Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

Published on 07 May 2026
Licensed under CC-BY 4.0

Check for updates

Cite this: *Digital Discovery*, 2026, **5**, 2629

How digital is chemical research? insights from the second NFDI4Chem community survey on research data and FAIR workflows

Jochen Ortmeyer,[a] Vitali Sidorin,[a] Daniela Adele Hausen,[b] Ann-Christin Andres,[c] John Jolliffe,[c] Theo Bender,[c] Giacomo Lanza,[d] Steffen Neumann,[e] Oliver Koepler,[f] Nicole Jung,[g] Christoph Steinbeck,[h] Johannes Liermann *[c] and Sonja Herres-Pawlis *[a]

Received 23rd December 2025
Accepted 29th April 2026

DOI: 10.1039/d5dd00584a

rsc.li/digitaldiscovery

Increasing digitalisation is revolutionising all scientific disciplines, but it poses a particular challenge in chemistry. Research data is generated, for example, during the synthesis of substances, the recording of spectra or the application of theoretical methods, and must therefore be appropriately documented, archived and made available for reuse. Traditionally, this happens in the form of paper-based laboratory notebooks. In order for the transition to the age of digital chemistry to succeed, a cultural change in the community is necessary. But how do chemists actually treat their research data? the recent NFDI4Chem survey provides answers on the progress of digitalisation. Moreover, we highlight how the consortium transported the community's needs into the second funding phase. This exemplifies a feedback-driven model for digital infrastructure design that could serve as a blueprint for other disciplines and national initiatives.

Introduction

The digital transformation of chemistry increasingly depends on systematic management, availability and reusability of research data. Experimental, computational and spectroscopic datasets underpin discovery, modelling and reproducibility, yet their long-term accessibility and interoperability remain uneven across subdisciplines and organisations. The FAIR (Findable, Accessible, Interoperable, and Reusable) principles[1] provide a widely accepted framework for improving research data management and have been adopted by funders,[2–4] infrastructures and many publishers.[5,6] Chemistry is more complicated in its digitalisation than other text- or number-based disciplines since chemistry is based on molecule drawings which contain topological information. In many cases, *e.g.*, in inorganic chemistry, this information cannot be fully represented in spite of 30 years of cheminformatics.[7]

Several domain-specific efforts have attempted to measure and accelerate FAIR uptake in chemistry. National and international initiatives have surveyed needs and practices to guide infrastructure development: the German NFDI4Chem consortium has run community surveys and produced guidance for chemistry research data stewardship as part of the NFDI effort, explicitly framing its work around FAIR implementation for chemical workflows.[8,9] The WorldFAIR project combined technical work with social surveys and policy briefs to identify cross-cutting obstacles (*e.g.*, metadata heterogeneity, the difficulty of linking samples and digital data, and the need for discipline-specific recommendations) and to recommend practical policy measures for chemistry and other domains.[10]

Complementary to large infrastructure projects, community and publisher-focused studies have highlighted practical pain points that limit FAIRness in chemical publishing and repository practices. For example, recent discussions have argued for standardised templates and mandatory, machine-readable deposition of chemical structures and associated metadata to make published chemical data genuinely reusable. Such community commentaries emphasise the gap between aspirational FAIR policies and routine author behaviour.[11–13] At the same time, cross-domain reviews[14–17] of FAIR assessment tools and studies of data repositories in material sciences and chemistry show a proliferation of technical solutions (assessment tools, metadata standards,

[a]Institute of Inorganic Chemistry, RWTH Aachen University, Landoltweg 1, 52074 Aachen, Germany. E-mail: sonja.herres-pawlis@ac.rwth-aachen.de

[b]Heinrich-Heine-University, Research Data Competence Centre, 40225 Düsseldorf, Germany

[c]JGU Mainz, Department of Chemistry, Duesbergweg 10-14, 55128 Mainz, Germany. E-mail: liermann@uni-mainz.de

[d]Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany

[e]Leibniz Institute of Plant Biochemistry, Program Center MetaCom, Weinberg 3, 06108 Halle, Germany

[f]Technische Informationsbibliothek (TIB), Welfengarten 1 B, 30167 Hannover, Germany

[g]Institute of Biological and Chemical Systems, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany

[h]Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller-University, Lessingstr. 8, 07743 Jena, Germany

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

domain repositories), while also noting variability in uptake and maturity across subfields.[18,19] These analyses underline that technical capability alone is insufficient: incentives, training, and community standards are equally important to increase FAIRness.[20,21] In principle, this is a socio-technical endeavour and a lot of experienced personnel (*e.g.*, data stewards and research data managers) is needed to make the cultural change happen.

Among infrastructure-level responses to the FAIR challenge in chemistry and the physical sciences is the UK-based Physical Sciences Data Infrastructure (PSDI).[22,23] PSDI aims to interconnect existing data systems, provide conversion and search services, and offer training and guidance to make data more findable, accessible, interoperable and reusable. By providing (i) a data conversion service tackling format interoperability; (ii) community data collections with enriched metadata schemas; and (iii) guidance on data sharing and FAIR practices, PSDI addresses several known barriers to FAIR adoption in chemistry (*e.g.*, legacy formats, heterogeneous metadata, lack of reuse tooling).

Taken together, the existing literature and survey work paint a consistent picture: awareness of FAIR concepts is growing in chemistry, and many infrastructural components (repositories, identifier systems, metadata initiatives) are emerging, but persistent cultural, technical and organisational barriers continue to limit routine FAIR implementation.[15,24] A recent study by Bloodworth *et al.*[13] highlights that current data-sharing practices in organic chemistry journals largely fail to meet FAIR Guiding Principles standards, with most authors sharing only minimal, non-machine-readable data and rarely depositing primary data in open repositories.

Moreover, the authors state that meaningful improvement in data accessibility and machine-readiness will require mandated journal policies, standardized formats, repository infrastructure, and cultural change to enable AI-driven discovery in chemistry. Especially, the PNNL report (ref. 15) reveals that chemistry lags materials science and physics in institutionalising FAIR data standards and that impediments to FAIR data generation in chemistry fall into two main groups: technical roadblocks (ontologies, metadata tools, repositories) and psychosocial barriers (incentives, uncertainty, questions of who funds and enforces compliance).[15]

The present user survey complements these prior initiatives by providing up-to-date empirical data from practicing chemists on awareness, concrete practices (data deposition, metadata capture and generation, use of persistent identifiers and ELNs), perceived obstacles (time, incentives, tooling), and priorities for training and infrastructure. By comparing our findings with the results and recommendations of NFDI4Chem, WorldFAIR, PSDI, FAIR assessment studies and community commentaries, we aim to provide actionable guidance for infrastructure providers, publishers and funders working to make chemical data more FAIR.

To set the framework, we would like to give some context on the National Research Data Infrastructure (NFDI): the NFDI is an initiative of the German Federal Government and the federal states. Its aim is to create a network across scientific disciplines that enables the sustainable handling of research data in accordance with the FAIR principles. The chemistry consortium NFDI4Chem strives to seamlessly digitalise the entire workflow in chemical research and thus close the analogue gaps and digital barriers in the digital data life cycle. To this end, the consortium provides the chemical community with various services, resources and training opportunities[25] in order to provide them with the best possible support in the current cultural change. The digital transformation includes, for example, the provision of open-source electronic lab notebooks (ELNs), which greatly simplify and facilitate the preparation of data for publication in data repositories, as well as the development of tools and standards to foster data integration and harmonization and a close exchange with the community. Chemotion ELN[26,27] together with its repo is an ideal example for FAIRification of data right from the beginning of the research process: the researchers plan the experiment in Chemotion, document every step, note observations in the software, collect the characterisation data, analyse them and use Chemotion to generate the synthesis parts of the SI (still needed for most chemical publications). The deposit in the Chemotion repository[28] is then seamlessly performed *via* one click.

Taking the community into account is essential in order to drive forward the development of the infrastructure in line with the needs of users. As early as in 2019, we performed a user survey to shape the working programme of NFDI4Chem in the first phase.[29] The survey showed that, although chemists broadly acknowledge the importance of research data management, its practical implementation remains limited. Data were often handled in an ad-hoc manner, with heterogeneous documentation practices and little use of standards or repositories. A major barrier identified in 2019 was the lack of user-friendly, well-integrated digital tools that fit smoothly into everyday laboratory workflows. Respondents expressed a strong need for standardized metadata, minimum information standards, and interoperable formats to enable data sharing and reuse in line with the FAIR principles. In addition to technical infrastructure, the survey highlighted the importance of training, guidance, and long-term institutional support to establish a sustainable culture of good data management. Overall, the results provided a baseline and clearly motivated the strategic focus of NFDI4Chem on infrastructure, standards, and education tailored to the chemistry community.

Results of the survey

NFDI4Chem regularly conducts surveys to determine the needs of users and the current status of research data handling in the community. Following the last major survey in 2019,[29–31] the consortium again surveyed members of the community from January to April 2023 in order to include the requirements into the consortium's next funding phase.

In total, 813 people participated in the survey, of which the most participants did their last research in Germany (84%), while the rest were spread out around the world. Approximately 86% of these people are working in a research related environment, namely universities (75%) and non-university research institutions (11%). Only a few participants stated

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

that they are working in the industry (6%). Focusing on universities, mostly people in positions requiring a master's degree or higher, for example PhD students (33%), postdoctoral researchers (24%) or professors (22%), took part in this survey. Master's (7%) and bachelor's (2%) students participated less, likely due to limited incorporation into the research data management policies of the working group they are employed in. The participants are relatively equally distributed across the three main chemical subdisciplines, inorganic (27%), organic (34%) and physical (21%) chemistry. The rest represented a large variety of other subdisciplines, such as chemical engineering (6%), theoretical (9%) or pharmaceutical (6%) chemistry. Selected results are compared with another survey, which was conducted by NFDI4Chem in 2019 where 623 people participated. Out of those, 541 came from Germany, therefore only the german datasets were analysed.

Design of the survey

Fig. 1 summarizes which focus areas were included in the survey. We aligned the new survey to the 2019 survey in order to see a temporal development. The survey allowed multiple answers in all cases (except from clear yes/no questions). All survey data can be downloaded here https://doi.org/10.25835/a8ih6h17.

Collecting, processing and storing of data

Most of the researchers who participated in the survey collect data from experimental synthesis (65%) and from analytical methods, such as spectroscopic data (75%) or crystallographic data (38%). Depending on data types and circumstances the level of digitalisation ranges from analog, *e.g.*, printed, to fully digital 45% stated that they collect data in a non-electronic format. Non-electronic data refers on the one hand to handwritten notes in a laboratory journal, such as observations or procedures, which is difficult to be directly digitised during the work in a laboratory. In these cases, the documentation of observations or procedures is also collected in an analogue format in the form of handwritten notes in a laboratory journal, which is difficult to be directly digitised during the work in a laboratory. On the other hand, spectroscopic data like NMR or MS is still received by many participants in a non-electronic format (*e.g.*, printed spectra). Digitalisation of those can be laborious and unnecessary especially with regard to the fact that modern devices always output electronic data. To counter inconsistent digitalisation throughout a working group, workflows can be employed where digital data is treated in an analogue way to become digital data again (valid for 24% of the participants who answered this question).

Long term data storage and archiving

During a project, or most importantly at its end, it is of relevance to provide services to store and archive produced data. In Germany, DFG regulations[32] require data from funded projects to be stored at least for a time period of ten years and recommend to organise them in a FAIR way. Therefore researchers should archive a version of their data. The survey reveals that 57% of the scientists store raw data, 52% store processed data, 55% store analyzed and 49% all data. As already introduced above, recommendations and requirements for long term data storage and archiving also exist in scientific working groups and the majority (62%) employ such methods in their working

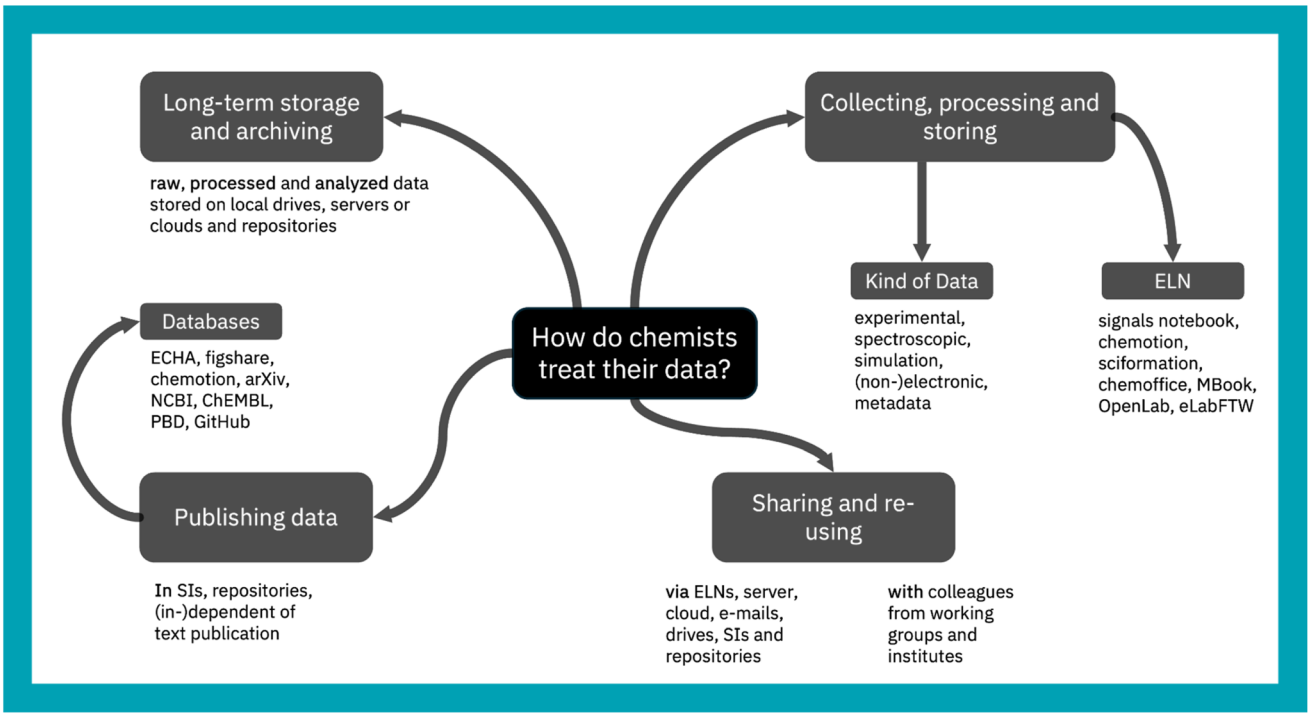

Fig. 1 Design of the survey covering all aspects of the data life cycle.

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

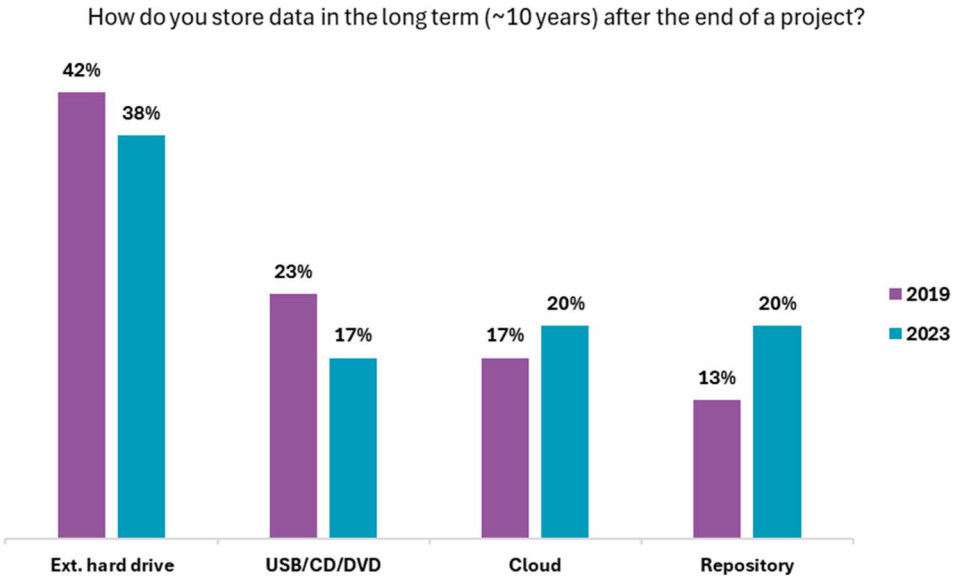

Fig. 2 Comparison of the usage of different systems for long term data storage as obtained from surveys in 2019 and 2023. (How do you store data in the long term (~10 years) after the end of a project?).

group. A smaller number of researchers stated that they are not aware of any rules (20%) or that a workflow is not yet implemented (18%). Apart from the state of the stored data, the medium on which the data is stored was queried (Fig. 2). The replies vary from storing data on local devices, like a computer (40%), external hard drive (38%) or on a USB/CD/DVD (17%), over to online services, like the server of the working group (64%) or university (39%), cloud systems (20%) or repositories (20%). In comparison to the questionnaire from 2019 an increase in the usage of online services is observed, namely cloud systems (from 17% to 20%) and data repositories (from 13% to 20%). Additionally a shift away from storage on hardware devices (hard drives, USB/CD/DVD) could be identified throughout the two surveys, showing a decrease from 65% to 55% respectively.

Metadata

Besides publishing or archiving raw data, data must be complemented with metadata, to describe it precisely. The main purpose of metadata is to help other researchers to understand

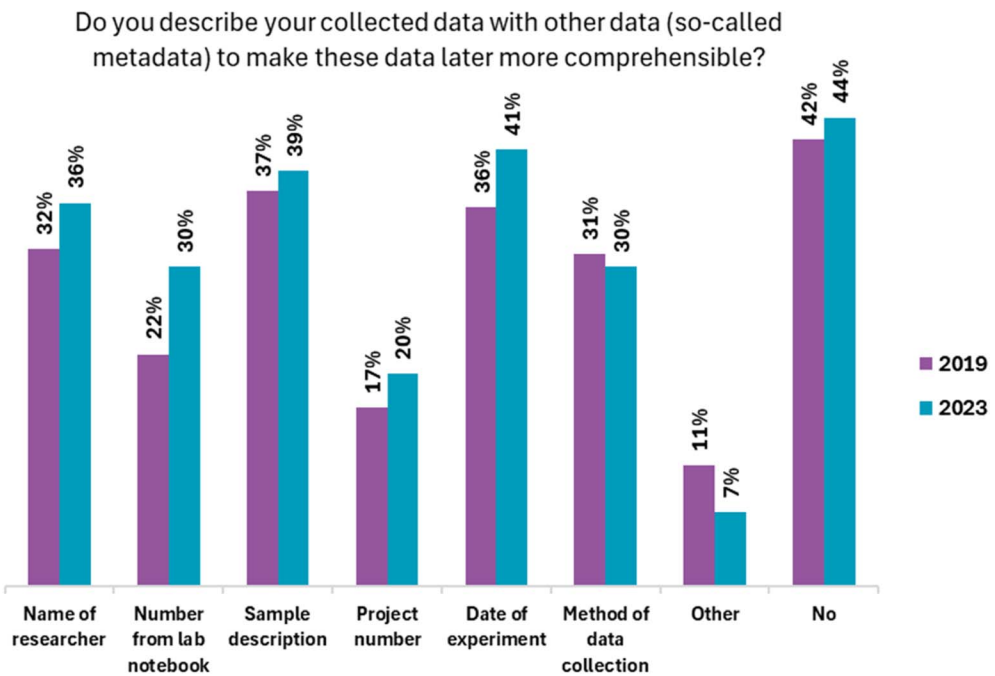

Fig. 3 Incorporation of different categories of metadata across both surveys in 2019 and 2023. (Do you describe your collected data with other data (so-called metadata) to make these data later more comprehensible?).

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

the presented data better. In line with this 45% of participants do think that their descriptive metadata helps other scientists within their working group and even 32% think the same within and beyond the working group. Metadata can be published in repositories or university servers, where colleagues and other researchers from outside can access them but in some cases they are contextually tied to a specific article or publication, in which 10% relate to. A minor part of 12% stated that their data can not be directly understood by other scientists without further explanation. In total 56% of participants use metadata for their collected data. The kind of added metadata (Fig. 3) varies from the name of the researcher (36%), the number from the lab notebook (30%), sample description (39%), project number (20%) to date of experiment (41%) or method of data collection (30%). Generation of the metadata occurs mainly manually (44%), *via* devices and software (11%) or through a combination of both (46%). Compared to the data from the survey in 2019, now more researchers tend to generate and collect metadata both manually and from devices and software (15% in 2019 *vs.* 46% in 2023). A large shift from manual metadata collection to a combination of manual and software based collection occurs. We derive from personal discussions that the software is now capable of providing more metadata and more software used for metadata generation. In comparison to previous responses, rules and workflows for the description of data with metadata exist, but only a small percentage of institutes or working groups (25%) apply them. An even smaller group has agreements on data structure with colleagues (11%). The main part stated that there are no standards in their working environment for metadata generation and collection (11%). Besides that, some people define their own system for describing their data with metadata because they perceive the standards as non-appropriate for the own data. Research data management services and individual scientists should strive to make metadata an easy understandable and helpful addition to data.

Electronic laboratory notebooks

Electronic laboratory notebooks (ELNs) help researchers shift away from recording their data and especially manual metadata in an analogue way, *e.g.*, paper notebooks, by directly creating entries digitally. Furthermore, an ELN allows the attachment of spectroscopic data to the chemical data (*e.g.*, reaction entry and procedure) which simplifies associating the data. Researchers should be encouraged to use an ELN for their documentation of experiments in laboratories for better research data management. Out of the participants in this survey, 30% claimed to use an ELN (*e.g.*, Signals Notebook, Chemotion, Sciformation, Chemoffice, MBook, OpenLab, eLabFTW, *etc.*). This is a significant increase in comparison to the survey from 2019, where only 18% of participants used an ELN. The use of ELNs are distributed across main chemical subdisciplines, where organic chemistry (34%) and material science (33%) lead and are followed by biological (26%), inorganic (23%) and physical (16%) chemistry. This distribution accords to the fundamental function of an ELN, namely planning, conducting, analyzing and archiving chemical reaction, which fits most especially for organic chemistry. According to this survey, Chemotion is the most popular ELN in use (26%) and is also mostly represented in organic (54%) and inorganic (28%) chemistry. A comparison between ELN users and non-users indicates that ELN adoption is associated with more research data management practices. ELN users are more likely to systematically add metadata to their datasets (67% *vs.* 51% of non-ELN users) and to share data *via* repositories (26% *vs.* 16%). This suggests that the use of ELNs promotes behaviors that facilitate data reuse and long-term accessibility. However, it is impaired by accessibility, time and infrastructure issues. Additionally, several ELNs are only commercially available, and most of them are optimized for biochemical or life-science usage rather than for the full range of chemical subdisciplines. Upgrading existing ELNs to the needs of general chemistry will therefore require dedicated development efforts and funding. A rise of popularity and interest in ELNs and research data management is expected, especially regarding that 86% of participants wish for an incorporation of the handling of research data and research data management into the official curriculum of future students.

Sharing and reusing data

Researchers share their data within and outside the working group they are employed in (Fig. 4). Most participants stated that they share data with colleagues by e-mail (52%) or using the server of the institute or working group (76%). A smaller part uses cloud services (35%), an ELN (22%) or a repository (10%). Furthermore, 25% use hardware devices like USB sticks, CD's or DVD's. After comparing with the survey from 2019, a shift towards using an ELN (15% to 22%) or cloud services (22% to 35%) and especially away from hardware devices (40% to 25%) can be observed. Most importantly, this trend shows the rising popularity of using an ELN. Outside the working group, e-mails (62%) and cloud services (40%) are the dominant way of sharing data. ELN (2%) and institute servers (14%) are highly under-represented whereas repositories (19%) are more used outside than within the working group due to facilitated accessibility by external users. There has been a clear reduction in the usage of hardware devices for data sharing in recent years, dropping from (24%) in 2019 to a mere 10% in 2023.

Additionally, participants were asked if they knew online databases or data repositories where they can find data for their domain. In this case, finding data is not only referred to viewing the data but also saving and opening it with the corresponding software, if the limitations due to licensing are not taken into consideration. 24% stated that they know such services. Most of the people who took the survey have reused data provided by other scientists (Fig. 5), mainly from colleagues in the same working group (50%) or SI from a publication (45%). A smaller part reuse data by colleagues from other institutes (28%) and *via* repositories (18%). Negated responses can be summarised with data being not available (8%), not sufficiently described (6%), distrust in others (3%) or no need to use other data (25%).

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

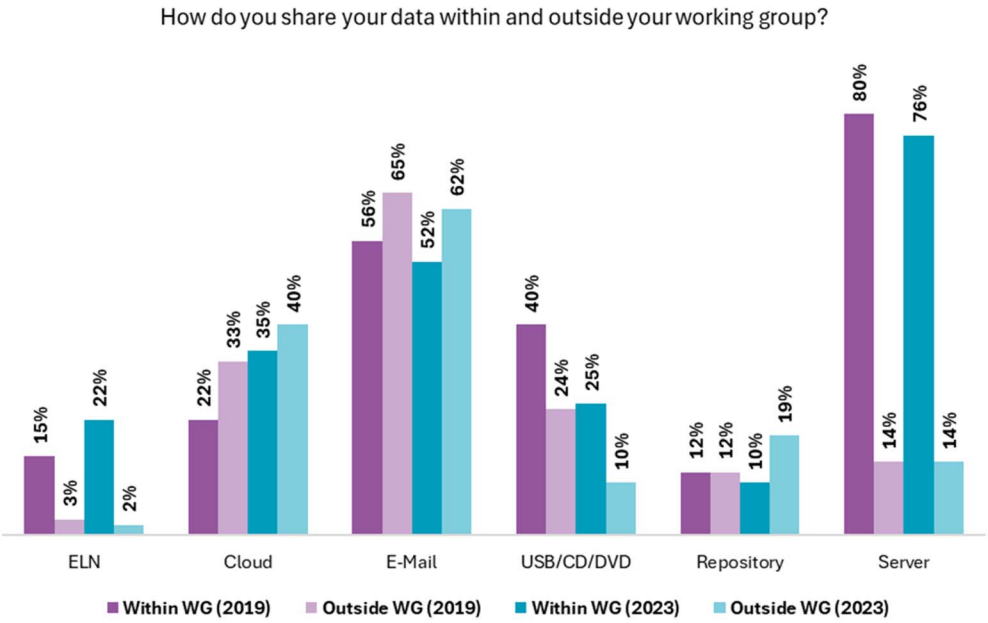

Fig. 4 Sharing data on different media in and outside the working group of the surveys in 2019 and 2023. (How do you share your data within and outside your working group?).

Especially the latter point can be explained by missing interdisciplinary overlap in some niche chemical fields.

Publishing data

Regarding the publication of research data (Fig. 6), most respondents in both surveys reported that their data are primarily contained in the text or the SI of a conventional journal article (59% and 57%). The proportion of scientists who additionally publish parts of their data in a repository increased from 16% in 2019 to 22% in 2023, indicating a gradual adoption of repository-based publication. In contrast, pure data publications in repositories, *i.e.*, datasets that are independent of an article, remain rare and were used by only 8% of participants in both surveys. At the same time, a substantial share of participants has not yet published any research data at all (32% and 36%) which correlates to the amount of young researchers as participants of the survey. Overall, the results reveal only a modest shift towards repository use, while journal-centric and

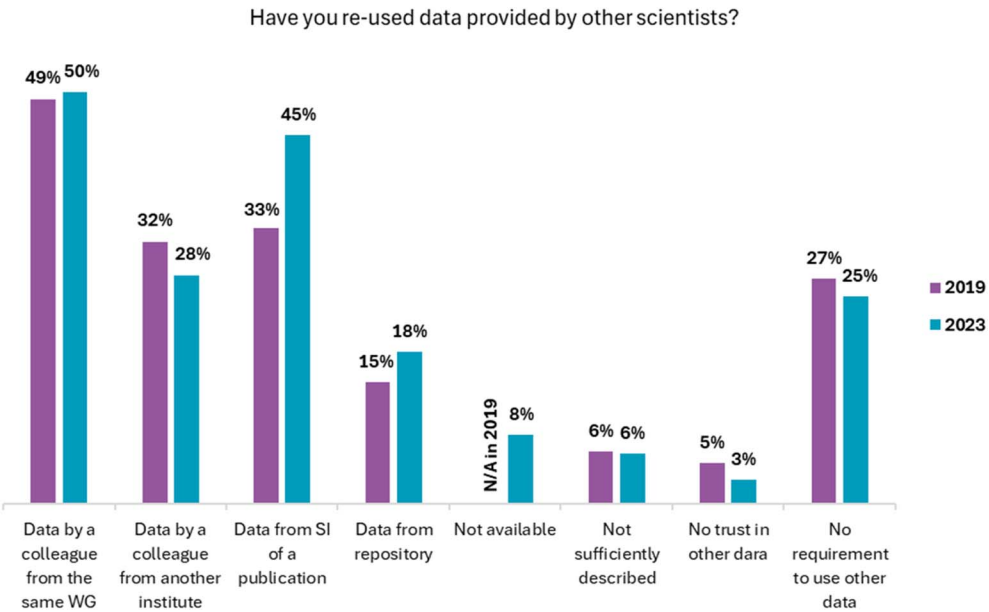

Fig. 5 Re-using data acquired by different means and negated responses of using external data. (Have you re-used data provided by other scientists?).

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

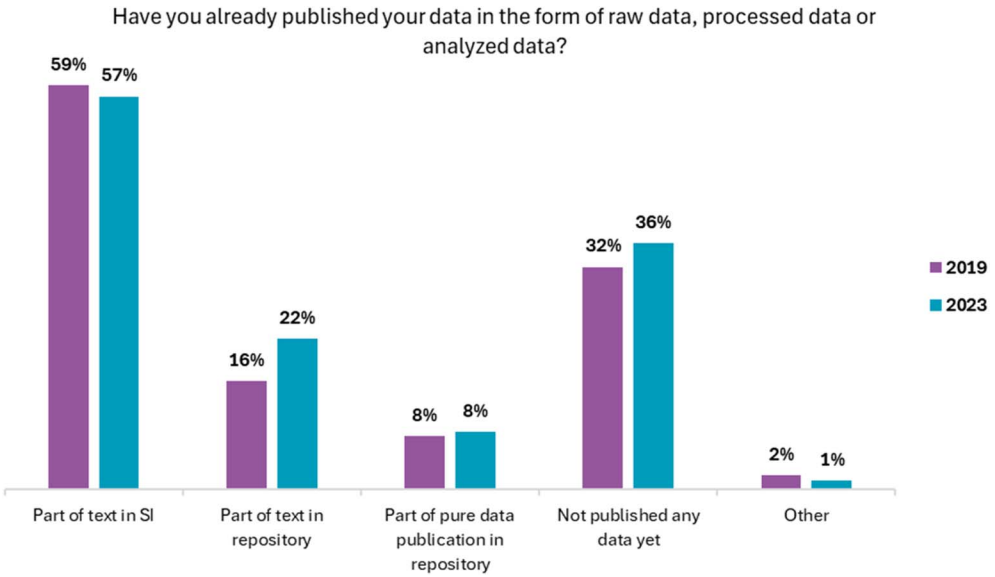

Fig. 6 Publishing data as part on different platforms. (Have you already published your data in the form of raw data, processed data or analyzed data?).

SI-based publication practices continue to dominate research data management.

Comparison of the NFDI4Chem and PSDI surveys: converging insights across domains

Both the NFDI4Chem community survey and the PSDI survey (in the year 2023, 44 participants)[33] illuminate how researchers in chemistry and the broader physical sciences are progressing in the adoption of FAIR data practices. While NFDI4Chem provides a domain-specific view grounded in the workflows of chemical research, PSDI adopts a cross-disciplinary perspective spanning chemistry, materials science, and physics. Together, the two datasets reveal converging trends in digital transformation and shared structural challenges that transcend disciplinary boundaries. In both surveys, general awareness of FAIR data has increased substantially since earlier studies. In the NFDI4Chem survey, nearly all respondents had heard of FAIR data, though fewer felt confident in implementing FAIR practices. Similarly, the PSDI survey reports a strong conceptual acceptance of FAIRness but indicates that practical application remains uneven across the physical sciences. Respondents in both communities cited uncertainty about institutional support and the complexity of FAIR compliance as major barriers. This alignment across countries suggests that, despite successful outreach, FAIR literacy still requires translation into day-to-day research practice.

The NFDI4Chem survey highlights growing but still limited adoption of ELNs: about 30% of respondents use an ELN, up from 18% in 2019. PSDI's cross-domain data reveal comparable adoption rates, with chemistry and materials science leading in digital documentation but substantial portions of the physical sciences still relying on analogue or semi-digital records. In both communities, ELN integration with analytical instruments and repositories remains a bottleneck. Whereas NFDI4Chem directly supports open-source ELN development (Chemotion) and fosters the linkage of ELNs to repositories, PSDI respondents emphasised a need for interoperable laboratory information systems and conversion services to harmonise data across tools and formats. The two surveys therefore converge in identifying the "last meter" of digitalisation, the laboratory interface, as a shared challenge. Bringing the ELN on a physical device into the laboratory would help in many cases but is not trivial.

A striking parallel emerges in the area of metadata. NFDI4Chem respondents report that 56% annotate data with metadata, mostly through manual or hybrid approaches, and only one quarter have institutional metadata standards. PSDI results show a similar fragmentation: respondents recognise metadata as essential but report inconsistent practices, limited automation, and a lack of agreed vocabularies. Across both surveys, metadata creation is perceived as time-consuming and poorly rewarded. Both communities explicitly request clearer guidance, domain-specific templates, and automated metadata extraction from instruments—findings that validate each other and highlight an urgent need for coordinated standardisation across consortia. This coordination between consortia and with IUPAC and CODATA is already active since international integration was already an important task in the first funding period of NFDI4Chem.[8]

Both surveys depict a gradual shift from local, hardware-based storage toward institutional and repository services. In the NFDI4Chem survey, 64% of respondents use institutional servers and 20% deposit data in repositories, reflecting progress compared to 2019 but still limited uptake. The PSDI survey reports similar figures for repository use and highlights that most physical scientists rely on project or group servers rather than formal repositories. Common obstacles include uncertainty about repository suitability, lack of long-term preservation guarantees, and unclear licensing. Both communities

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

express a clear wish for reliable, discipline-adapted repositories with intuitive submission workflows and machine-readable metadata – precisely the gaps NFDI4Chem's Chemotion repository and PSDI's community data collections aim to fill. On the long range, NFDI4Chem works on the federation of its repositories as well as fallback options and "exit-strategies" for single repositories.

A central outcome of both surveys is the recognition that infrastructure alone cannot achieve FAIRness without parallel cultural transformation. In the NFDI4Chem results, participants emphasise the need for institutional policies, incentives, and hands-on training; PSDI respondents echoed these concerns, adding that data management responsibilities are often unclear or undervalued in research careers. Both initiatives respond by embedding training components into their next phases: NFDI4Chem through the "FAIR for Chemists" curriculum support and an extensive set of training activities, PSDI through coordinated training resources and best-practice documentation. The parallel emphasis on community engagement indicates that sustainable change depends as much on people and skills as on technology. This is in high accordance with other studies on digital skills in chemistry in general.[34]

While NFDI4Chem focuses on the chemistry data lifecycle, its community has increasingly recognised the need for interoperability with adjacent domains such as catalysis, materials science, and engineering. PSDI's survey results reinforce this perspective by documenting widespread frustration with data silos and incompatible formats across the physical sciences. Both surveys call for persistent identifiers, harmonised metadata schemas, and linked ontologies as enablers of reuse. The alignment between chemistry-specific (NFDI4Chem) and cross-domain (PSDI) perspectives demonstrates that the FAIR transformation of chemistry cannot occur in isolation but must evolve within a federated, interoperable infrastructure landscape.

Although differing in scale and disciplinary focus, the NFDI4Chem and PSDI surveys are mutually reinforcing. NFDI4Chem offers granular insight into laboratory practices, repository interactions, and the chemistry community's digital culture; PSDI provides a broader systems-level view, identifying the same structural issues from a national infrastructure perspective. Where NFDI4Chem reveals the needs of individual chemists, PSDI situates those needs within a multi-domain ecosystem. Together, the two surveys provide a consistent empirical foundation for guiding European FAIR infrastructure development—emphasising interoperability, community engagement, and sustained training as the pillars of a successful digital transformation.

Mapping community needs to the next funding phase of NFDI4Chem

The 2023 survey confirms a steady but incomplete transition from analogue to digital data handling: nearly half of respondents still collect data non-electronically, and only 30% routinely use electronic laboratory notebooks (ELNs). In response, the NFDI4Chem extension proposal[35] makes the digital laboratory the focal point of its second phase. It prioritises (i) broad adoption of open-source ELNs, (ii) seamless ELN-to-repository publishing workflows, and (iii) enhanced device connectivity for automated data capture. These actions directly target the community's reported need for low-barrier digital documentation tools that integrate with everyday laboratory routines. New technical work packages include device integration frameworks, improved APIs, and cloud-ready deployment to facilitate installation and institutional uptake—thus addressing the survey's finding that local technical limitations still hinder digitalisation.

The survey reveals persistent uncertainty around metadata practices: only 56% of chemists enrich data with metadata, and just 25% report institutional rules for metadata creation. The extension proposal explicitly commits to standardising FAIR metadata workflows through harmonised ontologies and machine-actionable templates. With many NFDI4Chem tools we enable embedding ontologies in researchers' everyday tools which virtually "hides" their use (*e.g.*, Chemotion ELN). Building on outcomes of the first phase, it foresees expanded use of minimum information about chemical investigations (MIChI)[36,37] as metadata schemas, in alignment with IUPAC, and automated metadata harvesting from ELNs. Moreover, it introduces a "metadata readiness level" concept and quality assessment tools to encourage consistent, transparent metadata practices across institutions. In short, the proposal translates the community's desire for practical metadata guidance into structured, implementable standards supported by tooling and training.

According to the survey, only one-fifth of chemists currently use repositories for long-term storage, although awareness and usage are increasing. The extension proposal therefore strengthens the repository ecosystem as a key infrastructure component. It expands the capacity and interoperability of the Chemotion and RADAR4Chem repositories, introduces discipline-specific sub-repositories, and enhances DOI and ORCID integration for persistent identification. A further measure responds to the high reliance on institutional servers reported in the survey: NFDI4Chem will pilot interfaces between local data storage systems and central repositories, allowing researchers to deposit data directly from institutional infrastructures without duplication. This design supports both ease of use and FAIR compliance.

Cultural and educational barriers remain among the strongest obstacles identified in the community survey. Many participants cited uncertainty about FAIR data practices and insufficient institutional support. The freshly started second funding phase addresses this by expanding the training and outreach portfolio through:

- A modular FAIR for chemists training curriculum integrated into university courses,
- Certification schemes for FAIR data stewards,
- Targeted continuing-education formats for PIs and lab managers, and
- Online learning resources in the NFDI4Chem knowledge base.

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

The work package on "community engagement and training" explicitly embeds user co-creation and feedback cycles, ensuring that educational activities evolve alongside user needs. This mapping demonstrates that NFDI4Chem regards teaching RDM as equally vital as technical infrastructure.

Survey respondents indicated interest in broader discoverability and reuse of data beyond their own subdiscipline, yet only a quarter knew of relevant repositories (such as Chemotion Repo, nmrXiv *etc*). In response, the extension proposal emphasises cross-consortium collaboration within the NFDI and beyond. It plans integration with NFDI4Ing, NFDI4Cat, and FAIRmat, as well as with international partners such as the WorldFAIR Chemistry initiative. Through harmonised metadata profiles and shared vocabularies, NFDI4Chem aims to position chemistry within a connected FAIR ecosystem. For end users, this means enhanced data searchability, standard identifiers (*e.g.*, InChI[7]) and shared access portals – directly addressing the community's call for better data discoverability and interoperability.

A recurring theme in the survey is the lack of incentives and clear mandates for data publication. The extension proposal therefore introduces policy and governance measures to reward FAIR data practices. These include citation tracking for datasets, integration of data publication into academic evaluation, and persistent tracking of dataset reuse. Furthermore, the proposal's governance model strengthens community representation by creating an expanded community board and user advisory panels—ensuring that the consortium continues to translate community feedback into action, as demonstrated in the current renewal.

Conclusion and forward perspective

When viewed together, the 2023 survey and the extension proposal illustrate a coherent feedback loop: user feedback drives infrastructural and strategic priorities. The survey identified concrete obstacles (limited ELN use, inconsistent metadata, low repository uptake, and cultural hesitations), while the proposal outlines targeted responses (integrated ELN workflows, metadata standards, repository interconnection, and education programmes). Especially peer-to-peer teaching will become more relevant in the next funding phase to foster the cultural change in chemistry. Hence, NFDI4Chem will intensify its activity on conferences, workshops and dedicated training modules. The presented mapping demonstrates that NFDI4Chem has evolved from building individual technical components toward a mature, community-driven data ecosystem. The second funding phase[35] thus transforms the consortium from a provider of isolated FAIR tools into an orchestrator of digital transformation in chemistry, guided by quantitative evidence from its own community.

Ethical statement

The survey has been checked by the RWTH Aachen ethics committee and the participants had to read and agree to the data protection plan prior to the survey.

Author contributions

J. O., D. A. H., A.-C. A., J. J., T. B., G. L., S. N., O. K., N. J. C. S. J. L. and S. H.-P. designed the survey. J. O., V. S., O. K. and S. H.-P. analysed the survey results. J. O., V. S. and S. H.-P. wrote the manuscript and all coauthors edited and reviewed it.

Conflicts of interest

There are no conflicts to declare.

Data availability

Survey is available as data publication: Jochen Ortmeyer, Vitali Sidorin, Daniela Adele Hausen, Ann-Christin Andres, Theo Bender, Giacomo Lanza, Steffen Neumann, Oliver Koepler, Nicole Jung, Christoph Steinbeck, Johannes Liermann, Sonja Herres-Pawlis (2025). Second NFDI4Chem user survey [data set]. LUIS. https://doi.org/10.25835/a8ih6h17.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI4/1 – project number 441958208 (NFDI4Chem). Moreover, we would like to acknowledge the helpful comments of the referees of this work.

References

1 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, *Sci. Data*, 2016, **3**, 160018, DOI: 10.1038/sdata.2016.18.
2 Deutsche Forschungsgemeinschaft, Guidelines for Safeguarding Good Research Practice. Code of Conduct, Zenodo, 2025, DOI: 10.5281/zenodo.14281892.
3 Horizon Europe: Model Grant Agreement, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/agr-contr/unit-mga_he_en.pdf, (accessed December 2025).
4 Open Research Data and Data Management Plans: Information for ERC grantees, https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf, (accessed December 2025).
5 Beilstein Journal of Organic Chemistry: Instruction for Authors, 7.3 Data Deposition, https://www.beilstein-journals.org/bjoc/authorInstructions#7.3, (accessed December 2025).
6 Chemistry Europe: Notice to Authors, Appendix: Reporting Experimental Information and Data, https://chemistry-europe.onlinelibrary.wiley.com/hub/journal/1864564x/notice-to-authors#sectAppendixReportingExperimentalInformationandData, (accessed December 2025).
7 G. Blanke, J. Brammer, D. Baljozovic, *et al.*, *Faraday Discuss.*, 2025, **256**, 503–519, DOI: 10.1039/D4FD00145A.

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

8 C. Steinbeck, O. Koepler, F. Bach, *et al.*, *Res. Ideas Outcomes*, 2020, **6**, e55852, DOI: 10.3897/rio.6.e55852.

9 S. Herres-Pawlis, F. Bach, I. J. Bruno, S. J. Chalk, N. Jung, J. C. Liermann, L. R. McEwen, S. Neumann, C. Steinbeck, M. Razum and O. Koepler, *Angew. Chem., Int. Ed.*, 2022, **61**, e202203038, DOI: 10.1002/anie.202203038.

10 The WorldFAIR project, https://worldfair-project.eu/, (accessed December 2025).

11 E. L. Schymanski and E. E. Bolton, *J. Cheminf.*, 2021, **13**, 50, DOI: 10.1186/s13321-021-00520-4.

12 E. L. Schymanski and E. E. Bolton, *Exposome*, 2022, **2**, osab006, DOI: 10.1093/exposome/osab006.

13 S. Bloodworth, C. Willoughby and S. J. Coles, *Beilstein J. Org. Chem.*, 2025, **21**, 864–876, DOI: 10.3762/bjoc.21.70.

14 M. Martorana, T. Kuhn, R. Siebes and J. van Ossenbruggen, *PeerJ Comput. Sci.*, 2022, **8**, e1038, DOI: 10.7717/peerj-cs.1038.

15 Understanding Technical and Psychosocial Barriers to Realizing FAIR Data Process, https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-34942.pdf, (accessed December 2025).

16 FAIR Data Accelerator Pilot: Cultivating Cultures of Data Sharing. Project Overview, https://discovery.ucl.ac.uk/id/eprint/10209254/, (accessed December 2025).

17 L. Chisholm, F. Durán del Fierro, A. Littlejohn and E. Kennedy, *FAIR Data Accelerator Pilot: Cultivating Cultures of Data Sharing, Project Overview*, Zenodo, 2025, DOI: 10.5281/zenodo.15207800.

18 K. Stracke and J. D. Evans, *Commun. Chem.*, 2024, **7**, 63, DOI: 10.1038/s42004-024-01143-0.

19 M. Scheffler, M. Aeschlimann, M. Albrecht, *et al.*, *Nature*, 2022, **604**, 635–642, DOI: 10.1038/s41586-022-04501-x.

20 N. A. Krans, A. Ammar, P. Nymark, E. L. Willighagen, M. I. Bakker and J. T. K. Quik, *NanoImpact*, 2022, **27**, 100402, DOI: 10.1016/j.impact.2022.100402.

21 M. Suvarna and J. Pérez-Ramírez, *Nat. Catal.*, 2024, **7**, 624–635, DOI: 10.1038/s41929-024-01150-3.

22 Physical Sciences Data Infrastructure, https://www.psdi.ac.uk/, (accessed December 2025).

23 N. J. Knight, J. Bicarregui, S. J. Coles, W. Zhang, B. Matthews and J. G. Frey, *The Physical Sciences Data Infrastructure (PSDI) in the UK: Resources to accelerate physical sciences research*, Zenodo, 2025, DOI: 10.5281/zenodo.16736311.

24 L. D. Hughes, G. Tsueng, J. DiGiovanna, *et al.*, *Sci. Data*, 2023, **10**, 98, DOI: 10.1038/s41597-023-01969-8.

25 J. Ortmeyer and J. D. Jolliffe, *Nachr. Chem.*, 2022, **70**, 16–17, DOI: 10.1002/nadc.20224131398.

26 P. Tremouilhac, A. Nguyen, Y.-C. Huang, *et al.*, *J. Cheminf.*, 2017, **9**, 54, DOI: 10.1186/s13321-017-0240-0.

27 S. Kotov, P. Tremouilhac, N. Jung, *et al.*, *J. Cheminf.*, 2018, **10**, 38, DOI: 10.1186/s13321-018-0292-9.

28 P.-C. Huang, C.-L. Lin, P. Tremouilhac, *et al.*, *Nat. Protoc.*, 2025, **20**, 1097–1098, DOI: 10.1038/s41596-024-01074-z.

29 S. Herres-Pawlis, J. Liermann and O. Koepler, *Z. Anorg. Allg. Chem.*, 2020, **646**, 1748–1757, DOI: 10.1002/zaac.202000339.

30 S. Herres-Pawlis, O. Koepler and J. Liermann*Dataset: First NFDI4Chem User Survey*, LUIS, 2020, DOI: 10.25835/0077933.

31 O. Koepler, J. Liermann, F. Schön and S. Herres-Pawlis, *Nachr. Chem.*, 2020, **68**, 20–23, DOI: 10.1002/nadc.20204095910.

32 DFG Guidelines on the Handling of Research Data, https://www.dfg.de/resource/blob/172098/4ababf7a149da4247d018931587d76d6/guidelines-research-data-data.pdf, (accessed December 2025).

33 S. Kanza, C. Willoughby, N. J. Knight, C. L. Bird, J. G. Frey and S. J. Coles, *Digital Discovery*, 2023, **2**, 602–617, DOI: 10.1039/D2DD00121G.

34 A. R. McCluskey, M. Rivera and A. S. J. S. Mey, *Nat. Chem.*, 2024, **16**, 1383–1384, DOI: 10.1038/s41557-024-01613-x.

35 C. Steinbeck, N. Jung, F. Bach, *et al.*, *Res. Ideas Outcomes*, 2025, **11**, e177037, DOI: 10.3897/rio.11.e177037.

36 *MIChI Standard for Reporting Liquid-State NMR Experiments of Small Molecules (MARGARITAS)*, FAIRsharing, 2024, DOI: 10.25504/FAIRsharing.c29400.

37 UV-Vis MIChI Draft, https://www.nfdi4chem.de/uv-vis-michi-draft-is-ready/, (accessed December 2025).

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