stronger data sharing policies in future.

*Submitted* 31 January 2025 ~ *Accepted* 20 February 2025

Correspondence should be addressed to Rebecca Taylor-Grant, Email: Rebecca.taylorgrant@tandf.co.uk

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

# Introduction

While academic publishers are well placed to impact on the reproducibility of published research, adoption of new journal data sharing policies has slowed in recent years. Although the concept of research data sharing is widely supported, the existing evidence for the downstream value (for example data reuse or cost savings) may not be sufficient to motivate change. This creates challenges for stakeholders in reproducibility and data sharing, as supporting such activities can be resource intensive. Publishers, for example, must embed expertise within complex journal publishing workflows to support (and enforce) compliance with relevant Open Research policies. This paper describes a project which used a series of publisher workshops to develop improved approaches to data policy enforcement. By acknowledging the variability and complexity of academic publishing workflows, the project outcomes will help to reduce one of the key barriers to increased data policy implementation on academic journals.

## Assessing the Impact of Data Policies in the Academic Publishing Workflow

Although some academic journals have enforced data sharing policies for decades, an analysis of the journal data policy landscape conducted in 2012-14 concluded that the policy ecosystem was in "critical need of standardisation and harmonisation" (Naughton & Kernohan, 2016). In 2018, acknowledging the lack of consistency in journal data policy requirements and the potential negative impacts on authors and other stakeholders, the Data Policy Standardisation and Implementation Interest Group was formed at the Research Data Alliance (RDA).[1] This group published a framework of data policy requirements in 2020, which can be used to develop standardised data policies for all publishers and journals (Hrynaszkiewicz et al., 2020). Many academic publishers now implement journal data policies which align with a set of consistent data sharing requirements, for example Wiley,[2] Taylor & Francis[3] and Elsevier.[4] Journal data policy types have also been catalogued and made citeable on FAIRsharing.org[5], and as part of the work of the RDA FAIRsharing Working Group these policy descriptions have been enriched to improve their comparability.[6]

As journal data policy implementation has increased and become more consistent, some analysis has been conducted on the impact of such policies on the editorial workflow. For example, the Nature group of journals published an assessment of the editorial resource needed to check a data availability statement as a component of a submitted manuscript and evaluated the additional time and resource required to check for appropriate use of data repositories (Grant & Hrynaszkiewicz, 2018). Similar assessments of the impact of data policies on publishing workflows have been undertaken by other publishers including IOP (Holt, Walker & Jones, 2021) and Taylor & Francis (Cannon, Kelly & Freeman, 2022). Although these studies provide valuable analysis of the resources needed to implement data sharing policies, their focus has been on checks for the presence or absence of data availability statements and in some cases, data repository links. Open Data policies and the higher levels of the Research Data Alliance policy framework have many additional requirements however, including data citation, the use of open file formats for data, and application of open licences. These additional checks are critical to

[1] Research Data Policy Interest Group (formerly the Data Policy Standardisation and Implementation Group): https://www.rd-alliance.org/groups/research-data-policy-ig

[2] Wiley's Data Sharing Policies: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html

[3] Understanding our data sharing policies: https://authorservices.taylorandfrancis.com/data-sharing-policies

[4] Research data guidelines: https://www.elsevier.com/en-gb/researcher/author/tools-and-resources/research-data/data-guidelines

[5] FAIRsharing.org Search: https://fairsharing.org/search?recordType=journal

[6] Community Champions complete update on the FAIRsharing policy registry: https://blog.fairsharing.org/?p=765

supporting the publication of FAIR and reproducible research, as they are necessary to ensure that data is persistently accessible and reusable. To date, there has been a gap in the published literature assessing the resource needed to conduct more specific and detailed checks on data sharing, what kind of training or guidance might be needed, and when such checks should occur in a publishing workflow. Such evidence is necessary to allow publishers to understand the impact of strengthened data policies and to plan for their implementation and enforcement. There is evidence that the implementation of stronger data policies at journals has stalled in recent years, despite initiatives such as the publishing industry's STM Association's Research Data Program,[7] which established that only 52% of journals had a data sharing policy by 2020 (Smit & van Rossum, 2022. This may be a consequence of the lack of detailed guidance on the implementation and impacts of journal data policies.

### The TIER2 Project: Enhancing Trust, Integrity and Efficiency in Research Through Next-level Reproducibility

TIER2 is a Horizon Europe-funded research project which aims to increase reproducibility of scientific research results through a programme of pilot activities.[8] The academic publisher Taylor & Francis is an affiliate partner of the project and, with funded partner the University of Oxford, co-leads Pilot 8, "*An Editorial Reference Handbook for Reproducibility and FAIRness*"[9] (hereafter referred to as the *Handbook).* A first version of the *Handbook* was released in July 2024[10], and work on this pilot is set to be completed by December 2025. The development of a supporting workflow for the *Handbook*, described in this paper, demonstrates a collaborative approach across academic publishers to identify practical ways to increase the implementation of more stringent data sharing policies and support the publication of reproducible research. Through this work, the *Handbook* will provide publisher stakeholders with the tools needed to implement and enforce Open Data policies in the next ten years and beyond.

## FAIRness and Reproducibility Checks in the Publishing Workflow: A Workshop Series

Pilot 8 was initiated with a publisher workshop held in May 2023 which aimed to elicit feedback from publisher representatives on how the reproducibility of published research could be enhanced (Sansone et al., 2023). The workshop included representatives of publishers from Cambridge University Press, Cell Press, EMBO Press, F1000, Taylor & Francis, GigaScience Press, The Lancet, Oxford University Press, PLoS, Springer Nature and Wiley. The workshop identified key challenges to increasing the publication of reproducible research, including the cost and resource implications, the need for training for researchers and editorial teams, and difficulties in finding appropriate technical solutions that work across journals. The development of the *Handbook* was identified as a priority for the Pilot. The *Handbook's* scope includes a practical set of checks based on a harmonised version of existing reproducibility and FAIR-enabled community checklists, and a general framework to support improvement of internal editorial processes. These checks further develop the initial set of recommendations for journal data policies established via the RDA and incorporate elements of other relevant guidance resulting from the collaboration of FAIRsharing with the Digital Curation Centre (DCC)[11] and the UK Reproducibility Network (UKRN).[12] Sources which were reviewed and integrated into the

[7] The STM Research Data Program: https://www.stm-assoc.org/research-data-program
[8] TIER2 Pilots: https://tier2-project.eu/pilots
[9] Pilot 8 - An Editorial Reference Handbook for Reproducibility and FAIRness: https://tier2-project.eu/pilots/8
[10] TIER2 Editorial Reference Handbook: https://publishers.fairassist.org
[11] FAIRsharing and DCC collaborate to align policy metadata: https://blog.fairsharing.org/?p=451
[12] Policy enrichment initiative: new mapping to Concordat on Open Research Data – A collaboration with UKRN: https://blog.fairsharing.org/?p=691

checklist included Promoting Reusable and Open Methods and Protocols (PRO-MaP); The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences; FAIR4RS; ARRIVE Guidelines and editorial checklists from the Nature portfolio, F1000 and GigaScience.[13]

The manuscript checks proposed by the *Handbook* are as follows:

1. Are the availability statements for relevant digital objects present?
2. Are all digital objects and their contents clearly and correctly listed within the appropriate availability statement(s)?
3. How many digital objects are present across all availability statements?
4. Is the identifier provided for this digital object valid and recognised?
5. Is the licence for the digital object allowed by your journal?
6. Is the digital object openly available? If not, are there clearly stated and valid ethical or data protection reasons for access to be controlled?
7. If access is controlled, is the digital object available to peer reviewers?
8. Has the digital object been deposited in an appropriate repository recognised by your journal?
9. Has the digital object been anonymised if necessary?
10. Where applicable, is there evidence that the research has been approved by a specific, recognised committee?
11. Where applicable, has an appropriate domain-specific metadata format been used?
12. Is the accompanying metadata complete according to format requirements or community best practices?

Through initial discussions on the *Handbook*, the publishers reached consensus on the value and relevance of these checks and their potential to improve data sharing across journals. However, it was clear that workflows, resources, submission systems and policy requirements vary across academic publishers and that the *Handbook* therefore needed to acknowledge different publishers' needs. Addressing the practical aspects of implementation was key to ensuring that the *Handbook* could be successfully introduced by the maximum number of stakeholders. To address this, a series of collaborative publisher workshops was organised by the Pilot co-chairs[14] to identify which roles most commonly support editorial policy enforcement in publishing workflows and to reach consensus on who should conduct each data sharing check and when.

The process of defining a standard publication workflow, establishing a logical order for the data checks and proposing appropriate roles to undertake each check is described below. In undertaking this collaborative work, it was agreed that the contribution of each publisher should be anonymous, and that all outputs from this activity would be reviewed before being shared publicly in case they could be leveraged by "bad actors" to support paper mill activity or other manipulation of the publishing process (Alam, 2024). The final output of the workshops was the creation of a generic publishing workflow and set of standardised roles which can be used to make the *Handbook* implementable by any publisher or journal.

[13] A full list of sources including links is available here: https://publishers.fairassist.org/context-credits-and-references#included-sources.

[14] A list of the Pilot co-leads is available here: https://publishers.fairassist.org/context-credits-and-references.

## Who Can Check for FAIRness and Reproducibility? Identifying Publisher Roles

The first workshop task was to identify which roles or teams were most commonly present in academic journal publishing workflows and which tasks and skillsets were associated with them. Academic journals and their publishers vary in size, business models and staffing. While there is some consistency in the stages that a manuscript passes through from submission to publication, the systems and resources which support these workflows can differ widely across journals. The Scholarly Kitchen blog captured an extensive list of the tasks associated with journal publishing[15] but not all publishers will undertake each task, and how they are achieved is not consistent, even across journals within a single publisher's portfolio. Some tasks can be achieved using technology (for example manuscript submission systems), some by journal staff (whether paid or voluntary), some via the peer review process, and there is potential for newly emerging AI tools to address tasks such as image manipulation and plagiarism checks. The variance in approach and resourcing at an individual journal makes it more challenging to develop guidance like the *Handbook* which is relevant and useful across a range of journal publishers.

The workshop therefore intended to assess which roles were most prevalent across the publishers and gain consensus on those which were most likely to be present in a journal's workflow (and therefore available to conduct data policy compliance checks). To guide this activity, the workshop co-chairs brainstormed a list of roles ahead of time. During the workshop, participants used dot voting on the collaborative software Miro to identify which roles were most or least commonly present on their journal portfolios and gave feedback on the skills and tasks associated with each (Table 1). The publishers also had the opportunity to suggest additional roles which they believed were missing from the list.

**Table 1**. Prevalence of roles within the editorial publishing workflow.

| Role | Skills/tasks | Order of prevalence across publishers (most prevalent to least prevalent) |
|---|---|---|
| Production Editor | Conducting or overseeing the copyediting and typesetting process; involved in article metadata creation and curation | =1 |
| Peer Reviewer | External to journal team, subject expert, aiming to validate and/or improve manuscript quality | =1 |
| Editorial Office | May be internal or external to journal team; connected to the journal or publisher; may have PhD or subject expertise | =3 |
| Academic Editor | External to journal team, subject expert, journal leader, focused on article quality | =3 |
| Typesetter | Internal, involved in article metadata creation including data citations | =3 |
| Administrator | Often external, connected to many journals, may not be a subject expert | 6 |

[15] Focusing on Value – 102 Things Journal Publishers Do: https://scholarlykitchen.sspnet.org/2018/02/06/focusing-value-102-things-journal-publishers-2018-update/

| | | |
|---|---|---|
| Internal Editor | Internal journal team member, subject expert, may act as alternative to external academic editor role | 7 |
| AI (tools) | An Artificial Intelligence tool which may be used for checks or validation of manuscript components | 8 |
| API (tools) | Application Programming Interface which may be used to validate manuscript components against external data sources (e.g. PIDs) | 9 |
| Data curator | Internal or external, focused on dataset quality and validation | 10 |
| Academic Data Editor | Part of the academic editorial team, supporting authors with data sharing | 11 |

The most prevalent roles across publishers were Production Editors and Peer Reviewers, while fewer publishers had data-focused roles like Academic Data Editors or Data Curators available or used tools such as APIs or AI for checking data policy compliance. Having assessed the prevalence of roles, the participants understood that, for example, it would not be useful to assign substantial responsibility for data compliance checks to Academic Data Editors or Data Curators.

## Publisher Roles and Data Skills

Once the roles were identified, they were mapped according to whether their data skills and disciplinary or workflow knowledge were high or low—this was intended to be used as a quick reference point when assigning checks to each role. For example, if an Administrator role was mapped as having low data skills, then a complex data check could not logically be assigned to them in the final workflow.

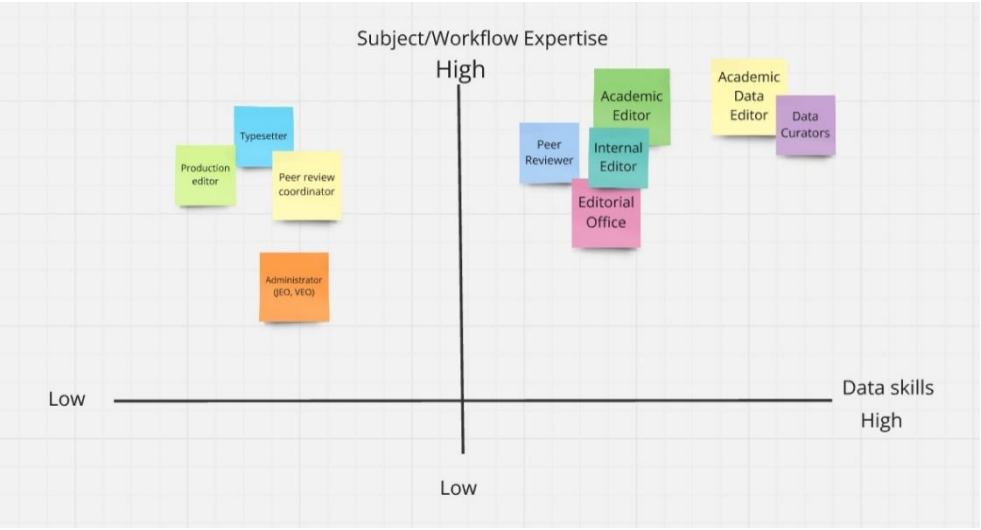


**Figure 1.** A mapping of publisher roles and their expertise.

While the skills for each role were mapped onto a quadrant (Figure 1), each of the roles had expertise in their subject or workflow, and the level of data-related skills could therefore be more usefully represented as a list: Data Curators and Academic Data Editors were most likely to have

high data skills, followed by Academic Editors, Internal Editors, the journal's Editorial Office and Peer Reviewers. More administrative roles in the workflow were designated as being least likely to have any data-related skills or expertise: these included Peer Review Coordinators, Typesetters, Production Editors and Journal Administrators. While this appears to be logical and unproblematic at first glance, it does create challenges: the roles with lower data skills most often undertake initial checks on incoming manuscripts or make final checks before publication. These checks often focus on the manuscript elements which include key data sharing information including data repository links and data citations.

With the challenges of variable data expertise in mind, the workshop participants then used sticky notes on the collaborative Miro board to assign a role to each proposed data check from the *Handbook*, and the results were collated to identify where there was consensus on which role could most appropriately undertake each check (Table 2). Note that the wording to describe each check was simplified slightly in comparison to the final version included in the *Handbook*, for ease of understanding during the workshop.

In proposing which role should undertake each check, the participants considered both the complexity of the intended check, and the likelihood that a journal would have such a role in place. For example, checking whether a Data Availability Statement is present in a manuscript is relatively simple, and could easily be included the usual checks of incoming manuscripts which are undertaken by Journal Administrators. As there was not always consensus on every check, second and third choices which were ranked highly are included in Table 2.

**Table 2**: Example mapping of roles for each data check

| | Is Data Availability Statement present? | Is the Data Availability Statement complete? | Is data in a repository? | Is a link to data included? | Is the repository appropriate for the data? | Is the data open/is there an exemption? |
|---|---|---|---|---|---|---|
| 1st choice | Administrator | Editorial Office | Editorial Office | Editorial Office | Internal Editor | Editorial Office |
| 2nd choice | None suggested | Internal editor | Academic editor | Academic editor | Data editor | Internal editor |
| 3rd choice | None suggested | Academic editor | None suggested | None suggested | Peer reviewer | Academic editor |

## Mapping FAIRness and Reproducibility Checks to a Publishing Workflow

Finally, the checks were mapped into a generalised manuscript submission and publication workflow which had been created by the co-chairs ahead of the workshop. This high-level workflow was intended to capture the stages that most manuscripts follow at most publishers. The stages suggested were: 1. Manuscript Submission; 2. Initial Quality Control Checks; 3. Editor Assessment; 4. Peer Review and Revisions; 5. Pre-acceptance Editorial Checks; 6. Production and Typesetting; and 7. Publication. The workshop participants agreed that this workflow accurately captured the key stages of publication in a logical order and could be used for the mapping activity.

Using the submission and publication workflow, the next activity was intended to sense check the timing and logic of the checks and ensure that the correct team or individual could conduct the check at the point where they were likely to interact with the manuscript. For example, it would not be logical to ask that an Editor assesses the author's choice of data repository before an

Administrator had checked that a Data Availability Statement had been provided. Where the group had created a clash of role versus timings (for example, suggesting that a particular role should undertake a check, but placing the check at a point of the workflow when the individual or team was unlikely to handle the manuscript) this was resolved by the Pilot's co-chairs by consensus.

The final output of the workshop series is an idealised publishing workflow where each check has been placed at the appropriate workflow stage and assigned to an individual or team who would be capable of undertaking it (Figure 2). The full workflow diagram[16] supports the more comprehensive *Handbook* which provides additional guidance on how each check should be performed. Due to the work undertaken to map a realistic workflow with suggested editorial and production roles assigned to it, the *Handbook* is both an overview of best practice and a practical guide to implementing comprehensive data checking in a manuscript submission workflow. As of January 2025, several of the publishers involved in the workshop have agreed to pilot the *Handbook* and provide further validation of its suggested implementation. The Pilot team intend to publish the results of these test implementations as well as qualitative feedback from publishers on their experiences of enforcing stronger Open Data policies.

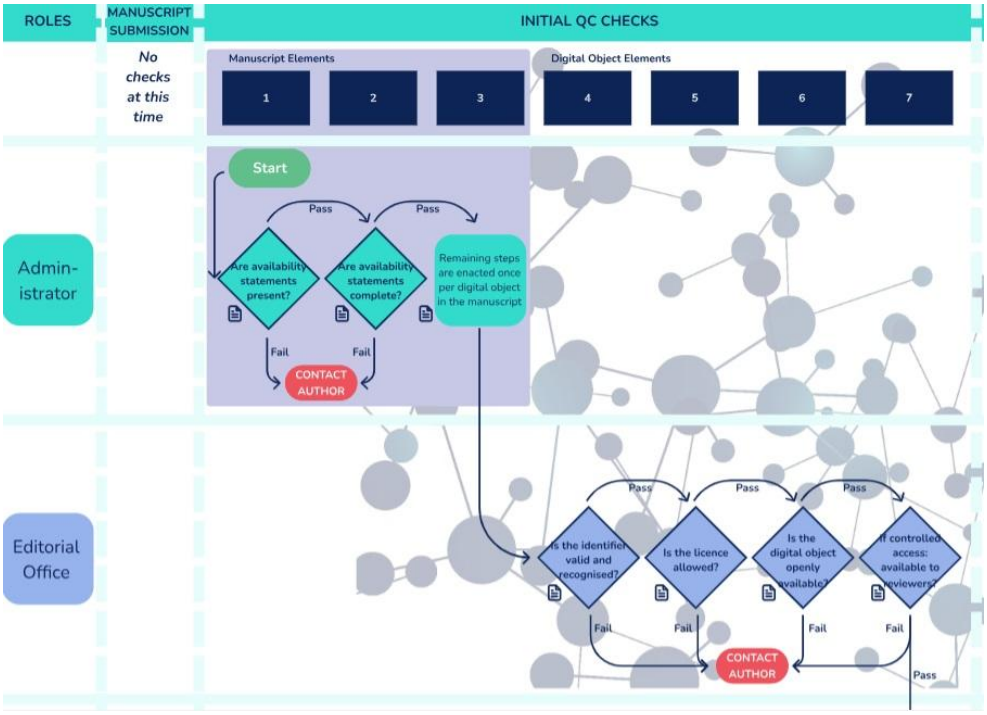


**Figure 2.** An extract of the final workflow diagram including roles and publication workflow stage.

# Conclusion

While the mapping of the data policy checks to the publishing workflow initially appeared to be a straightforward task, it required multiple rounds of collaborative brainstorming and consensus-building with publishing representatives. This work has the potential to significantly improve understanding of data policy implementation requirements at academic publishers, facilitating assessments of resourcing and workflow planning to ensure that Open Data policies can be enforced.

[16] Flowchart - Editorial Reference Handbook: https://osf.io/knqzu/

### Future Implications

In the development of the *Handbook* Pilot proposal, it was apparent that there was an information gap surrounding the practicalities of the publishing workflow and the range of roles and teams involved in it. While the research and data curation lifecycles have been documented in conceptual models such as the Data Curation Lifecycle[17] or Research Data Management Lifecycle[18], no standardised model exists to capture the stages and roles involved in manuscript submission, revision and publication. This gap could potentially hinder collaborative research projects with stakeholders external to academic publishing. The development of a publisher-agnostic, idealised publishing workflow with a set of generic roles has the potential to support the implementation of numerous future innovations, connected to reproducibility or other areas such as research integrity.

While the RDA Data Policy framework provides a template of policy features, the value of data policies is in their implementation and enforcement. By working collaboratively with a group of publishers, the publishing workflow output has created supporting guidance to clarify the requirements of each element of a chosen policy and the additional staffing or activity needed to support it. The workflow is therefore a complementary resource which can be used to support the future implementation of the RDA's existing policy framework.

2025 is an appropriate time to reflect on the success of publisher data sharing policies, marking 10 years since the journal policy landscape was assessed as being inconsistent and fragmented. In the intervening period, a range of initiatives have been undertaken to align the policy requirements and author guidance provided by thousands of academic journals. To prepare for the next ten years, policy enforcement, not just implementation, must be key—whether that is achieved by editorial checks, AI, or other tools. The work described in this paper aims to move academic journals towards this goal to support increasing the FAIRness of published data and the reproducibility of research towards 2035.

## Acknowledgements

The authors would like to thank all of the publisher representatives who gave their time to work collaboratively on the Pilot and its outputs.

## References

Alam, S. (2024). Trends in research integrity concerns and the evolving role of the publisher. *Insights* 37: 13, 1–6. doi:10.1629/uksg.663

Cannon, M., Kelly, A. & Freeman, C. 2022. Implementing an Open & FAIR data sharing policy—A case study in the earth and environmental sciences. *Learned Publishing* 35, 56–66. doi:10.1002/leap.1442

Grant, R., & Hrynaszkiewicz, I. (2018). The Impact on Authors and Editors of Introducing Data Availability Statements at Nature Journals. *International Journal of Digital Curation* 13: 1, 195–203. doi:10.2218/ijdc.v13i1.614

Holt, J., Walker, A. & Jones, P. (2021). Introducing a data availability policy for journals at IOP Publishing: Measuring the impact on authors and editorial teams. *Learned Publishing* 34: 478–486. doi:10.1002/leap.1386

[17] Curation Lifecycle Model: https://www.dcc.ac.uk/guidance/curation-lifecycle-model
[18] Harvard Biomedical Data Management: https://datamanagement.hms.harvard.edu

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Insights* 19. doi:10.5334/dsj-2020-005

Naughton, L. & Kernohan, D. (2016). Making sense of journal research data policies. *Insights* 29:1. https://insights.uksg.org/articles/10.1629/uksg.284

Sansone, S.-A., Lister, A., Klebel, T. & Ross-Hellauer, T. (2023). Summary of the first TIER2 Publisher Workshop. doi:10.17605/OSF.IO/TGUXZ

Smit, S. & van Rossum, J. (2022). STM's Research Data Year: Results, lessons learnt, and next steps. *Learned Publishing* 35: 1, 1-87. https://doi.org/10.1002/leap.1434