Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

Published on 08 June 2026
Licensed under CC-BY 4.0

Check for updates

Cite this: *Digital Discovery*, 2026, **5**, 2795

Journal research data policies in materials science

Lukas Hörmann,[ab] Hemanadhan Myneni,[c] Rwayda Kh. S. Al-Hamd,[d] Katarina Batalović,[e] Silvia Bonfanti,[f] Federico Grasselli,[gh] Saulius Gražulis,[i] Bahattin Koç,[j] Konstantinos Konstantinou,[k] Ivor Lončarić,[l] Nataliya Lopanitsyna,[m] José Manuel Oliveira,[no] Paolo Pegolo,[p] Patrícia Ramos,[nq] Kevin Rossi,[rs] Sebastian P. Schwaminger,[tu] Edith Simmen,[v] Milica Todorović,[k] Markus Stricker*[w] and Jonathan Schmidt[vp]

Open and reproducible research in materials science relies on the availability of data, code, and established metadata standards. Journal research data policies (RDPs) are a primary mechanism by which these community norms are enforced. We survey RDPs for 171 materials science journals spanning 17 publishers, using an expanded coding framework that captures both data-and-code sharing behavior as well as refereeing standards. We find clear signs of progress in comparison to earlier research on RDPs: nearly all journals provide an RDP, and most mention data availability statements. However, enforceable requirements remain uncommon, public deposition of underlying data is rarely mandatory, and FAIR publication is typically encouraged rather than required. Expectations for research software are substantially less developed than those for data, with limited attention to versioning and persistent identifiers, dependency disclosure, reproducible execution environments, or software quality practices. Aggregating the findings on policy features into an open research data score reveals pronounced heterogeneity across journals. Neither impact factor nor access model reliably predicts policy strength. Double-coding further shows that more complex policies and stricter policies can be more challenging to interpret consistently, and we highlight challenges in consistent RDP encoding across studies. Lastly, we conclude with recommended best practice directions for the future.

Received 11th March 2026
Accepted 5th June 2026

DOI: 10.1039/d6dd00111d

rsc.li/digitaldiscovery

1 Introduction

The field of materials science has witnessed a transformative acceleration of research driven by the integration of data-driven approaches and machine learning.[1–4] As researchers strive to develop innovative materials with tailored properties for applications ranging from energy storage and conversion to electronics by relying on data-heavy workflows, the need for data-availability has never been more critical. Achieving verifiable results requires systematic management and sharing of the underlying research artifacts, including all datasets and the software and workflows used to generate and analyze them,

[a]Faculty of Physics, University of Vienna, Vienna, 1090, Austria
[b]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK
[c]The Faculty of Industrial Engineering, Mechanical Engineering, Computer Science, University of Iceland, Reykjavík, Iceland
[d]Department of Civil Engineering and Management, Faculty of Science and Engineering, The University of Manchester, Manchester, M13 9PL, UK
[e]Vinča Institute of nuclear sciences-national institute of the Republic of Serbia, University of Belgrade, Serbia
[f]NOMATEN Centre of Excellence, National Center for Nuclear Research, ul. A. Sołtana 7, 05-400, Swierk/Otwock, Poland
[g]Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli Studi di Modena e Reggio Emilia, 41125 Modena, Italy
[h]CNR-NANO S3, 41125, Modena, Italy
[i]Vilnius University, Life Sciences Center, Institute of Biotechnology, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania
[j]Sabanci University, Orta Mah, FENS 1023, Tuzla, Istanbul, Turkey
[k]Department of Mechanical and Materials Engineering, University of Turku, Turku, 20014, Finland
[l]Ruđer Bošković Institute, HR-10000 Zagreb, Croatia
[m]Syngenta Crop Protection AG, Schaffhauserstrasse, Stein, 4332, AG, Switzerland
[n]Institute for Systems and Computer Engineering, Technology and Science, 4200-465, Porto, Portugal
[o]School of Economics and Management, University of Porto, 4200-464, Porto, Portugal
[p]Laboratory of Computational Science and Modeling, Institut des Matériaux, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[q]CEOS.PP, ISCAP, Polytechnic of Porto, 4465-004 S, Mamede de Infesta, Portugal
[r]Department of Materials Science and Engineering, Delft University of Technology, Delft, Netherlands
[s]Climate Safety and Security Centre, TU Delft The Hague Campus, Delft University of Technology, 2594 AC, The Hague, The Netherlands
[t]NanoLab, Division of Medicinal Chemistry, Otto Loewi Research Center, Medical University of Graz, Neue Stiftingtalstraße 6, 8010, Graz, Austria
[u]BioTechMed-Graz, Mozartgasse 12, 8010 Graz, Austria
[v]Department of Materials, ETH Zürich, Zürich, CH-8093, Switzerland
[w]Interdisciplinary Centre for Advanced Materials Simulation, Ruhr University Bochum, D-44799 Bochum, Germany. E-mail: markus.stricker@rub.de

© 2026 The Author(s). Published by the Royal Society of Chemistry

beyond what can be reported in the manuscript alone.[5–7] An environment of data sharing, code sharing and competitive benchmarks ideally lead to a state of *frictionless reproducibility* which has been a key driver of progress in machine learning[8] and its application to materials science.

A commitment to open research practices fosters collaboration, accelerates discovery, enables validation of results, and facilitates the reuse of tools across disciplines.[3,9] For data-driven methods in particular, the availability of high-quality, machine-readable datasets forms the essential foundation for building robust models.[10]

Data sharing practices vary substantially across scientific disciplines. In crystallography and structural biology, data sharing is well established: the Protein Data Bank (PDB) founded in the 1970s[11] evolved into the global archive for macromolecular structures.[12,13] By 2008, deposition of not only structure coordinates but also structure factors and Nuclear Magnetic Resonance restraints became mandatory,[14] enabling validation and, in some cases, improvement of the original findings. This policy, coordinated by the Worldwide PDB consortium[15] and supported by major publishers and funding agencies, substantially improved the availability of structural data and enabled machine learning success stories, such as in protein folding prediction.[16–19] More recently, advances in computing and storage have enabled repositories for raw diffraction and microscopy data.[20]

In the materials science community, data sharing is likewise recognized as essential to advancing the field, as highlighted by the adoption of international frameworks such as the FAIR (Findable, Accessible, Interoperable, Reusable) principles for data[7] and code[21] in materials science.[22] Although no single centralized repository exists for crystal structure data, maintainers of major computational databases[23–28] have integrated the OPTIMADE API standard,[29] effectively creating a common data space for basic, simulated crystal structure data. Although crystal structures are only a small subspace of the diverse materials science research data, these successful data-sharing initiatives demonstrate that best practices can be adapted to address gaps in materials-data sharing standards.

Various ongoing projects and communities work on the development of data and metadata standards, such as NOMAD,[26] the Battery Data Genome,[30] the Materials Genome Initiative,[31] and the Materials Research Data Alliance,[32] and strategies have been proposed to encourage authors to share data at all levels. This includes initiatives by funding agencies,[33–38] universities,[39,40] research communities such as NFDI FAIRmat and NFDI MatWerk in Germany,[41–43] and individual research groups.

Despite global community efforts, issues such as the lack of standardization of data formats, reluctance to share proprietary or sensitive data, insecurities about sharing data and code, and insufficient infrastructure for storage and access continue to hinder progress.[30,44,45] Furthermore, the complexity, heterogeneity, and sheer volume of materials science data—much of it generated without established curation frameworks—pose persistent obstacles to effective data management.[42,43] In combination with the rapid expansion of machine-learning-driven research, these limitations raise a second, more systemic risk: low-quality, weakly documented, or insufficiently validated datasets and workflows can be reused at scale, silently propagating errors through follow-on studies and into widely used databases, generating noise and erroneous conclusions. Strengthening data stewardship and reproducibility practices is therefore not only a matter of more efficient data management, but also a prerequisite for maintaining credibility as machine learning (ML) becomes increasingly embedded in materials science.

Research across STEM disciplines (science, technology, engineering, mathematics) highlights that journals are the arbiters of community practice and can, even more than funding bodies, incentivize authors to follow community-specific best practices[46] that move the field forward.[47–50] However, such journal-specific policies have often been found inadequate in their standards, and in some cases, entirely absent.[48,51–63] For example, Rousi *et al.* (2020)[52] reported that 35% of the highly cited physics journals they examined had no research data policy (RDP), while Resnik *et al.* (2019)[51] found that 43.6% of the examined journals did not address data sharing in their policies at all.

There are systemic challenges to adopting more comprehensive and stringent RDPs. First, such policies inevitably impose additional workload on researchers, reviewers, and editors,[64] which can conflict with the incentives of profit-driven journals. Stricter standards may slow publication processes and reduce the number of accepted manuscripts, potentially lowering revenue, particularly in open-access models where income is tied to publication volume. Second, in many materials science domains, companies are significant contributors to research but are often unable to share data.[65] Enforcing strict data publication standards could reduce their scientific contributions to public research. On the other hand, access to results and code is necessary for referees and editors to make informed judgments about reproducibility. Currently, it is largely up to individual editors and reviewers to insist on good practices prior to acceptance, but their mandate is limited without explicit journal policies related to data and code sharing. On a positive note, there are already activities aiming at automating the process of good code sharing practices, such as FACILE-RS.[66]

Nevertheless, longitudinal studies in other fields indicate progress in the adaptation of RDPs: in mathematical and multidisciplinary science journals, the fraction without data policies was 66% in 2011 and 62% in 2012,[47] while in the social sciences it decreased from 90% in 2003 to 61% in 2015. A comprehensive overview of these and related studies is provided in ref. 52.

In this article, we extend existing coding frameworks for journal research data policies[51,52] to quantitatively assess the current policy landscape in materials science. Here, "coding" follows the social science meaning of systematically categorizing qualitative and quantitative information to enable structured analysis. Our adapted framework incorporates materials science-specific considerations, explicitly evaluates research software beyond code sharing (including expectations for code standards and computational workflow reproducibility), and captures enforcement mechanisms such as guidance for editors

 © 2026 The Author(s). Published by the Royal Society of Chemistry

and referees (full framework in Sec. 5). Using the Dimensions[67] API and selected research field definitions, we compiled a dataset of 171 journals from 17 publishers that is broadly representative of materials science publication venues. For each publisher, we sampled—where available—the three highest impact journals (by journal impact factor), the three most prolific journals (by publication volume), and three randomly selected journals to test whether journal characteristics such as impact or popularity relate to policy strictness. We additionally included a small number of data-focused journals (*e.g.*, *CODATA Data Science Journal*) to benchmark best-practice-oriented policies. Lastly, we examine the challenges of consistently coding RDPs, both within a single study and when comparing results across different studies.

2 Results

An overview of the results concerning the fundamentals of the RDPs is given in Fig. 1. These results include (a) the existence of an RDP in the first place, (b) the rules concerning data availability statements (DAS), (c) specifics on data sharing requirements, (d) specifics on code sharing requirements, (e) timing of data release, and (f) whether data and/or code are mentioned in the journals' referee guidelines.

2.1 Data policies

To start, RDPs show a positive trend: compared to earlier studies,[47,51,52] 98% of the journals we surveyed have an RDP (Fig. 1a), and 95% mention a DAS (Fig. 1b). However, nearly half of all journals still do not require a DAS. Although it is known that DAS alone are insufficient for establishing a strong data publishing culture,[68,69] we still consider it one of the foundations of an RDP, provided that the truthfulness of the DAS is verified during the review process before publication.

Most RDPs mention data sharing requirements, but typically as 'optional' (Fig. 1c): only 7% of journals require public sharing of all research data in all cases. Requirements for specific data types (*e.g.*, crystal structures, genomics) are more common for

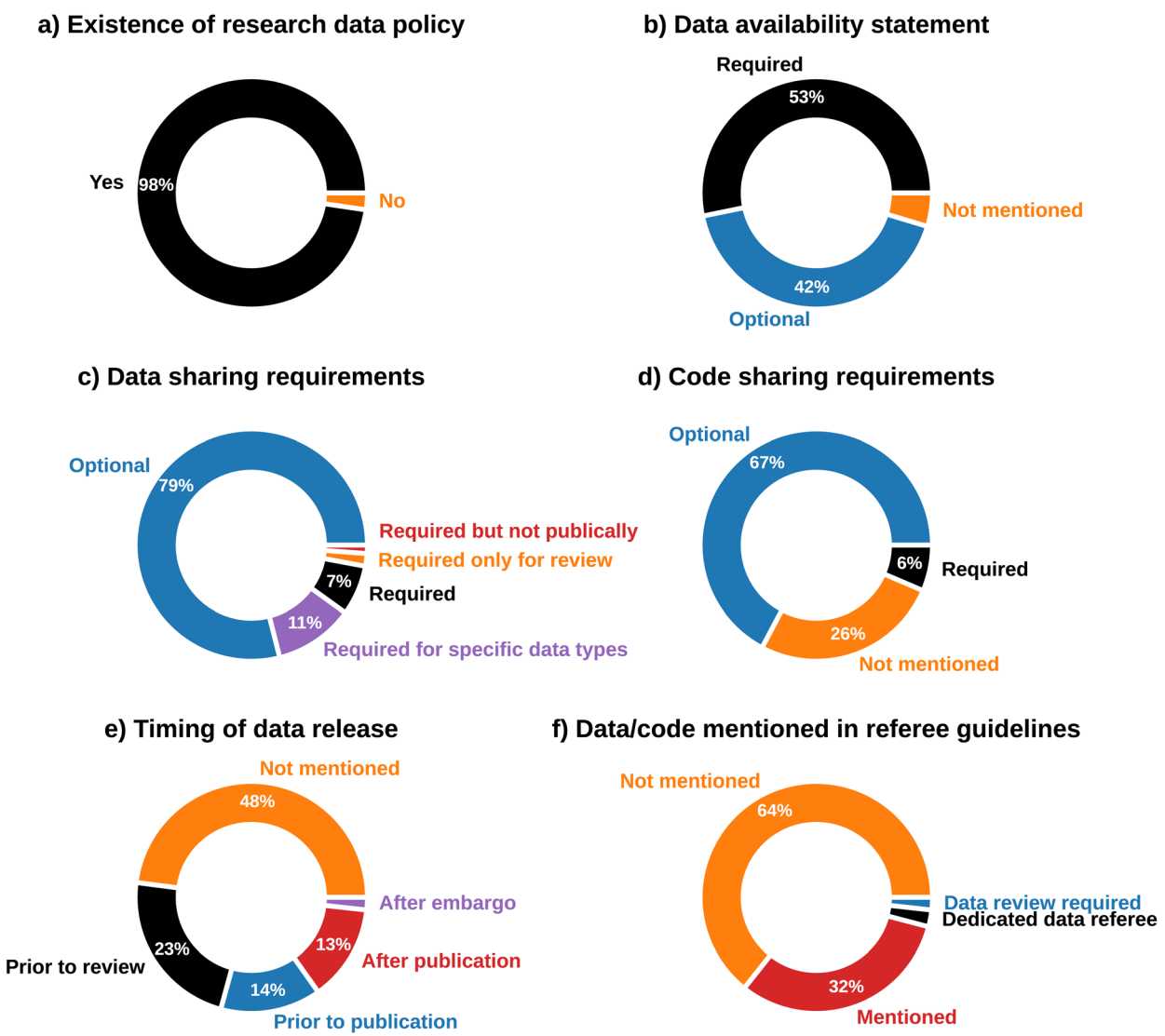


Fig. 1 Overview of requirements for sharing data and code.

© 2026 The Author(s). Published by the Royal Society of Chemistry

11% of the journals. An additional 1.8% require sharing of all data during peer review, and 1.2% require sharing upon reasonable request.

Since previous research has found that a DAS alone is not sufficient to foster good data-sharing behavior,[68,69] we also survey the timing of data release (*cf.* Fig. 1e) and the requirements to make data publicly available. Rousi *et al.*[52] found that 15 out of 40 surveyed physics journals required data to be released upon publication. In our materials-science sample, combining the categories "required data available prior to the review process" and "required data available after official publication", we find that 36.3% of journals include an explicit requirement for data availability by review and/or by publication. This is broadly comparable to, but slightly lower than, the corresponding fraction reported in ref. 52.

So far, we have only discussed whether and when research data are shared. Equally important is the location of the shared data, as this affects both its findability, long-term preservation, and adherence to metadata standards. Specifically, two-thirds of journals encourage sharing data on FAIR repositories, 11.9% require the sharing of certain data types on FAIR repositories, and only 1.2% require all data to be shared on FAIR repositories. Conversely, 77.4% of journals do not explicitly mention the licensing of data in their open data policies. Following the FAIR definition,[7] 'Reusable' is defined as "(*meta*) *data are released with a clear and accessible data usage license*". In any case, the mere existence of a license does not imply a permissible license and reusability *per se*. Hence, journal policies aimed at data sharing should support permissible licenses.

Strict policies provide little benefit without enforcement, which usually requires the expertise of referees. However, few journals consider the role of data in the peer review process, and 61% of journals do not mention the RDP in the referee guidelines (Fig. 1f). Fewer than one-third make any reference to it, and only a very small number explicitly require data and/or code sharing in their referee guidelines or indicate that a dedicated data referee is appointed.

2.2 Code policies

In stark contrast to data policies, software code policies are weaker across all surveyed journals. Code sharing is encouraged but optional for 70% of journals and required in only 1.8%. Even among journals that encourage code sharing, more than 80% do not require authors to specify software dependencies or execution environments. Over 80% of policies omit guidance on persistent identifiers for code. Mentions of coding standards (*e.g.*, style or safe constructs) are uncommon – around 80% of policies do not address them – and fewer than 10% require any minimal standards and references to testing and linting are rarer still.

2.3 Open Data score

To summarize policy strictness, we report an *open data score* (ODS; $s \in [0, 1]$, higher = stricter; full definition in Section 5.3). Intuitively, we interpret "strictness" as a larger amount of data being more openly shared. We evaluate the average ODS across journals, while the variability across journals for each question is captured by a per-question standard deviation. These results are presented in Fig. 2.

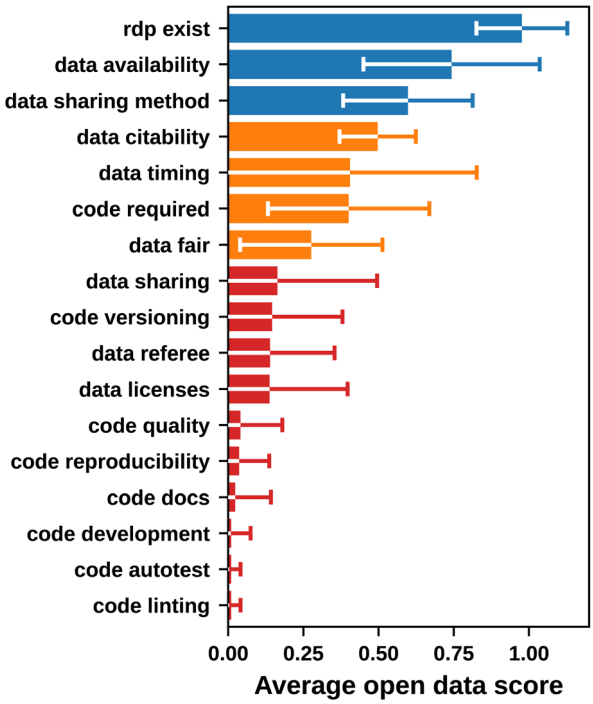


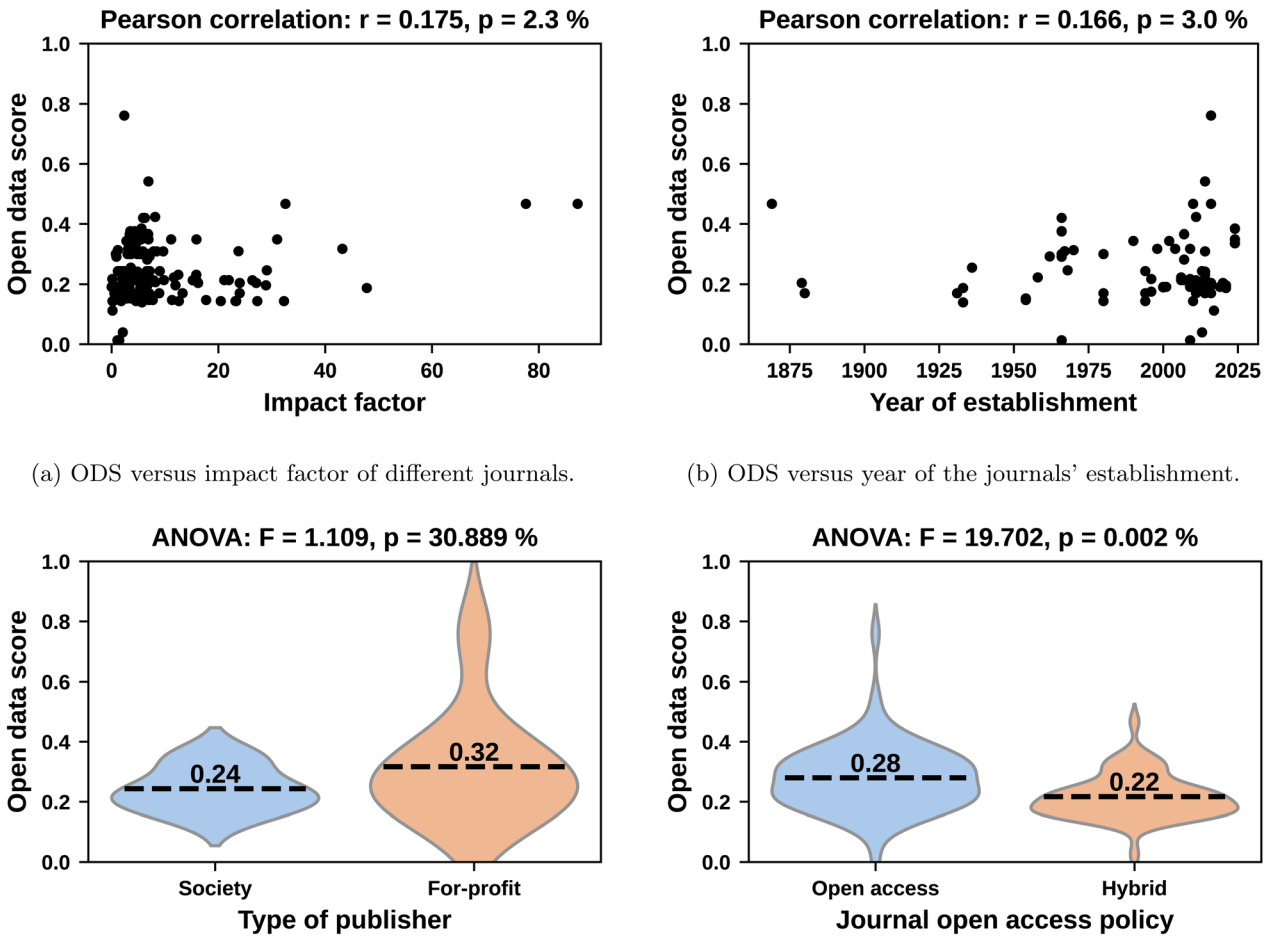
Fig. 2 Average open data score by question; error bars indicate standard deviation of open data scores. The colors indicate questions with ODS below 0.2 (red), below 0.5 (orange) and above 0.5 (blue).

Journals achieve relatively high ODSs for the existence of an RDP, the requirements for a *data availability* statement, and policies on *data sharing methods*, with many journals recommending FAIR data repositories and *data citability*. Conversely, concrete requirements for sharing data achieve lower ODSs. For instance, *data sharing* requirements achieve an ODS of below 0.2. Overall, we find large standard deviations. For instance, the large standard deviations for *data availability* and *data timing* indicate substantial differences in whether the respective DASs are required and when data must be released.

We tested several hypotheses about factors associated with stricter RDPs, focusing on the relationships between the ODS and (i) journal impact factor, (ii) society *vs.* non-society publishers, (iii) open-*vs.* closed-access journals, and (iv) the year of the journal's establishment.

Fig. 3a illustrates the ODS *versus* journal impact factor. Although some high-impact journals have higher ODS, there is no clear monotonic trend: journals across the impact spectrum span the full ODS range, suggesting that prestige (impact factor) does not reliably correlate with policy strictness. The Pearson correlation is $r = 0.175$ ($n = 169$, $p = 2.3\%$), indicating a weak but statistically still significant association. To assess sensitivity to extreme values, we repeated the analysis after excluding the two highest-impact journals (pre-specified rule: impact factor > *rbin*75). The correlation disappeared ($r = -0.007$, $p = 93.9\%$), indicating that the apparent association is driven by these outliers rather than a general trend.

 © 2026 The Author(s). Published by the Royal Society of Chemistry



(a) ODS versus impact factor of different journals.

(b) ODS versus year of the journals' establishment.

(c) ODS versus type of publisher; the average ODS of each publisher is considered as one sample.

(d) ODS versus open access policy of journals.

Fig. 3 Open data score (ODS) *versus* journal characteristics. (a) ODS *versus* impact factor of different journals. (b) ODS *versus* year of the journals' establishment. (c) ODS *versus* type of publisher; the average ODS of each publisher is considered as one sample. (d) ODS *versus* open access policy of journals.

Fig. 3b depicts the ODS as a function of the year of establishment. The Pearson correlation is $r = 0.166$ ($n = 171$, $p = 3.0\%$), indicating a weak but statistically significant association.

Fig. 3c shows the relationship between the publisher type (society or for-profit) and the ODS. While we find that the average ODS of for-profit publishers is higher than that of society publishers, the ANOVA F-test as well as a *t*-test find no statistically significant correlations.

Fig. 3d shows the relationship between the open access policy of a journal (full open access *versus* hybrid) and the ODS. We find that the average ODS of open-access journals is higher than that of journals with hybrid policies. To gauge significance, we perform an ANOVA F-test as well as a *t*-test. We find an *F*-statistic of 19.702. The *p*-score is 0.002%, indicating that this result is statistically significant.

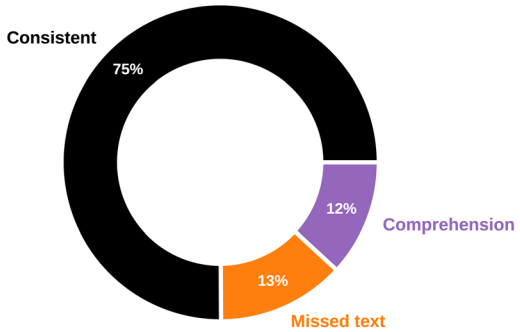
2.4 Consistency of encoding process

As detailed in Section 5.2, two researchers independently coded each RDP and then compared their results to ensure consistency. Typical inconsistencies arose when the same policy used conflicting language in different sections. For example, requiring deposition in a FAIR repository in one place but only *encouraging* it elsewhere. Under our conservative rule, such cases were coded as *encouraged*. Publishers with a single brief, uniform policy across journals (*e.g.*, APS at the time of coding) yielded the highest agreement, whereas more complex, multi-page frameworks with journal-specific appendices (*e.g.*, Frontiers, MDPI, RSC) produced more discrepancies.

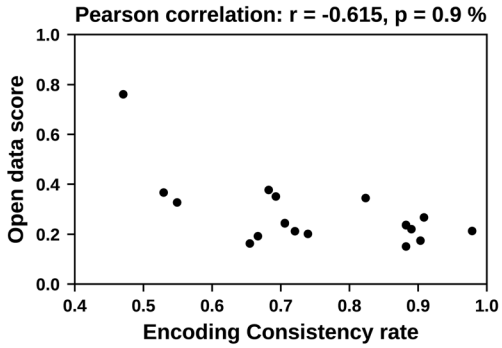
In Fig. 4a we provide an overview of the coding process. 75% of the questions were initially coded consistently, while initially inconsistent encoded questions could be attributed to RDP text that was missed (13% of all questions) and a misunderstanding of the meaning of the policy (12%).

The most inconsistently-coded items were questions 10 & 11 (data types for which unique policies are recommended/required), question 7 (data sharing method), and question 6 (timing of data release). Higher inconsistency for questions 10 & 11 is expected because these required free-text extraction of data-type-specific rules. For question 7, the option "*Multiple*

© 2026 The Author(s). Published by the Royal Society of Chemistry



(a) Summary of the percentage of coding questions that were answered consistently by the two encoders per journal and the reason for the inconsistency.



(b) Open data score versus the consistency of encoding, excluding questions 10 & 11.

Fig. 4 Consistency of encoding and its relationship to open data score. (a) Summary of the percentage of coding questions that were answered consistently by the two encoders per journal and the reason for the inconsistency. (b) Open data score *versus* the consistency of encoding, excluding questions 10 & 11.

data sharing methods equally recommended in the RDP" accounted for 75% of its inconsistencies, indicating that this question is ill-posed and should be revised in future studies. For Question 6, distinctions between "*required before publication*" and "*required after publication*" were often unclear in policy text, suggesting that journals could improve the explicitness of release timelines.

These results led to a *post hoc* hypothesis that policy strictness impacts coding consistency (Fig. 4b). In short: if encoders struggle to code an RDP, it seems likely that authors submitting to that journal and respective reviewers will struggle to understand the RDP. Using the open data score as a proxy for policy comprehensiveness and excluding questions 10 & 11 from the consistency rating, we observe a strong negative Pearson correlation of $r = -0.615$ with $p = 0.9\%$: more comprehensive policies tend to be harder to code consistently. As the hypothesis emerged during the analysis phase, these findings should be considered in a hypothesis generating fashion and require confirmation in future work.

3 Discussion

Our results show that materials science journals have largely converged on the *language* of open research (*e.g.*, policies and data availability statements), but not on a shared, enforceable baseline that reliably yields reusable datasets and software. The consequence is a fragmented policy landscape in which the practical burden of interpretation and compliance is shifted to authors, editors, and referees. Practically, this means the reproducibility of published work can still hinge on venue-specific norms rather than community-wide expectations.

3.1 Data and code policies

A sizable fraction of journals in materials science (36.3%) mandate data availability by publication, a somewhat higher proportion than reported previously for physics journals.[52] Nevertheless, a majority of journals still permit delayed or conditional data release, which undermines the goal of immediate reproducibility. Perhaps most concerning are the results regarding *data sharing* itself. Fewer than 7% of journals require public deposition of all data, while most journals merely encourage the sharing of data. This disconnect between aspirational policy language and enforceable requirements likely contributes to the persistent gap between journal policy and author practice.[68,70–72]

The method of data sharing is an area where improvements can be made at little cost. Encouraging data sharing on FAIR repositories does not require additional resources, and for data types where repositories and standards already exist, journals without a respective policy can easily adopt existing recommendations. Here we observe positive trends. FAIR repositories are frequently encouraged, and citeability is often mentioned, reflecting an awareness of best practices for data management. Still, requirements for FAIR-compliant publication remain rare, and the lack of licensing guidance in the majority of policies (77.4%) raises questions about the reusability of shared data.

A policy's practical effect depends on whether it is embedded in editorial routines. While references to data/code in referee guidelines are more common than previously reported,[51] most journals still do not treat data and software as objects that are routinely checked during peer review. Evaluating whether a dataset actually enables the reproducibility of an article is difficult and unless this is verified during the peer review process, it is unlikely that this can be confirmed later. In our study, we only identified three journals that use a data and/or code referee who is responsible for checking the completeness of the data and confirming the basic functionality of the code.

Coding standards and quality assurance are the cornerstones of industrial software reliability, especially in high-responsibility domains such as the automotive industry,[73] avionics,[74] or chemistry.[72,75] It is evident that scientific software developers still have a lot to learn from industrial software developers.[76] Compared to previous studies, Stodden *et al.*[47] found a code policy in 22% of computational and mathematical

 © 2026 The Author(s). Published by the Royal Society of Chemistry

biology journals, similar to 19.1% for physical science journals in ref. 51. Consequently, 70% appears as an improvement. However, Resnik *et al.*[51] also found that 20.1% of journals required depositing code into repositories, in stark contrast to 6.4% in our dataset. We analyzed the research policies of 17 of the physics science journals that were identified as requiring code deposition in ref. 51 and found no requirement for code deposition in 11 of the journals. We see two possible explanations for this discrepancy. First, RDPs concerning code could have deteriorated. This seems unlikely given the general strengthening of open data requirements reported in our study and others.[52] A more plausible explanation lies in the challenges of consistently encoding RDPs across and within studies. For example, our study applies a strict interpretation of the term "required", as outlined in Section 5, whereas other studies may adopt a more flexible definition. Such differences in interpretation can readily lead to diverging results. In this context, we suggest that future studies should ideally include, as part of their data, the specific text excerpts from RDPs that informed their coding decisions, to better address these types of questions in the future. In addition, we find that even within our own coding process, results varied between encoders, as detailed in Section 5.2 and Section SI in the SI. This variability likely reflects not only deficiencies in the coding framework but also ambiguity in the wording of certain open data policies themselves.

Code quality is almost never mentioned in RDPs, as shown by near-zero *research data scores* of linting and automatic testing. The lack of attention to code quality standards may partly stem from the fact that most materials researchers are not formally trained in software development, and that this is often not the focus of studies. As good software practices require significant time and resources, they are rarely prioritized by funding agencies in materials science. Given the wide diversity of research code—from short analysis scripts to large-scale simulation packages—strict and universal standards are difficult to enforce at the journal level. Nevertheless, for journals with a strong focus on method and software development, there is a clear case for requiring higher standards of code quality.

Finally, the inverse correlation between the open data score and encoding consistency (Fig. 4b) highlights a tradeoff in policy design: policies that attempt to cover many cases and data types can become harder to interpret and implement. This suggests that strengthening RDPs is not only a question of adding more requirements, but of making requirements *testable* and *unambiguous*. A practical direction is therefore to align policies with existing standards and infrastructure while expressing them as a small set of checks to be verified during the editorial and refereeing process. Concretely, future policy templates could prioritise (i) a concise core with non-conflicting language; (ii) clearly tiered obligations (*e.g.*, *required* → *expected* (may affect editorial decisions) → *mentioned*), and (iii) a canonical decision path for timing, location (FAIR repositories), persistent identifiers, and licensing, with journal-specific annexes that cannot override the core without explicit precedence rules.

Our results point to a small number of recurrent gaps that are well covered by mature community standards but are rarely made enforceable in journal policies. However, there are limitations of this study concerning which journals have been chosen for RDP evaluation. Below, we map these gaps to policy clauses and checks that journals could adopt without requiring discipline-specific reinvention.

3.2 From observed gaps to policy templates

3.2.1 From "encouraged" to verifiable deposition. While many journals encourage repository deposition, only a minority require FAIR-compliant deposition, and licensing guidance is absent in most policies (77.4%). A minimal enforceable baseline is therefore: (i) deposition of all data needed to reproduce the main claims in a community-recognized repository (FAIR where available), (ii) a persistent identifier (typically a DOI), and (iii) an explicit reuse license. These elements are supported by established infrastructure such as DataCite and DOIs.[77,78] Editorial verification can be lightweight: the data availability statement must contain a resolvable identifier and license, and referees must have confirmed availability of the deposited artifacts at review time. The existence of a persistent, resolvable identifier, such as a DOI can further be used to link data in machine-readable form and enable automated or semi-automated meta analyses or quality control checks.

3.2.2 Treating research software as a first-class research output. Our survey shows that code requirements lag far behind data requirements: code sharing is usually optional, and most policies omit dependencies, versions, and reproducible execution environments. A corresponding baseline, aligned with FAIR4RS,[79] requires an archived versioned code release with a persistent identifier, together with an explicit specification of the execution environment. Here, *e.g.*, `requirements.txt`/`environment.yml` are a minimal example for python code that does not enable full reproducibility, while more modern environment managers[80] or a container recipe allow for full reproducibility. Where feasible, journals can require a single "reproduce" entry point that regenerates key figures or tables. This is supported by existing repositories and archival services (*e.g.* Zenodo[81]) and by common environment managers (*e.g.* Conda/Spack/Pixi) and containerization approaches. Importantly, these requirements can be adjusted for scope: journals may apply stricter expectations for methods/software papers than for primarily experimental studies.

3.2.3 Code quality: focusing on minimal signals rather than full industrial compliance. Because scientific software varies from short scripts to large simulation packages, universal industry-grade standards are unrealistic at the journal level. However, journals can still require minimal quality signals that improve maintainability and reusability: basic documentation of inputs/outputs, a license, and at least one automated sanity check (unit or functional test) for core routines when software is central to the contribution. Such requirements are consistent with established scientific software engineering guidance[82,83] and help address the near absence of current policies.

© 2026 The Author(s). Published by the Royal Society of Chemistry

Given that most journals still do not embed RDP requirements in referee guidance, a key step is procedural: requiring a short reproducibility checklist at submission (data DOI + license, code release + version, environment specification, mapping of figures to artifacts). Where workloads permit, journals can adopt specialized data/code editors for selected submissions, following emerging practice in a small number of venues. This targets the observed gap between policy language and actual verifiability[68,70–72] without placing the full burden on individual referees.

3.2.4 Beyond repository-by-repository compliance: interoperability as a future policy dimension. Finally, our survey suggests that current RDPs rarely address interoperability across repositories and disciplines. Emerging concepts such as FAIR Digital Objects (FDOs), discussed within the Research Data Alliance and the FDO Forum,[84–87] aim to support machine-actionable resolution of identifiers and access to heterogeneous data across domains. While FDO systems are still developing, the absence of any policy language on cross-domain interoperability indicates an opportunity for forward looking RDP templates to anticipate the increasing reuse of materials data across domains.

3.2.5 Capacity building as part of policy design. Finally, stricter and more software-aware policies will only be effective if researchers can realistically comply. Because many materials scientists receive limited formal training in data stewardship and software engineering, journals and publishers can reduce friction by explicitly pointing authors to community training resources (*e.g.*, The Carpentries and CodeRefinery) that instruct on version control, licensing, documentation, testing, continuous integration, and reproducible environments.[88,89] Referencing such resources within RDPs complements enforcement with enablement and directly targets the skills gaps that likely contribute to weak code-policy uptake.

4 Conclusions

We expanded existing coding frameworks to cover data and code sharing requirements and applied them to create a quantitative survey of materials science research data policies across 171 journals. Overall, the landscape shows clear progress in policy existence and intention, while enforceable requirements remain rare, with research software policies lagging even further behind. Aggregating policy features into an open data score reveals substantial heterogeneity across journals, with journal impact factor and access model providing limited predictive power for policy strength. Finally, double-coding demonstrates that policy complexity and strictness might reduce interpretability, highlighting the need for concise, internally consistent language and clearly tiered obligations.

Together, these findings provide a baseline of current RDP practice in materials science. They motivate evidence-based policy templates that translate community standards into a small set of unambiguous, verifiable checks during submission and peer review, while anticipating future needs such as cross-repository interoperability.

5 Methods

5.1 Journal selection

To select a representative set of journals, we used the DIMENSIONS[67] Python API. As a starting point, we limited our selection to 9 fields of research as defined by the Australian and New Zealand Standard Research Classification (ANZSRC 2020):

- 3047 Theoretical and computational chemistry
- 3403 Macromolecular and materials chemistry
- 3402 Inorganic chemistry
- 3405 Organic chemistry.
- 3406 Physical chemistry
- 4016 Materials engineering
- 4018 Nanotechnology
- 5102 Atomic, molecular, and optical physics
- 5104 Condensed matter physics

Searching for the most popular publishers in terms of the number of articles, we arrived at a preliminary selection. For each publisher, we then selected the three journals with the highest average citation count, the three most popular journals, and three random journals. All available journals were selected for publishers with fewer than nine journals. The detailed query used to obtain the data, along with the resulting dataset, is published with the article. Due to the high time cost of encoding journals, we limited ourselves to the 17 most popular publishers, resulting in a total of 171 journals.

5.2 Encoding process

Two researchers independently coded each publisher's RDP. For every answer, coders saved the exact policy text that supported their choice (when no relevant policy text existed, this field was necessarily empty). After independent coding, one of the coders compared the two sets to ensure consistency. In cases of discrepancies, the reviewer selected, if possible, the correct choice based on the collected texts, or, if this was not possible, discussed a correct choice with the rest of the research team. The reason for the discrepancy is also noted and analyzed to improve both the coding framework and RDPs for the future. When multiple interpretations were possible or different text sections provided conflicting information, the most conservative interpretation was used. The coding process took place between June 2024 and May 2025. For each journal, the date of the encoding is included in the published data. We note that during the study, RDPs already improved, *e.g.*, the research policies of APS became significantly stricter, and we hope to make a systematic comparison of the progress in a few years.

5.3 Numerical scoring of data policies

The *open data score* relies on the fact that our encoding framework is exclusive and hierarchically ranked. Each answer α to every question n is assigned a strictness value $w_{n\alpha}$ between 0 (most relaxed) and 1 (most strict). We denote the selected answer of each question with α_n. To attain the *open data score* (s), we sum over all questions and normalize by the number of questions N.

 © 2026 The Author(s). Published by the Royal Society of Chemistry

$$s = \frac{1}{N}\sum_{n=1}^{N}\sum_{\alpha=1} w_{n\alpha}\delta_{\alpha,\alpha_n} \quad (1)$$

Here, $\delta_{\alpha,\alpha n}$ is the Kronecker delta, indicating whether a given answer was selected.

5.4 Coding framework

To objectively and reproducibly evaluate open data policies, we have developed a coding framework in the form of mutually exclusive multiple-choice questions. The possible choices are ranked by how strongly they reflect open data principles, with later answers considered stronger policies. This allows for the association of each question with a numerical value. Our coding framework builds upon previous efforts,[51,52] emphasizing clear rankability and incorporating additional relevant elements, such as scientific code. We mark questions overlapping with ref. 51 and 52 with the respective citations. The coding framework is defined as follows:

1. Existence of research data policy[52]
(a) No
(b) Yes
2. Data sharing requirements (adapted from ref. 51)
(a) Data sharing only with editors and referees (no public sharing)
(b) Data sharing encouraged but optional
(c) Data sharing required but not public (*e.g.* available upon request).
(d) Public data sharing required for specific data types
(e) Public data sharing of all data required
3. FAIR data sharing
(a) Public data sharing on a FAIR repository not mentioned
(b) Public data sharing on a FAIR repository encouraged
(c) Public data sharing on a FAIR repository required for specific data types
(d) Public data sharing of all data on a FAIR repository required
4. Data availability statement
(a) Not mentioned
(b) Mentioned but optional
(c) Required
5. Citability and findability of data
(a) No mention of DOIs or other persistent identifiers (Handle, Archival Resource Key (ARK), *etc.*)
(b) DOIs or other persistent identifiers recommended for datasets or code
(c) DOIs or other persistent identifiers required for datasets or code
6. Timing of data release[52]
(a) Timing of data availability not addressed
(b) Required data available before review
(c) Required data available before publication
(d) Required data available after publication
(e) Required data available after an embargo period
7. Recommended data-sharing method (adapted from ref. 51)
(a) No method recommended
(b) Multiple methods equally recommended
(c)Data sharing upon request to authors
(d) Supplementary material or journal-hosted sharing recommended
(e) Public online repositories recommended
(f) FAIR repositories recommended
8. Recommended/required licenses
(a)No mention of licenses
(b) Specific license type mentioned
(c) Open access license required for shared data
(d) Open source license required
9. Referee guidelines concerning research data
(a) Data sharing policy not mentioned in refereeing guidelines
(b) Data sharing policy mentioned in refereeing guidelines
(c) Referees confirm shared data/code
(d) Additional data/code-specific referee required
10. List data types for which unique policies are recommended (adapted from ref. 51)
11. List data types for which unique policies are required (adapted from ref. 51)
12. Code sharing requirements
(a) Code sharing not mentioned
(b) Code sharing encouraged but optional
(c) Code sharing required
13. Code reproducibility
(a) No mention of dependencies
(b) Listing dependencies and versions encouraged
(c) Listing dependencies and versions required
(d) Container/installation script required
(e) Full reproducibility of figures/tables with container/scripts required
14. Versioning and persistent identifiers
(a) No mention of persistent identifiers or versions
(b) Persistent identifier or version encouraged
(c) Persistent identifier or version required
15. Code quality standards
(a) No mention of code quality standards
(b) Standards encouraged
(c) Standards required
(d) Specific criteria for referees on code quality
16. Automatic testing
(a) No mention of automatic testing
(b) Automatic testing encouraged
(c) Automatic testing required
17. Code documentation
(a) No mention of documentation standards
(b) Documentation encouraged
(c) Documentation required
18. Linting standards
(a) No mention of linting standards
(b) Linting standards encouraged
(c) Linting standards required
19. Code development and continuous integration
(a) No mention of development standards
(b) Development standards encouraged
(c) Development standards required

© 2026 The Author(s). Published by the Royal Society of Chemistry

Author contributions

Conceptualization: F. G., H. M., J. S., L. H., M. S., R. K. S. A. H., S. G.; methodology: H, M., J. S., L. H., M. S., S. G.; software: H. M., I. L., J. S., L. H., M. S., N. L., S. G.; validation: F. G., J. S., L. H., M. S., R. K. S. A. H., S. G.; formal analysis: F. G., H. M., J. S., L. H., M. S., S. G.; investigation: B. K., E. S., F. G., H. M., I. L., J. M. O., J. S., K. B., K. K., K. R., L. H., M. S., M. T., P. P., P. R., R. K. S. A. H., S. B., S. G., S. P. S.; data curation: F. G., H. M., J. S., L. H., N. L., S. G.; writing – original draft: F. G., H. M., J. S., L. H., M. S., S. G.; writing – review & editing: B. K., E. S., F. G., H. M., I. L., J. M. O., J. S., K. B., K. K., K. R., L. H., M. S., M. T., P. P., P. R., R. K. S. A. H., S. B., S. G., S. P. S.; visualization: F. G., H. M., J. S., L. H., M. S.; supervision: F. G., J. S., L. H., M. S.; project administration: J. S., L. H., M. S.; funding acquisition: F. G., K. R., M. T.

Conflicts of interest

K. R. is part of the editorial board of Journal of Physics:Materials (IOP Publishing).

Data availability

All data and code to reproduce the research are available at https://github.com/daemoncost/Journal-Research-Data-Policy. The code version at publication time is on Zenodo with the persistent identifier https://doi.org/10.5281/zenodo.20529191.[90]

Supplementary information (SI) is available. See DOI: https://doi.org/10.1039/d6dd00111d.

Acknowledgements

This article is a result of joint work in COST Action CA22154 – Data-driven Applications towards the Engineering of functional Materials: an Open Network (DAEMON) supported by COST (European Cooperation in Science and Technology). JS was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program project HERO Grant Agreement No. 810451 and funded by the SNSF Ambizione grant number 233444. LH was supported by a UKRI Horizon grant (MSCA, EP/Y024923/1). KB acknowledges the support by Ministry of Science, Technology and Innovation of the Republic of Serbia *via* contract no. 451-03-33/2026-03/200017. MS acknowledges support by the European Union by ERC grant, project no. 101161287. The views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. SB acknowledges support from the National Science Center in Poland through the SONATA BIS grant DEC-2023/50/E/ST3/00569 and from the Foundation for Polish Science in Poland through the FENG.02.02-IP.05-0177/23 project. This work was carried out within the "Projektowanie Ulepszonych Szkieł Metalicznych" project (FENG.02.02-IP.05-0177/23) under the 2.2 First Team programme of the Foundation for Polish Science co-financed by the European Union from the European Funds for Smart Economy 2021–2027 (FENG). All authors gratefully acknowledge Nicola Spaldin, Julian Dederke, and Nicolai Bissantz for helpful and insightful discussions, and the ETH Library for their support.

References

1 I. Tanaka, K. Rajan and C. Wolverton, Data-centric science for materials innovation, *MRS Bull.*, 2018, **43**, 659.

2 S. Sandfeld, T. Dahmen, F. O. R. Fischer, C. Eberl, S. Klein, M. Selzer, B. Nestler, J. Möller, F. Mücklich, M. Engstler, S. Diebels, R. Tschuncky, A. Prakash, D. Steinberger, C. Kübel, H.-G. Herrmann and R. Schubotz, Digitale Transformation in der Materialwissenschaft und Werkstofftechnik DGM Strategiepapier (strategy whitepaper of the German Materials Society), Deutsche Gesellschaft für Materialkunde e.V., 2018.

3 M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C. T. Koch, K. Kremer, *et al.*, FAIR data enabling new horizons for materials research, *Nature*, 2022, **604**, 635.

4 S. Bauer, P. Benner, T. Bereau, V. Blum, M. Boley, C. Carbogno, R. Catlow, G. Dehm, S. Eibl, R. Ernstorfer, Á. Fekete, L. Foppa, P. Fratzl, C. Freysoldt, B. Gault, L. M. Ghiringhelli, S. K. Giri, A. Gladyshev, P. Goyal, J. Hattrick-Simpers, L. Kabalan, P. Karpov, M. S. Khorrami, C. T. Koch, S. Kokott, T. Kosch, I. Kowalec, K. Kremer, A. Leitherer, Y. Li, C. H. Liebscher, A. Logsdail, Z. Lu, F. Luong, A. Marek, F. Merz, J. Rezaei Mianroodi, J. Neugebauer, Z. Pei, T. A. R. Purcell, D. Raabe, M. Rampp, M. Rossi, J. M. M. Rost, J. E. Saal, U. Saalmann, K. N. Sasidhar, A. Saxena, L. Sbailo, M. Scheidgen, M. Schloz, D. Schmidt, S. Teshuva, A. Trunschke, Y. Wei, G. Weikum, R. P. Xian, Y. Yao, J. Yin, M. Zhao and M. Scheffler, Roadmap on data-centric materials science, *Model. Simul. Mater. Sci. Eng.*, 2024, **32**, 063301.

5 O. Kononova, T. He, H. Huo, A. Trewartha, E. A. Olivetti and G. Ceder, Opportunities and challenges of text mining in materials research, *iScience*, 2021, **24**, 102155.

6 M. Schilling-Wilhelmi, M. Ríos-García, S. Shabih, M. V. Gil, S. Miret, C. T. Koch, J. A. Márquez and K. M. Jablonka, From text to insight: large language models for chemical data extraction, *Chem. Soc. Rev.*, 2025, **54**, 1125.

7 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, The FAIR guiding

 © 2026 The Author(s). Published by the Royal Society of Chemistry

principles for scientific data management and stewardship, *Sci. Data*, 2016, **3**, 160018.
8 D. Donoho, Data science at the singularity, *arXiv*, 2023, preprint arXiv:2310.00865.
9 R. Ramachandran, K. Bugbee and K. Murphy, From open data to open science, *Earth Space Sci.*, 2021, **8**, e2020EA001562.
10 I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson and Y. Gal, AI models collapse when trained on recursively generated data, *Nature*, 2024, **631**, 755.
11 Crystallography: Protein Data Bank, *Nat. New Biol.* **233**, 223 (1971).
12 F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. M Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, The protein data bank: A computer-based archival file for macromolecular structures, *J. Mol. Biol.*, 1977, **112**, 535.
13 J. D. Westbrook, J. Y. Young, C. Shao, Z. Feng, V. Guranovic, C. L. Lawson, B. Vallat, P. D. Adams, J. M. Berrisford, G. Bricogne, K. Diederichs, R. P. Joosten, P. Keller, N. W. Moriarty, O. V. Sobolev, S. Velankar, C. Vonrhein, D. G. Waterman, G. Kurisu, H. M. Berman, S. K. Burley and E. Peisach, PDBx/mmCIF ecosystem: foundational semantic tools for structural biology, *J. Mol. Biol.*, 2022, **434**, 167599.
14 T. Terwilliger and G. Bricogne, Continuous mutual improvement of macromolecular structure models in the pdb and of x-ray crystallographic software: the dual role of deposited experimental data, *Acta Crystallogr. D*, 2014, **70**, 2533.
15 wwPDB, wwPDB policies and processing procedures document. Section A: wwPDB policies, *Tech. Rep.* (wwPDB, 2024).
16 M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Žídek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis and S. Velankar, Alphafold protein structure database in 2024: providing structure coverage for over 214 million protein sequences, *Nucleic Acids Res.*, 2024, **52**, D368.
17 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, Highly accurate protein structure prediction with alphafold, *Nature*, 2021, **596**, 583.
18 M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Žídek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis and S. Velankar, Alphafold protein structure database in 2024: providing structure coverage for over 214 million protein sequences, *Nucleic Acids Res.*, 2023, **52**, D368–D375.
19 F. Zhao and J. Xu, A position-specific distance-dependent statistical potential for protein structure and functional study, *Structure*, 2012, **20**, 1118.
20 N. Methods, Data sharing comes to structural biology, *Nat. Methods*, 2016, **13**, 381.
21 M. Barker, N. P. Chue Hong, D. S. Katz, A.-L. Lamprecht, C. Martinez-Ortiz, F. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez and T. Honeyman, Introducing the FAIR principles for research software, *Sci. Data*, 2022, **9**, 622.
22 L. M. Ghiringhelli, C. Baldauf, T. Bereau, S. Brockhauser, C. Carbogno, J. Chamanara, S. Cozzini, S. Curtarolo, C. Draxl, S. Dwaraknath, Á. Fekete, J. Kermode, C. T. Koch, M. Kühbach, A. N. Ladines, P. Lambrix, M.-O. Himmer, S. V. Levchenko, M. Oliveira, A. Michalchuk, R. E. Miller, B. Onat, P. Pavone, G. Pizzi, B. Regler, G.-M. Rignanese, J. Schaarschmidt, M. Scheidgen, A. Schneidewind, T. Sheveleva, C. Su, D. Usvyat, O. Valsson, C. Wöll and M. Scheffler, Shared metadata for data-centric materials science, *Sci. Data*, 2023, **10**, 626.
23 S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko and D. Morgan, AFLOW: An automatic framework for high-throughput materials discovery, *Comput. Mater. Sci.*, 2012, **58**, 218.
24 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.*, 2013, **1**, 011002.
25 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, *npj Comput. Mater.*, 2015, **1**, 15010.
26 C. Draxl and M. Scheffler, The NOMAD laboratory: from data sharing to artificial intelligence, *Int. J. Mater. Phys.*, 2019, **2**, 036001.
27 T. Cavignac, J. Schmidt, P.-P. De Breuck, A. Loew, T. F. T. Cerqueira, H.-C. Wang, A. Bochkarev, Y. Lysogorskiy, A. H. Romero, R. Drautz, S. Botti, and M. A. L. Marques, *Ai-driven expansion and application of the alexandria database* (2025).
28 L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, C. S. Adorf, C. W. Andersen, O. Schütt, C. A. Pignedoli, D. Passerone, J. VandeVondele, T. C. Schulthess, B. Smit, G. Pizzi and N. Marzari, Materials Cloud, a platform for open computational science, *Sci. Data*, 2020, **7**, 299.
29 M. L. Evans, J. Bergsma, A. Merkys, C. W. Andersen, O. B. Andersson, D. Beltrán, E. Blokhin, T. M. Boland, R. Castañeda Balderas, K. Choudhary, A. Díaz Díaz, R. Domínguez García, H. Eckert, K. Eimre, M. E. Fuentes Montero, A. M. Krajewski, J. J. Mortensen, J. M. Nápoles Duarte, J. Pietryga, J. Qi, F. d. J. Trejo Carrillo, A. Vaitkus, J. Yu, A. Zettel, P. B. de Castro, J. Carlsson, T. F. T. Cerqueira, S. Divilov, H. Hajiyani, F. Hanke, K. Jose, C. Oses, J. Riebesell, J. Schmidt, D. Winston, C. Xie, X. Yang, S. Bonella, S. Botti, S. Curtarolo, C. Draxl,

© 2026 The Author(s). Published by the Royal Society of Chemistry

L. E. Fuentes Cobas, A. Hospital, Z.-K. Liu, M. A. L. Marques, N. Marzari, A. J. Morris, S. P. Ong, M. Orozco, K. A. Persson, K. S. Thygesen, C. Wolverton, M. Scheidgen, C. Toher, G. J. Conduit, G. Pizzi, S. Gražulis, G.-M. Rignanese and R. Armiento, Developments and applications of the OPTIMADE API for materials discovery, design, and data exchange, *Digit. Discov.*, 2024, **3**, 1509.

30 L. Ward, S. Babinec, E. J. Dufek, D. A. Howey, V. Viswanathan, M. Aykol, D. A. Beck, B. Blaiszik, B.-R. Chen, G. Crabtree, S. Clark, V. De Angelis, P. Dechent, M. Dubarry, E. E. Eggleton, D. P. Finegan, I. Foster, C. B. Gopal, P. K. Herring, V. W. Hu, N. H. Paulson, Y. Preger, D. Uwe-Sauer, K. Smith, S. W. Snyder, S. Sripad, T. R. Tanim and L. Teo, Principles of the battery data genome, *Joule*, 2022, **6**, 2253.

31 A. White, The Materials Genome Initiative: One year on, *MRS Bull.*, 2012, **37**, 715.

32 Materials Research Data Alliance (MaRDA), Home, https://www.marda-alliance.org/2025, accessed: 2025-08-15.

33 European Research Council, Open research data and data management plans, https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf 2022.

34 Deutsche Forschungsgemeinschaft (DFG), Handling of research data, https://www.dfg.de/en/research-funding/funding-initiative/research-data 2025.

35 Swiss National Science Foundation, Open research data, https://www.snf.ch/en/dMILj9t4LNk8NwyR/topic/open-research-data 2025.

36 Croatian Science Foundation, Research data management plan, https://hrzz.hr/plan-upravljanja-istrazivackim-podacima-za-projekte-hrvatske-zaklade-za-znanost/2022.

37 The Icelandic Centre for Research, Scientific publication: Policy on open access, https://en.rannis.is/activities/open-access/2025, accessed: June 18, 2025.

38 Thematic Digital Competence Centre (TDCC), Natural & engineering sciences (tdcc-nes), https://tdcc.nl/about-tddc/nes/, 2025, accessed: 2026-02-02.

39 Ruhr-University Bochum, Leitlinien zum forschungsdatenmanagement, https://researchdata.ruhr-uni-bochum.de/wp-content/uploads/2024/07/LeitlinienFDM.pdf 2018.

40 Vilnius University, Vilnius university open science policy guidelines, https://www.vu.lt/site_files/Vertimai/EN_Translation_VU_atvirojo_mokslo_politikos_gaires.pdf, 2022.

41 N. Hartl, E. Wössner and Y. Sure-Vetter, Nationale forschungsdateninfrastruktur (nfdi), *Informatik-Spektrum*, 2021, **44**, 370.

42 C. Eberl, M. Niebel, E. Bitzek, T. Dahmen, F. Fritzen, P. Gumbsch, T. Hickel, S. Klein, F. Mücklich, M. S. Müller and *et al.*, *Consortium proposal NFDI-MatWerk* 2021.

43 A. E. Mansour, L. Rotheray, K. Helbig, S. Botti, H. B. Weber, M. Aeschlimann and C. Draxl, Fairmat guide to writing data management plans: A practical guide for the condensed-matter physics and materials-science communities, CoRDI2023, *Proc. Conf. Res. Data Infrastr.*, 2023, **1**, 1–4.

44 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, Data-driven materials science: Status, challenges, and perspectives, *Adv. Sci.*, 2019, **6**, 1900808.

45 D. G. E. Gomes, P. Pottier, R. Crystal-Ornelas, E. J. Hudgins, V. Foroughirad, L. L. Sánchez-Reyes, R. Turba, P. A. Martinez, D. Moreau, M. G. Bertram, C. A. Smout and K. M. Gaynor, Why don't we share data and code? perceived barriers and benefits to public archiving practices, *Proc. Biol. Sci.*, 2022, **289**, 20221113.

46 Y. Kim and J. M. Stanton, Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis, *J. Assoc. Inf. Sci. Technol.*, 2016, **67**, 776.

47 V. Stodden, P. Guo and Z. Ma, Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals, *PLoS One*, 2013, **8**, e67111.

48 N. A. Vasilevsky, J. Minnier, M. A. Haendel and R. E. Champieux, Reproducible and reusable research: are journal data sharing policies meeting the mark?, *PeerJ*, 2017, **5**, e3208.

49 K. W. McCain, Mandating sharing: Journal policies in the natural sciences, *Sci. Commun.*, 1995, **16**, 403.

50 C. Barbui, Sharing all types of clinical data and harmonizing journal standards, *BMC Med.*, 2016, **14**, 63.

51 D. B. Resnik, M. Morales, R. Landrum, M. Shi, J. Minnier, N. A. Vasilevsky and R. E. Champieux, Effect of impact factor and discipline on journal data sharing policies, *Account. Res.*, 2019, **26**, 139.

52 A. M. Rousi and M. Laakso, Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research, *Scientometrics*, 2020, **124**, 131.

53 M. Crosas, J. Gautier, S. Karcher, D. Kirilova, G. Otalora, A. Schwartz, Data policies of highly-ranked social science journals, *SocArXiv*, preprint, 2018, DOI: 10.31235/osf.io/9h7ay.

54 E. Castro, M. Crosas, A. Garnett, K. Sheridan and M. Altman, Evaluating and promoting open data practices in open access journals, *J. Sch. Publish.*, 2017, **49**, 66.

55 L. Naughton and D. Kernohan, Making sense of journal research data policies, *UKSG Insights*, 2016, **29**, 84.

56 B. Blahous, J. Gorraiz, C. Gumpenberger, O. Lehner, and U. Ulrych, Data policies in journals under scrutiny: their strength, scope and impact, Bibliometrie-Praxis und Forschung 5, DOI: 10.5283/BPF.269 (2016).

57 J. Herndon and R. O'Reilly, Data sharing policies in social sciences academic journals: Evolving expectations of data sharing as a form of scholarly communication, in *Databrarianship: The academic data librarian in theory and practice*, ed. L. Kellam and K. Thompson, American Library Association, Chicago, 2016, pp. 219–242.

58 P. Sturges, M. Bamkin, J. H. Anders, B. Hubbard, A. Hussain and M. Heeley, Research data sharing: Developing a stakeholder-driven model for journal policies, *J. Assoc. Inf. Sci. Technol.*, 2015, **66**, 2445.

 © 2026 The Author(s). Published by the Royal Society of Chemistry

59 W. Zenk-Möltgen and G. Lepthien, Data sharing in sociology journals, *Online Inf. Rev.*, 2014, **38**, 709.

60 N. Moles, Data-pe: A framework for evaluating data publication policies at scholarly journals, *Data Sci. J.*, 2015, **13**, 192.

61 s. gherghina and a. katsanidou, Data availability in political science journals, *Eur. Polit. Sci.*, 2013, **12**, 333.

62 N. M. Weber, H. A. Piwowar, and T. J. Vision, Evaluating data citation and sharing policies in the environmental sciences: Evaluating data citation and sharing policies in the environmental sciences, *Proceedings of the American Society for Information Science and Technology* 47, 1 (2010).

63 H. Piwowar and W. Chapman, A review of journal policies for sharing research data, *Nature Precedings*, 2008, DOI: 10.1038/npre.2008.1700.1.

64 R. Grant and I. Hrynaszkiewicz, The impact on authors and editors of introducing data availability statements at nature journals, in *Proceedings of the International Digital Curation Conference*, 2018, submitted to the International Journal of Digital Curation.

65 B. Suhr, J. Dungl and A. Stocker, Search, reuse and sharing of research data in materials science and engineering—a qualitative interview study, *PLoS One*, 2020, **15**, 1.

66 M. Houillon, J. Klar, Z. Boutanios, T. Stary, T. Cojean, H. Anzt and A. Loewe, FACILE-RS: archiving and long-term preservation of research software repositories made easy, *J. Open Source Softw.*, 2025, **10**, 7330.

67 D. Science, Dimensions api, https://api-lab.dimensions.ai/, accessed: April, 2024.

68 M. Gabelica, R. Bojčić and L. Puljak, Many researchers were not compliant with their published data sharing statement: a mixed-methods study, *J. Clin. Epidemiol.*, 2022, **150**, 33.

69 D. G. Hamilton, M. J. Page, S. Finch, S. Everitt and F. Fidler, How often do cancer researchers make their data and code available and what factors are associated with sharing?, *BMC Med.*, 2022, **20**, 438.

70 L. Tedersoo, R. Küngas, E. Oras, K. Köster, H. Eenmaa, A. Leijen, M. Pedaste, M. Raju, A. Astapova, H. Lukner, K. Kogermann and T. Sepp, Data sharing practices and data availability upon request differ across scientific disciplines, *Sci. Data*, 2021, **8**, 192.

71 V. Stodden, J. Seiler and Z. Ma, An empirical analysis of journal policy effectiveness for computational reproducibility, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 2584–2589.

72 S. Bloodworth, C. Willoughby and S. J. Coles, Data accessibility in the chemical sciences: an analysis of recent practice in organic chemistry journals, *Beilstein J. Org. Chem.*, 2025, **21**, 864.

73 Parasoft, ISO 26262 software compliance in the automotive industry, *Tech. Rep.*, 2025.

74 NASA, NASA software safety guidebook, Standard NASA-GB-8719**13** NASA, 2004.

75 N. A. Parks, T. G. Fischer, C. Blankenburg, V. F. Scalfani, L. R. McEwen, S. Herres-Pawlis and S. Neumann, The current landscape of author guidelines in chemistry through the lens of research data sharing, *Pure Appl. Chem.*, 2023, **95**, 439.

76 L. N. Joppa, G. McInerny, R. Harper, L. Salido, K. Takeda, K. O'Hara, D. Gavaghan and S. Emmott, Troubling trends in scientific software use, *Science*, 2013, **340**, 814.

77 J. Brase, Datacite - a global registration agency for research data, in *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, 2009 pp. 257–261.

78 J. Neumann and J. Brase, Datacite and doi names for research data, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 1035.

79 M. Barker, N. P. Chue Hong, D. S. Katz, A.-L. Lamprecht, C. Martinez-Ortiz, F. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez and T. Honeyman, Introducing the FAIR Principles for research software, *Sci. Data*, 2022, **9**, 622.

80 T. Fischer, W. Vollprecht, B. Zalmstra, R. Arts, T. de Jager, A. Fontan, A. D. Hines, M. Milford, S. Traversaro, D. Claes, and S. Raine, Pixi: Unified software development and distribution for robotics and ai 2025.

81 European Organization For Nuclear Research and OpenAIRE, Zenodo 2013.

82 G. Wilson, D. A. Aruliah, C. T. Brown, N. P. Chue Hong, M. Davis, R. T. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White and P. Wilson, Best practices for scientific computing, *PLoS Biol.*, 2014, **12**, e1001745.

83 R. C. Jiménez, M. Kuzak, M. Alhamdoosh, M. Barker, B. Batut, M. Borg, S. Capella-Gutierrez, N. C. Hong, M. Cook, M. Corpas, M. Flannery, L. Garcia, J. Ll. Gelpí, S. Gladman, C. Goble, M. G. Ferreiro, A. Gonzalez-Beltran, P. C. Griffin, B. Grüning, J. Hagberg, P. Holub, R. Hooft, J. Ison, D. S. Katz, B. Leskošek, F. L. Gómez, L. J. Oliveira, D. Mellor, R. Mosbergen, N. Mulder, Y. Perez-Riverol, R. Pergl, H. Pichler, B. Pope, F. Sanz, M. V. Schneider, V. Stodden, R. Suchecki, R. S. Vařeková, H.-A. Talvik, I. Todorov, A. Treloar, S. Tyagi, M. van Gompel, D. Vaughan, A. Via, X. Wang, N. S. Watson-Haigh and S. Crouch, Four simple recommendations to encourage best practices in research software, *F1000Research*, 2017, **6**, 876.

84 N. Blumenröhr, P.-J. Ost, F. Kraus, and A. Streit, FAIR Digital Objects for the realization of globally aligned data spaces, in *2024 IEEE International Conference on Big Data (BigData)*, IEEE, 2024, pp. 374–383.

85 S. Soiland-Reyes, C. Goble and P. Groth, Evaluating FAIR Digital Object and Linked Data as distributed object systems, *PeerJ Comput. Sci.*, 2024, **10**, e1781.

86 I. Anders, C. Blanchi, D. Broder, M. Hellström, S. Islam, T. Jejkal, L. Lannom, K. P. von Gehlen, R. Quick, A. Schlemmer, U. Schwardmann, S. Soiland-Reyes, S. George, D. van Uytvanck, C. Weiland, P. Wittenburg, and C. Zwölf, FDO Forum FDO requirement specifications, v3.0, Zenodo DOI: 10.5281/ZENODO.7782262 (2023.

87 H. Koers, D. Bangert, E. Hermans, R. van Horik, M. de Jong and M. Mokrane, Recommendations for services in a FAIR data ecosystem, *Patterns*, 2020, **1**, 100058.

© 2026 The Author(s). Published by the Royal Society of Chemistry

88 The Carpentries2025, Online; accessed 20. Nov. 2025.
89 CodeRefinery2025, Online; accessed 20. Nov. 2025.
90 L. Hörmann, H. Myneni, R. K. S. Al-Hamd, K. Batalović, S. Bonfanti, F. Grasselli, S. Gražulis, B. Koç, K. Konstantinou, I. Lončarić, N. Lopanitsyna, J. M. Oliveira, P. Pegolo, P. Ramos, K. Rossi, S. P. Schwaminger, E. Simmen, M. Todorović, M. Stricker, and J. Schmidt, Code and Data for Article "Journal Research Data Policies in Materials Science" (Digital Discovery), Zenodo DOI: 10.5281/zenodo.20529191, 2026.

 © 2026 The Author(s). Published by the Royal Society of Chemistry