

Nicole A. Parks, Tillmann G. Fischer, Claudia Blankenburg, Vincent F. Scalfani, Leah R. McEwen, Sonja Herres-Pawlis and Steffen Neumann*

The current landscape of author guidelines in chemistry through the lens of research data sharing

<https://doi.org/10.18154/RWTH-2023-02772>

Abstract: As the primary method of communicating research results, journals garner an enormous impact on community behavior. Publishing the underlying research data alongside journal articles is widely considered good scientific practice. Ideally, journals and their publishers place these recommendations or requirements in their author guidelines and data policies. Several efforts are working to improve the infrastructure, processes, and uptake of research data sharing, including the NFDI4Chem consortium, working groups within the RDA, and IUPAC, including the WorldFAIR Chemistry project. In this article, we present the results of a large-scale analysis of author guidelines from several publishers and journals active in chemistry research, showing how well the publishing landscape supports different criteria and where there is room for improvement. While the requirement for deposition of X-ray diffraction data is commonplace, guidelines rarely mention machine-readable chemical structures and metadata/minimum information standards. Further evaluation criteria included recommendations on persistent identifiers, data availability statements, data deposition into repositories as well as of open analytical data formats. Our survey shows that publishers and journals are starting to include aspects of research data in their guidelines. We as authors should accept and embrace the guidelines with increasing requirements for data availability, data interoperability, and re-usability to improve chemistry research.

Keywords: Academic publishing; cheminformatics; chemistry; data repositories; education; FAIR research data; interoperability; metadata; research data; scientific journals; standards; validation; workflows.

Introduction

When asked whether they have read an article in the last year, e.g., in a graduate course on scientific writing, every researcher will raise their hand. When inquiring who then was satisfied with the completeness of information provided, some hands are lowered. Only few remain raised when continuing to ask who was able to find and access the actual underlying research data so that they can reproduce the work or build upon it.

*Corresponding author: Steffen Neumann, <https://orcid.org/0000-0002-7899-7192>, Computational Plant Biochemistry, Leibniz Institute of Plant Biochemistry, Halle, Germany, e-mail: sneumann@ipb-halle.de, <https://ror.org/01mzk5576>.

Nicole A. Parks, <https://orcid.org/0000-0002-6243-2840>, IT Center, RWTH Aachen University, Aachen, Germany, <https://ror.org/04xfq0f34>.

Tillmann G. Fischer, <https://orcid.org/0000-0003-4480-8661>, Claudia Blankenburg, <https://orcid.org/0000-0002-7955-8378>, Computational Plant Biochemistry, Leibniz Institute of Plant Biochemistry, Halle, Germany, <https://ror.org/01mzk5576>.

Vincent F. Scalfani, <https://orcid.org/0000-0002-7363-531X>, University Libraries, Rodgers Library for Science and Engineering, University of Alabama, Tuscaloosa, AL, USA, <https://ror.org/03xrjrk67>.

Leah R. McEwen, <https://orcid.org/0000-0003-2968-1674>, Cornell University Library, Cornell University, Ithaca, NY, USA, <https://ror.org/05bnh6r87>.

Sonja Herres-Pawlis, <https://orcid.org/0000-0002-4354-4353>, Institute of Inorganic Chemistry, RWTH Aachen University, Aachen, Germany, <https://ror.org/04xfq0f34>.

Publishing the underlying research data alongside journal articles is widely considered good scientific practice [1, 2]. In simplest terms, providing these data ensures transparency and is crucial for others to be able to build upon the published results. However, sharing data is not yet as common as sharing results, and more sharing requires a cultural shift [3]. As the primary method of communicating research results, journals garner a great amount of influence on community behavior. This includes how researchers handle the data on which their published results are based. For example, data availability statements included in the back matter of a manuscript not only provide information on how a reader may find and access the data—requiring such statements also encourages the author to think about whether and how they will make their research data available to others in the first place.

Journals are able to guide researchers in making their data discoverable by recommending that the data underlying the published research results should be deposited in an appropriate recommended repository. Ideally, journals and their publishers place these recommendations in their author guidelines and data policies.

For several decades, publishers and journals have been providing instructions to scientists in the form of author guidelines, with the aim that articles published in a journal have a consistent style, length, or, e.g., a maximum number of figures. Predominantly chemistry journals have also included specific recommendations and directives concerning chemical information. As of 1953, *The Analyst* provided instructions [4] on abbreviations and noted that units and formulae may only be used in addition to the names of elements and compounds. Guidelines from the 1970s in *Chemical Communications* [5] required authors to follow symbols or nomenclature as established by, e.g., the International Union of Pure and Applied Chemistry (IUPAC).

Early examples of guidelines specifically addressing data, such as deposition into the Cambridge Structural Database (CSD) with the Cambridge Crystallographic Data Centre (CCDC) for small structures as required by *Angewandte Chemie* in 1992, were less common [6]. More detailed instructions also began to appear, including how to list physical measurement data such as boiling and melting point as well as spectroscopic data within the text [7].

Data repositories have also provided guidance for research data. For example, the Protein Data Bank (PDB) for protein structures provides instructions for journals [8] on how deposition to PDB should be requested in their respective author guidelines and the CCDC has worked with many publishers to improve the guidance for crystal structure deposition [9]. Therefore, sharing data along with research results has started to go hand in hand.

While titled “author” guidelines, these are important for different stakeholders, e.g., for editorial staff and reviewers, who check and can request adherence to the guidelines, including the scientific and quality aspects of the research data. Fig. 1 illustrates the relevant stakeholders with respect to author guidelines, as well as aspects covered in author guidelines and their subsequent impact on the publication of research outputs.

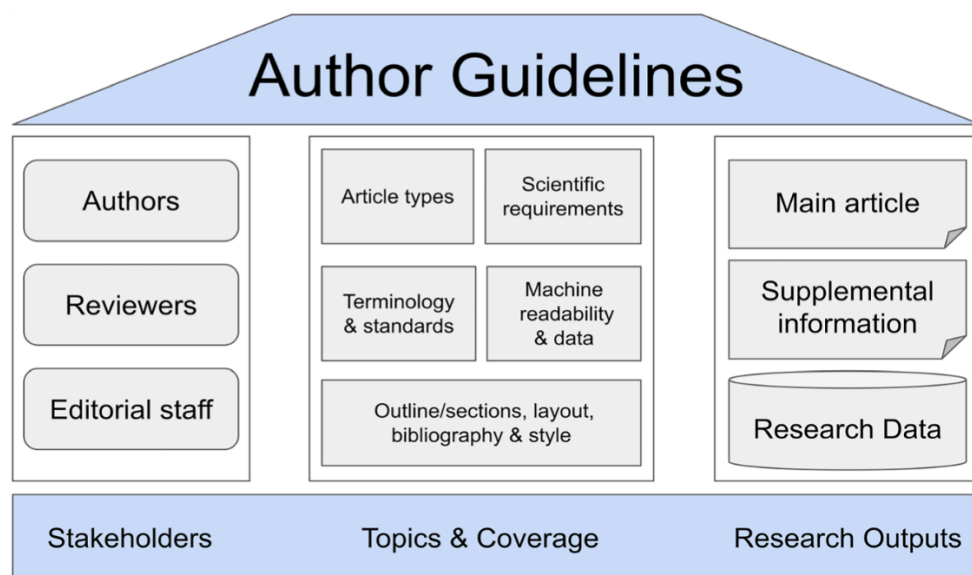


Fig. 1: Author guidelines are relevant for different stakeholders who can obtain information on topics and scope of a journal to present, review or edit research outputs compliant with a journal’s guideline.

An important aspect in the publishing of today and the future is an increased level of machine-readability to support finding, navigating, and analyzing the publication space, a concept in line with the FAIR Data Principles [10]. This requires unique and persistent identifiers (PIDs) to exactly pinpoint digital items in that (vast) space. The most prominent example is the Digital Object Identifier (DOI) [11], the use of which began at the end of the last century and is now commonplace in academic publishing. The ORCID iD [12] identifies researchers (i.e., authors) and is on its way to reach a similar uptake as the DOI in just 10 years. Similarly, the Research Organization Registry (ROR) may be used as an identifier for institutions [13].

Several efforts are underway to improve the infrastructure, processes, and uptake of research data sharing in chemistry. To set the stage, a workshop on “FAIR Publishing Guidelines for Spectral Data and Chemical Structures” was organized by some of the authors (LRM and VFS) and held on March 29–30, 2019 in Orlando, FL (USA). Among other topics [14, 15], a survey and checklist on Chemistry Journal Data Submission and Sharing Policies was presented [16]. This checklist was created for a Research Data Alliance (RDA) Publisher Forum presentation [17], a broad community venue for discussion on RDM efforts across domains and publishing, such as the respective Chemistry Research Data Interest Group [18] and the Data Policy Standardisation and Implementation Interest Group [19].

The NFDI4Chem [20] consortium aims to develop infrastructure to support chemists in making their research data findable, available, interoperable, and re-usable, or FAIR (see FAIR Data Principles [10]). The adoption of the FAIR principles requires a shift throughout the community, where research data is readily shared on established and appropriate repositories and standards on content, metadata, and data formats are adhered to. Open research practices, such as the use of research registries and preprint servers, are encouraged to increase transparency throughout the community.

IUPAC is leading a project on FAIR chemical data standards under the global WorldFAIR Initiative led by CODATA and the Research Data Alliance to advance FAIR data sharing. The goal of WorldFAIR Chemistry is to review gaps in FAIR implementation of chemical data and enable the development and adoption of chemical data standards in research workflows to enable downstream data reuse [21, 22].

In this article, we are building on the initial chemistry data checklist work [16], and present the results of a large-scale analysis of author guidelines from several publishers and journals active in chemistry research. The main aim is to provide an overview of how well the publishing landscape currently supports criteria needed to enable FAIR data sharing and how it supports an interoperable data ecosystem and considers FAIR digital research data as first-class research output in publications.

Methodology

To assess the current journal requirements for research data in chemistry, the author guidelines and policies from 42 journals, released by 13 publishers, were reviewed. The selected journals were sourced based on where the various chemists in the NFDI4Chem consortium commonly publish or source information and are considered to be common and established and therefore representative of the prevailing publishing landscape. The full list of journals are available as table `JournalGuidelinesDataCollection` in the supplementary data published on RADAR4Chem [23].

Through the lens of research data, our review criteria intend to survey how journals support authors in adhering to the FAIR Data principles and Open Science [24] practices during the publication process. Therefore, the review incorporated categories which focus on the data underlying the article, such as data publication through deposition into repositories. It also included aspects of the article (prior publication on preprint servers), its metadata (use of ORCID iDs), and the backmatter (implementation of data availability statements). Many of these criteria, e.g., minimum information standards and publishing the underlying data to journal articles in subject-specific repositories, reflect efforts being put forth by NFDI4Chem [20]. While many also reflect aspects of the FAIR Data principles, the publication process includes aspects that go beyond the scope of FAIR. Table 1 summarizes these criteria.

Table 1: Criteria for author guidelines evaluation. The Category “Article” includes aspects to build a publication map or network connecting papers, authors and the data. Category “Information and Data” includes requirements whether and where to deposit data, as well as guidelines on minimum levels of information and interoperability through (open) and machine readable data formats.

Pertains to	Category	Criterium	Relevance
Article	ORCID iD [12]	Whether authors and/or coauthors must submit their ORCID iD.	Promotes unambiguous and persistent author identification, which is part of rich and descriptive metadata for the scientific article and the associated dataset.
	Preprint servers	Whether manuscripts posted on open access preprint servers such as arXiv, bioRxiv, or ChemRxiv prior to publication in the journal will be considered/whether posting on preprint servers is encouraged.	Publishing on preprint servers prior to journal publication furthers transparency and accessibility within the research community.
	Data availability statement	Whether a journal expects a data availability statement. The exact wording of the statement was not taken into account.	Provides information on how underlying research data can be found and accessed.
Information and data	Data deposition into repository	Whether journals expect underlying data to be published in a repository. This was split into all data, nuclear magnetic resonance (NMR) data, and X-ray diffraction (XRD) data.	Ensures research data are findable and accessible. XRD and NMR represent two of the most common analytical methods in chemistry.
	Recommended repositories	Whether the journal explicitly suggests subject specific and/or generic repositories. Subject specific repositories include limited types of data but often have advanced features for data analysis or visualization; generic repositories may accept any type of research data regardless of subject, method, or format. ^a	Assists authors in choosing suitable repositories for their data, thus enhancing accessibility and findability, while ensuring adherence to community standards.
	Metadata/minimum information requirements	Whether the journal expects authors to follow (field-)specific guidelines and standards regarding minimal descriptive information with respect to the context of the research data.	Enhances findability and reusability of datasets.
	Open analytical data formats	Whether the journal lists specific open file formats for analytical data. ^b	Ensures the datasets are interoperable and reusable.
	Machine readable structures (InChI, SMILES)	Whether a journal requires chemical structures to be reported in machine-readable formats	Ensures data provided on chemical structures is findable, interoperable and can be interpreted by machines.

^aThe publication of crystallographic data is well-established. Author guidelines which solely mentioned the Cambridge Structural Database (CSD) as well as the Crystallography Open Database (COD) were not considered as guidelines clearly considering and pointing to the concept of field-specific repositories, also for other fields of research. These guidelines are covered by the sub-category only XRD under Data deposition into repository, as shown in the shared data [23]. ^bThe Crystallographic Information File (CIF) is a well-established open analytical data format standard in crystallography. Author guidelines which exclusively mention CIF files were not considered as taking a clear stance on the concept of open analytical data formats for other areas. Use of CIF files is covered by the sub-category only XRD under Data deposition into repository, as shown in the data [23]. Please also see footnote a.

A team consisting of three scientists with expertise in chemistry and data management divided the journals up amongst themselves, grouped by the associated publisher. Each team member reviewed their groups of journals individually, then gathered feedback in group discussions. Furthermore, the reviews were cross-checked by the team members to ensure consistent rating and increase objectivity. The first round of reviews were carried out in September 2021, the results of which were presented at the Editors4Chem Workshop organized jointly by NFDI4Chem and IUPAC in November 2021. Following this workshop, a second review took place in January of 2022 to correct possible inconsistencies noted by the workshop participants and to incorporate changes to guidelines that occurred after the workshop. The results reported here and depicted in Fig. 2 pertain to the second and final review.

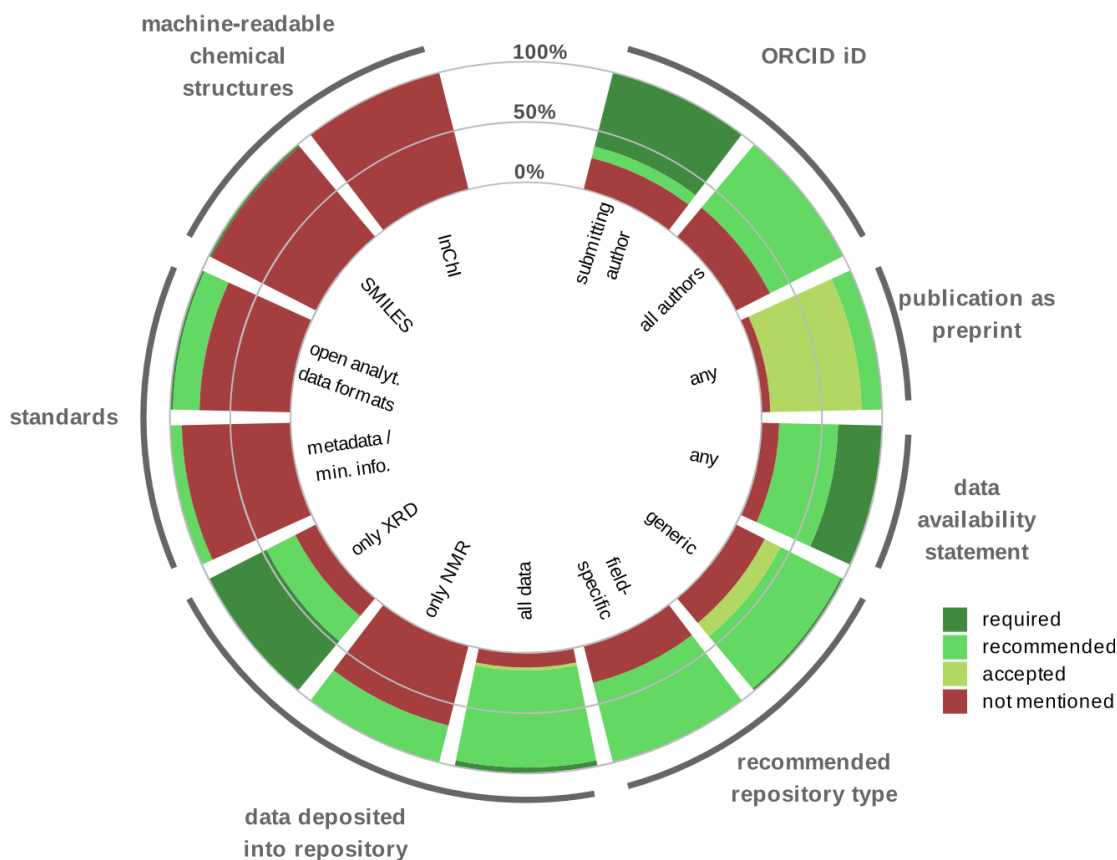


Fig. 2: Overview of requirements and recommendations for FAIR and open data sharing in author guidelines. This circular plot shows the selected categories (outer legend) and criteria (inner legend) of the guidelines. Segments were classified into four categories: from “required” (dark green), “recommended” (green), “accepted” (light green) to “not mentioned” (red) pertaining to the article (upper right) and digital research data. E.g., the ORCID iD for the submitting author was required by 63 % of journals, recommended by another 10 %, and not mentioned by the remaining 27 %.

The selected author guidelines were reviewed for each category and assessed as to their level of requirement as shown in Fig. 2. The color scale indicates which criteria were determined to align with FAIR Data and Open Science practices from most (dark green) to least (red). Specifically, dark green signifies the portion of the journals that required a certain practice or stated that a practice is mandatory, light green signifies the portion that recommended a practice, also described as a practice authors are requested to follow or which is supported by the journal’s guideline. Yellow-green indicates the portion which stated that they accepted a practice, and red the portion that did not mention a practice at all. The absence of categories such as “is not recommended” or “is not accepted” is intentional, as no criteria fell into these categories.

No further input was collected from the journals aside from their online, publicly available author guidelines. Publisher guidelines were included in this evaluation only when referenced specifically within a journal’s guideline. The term “author guidelines” is interpreted to be inclusive, as linked or accompanying “data policies” or “publishing policies” were also considered.

Since author guidelines are continuously evolving, we collected snapshots of the web pages to document the state at the beginning of 2022 and provide these as supplementary information guideline_screenshots published on RADAR4Chem [23]. Perceptible changes of author guidelines towards FAIR research data, compared to before the Editors4Chem Workshop in November 2021 was held, were also noted and documented.

The review results were collected in the spreadsheet [23] JournalGuidelinesDataCollection, summarized and visualized via the accompanying Jupyter notebook using Python and the statistics language R.

As mentioned, this review aims to gain an understanding of the current state regarding research data in academic publishing in chemistry. Therefore, the data were evaluated to give percentages of journals that fell into the respective criteria for each category. This review does not intend to single out or rate individual journals or publishers according to the criteria.

Results and discussion

In the following we summarize and discuss the results of the survey evaluation data. Fig. 2 visualizes the results. The code to reproduce the figure is available as a Jupyter notebook online and also as part of the supplementary data, see the README in the supplementary data [23].

Categories and criteria pertaining to the article

DOIs are assigned to basically all scientific articles [23], as well as an increasing number of datasets in data repositories. Nonetheless, 63 % of journals required ORCID iDs solely for submitting authors, while 10 % recommended its use. There was no requirement for all authors to submit an ORCID iD, while 68 % recommended that they do. About a quarter of the journals did not mention ORCID.

Notably, journals generally accepted manuscripts previously published on preprint servers, with only 7 % not mentioning them in their guidelines, while 76 % accepted them and an additional 17 % recommended them. This acceptance by the majority of journals shows a promising push towards transparency in the publishing landscape.

The results for data availability statements show a clear tendency towards their acceptance, as 51 % recommended including these and 31 % required that authors submit such statements, while 17 % did not mention them. In other words, over 80 % of journals either required or recommended the submission of data availability statements. This is promising, as it encourages authors to reflect upon how they make their data available to others. Moreover, publishers which previously recommended data availability statements have started to require these [23, 25]. However, it should be noted that the criteria, as mentioned in Table 1, do not necessarily account for the content of the data availability statement. Therefore, these may include vague statements such as “Underlying research data are available upon reasonable request to the authors” or “Underlying research data are available from the online supplementary material”. The latter often provides analytical data only as pictures in PDF files rather than depositing the actual analytical data files into repositories.

Perceivable changes in author guidelines on data availability statements were noted between our internal review in September 2021 and re-examination in January 2022, turning the overall picture from nearly half red to mostly green. This observation is partially due to two publishers adding links within most of their surveyed journal author guidelines to previously hardly findable publisher-wide data policies, which also provide examples of these statements. Journals from another publisher tightened their guidelines, moving from recommending data availability statements to requiring them. While 43 % did not mention data availability statements in September 2021, this dropped to only 15 % in January 2022. These changes indicate a move in the direction of FAIR research data publication as also previously described for other publishers [26], albeit with a less pronounced focus on chemistry. This signifies a commendable awareness on behalf of the academic publishing community.

Categories and criteria pertaining to the research data

Only a few journals (5 %) required that authors make all underlying data available by deposition into a repository, with 85 % of the journals recommending data sharing. 2 % stated that they accept it and 12 % did not mention this practice. It should be stated here that during the time period between our two reviews, some journals also updated their guidelines to include data sharing via repositories, which, again, underlines the shift towards supporting FAIR and open data practices in publishing.

With Nuclear magnetic resonance (NMR) spectroscopy and X-ray diffraction (XRD) crystallographic data being the most common analytical methods in chemistry, requirements for these were reviewed specifically. As expected, deposition of XRD data in repositories such as the CSD is commonplace, as 54 % require and 27 % recommend the

deposition of this type of data. About a fifth of the reviewed journals do not mention this data type, however, this is often simply due to the nature of these journals' subject areas.

On the other hand, journals mentioned depositing NMR data rather infrequently, with merely 32 % of journals listing it specifically, in almost all cases without any recommendations on repositories. This reflects the lack of an established repository for this type of data, while the handling of crystallographic data, in comparison, is well established and could be used as a template for this (as well as other) chemical data types.

A shift in the requirements is generally expected as more repositories become both available and accepted by the respective communities and since funders such as the Deutsche Forschungsgemeinschaft (DFG [German Research Foundation]) have started to encourage their use. The results for repositories explicitly listed in the author guidelines best represent this. Many generic repositories, such as Zenodo, Dryad, and Figshare, exist and were required, recommended, or accepted by 73 % of journals surveyed. In terms of field-specific repositories, 32 % did not give any recommendations. As for the remainder, suggestions included the PDB, Standards for reporting Enzymology Data Database (STRENDa DB), and Novel Materials Discovery (NOMAD) repository, to name a few, with some gaps for certain data types, especially in the area of spectroscopic data. A clear push towards data sharing exists and many publishers and journals are keen on assisting authors in finding an appropriate platform for data deposition. To this end, catalogs such as FAIRsharing [27–29] are including community curated collections of chemistry-friendly repositories (as well as standards and policies relevant in chemistry), and curated lists [30], also provided by initiatives such as NFDI4Chem [31] assist publishers in suggesting repositories (and authors to choose repositories). Following Fig. 2 from the right-hand side of the circle clockwise to the left reveals a trend from green to red, with journal guidelines rarely including the more technical aspects of FAIR and open data practices. In terms of standards, recommendations for metadata or minimum information standards only occur in 10 % of journals surveyed, with no mention of the topic by the remaining 90 %. The minimum information guidelines mentioned include standards such as Minimum Information for Biological and Biomedical Investigations (MIBBI, which are now a collection of guidelines under FAIRsharing), Minimum Information About a Microarray Experiment Involving Plants (MIAME), and Standards for Reporting Enzymology Data (STRENDa guidelines). This again reflects the lack of chemistry-specific guidelines also on minimum information standards.

Similarly, only 26 % of journals recommend or require open analytical data formats to be used for deposited datasets. These recommendations in almost all cases mentioned, besides CIF (see Table 1, footnote b), generic formats such as comma-separated values (CSV) and tab-separated values (TSV). Some journals gave indirect recommendations by listing file extensions of analytical data files, e.g., the JCAMP-DX [32] for NMR data. Machine-readable chemical structures fare even worse and represent the reddest portion of the figure [33]. Only one surveyed journal requires the SMILES line notation for chemical structures, while the rest mentioned neither SMILES nor InChI, as two examples of long-standing chemical structure notations and identifiers for small organic molecules in their guidelines. Furthermore, there has been little advancement or adoption of robust machine-readable representations for other types of molecules, leaving few options for author guidelines to reference; see, for example, the known limitations regarding inorganic structures [34].

Usability of journal and publisher guidelines from an author's perspective

One main observation our review team made was that it was much more challenging than initially expected to collect the data on the landscape of author guidelines. In some cases, information was presented in a single place or even a single page. In others, information was hidden by the nested, hence, confusing structure of the author guidelines, rendering information unfindable. Beside author-guidelines, publisher-wide policies were also available and needed to be considered. In most cases, these publisher-wide policies were linked from the journal's author guidelines, while a conceivable number of guidelines did not link to additional policies provided by their publisher. Furthermore, we also noted inconsistent information on publisher-wide pages and journal's author guidelines.

Considering that the submitting author might not be a senior researcher, or a first-time submitter to the journal, the publishing houses should enhance the findability and accessibility, and hence, useability of their guidelines. One brief checklist could collect all recommendations and requirements relevant for a journal, with more detailed instructions linked from there, but nested structure of author guidelines should be avoided. Publishers should also engage with authors to obtain feedback on the entire publication process to identify usability patterns and issues.

To ease and streamline this effort for publishers and journals, the Research Data Alliance has defined and published features for various data policy types [35], while the European FAIRsFAIR project's data policy checklist [36] incorporates these features. A recent collaboration between FAIRsharing and the Digital Curation Center (DCC) has led FAIRsharing.org to include these features and the checklist into their data policy registry [37]. This new registry structure aims to provide a more comprehensive description of the content and may function both as a point of reference for policy creators as well as an overview for authors.

Indications for stakeholders

The above results provide indications for various stakeholders. These pertain on the one hand to the publishing side, namely authors, reviewers and publishers, and on the other hand to those who are providing infrastructure and technical resources and solutions.

Authors

It is not expected that authors choose a journal mainly due to its more stringent guidelines. Journal scope will, in most cases, remain the basis for that choice. Authors should not interpret these guidelines as a “minimum requirement to pass the bar”. Instead, they are an indicator of good scientific practice. If other journals include more criteria or if some aspect is only “allowed” or “recommended” for one's target journal, all reasonable efforts should be made to match and even surpass these. Briefly mentioning this aspect in the manuscript will be appreciated by the readers, by the reviewers, and may lead to a citation advantage [26]. Having an ORCID iD and including a data availability statement that clearly outlines in which repositories the underlying research data can be found and how they may be accessed, even if not mentioned in the journal guidelines, underlines the author's adherence to good scientific practice. This also enhances transparency, and encourages others to build upon the work.

NFDI4Chem provides various supporting resources for authors such as a Knowledge Base [31], electronic helpdesk, and YouTube channel [38] with extensive background information, support, and tutorials. Furthermore, (early career) training is provided and can help to stress the importance of publishing along with high quality research data. As working with machine-readable data may be unfamiliar for researchers and others involved in publishing workflows, the WorldFAIR Chemistry project is developing an online community resources of practical training demos for working with digital data files [21, 22].

Publishers and editors

Publishers and journal editors always have to strike a balance between the good intentions of harmonizing article styles, ensuring quality reporting, and imposing burden and time-consuming work on top of the scientific achievements. This balancing act might become more delicate if the publisher's business model builds upon article processing charges (APCs). Losing authors to journals with less strict guidelines needs to be weighed against losing readers, citations, and eventually authors due to perceived lower quality and/or lower usability.

For editors and publishers, the results presented here could be considered as priority guidance. Journals are able to guide researchers in making their data more FAIR by giving recommendations within their guidelines. These recommendations may be updated at a later point in time to become a requirement, especially if they have evolved to a community standard.

Widely accepted guidelines, such as to recommend or require the submission of the author's ORCID iD, would represent a good starting point for updates to the guidelines, if not already present. This would also harmonize the publishing landscape. Tightening the recommendation that authors, articles, data, software etc. are identified and linked via unique and persistent identifiers, such as ORCID iDs or DOIs, is a high-impact gain for the publishing community and builds on existing infrastructure. Together, the use of these identifiers enables a network of knowledge in the form of research data and citation graphs to be built. Prominent examples are the Scholix [39] framework or the Open Research Knowledge Graph (ORKG) [40]. Such graphs and relationships can also be included on the publisher and article pages, providing added value to authors and readers.

As indicated above, if data availability statements were mentioned, we did not further consider what kinds of statements were considered acceptable. The FAIR data principles exist as a continuum and do not require data to be fully open to be considered accessible. In the scope of this study, we placed our focus on whether journals require such statements in the first place. While a requirement alone may cause authors to consider their data sharing practices, it should also be noted that studies [41, 42] have shown that "Data are available on request from the authors" commonly leads to loss of that data if the authors are unresponsive, have no time to search for the data, have moved on, or even passed away. Success rates to obtain data upon request can drop by 17 % per year since the time of publication [43]. Consequently, the FAIR data principles strongly discourage data access requiring manual human intervention. Hence, we would suggest such statements only in cases where data cannot be publicly shared, for which these circumstances should be stated as well. Even for sensitive patient data, repositories have established procedures [43] that evaluate data access requests.¹ Some journals have a no exceptions policy on sharing even for human data [44].

We also suggest the deposition of data in suitable repositories, rather than having data in supplementary PDFs, where data in tables are difficult to extract and spectral data are buried in pictures. Beside generic repositories, which accept all data regardless of field, method, or format, the concept of (and preference for) field-specific repositories should also be mentioned. These repositories enable and facilitate the implementation of advanced and method-specific features such as spectral viewers, the ability to search through analytical data file metadata, and data annotation using ontologies to enhance a data set's machine-readability, to mention a few. However, there is still a lack of these field-specific repositories within the domain of chemistry, with the Cambridge Structural Database (CSD) being one of the pioneers. Hence, further field-specific repositories, e.g., nmrXiv for NMR [45], and others for UV/VIS and Raman data are needed, while mass spectrometry databases such as the reference database MassBank [46] are already at the author's service. NFDI4Chem maintains a list of chemistry-friendly repositories within its Knowledge Base [31], which will be expanded in the future and may also act as a resource for publishers for repository recommendations.

We encourage and welcome publishers to (continue to) engage with their community, following earlier events such as the above-mentioned NSF OAC Workshop 2019, the first Editors4Chem workshop in 2021, or events hosted by the RDA [47].

Infrastructure and resource providers

Some of the above results show a need for infrastructures and technical solutions to enable data sharing in a FAIR manner. Categories, such as the machine-readability of structures and reactions, will require further work and developments from the community. Software for chemical structures and especially electronic laboratory notebooks (ELNs; see, for example, the Chemotion ELN [48]) should and can facilitate the inclusion of machine-readable structures, rather than authors manually compiling a list of compounds with their SMILES and InChI. Representation of structures and substances beyond organic chemistry have been challenging for a long time and are the focus of recent efforts [34].

Similarly, an increased availability of subject-specific repositories which adhere to metadata and open analytical format standards will provide publishers the resources needed to recommend and require FAIR data sharing. Without standards on data formats and (their) metadata, research data may be supplied in a number of proprietary formats and

¹ While not publicly open to all, this ensures accessibility to the data, and satisfies the A in the FAIR criteria (specifically, A1.2).

without sufficient accompanying information on how the data was acquired and how it can be (re)used. Open data formats and (their) metadata are important to the aspects of interoperability and reusability in FAIR data. For data annotation, repositories should provide recommendations, ask authors for optional information, but should only require these (meta)data if evolved as a community standard. A shift is also expected here as awareness on these matters grows due to the increase in recommendations for metadata, standard data formats, and the implementation of simple and efficient conversion tools by the scientific community.

Outlook

This survey of the author guideline landscape through a chemistry lens has shown that publishers and journals are starting to include further aspects of research data sharing in their guidelines. The example of (missing) requirements on machine-readable chemical structures, minimum information standards, and open analytical data format standards shows that there is plenty of room for improvement, not only by publishers but also by infrastructure and resource providers. Promising developments were observed in terms of data findability by journals and publishers calling for the submission of data availability statements. Similarly, recommendations on research repositories increase data accessibility.

Many of these improvements will be supported by efforts such as the NFDI4Chem [3, 20] and IUPAC [49] establishing further repositories for data produced in the different sub-disciplines of chemistry, standardizing metadata and data formats for a majority of the analytical data coming out of chemical research, and spreading awareness through information and training sessions. Further guidance for implementing chemical data guidelines from policy to practice from IUPAC is under development in the WorldFAIR Chemistry project [21, 22], drawing from community practices in chemistry and more broadly [49]. Activities include FAIR analysis methods, recipes for incorporating chemistry standards, and resources for working with FAIR-enabled chemistry data [21].

As a continuum, FAIR has the unique ability to draw the research community and all of its varied stakeholders to its principles, with gradual incorporation of its recommendations in a stepwise manner that makes implementation easy. All it needs to succeed is a community that desires to be able to understand and re-use the results and knowledge of chemical research. Let us work towards a future where the participants' hands in the room remain raised when asked "who read an article, went to the corresponding data, and was able to base new science upon it?"

Supplementary information

The supplementary data package [23] on RADAR4Chem contains a README with further instructions and the Jupyter notebook to summarize the guideline survey to reproduce Fig. 2. The dataset also includes screenshots of the author guidelines. Finally, the notebook rendered to HTML and PDF is included.

Acknowledgments: NAP, TGF, CB, SHP and SN acknowledge DFG funding under the project number 441958208 (NFDI4Chem). LRM acknowledges European Commission WorldFAIR Grant Agreement No. 101058393 and IUPAC Project No. 2022-012-1-024. LRM and VFS acknowledge NSF OAC, Award No. 1838958 and 1838960.

Author contributions: NAP, TGF and CB collected and evaluated the guideline data. TGF conceptually managed the data collection process. NAP and CB performed data analysis and visualization. NAP, TGF and SN prepared and submitted the supplementary data to RADAR4Chem. NAP and SN wrote the majority of the manuscript. LRM and TGF reviewed, edited and wrote substantial parts of the manuscript. VFS and SHP reviewed and edited the manuscript. SN conceptualized the manuscript and led this study. NAP and TGF contributed equally to the overall study.

Declarations: LRM is Chair of IUPAC Committee on Publications and Cheminformatics Data Standards (CPCDS). VFS is Chair of the IUPAC Subcommittee on Cheminformatics Data Standards (SCDS).

References

- [1] Deutsche Forschungsgemeinschaft. Zenodo (2022), <https://doi.org/10.5281/zenodo.6472827>.
- [2] NSF – National Science Foundation, Dissemination and Sharing of Research Results. Website. URL: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp> (visited with snapshot September 29, 2022).
- [3] S. Herres-Pawlis, O. Koepler, C. Steinbeck. *Angew. Chem. Int. Ed.* 58, 10766 (2019), <https://doi.org/10.1002/anie.201907260>.
- [4] Journal Editors. *The Analyst* 78, 507 (1953), <https://doi.org/10.1039/an9537800507>.
- [5] Journal Editors. *J. Chem. Soc. Chem. Commun.* 5, 299 (1972), <https://doi.org/10.1039/C39720000299>.
- [6] Journal Editors. *Angew. Chem. Int. Ed. Engl.* 31, A15 (1992), no DOI registered.
- [7] The ACS Style Guide: A Manual for Authors and Editors (J. S. Dodd, ed.). American Chemical Society, Washington D.C., (1997).
- [8] World Protein Data Bank, Instructions to Journals. Website. URL: <https://www.wwpdb.org/documentation/journals> (visited with snapshot April 22, 2022).
- [9] D. G. Watson. *J. Res. Natl. Inst. Stand. Technol.* 101, 361 (1996), <https://doi.org/10.6028/jres.101.038>.
- [10] 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. 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, B. Mons. *Sci. Data* 3, 160018 (2016), <https://doi.org/10.1038/sdata.2016.18>.
- [11] Digital Object Identifier System. Website. URL: <https://www.doi.org/index.html> (visited with snapshot April 22, 2022).
- [12] L. L. Haak, M. Fenner, L. Paglione, E. Pentz, H. Ratner. *Learn. Publ.* 25, 259 (2012), <https://doi.org/10.1087/20120404>.
- [13] R. Lammey. *Sci. Ed.* 7, 65 (2020), <https://doi.org/10.6087/kcse.192>.
- [14] V. F. Scalfani, L. McEwen. NSF OAC 2019 Workshop: FAIR Publishing Guidelines for Spectral Data and Chemical Structures. Website. URL: <https://osf.io/psq7k/> (visited with snapshot April 22, 2022).
- [15] V. Scalfani, L. McEwen. eCommons (2020), <https://doi.org/10.7298/fs2d-hx95>.
- [16] V. F. Scalfani. Figshare (2019), <https://doi.org/10.6084/m9.figshare.8870144.v1>.
- [17] Research Data Alliance, Research Data Sharing without barriers. Website. URL: <https://www.rd-alliance.org/> (visited with snapshot December 9, 2022).
- [18] Chemistry Research Data IG. Website. URL: <https://www.rd-alliance.org/groups/chemistry-research-data-interest-group.html> (visited with snapshot December 9, 2022).
- [19] Data policy standardisation and implementation IG. Website. URL: <https://www.rd-alliance.org/groups/data-policy-standardisationand-implementation-ig> (visited with snapshot December 9, 2022).
- [20] C. Steinbeck, O. Koepler, F. Bach, S. Herres-Pawlis, N. Jung, J. Liermann, S. Neumann, M. Razum, C. Baldauf, F. Biedermann, T. Bocklitz, F. Boehm, F. Broda, P. Czodrowski, T. Engel, M. Hicks, S. Kast, C. Kettner, W. Koch, G. Lanza, A. Link, R. Mata, W. Nagel, A. Porzel, N. Schlörner, T. Schulze, H.-G. Weinig, W. Wenzel, L. Wessjohann, S. Wulle. *Res. Ideas Outcomes* 6, e55852 (2020), <https://doi.org/10.3897/rio.6.e55852>.
- [21] WorldFAIR: Global cooperation on FAIR data policy and practice. Website. URL: <https://iupac.org/worldfair-global-cooperation-on-fairdata-policy-and-practice/> (visited with snapshot August 23, 2022).
- [22] WorldFAIR Chemistry: making IUPAC assets FAIR. Website. URL: <https://iupac.org/project/2022-012-1-024/> (visited with snapshot January 16, 2023).
- [23] N. A. Parks, T. G. Fischer, C. Blankenburg, V. F. Scalfani, L. McEwen, S. Neumann, S. Herres-Pawlis. *RADAR* (2022), <https://doi.org/10.22000/702>.
- [24] P. Kraker, D. Leony, W. Reinhardt, G. Beham. *Int. J. Technol. Enhanc. Learn.* 3, 643 (2011), <https://doi.org/10.1504/IJTEL.2011.045454>.
- [25] A. Hunter. Announcing the Launch of ACS Publications' Data Availability Statement Pilot at The Journal of Organic Chemistry, Organic Letters, and ACS Organic & Inorganic Au. Website. URL: <https://axial.acs.org/2022/09/14/data-availability-statement-pilot-2022/> (visited with snapshot September 30, 2022).
- [26] G. Colavizza, I. Hrynaskiewicz, I. Staden, K. Whitaker, B. McGillivray. *PLOS ONE* 15, e0230416 (2020), <https://doi.org/10.1371/journal.pone.0230416>.
- [27] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A. L. Lister, M. Thurston. *Nat. Biotechnol.* 37, 358 (2019), <https://doi.org/10.1038/s41587-019-0080-8>.
- [28] FAIRsharing Team. FAIRsharing (2011), <https://doi.org/10.25504/FAIRsharing.2abjs5>.
- [29] E. Willighagen, FAIRsharing, Chemistry. Website. URL: <https://fairsharing.org/3524> (visited with snapshot January 16, 2023).
- [30] L. McEwen. *Am. Chem. Soc.* (2019), <https://doi.org/10.1021/acsguide.30105>.
- [31] NFDI4Chem, Knowledge Base. Website. URL: <https://knowledgebase.nfdi4chem.de/> (visited with snapshot April 24, 2022).
- [32] A. N. Davies, P. Lampen. *Appl. Spectrosc.* 47, 1093 (1993), <https://doi.org/10.1366/0003702934067874>.
- [33] D. S. Wigh, J. M. Goodman, A. A. Lapkin. *WIREs Comput. Mol. Sci.* 12, e1603 (2020), <https://doi.org/10.1002/wcms.1603>.

- [34] J. C. Brammer, G. Blanke, C. Kellner, A. Hoffmann, S. Herres-Pawlis, U. Schatzschneider. *J. Cheminform.* 14, 66 (2022), <https://doi.org/10.1186/s13321-022-00640-5>.
- [35] I. Hrynaskiewicz, N. Simons, A. Hussain, R. Grant, S. Goudie. *Data Sci. J.* 19, 5 (2020), <https://doi.org/10.5334/dsj-2020-005>.
- [36] J. Davidson, M. Grootveld, M. Verburg, R. van Horik, R. O'Connor, C. Engelhardt, F. Garbuglia, A. Vieira, E. Newbold, V. Proudman, L. Horton. Zenodo (2022), <https://doi.org/10.5281/zenodo.6225775>.
- [37] A. Lister, FAIRsharing and DCC collaborate to align policy metadata. Website. URL: <https://blog.fairsharing.org/?p=451> (visited with snapshot December 6, 2022).
- [38] NFDI4Chem – YouTube. Website. URL: <https://www.youtube.com/channel/UCQIKQDjyYFzIUFrDfR9vVJg> (visited with snapshot April 24, 2022).
- [39] A. Burton, A. Aryani, H. Koers, P. Manghi, S. La Bruzzo, M. Stocker, M. Diepenbroek, U. Schindler, M. Fenner. *Lib. Mag.* 23, 1/2 (2017), <https://doi.org/10.1045/january2017-burton>.
- [40] S. Auer, A. Oelen, M. Haris, M. Stocker, J. D'Souza, K. E. Farfar, L. Vogt, M. Prinz, V. Wiens, M. Y. Jaradeh. *Bib. Forsch. Prax.* 44, 516 (2020), <https://doi.org/10.1515/bfp-2020-2042>.
- [41] L. Tedersoo, R. Kungas, E. Oras, K. Köster, H. Eenmaa, Ä. Leijen, M. Pedaste, M. Raju, A. Astapova, H. Lukner, K. Kogermann, T. Sepp. *Sci. Data* 8, 192 (2021), <https://doi.org/10.1038/s41597-021-00981-0>.
- [42] M. Gabelica, R. Bojčić, L. Puljak. *J. Clin. Epidemiol.* 150, 33 (2022), <https://doi.org/10.1016/j.jclinepi.2022.05.019>.
- [43] T. H. Vines, A. Y. K. Albert, R. L. Andrew, F. Débarre, D. G. Bock, M. T. Franklin, K. J. Gilbert, J.-S. Moore, S. Renaut, D. J. Rennison. *Curr. Biol.* 24, 94 (2014), <https://doi.org/10.1016/j.cub.2013.11.014>.
- [44] K. Powell. *Nature* 590, 198 (2021), <https://doi.org/10.1038/d41586-021-00331-5>.
- [45] nmrXiv. Website. URL: <https://nmrxiv.org/> (visited with snapshot September 30, 2022).
- [46] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka. *J. Mass Spectrom.* 45, 703 (2010), <https://doi.org/10.1002/jms.1777>.
- [47] Research Data Alliance, RDA and Chemistry. Website. URL: <https://www.rd-alliance.org/rda-disciplines/rda-and-chemistry> (visited with snapshot April 24, 2022).
- [48] P. Tremouilhac, A. Nguyen, Y.-C. Huang, S. Kotov, D. S. Lütjohann, F. Hübsch, N. Jung, S. Bräse. *J. Cheminformatics* 9, 54 (2017), <https://doi.org/10.1186/s13321-017-0240-0>.
- [49] I. Bruno, S. Coles, W. Koch, L. McEwen, F. Meyers, S. Stall. *Chem. Int.* 43, 12 (2021), <https://doi.org/10.1515/ci-2021-0304>.