

OPEN  
ARTICLE

# re3data – Indexing the Global Research Data Repository Landscape Since 2012

Heinz Pampel<sup>1,2</sup>✉, Nina Leonie Weisweiler<sup>1,2</sup>✉, Dorothea Strecker<sup>1</sup>, Michael Witt<sup>3</sup>, Paul Vierkant<sup>4</sup>, Kirsten Elger<sup>5</sup>, Roland Bertelmann<sup>2</sup>, Matthew Buys<sup>4</sup>, Lea Maria Ferguson<sup>2</sup>, Maxi Kindling<sup>6</sup>, Rachael Kotarski<sup>7</sup> & Vivien Petras<sup>1</sup>

For more than ten years, re3data, a global registry of research data repositories (RDRs), has been helping scientists, funding agencies, libraries, and data centers with finding, identifying, and referencing RDRs. As the world's largest directory of RDRs, re3data currently describes over 3,000 RDRs on the basis of a comprehensive metadata schema. The service allows searching for RDRs of any type and from all disciplines, and users can filter results based on a wide range of characteristics. The re3data RDR descriptions are available as Open Data accessible through an API and are utilized by numerous Open Science services. re3data is engaged in various initiatives and projects concerning data management and is mentioned in the policies of many scientific institutions, funding organizations, and publishers. This article reflects on the ten-year experience of running re3data and discusses ten key issues related to the management of an Open Science service that caters to RDRs worldwide.

## Introduction

In the 2010s, making research data publicly accessible gained importance: Terms such as e-science<sup>1</sup> and cyber-science<sup>2</sup> were shaping discourses about scientific work in the digital age. Various discussions within the scientific community<sup>3–8</sup> resulted in an increased awareness of the value of permanent access to research data. Policy recommendations of the Organization for Economic Co-operation and Development (OECD)<sup>9</sup> or the European Commission<sup>10</sup> reflected this shift.

The need for professional data management was increasingly emphasized with the publication of the now widely recognized *FAIR Data Principles*<sup>11</sup>. Researchers, academic institutions, and funders started to address this issue in policies<sup>12</sup>, initiatives and networks<sup>13–15</sup>, and infrastructures<sup>16–19</sup>. For example, the National Science Foundation (NSF) in the United States published a Data Sharing Policy in 2011, in which the funding agency required beneficiaries to provide information about data handling in a Data Management Plan<sup>20</sup>. In Germany, the German Research Foundation (DFG) published a similar statement regarding access to research data in the 2010s<sup>21,22</sup>.

The handling of research data was also discussed in library and computing center communities: In 2009, the German Initiative for Networked Information (DINI), a network of information infrastructure providers, published a position paper on the need for research data management (RDM) at higher education institutions<sup>23</sup>. Through the discussions within DINI, the need for a registry of RDRs became evident. At the time, the Directory of Open Access Repositories (OpenDOAR)<sup>24</sup> had already established itself as a directory of subject and institutional Open Access repositories. However, there was no comparable directory for RDRs, and it remained unclear how many repositories dedicated to research data existed.

In 2011, a consortium of research institutions in Germany submitted a proposal to the German Research Foundation (DFG), asking for funding to develop 're3data – Registry of Research Data Repositories'<sup>25</sup>. Members of the consortium were the Karlsruhe Institute of Technology (KIT), the Humboldt-Universität zu Berlin, and

<sup>1</sup>Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Berlin, Germany. <sup>2</sup>Helmholtz Association, Helmholtz Open Science Office, Potsdam, Germany. <sup>3</sup>University of Purdue, Distributed Data Curation Center, West Lafayette, IN, USA. <sup>4</sup>DataCite - International Data Citation Initiative e. V, Hannover, Germany. <sup>5</sup>GFZ German Research Centre for Geosciences, Library and Information Services, Potsdam, Germany. <sup>6</sup>Freie Universität Berlin, Open-Access-Büro Berlin, Berlin, Germany. <sup>7</sup>University of Bath, Library, Bath, UK. ✉e-mail: [heinz.pampel@hu-berlin.de](mailto:heinz.pampel@hu-berlin.de); [nina.weisweiler@os.helmholtz.de](mailto:nina.weisweiler@os.helmholtz.de)

the Helmholtz Open Science Office at the GFZ German Research Centre for Geosciences. The DFG approved the proposal in the same year. The project aimed to develop a service that would help researchers identify suitable RDRs to store their research data. re3data went online in 2012, and already listed 400 RDRs one year later<sup>26</sup>.

While working on the registry, the project team in Germany became aware of a similar initiative in the USA. With support from the Institute of Museum and Library Services, Purdue and Pennsylvania State University libraries developed Databib, a 'curated, global, online catalog of research data repositories'<sup>27</sup>. Databib went online in the same year<sup>28</sup>. At the time, RDRs were indexed and curated by library staff at re3data partner institutions, whereas Databib had established an international editorial board to curate RDR descriptions<sup>27</sup>. Databib and re3data signed a Memorandum of Understanding in 2012, and, following excellent cooperation, the two services merged in 2014<sup>29</sup>. The merger brought together successful ideas from each service: The metadata schemas were combined, resulting in version 2.2 of the re3data metadata schema<sup>30</sup>, and the sets of RDR descriptions were merged. The international editorial board of Databib was expanded to include re3data editors. Development of the IT infrastructure of re3data continued, combining the expertise both services had built. For operating the service, a management duo was installed, comprising a member each from institutions representing re3data and Databib.

The two services have always been closely corresponding with DataCite, an international not-for-profit organization that aims to ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future<sup>31</sup>. In this process, the main objective was to cover the interests of the global community of operators more comprehensively. In 2015, the DataCite Executive Board and the General Assembly decided to enter into an agreement with re3data, making re3data a DataCite partner service<sup>29</sup>. In 2017, re3data won the Oberly Award for Bibliography in the Agricultural or Natural Sciences from the American Libraries Association<sup>32</sup>.

Today, re3data is the largest directory of RDRs worldwide, indexing over 3,000 RDRs as of March 2023. re3data is widely used by academic institutions, funding organizations, publishers, journals, and various other stakeholders, such as the European Open Science Cloud (EOSC) and the National Research Data Infrastructure in Germany (NFDI). re3data metadata is also used to monitor and study the landscape of RDRs, and it is reused by numerous tools and services. Third-party-funded projects support the continuous development of the service. Currently, the DFG is funding the development of the service within the project re3data COREP<sup>33,34</sup>. In addition, the project partners DataCite and KIT bring the re3data perspective into EOSC projects such as FAIRsFAIR (completed)<sup>35</sup> and FAIR-IMPACT<sup>29</sup>.

This article outlines the decade-long experience of managing a widely used registry that supports a diverse and global community of stakeholders. The article is clustered around ten key issues that have emerged over time. For each of the ten issues, we first present a brief definition from the perspective of re3data. We then describe our approach to addressing the issue, and finally, we offer a reflection on our work.

## Results

The section outlines ten key issues that have emerged in the last ten years of operating re3data.

**Openness.** *Scope.* For re3data, Open Science means providing unrestricted access to the re3data metadata and schema, transparency of the indexing process, as well as open communication with the community of global RDRs.

*Approach.* At all times, re3data has been committed to Open Science by striving to be transparent and by sharing metadata. The openness of re3data pertains not only to the handling of its metadata and the associated infrastructure, but also to collaborative engagements with the community of research data stewards and other stakeholders in the field of research data management.

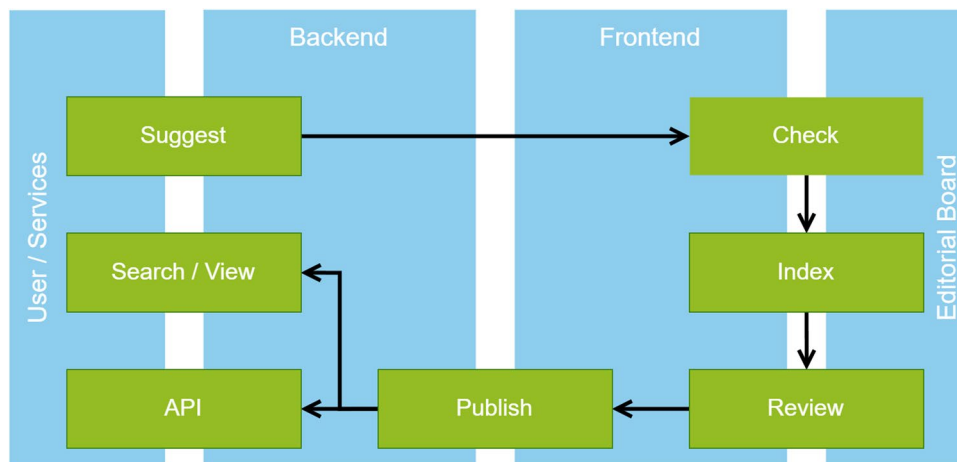
An example of this is the development of the re3data metadata schema: The initial version of the schema integrated a request for comments that allowed stakeholders to offer suggestions and improvements<sup>26</sup>. This participatory approach, accompanied by a public relations campaign, has yielded positive outcomes. Numerous experts engaged in the request for comments and contributed their perspective and expertise. Based on the positive feedback, we subsequently integrated a participatory phase in further updates of the metadata schema<sup>30,36</sup>.

In addition to this general commitment to openness, re3data has made its metadata available under the Creative Commons deed CC0. Due to adopting this highly permissive license, re3data metadata is strongly utilized by other parties, thereby enabling the development of new and innovative services and tools. Moreover, adaptable Jupyter Notebooks<sup>37</sup> have been published to facilitate the use of the re3data metadata. Additionally, workshops<sup>38</sup> have been arranged to support individuals in working with the notebooks and re3data data in general.

As a registry of RDRs, re3data also promotes Open Science by helping researchers find suitable repositories for publishing their data. For researchers who are looking for a repository that supports Open Science practices, re3data offers concise information on repository openness via its icon system. A recent analysis showed that most repositories indexed in re3data are considered 'open'<sup>39</sup>.

*Lessons learned.* The extensive reuse of re3data metadata increases its overall value, and participatory phases allow for incorporating different perspectives and experiences.

**Quality assurance.** *Scope.* For re3data, quality assurance encompasses all processes to ensure a service that meets the needs of a global community, as well as verifiably high-quality information.



**Fig. 1** Schematic overview of the editorial process in re3data.

**Approach.** High-quality RDR descriptions are at the core of re3data. Therefore, continuous efforts ensure that re3data metadata describes appropriately and correctly. Figure 1 shows the editorial process in re3data. Anyone, for example RDR operators, can submit repositories to be indexed in re3data by providing the repository name, URL, and some other core properties via a web form<sup>40</sup>. The re3data editorial board analyzes if the suggested RDR conforms with the re3data registration policy<sup>40</sup>. The policy requires that the RDR is operated by a legal entity, such as a library or university, and that the terms of use are clearly communicated. Additionally, the RDR must have a focus on storing and providing access to research data. If an RDR meets these requirements, it is indexed based on the re3data metadata schema. A member of the editorial board creates an initial RDR description, which is then reviewed by another editor. This approach has proven effective in resolving any inconsistencies in interpreting RDR characteristics. An indexing manual explains how the schema is to be applied and helps to ensure consistency between RDR descriptions. Once this review is complete, the RDR description is made publicly visible.

re3data applies a number of measures to ensure the long-term quality and consistency of RDR descriptions, including automated quality checks. For example, it is periodically checked whether the URLs of the RDR still resolve – if not, the entry of a RDR is reexamined. Figure 2 shows a screenshot of a re3data RDR description.

The re3data metadata schema on which RDR descriptions are based is reviewed and updated regularly to ensure that users' changing information needs are met. Operators of an RDR, as well as any other person, can suggest changes to RDR descriptions by submitting a change request. A link for filing a change request can be found at the bottom of each RDR description in re3data. Once a change request has been submitted, a member of the editorial board will review the proposed changes and verify them against information on the RDR website. If the change request is deemed valid, the RDR description will be adapted accordingly.

As part of the project re3data COREF, quality assurance practices at RDRs were systematically investigated. The aim was to understand how RDRs ensure high-quality data, and to better reflect these measures in the metadata schema. The results of the study<sup>41</sup>, which were based on a survey among RDR operators, show that approaches to quality assurance are diverse and depend on the mission and scope of the RDR. However, RDRs are key actors in enabling quality assurance. Furthermore, there is a path dependence of data review on the review process of textual publications. In addition to the study, a workshop<sup>42,43</sup> was held with CoreTrustSeal that focused on quality assurance measures RDRs have implemented. CoreTrustSeal is a RDR certification organization launched in 2017 that defines requirements for base-level certification for RDRs<sup>44</sup>.

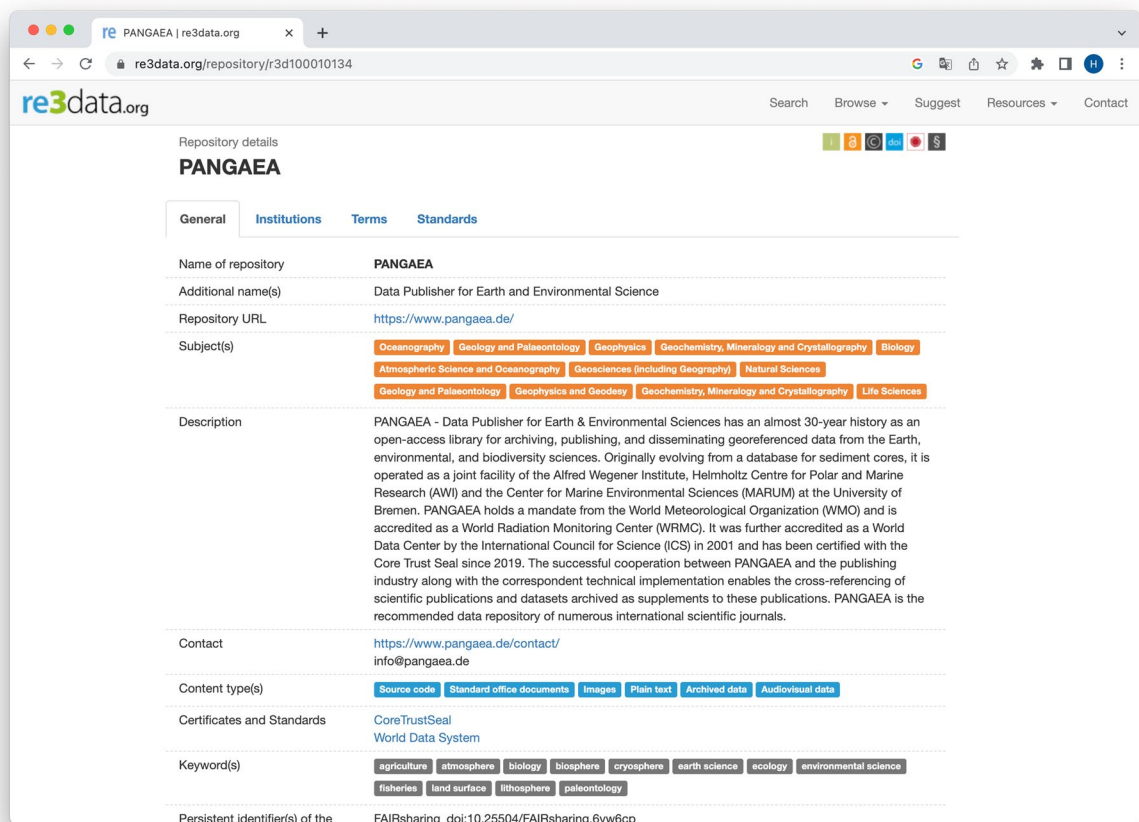
**Lessons learned.** Combining manual and automated verification was shown to be most effective in ensuring that RDR descriptions remain consistent while meeting users' diverse information needs.

**Community engagement.** *Scope.* For re3data, community engagement encompasses all activities that ensure interaction with the global RDR community in a participatory process.

**Approach.** Collaboration has always been a central principle for re3data. This is reflected in the fact that research communities, RDR providers, and other relevant stakeholders contribute significantly to the completeness and accuracy of the re3data metadata as well as its further technical and conceptual development. Examples include the participatory phase during the revision of the metadata schema, the involvement of important stakeholders in the development of the *re3data Conceptual Model for User Stories*<sup>45,46</sup>, or the activities that investigate data quality assurance at RDRs.

re3data engages in collaborations in various forms with diverse stakeholders, for example:

- In collaboration with the Canadian Data Repositories Inventory Project and later with the Digital Research Alliance of Canada, both initiatives aiming at describing the Canadian landscape of RDRs comprehensively, descriptions of Canadian RDRs in re3data were improved, and additional RDRs were indexed<sup>47,48</sup>.



**Fig. 2** Screenshot of the re3data description of the research data repository PANGAEA<sup>97</sup>.

- A collaboration initiative was initiated in Germany with the Helmholtz Metadata Collaboration (HMC). In this initiative, the descriptions of research data infrastructures within the Helmholtz Association are being reviewed and enhanced<sup>49</sup>.
- re3data also engages in international networks, particularly within the Research Data Alliance (RDA). Activities focus on several RDA working and interest groups<sup>50–52</sup> that touch on topics relevant to RDR registries.

**Lessons learned.** Combining strategies of engagement connects the service to its stakeholders and creates opportunities for collaboration and innovation.

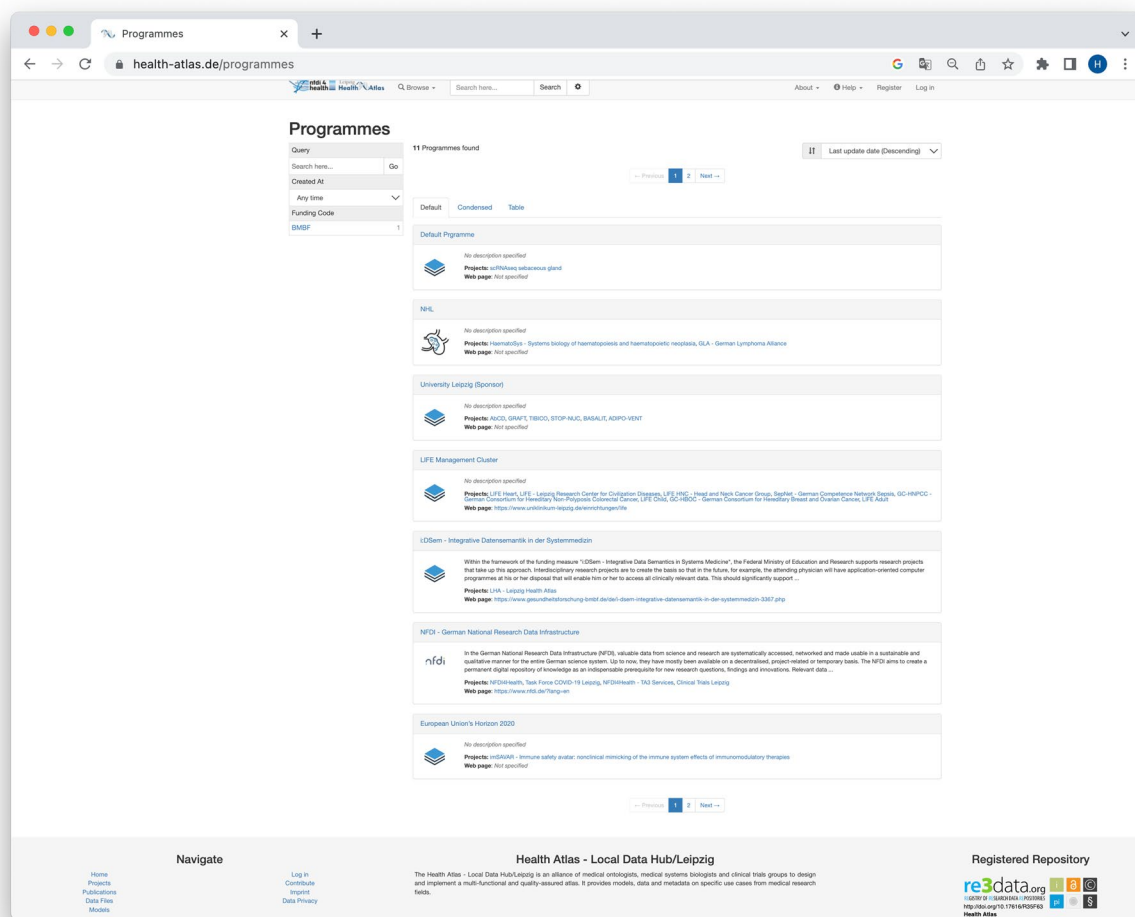
**Interoperability.** *Scope.* For re3data, interoperability means facilitating interactions and metadata exchange with the global RDR community by relying on established standards.

*Approach.* Interoperability is a necessary condition to integrate a service into a global network of diverse stakeholders. International standards must be implemented to achieve this, for example with the re3data API<sup>53</sup>. The API can be used to query various parameters of an RDR as expressed in the metadata schema. The API enables the machine readability and integration of re3data metadata into other services. The re3data API is based on the RESTful API concept and is well-documented. Applying the HATEOAS principles<sup>54</sup> enables the decoupling of clients and servers, and thus allows for independent development of server functionality. This results in a robust interface that promotes interoperability and reduces barriers to future use. Also, re3data supports OpenSearch, a standard that enables interaction with search results in a format suitable for syndication and aggregation.

Interoperability also guides the development of the metadata schema: Established vocabularies and standards are used to describe RDRs wherever possible. Examples of standards used in the metadata schema include:

- ISO 639-3 for language information, for example a RDR name
- ISO 8601 for the use of date information on a RDR
- DFG Classification of Subject Areas for subject information on a RDR

In addition, re3data pursues interoperability by jointly working on a mapping between the DFG Classification of Subject Areas used by re3data and the OECD Fields of Science classification used by DataCite<sup>55</sup>.



**Fig. 3** The re3data badge integrated in the research data repository Health Atlas.

re3data records whether an RDR has obtained formal certification, for example by World Data System (WDS) or CoreTrustSeal. The certification status, along with other properties, is visualized by the re3data icon system that makes the core properties of RDRs easily accessible visually. The icon system provides information about the openness of the RDR and its data collection, the use of PID systems, as well as the certification status. The icon system can also be integrated into RDR websites via badges<sup>56</sup>. Figure 3 shows an example of a re3data badge.

re3data captures information that might be relevant to metadata aggregator services, including API URLs, as well as the metadata standard(s) used. In offering this information in a standardized form, re3data fosters the development of services that span multiple collections, such as data portals. For example, as part of the FAIRFAIR project work, re3data metadata has been integrated into DataCite Commons<sup>57</sup> to embed repository information in the DataCite PID Graph. This step not only improves the discoverability of repositories that support research data management in accordance with the FAIR principles but also serves as a basis for the development of new services such as the FAIR assessment tool F-UJI<sup>35,58</sup>.

**Lessons learned.** The adherence to established standards facilitates the reuse of re3data metadata and increases the integration of the service into the broader Open Science landscape.

**Development.** *Scope.* For re3data, continuous development ensures that the service is able to respond dynamically to evolving requirements of the global RDR community.

**Approach.** Maintaining a registry for an international community poses a significant challenge, particularly the continued provision of reliable technical operations and a governance structure capable of responding adequately to user demands. re3data has found suitable solutions to these challenges, which have enabled the service to be in operation for more than ten years. The long-standing collaboration with DataCite has contributed to this success. Participation in third-party-funded projects has facilitated the collaborative development of core service elements together with partners. Participation in committees such as those surrounding EOSC

and RDA, as well as active engagement with the RDR community, have motivated discussions about changing requirements and led to the continuous evolution of the registry.

Responsibilities for specific tasks are divided among several entities, such as a working group responsible for guiding future directions of the service and the editorial board responsible for maintaining re3data metadata. In addition, there are teams responsible for technology as well as for outreach and communication. The working group includes experts from DataCite and other stakeholders, who discuss current requirements, prioritize developments, and ensure coordination with RDR operators worldwide. In addition to these entities, coordination with third-party-funded projects involving re3data is ongoing.

*Lessons learned.* Continuous and agile development addresses the users' constantly evolving needs. Operating a registry that meets those needs in the long term requires flexibility.

**Sustainability.** *Scope.* For re3data, sustainability means ensuring a long-term and reliable service to the global RDR community.

*Approach.* Maintaining the sustainable operation of a service like re3data beyond an initial project phase is a challenge. For re3data, the consortium model has proven effective, as the service is supported by a wide range of scientific institutions. This model, which is embedded in the governance of re3data, allows the operation of the service to be sustained through self-funding while also enabling important developments to be undertaken within the scope of third-party projects. Thanks to funding received from the DFG (re3data COREF project) and the European Union's Horizon 2020 program (FAIRsFAIR project), significant investments have been made in the IT infrastructure and overall advancement of the service in recent years.

*Lessons learned.* A strategy based on diverse revenue streams contributes to securing funding for the service long-term.

**Policies.** *Scope.* For re3data, being mentioned in policies comes with a responsibility for operating a reliable service and maintaining high-quality metadata for the global RDR community.

*Approach.* During the development of the re3data service, the partners engaged in dialogues with various stakeholders that were interested in using the registry to refer to RDRs in their policies. They might do this, for example, to recommend or mandate the use of RDRs in general for publishing research data, or the use of a specific RDR. Today, re3data is mentioned in the policies of several funding agencies, scientific institutions, and journals. These actors use re3data to identify RDRs operated by specific academic institutions that were developed using funding from a funding organization, or that store data that are the basis of a journal article. Examples of policies and policy guidance documents that refer to re3data:

- Academic institutions:
  - Brandon University, Canada<sup>59</sup>
  - Technische Universität Berlin, Germany<sup>60</sup>
  - University of Edinburgh, United Kingdom<sup>61</sup>
  - University of Eastern Finland<sup>62</sup>
  - Western Norway University of Applied Sciences<sup>63</sup>
- Funders:
  - Bill & Melinda Gates Foundation, USA<sup>64</sup>
  - European Commission<sup>65</sup> and ERC, EU<sup>66</sup>
  - National Science Foundation (NSF), USA<sup>67</sup>
  - NIH, USA<sup>68</sup>
- Journals and Publishers:
  - Taylor & Francis, United Kingdom<sup>69</sup>
  - Springer Nature, United Kingdom<sup>70</sup>
  - Sage, United Kingdom<sup>71</sup>
  - Wiley, Germany<sup>72</sup>

Regular searches are conducted to track mentions of re3data in policies. On the re3data website, a list of policies referring to re3data is maintained and regularly updated<sup>73</sup>.

As a result of being mentioned in policies so frequently, re3data receives inquiries from researchers for information on listed RDRs almost daily. These inquiries are usually forwarded to the RDR directly.

*Lessons learned.* Policies represent firm support for research data management by academic institutions, funders, and journals and publishers. By facilitating the search for and referencing of RDRs in policies, re3data further promotes Open Science practices.

**Data reuse.** *Scope.* For re3data, data reuse is one of the main objectives, ensuring that third parties can rely on re3data metadata to build services that support the global RDR community.

*Approach.* Because re3data metadata are published as open data, third parties are free to integrate it into their systems. Several service operators have already taken advantage of this opportunity. In general, there are three types of services that work with re3data data:

- (1) Services for finding and describing RDRs: These services usually work with a subset of re3data metadata. Sometimes, the data is manually curated, and then integrated into external services based on specific parameters. Examples include:
  - DARIAH-EU has developed its Data Deposit Recommendation Service based on a subset of re3data metadata, which helps humanities researchers find suitable RDRs<sup>74,75</sup>.
  - The American Geophysical Union (AGU) has utilized re3data metadata to create a dedicated gateway for RDRs in the geosciences with its Repository Finder tool<sup>76,77</sup>, which was later incorporated into the DataCite Commons web search interface.
- (2) Services for monitoring the landscape of RDRs: These services analyze re3data metadata using specific parameters and visualize the results. Examples include:
  - OpenAIRE has integrated re3data metadata into its Open Science Observatory to provide information on RDRs that are part of OpenAIRE<sup>78</sup>.
  - The European Commission operates the Open Science Monitor, a dashboard that analyzes re3data metadata. The following metrics are displayed: number of RDRs by subject, number of RDRs by access type, and number of RDRs by country<sup>79,80</sup>.
- (3) Services for assessing RDRs: These services use re3data metadata and other data sources to evaluate RDRs more comprehensively. Examples include:
  - The F-UJI Automated FAIR Data Assessment Tool is a web-based service that assesses the degree to which individual datasets conform to the FAIR Data principles. The tool utilizes re3data metadata to evaluate characteristics of the RDR that store the datasets<sup>81</sup>.
  - Charité Metrics Dashboard, a dashboard on responsible research practices from the Berlin Institute of Health at Charité in Berlin, Germany, builds on F-UJI data and combines this information with additional re3data metadata<sup>82</sup>.

These examples underscore the value Open Science tools like re3data generate by making their data openly available without restrictions. As a result of the permissive licensing, re3data metadata can be used for new and innovative applications, establishing re3data as a vital data provider for the global Open Science community.

*Lessons learned.* Permissive licensing and extensive collaboration have turned re3data into a key data provider in the Open Science ecosystem.

**Metadata for research.** *Scope.* For re3data, providing RDR descriptions also means offering metadata that enables analyses of the global RDR community.

*Approach.* In research disciplines studying data infrastructures, for example library and information science or science and technology studies, re3data is regularly used for information on the state of research infrastructures. As re3data has been mapping the landscape of data infrastructures for ten years, it has evolved into a tool that is used for monitoring Open Science activities, research data management, and other topics. Studies reusing re3data metadata include analyses of the overall RDR landscape, the landscape of RDRs in a specific domain, or the RDR landscape of a region or country. Some examples of studies reusing re3data metadata for research are:

- Overall studies: Boyd<sup>83</sup> examined the extent to which RDR exhibit properties of infrastructures. Khan & Ahangar<sup>84</sup> and Hansson & Dahlgren<sup>85</sup> focused on the openness of RDRs from a global perspective.
- Regional studies: Bauer *et al.*<sup>86</sup> examined Austrian RDRs, Cho<sup>87</sup> Asian RDRs, Milzow *et al.*<sup>88</sup> Swiss RDRs, and Schöpfel<sup>89</sup> French RDRs.
- Domain studies: Gómez *et al.*<sup>90</sup> and Li & Liu<sup>91</sup> investigated the landscape of RDRs in humanities and social science. Prashar & Chander<sup>92</sup> focused on computer science.

Members of the re3data team have also published studies reusing re3data metadata, including studies of the global state of RDR<sup>93</sup>, openness<sup>39</sup>, and quality assurance of RDRs<sup>41</sup>.

In response to the demand for information on the RDR landscape, the re3data graphical user interface provides various visualizations of the current state of RDRs. For example, re3data metadata can be browsed visually by subject category and on a map. In addition, the metrics page of re3data shows how RDRs are distributed across central properties of the metadata schema<sup>94</sup>.

The start page of re3data includes a recommendation for how to cite the service if it was used as a source in papers:

re3data - Registry of Research Data Repositories. <https://doi.org/10.17616/R3D> last accessed: [date].

In citing the service, the use of re3data as a data source in research and the service in general becomes more visible.

*Lessons learned.* The increasing number of studies reusing re3data metadata shows a real demand for reliable information on the global RDR landscape.

**Communications.** *Scope.* For re3data, communication means engaging in dialogue with relevant stakeholders in the global RDR community.

*Approach.* Broad-based public relations are very important for a service catering to a global community. In recent years, re3data has pursued a communication strategy that includes the following elements:

- Conference presentations: It has been proven effective to represent the service at conferences, paving new ways to engage with the community.
- Mailing lists: The re3data team regularly informs members of a variety of mailing lists about news from the service.
- Social media: re3data communicates current developments via Mastodon (<https://openbiblio.social/@re3data>) and Twitter (<https://twitter.com/re3data>).
- Help desk: Communication via the help desk is essential for the re3data service. The help desk team answers questions about RDR descriptions, as well as general questions about data management. The number of general inquiries, e.g., for finding a suitable RDR, has increased over the years.
- Blog: The project re3data COREF operates a blog that informs about developments in the project<sup>95</sup>. Some blog posts are also published in the DataCite Blog<sup>96</sup>.

*Lessons learned.* Establishing broad-based communication channels enables the service to reach and engage with relevant stakeholders in a variety of formats.

## Discussion

Over the past ten years, re3data has evolved into a reliable and valuable Open Science service. The service offers high-quality RDR descriptions from all disciplines and regions. re3data is managed cooperatively; new features are developed in third-party projects.

Four basic principles guide the development of re3data: openness, community engagement, high-quality metadata, and ongoing consideration of users' needs. These principles ensure that the activities of the service align with the values and interests of its stakeholders. In the context of these principles, ten key issues for the operation of the service have emerged over the last ten years.

In the past two years, following in-depth conversations with diverse parties, a new conceptual model for re3data was developed<sup>45</sup>. This process contributed to a better understanding of the needs of RDR operators and other stakeholders. The conceptual model will guide developments of re3data, embedding the service further in the evolving ecosystem of Open Science services with the intention to support researchers, scientific institutions, funding organizations, publishers, and journals in implementing the FAIR principles and realizing an interconnected global research data ecosystem.

## Methods

This article describes the history and current status of the global registry re3data. Based on operational experience, it reflects on some of the basic principles that have shaped the service since its inception.

Having been launched more than ten years ago, re3data is now the most comprehensive registry of RDRs. The service currently describes more than 3,000 RDRs and caters to a diverse user base including RDR operators, researchers, funding agencies, and publishers. Ten key issues that are relevant for operating an Open Science service like re3data are identified, discussed, and reflected: openness, quality assurance, community engagement, interoperability, development, sustainability, policies, data reuse, metadata for research, and communications. For each of the key issues, we provide a definition, explain the approach applied by the re3data service, and describe what the re3data team learned from working on each issue.

Among other aspects, the paper outlines the design, governance, and objectives of re3data, providing important background information on a service that has evolved into a central data source on the global RDR landscape.

## Data availability

The re3data RDR descriptions are openly available via <https://re3data.org> under a CC0 deed.

## Code availability

The source code of the directory is not publicly released. The re3data subject ontology and several Jupyter notebooks with examples for using the re3data API can be found at: <https://github.com/re3data>.

Received: 13 July 2023; Accepted: 9 August 2023;

Published online: 29 August 2023



## References

- National Science Foundation Cyberinfrastructure Council. *Cyberinfrastructure Vision for 21st Century Discovery 2007*. <https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf> (2023).
- National Science Foundation. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure 2003*. <https://www.nsf.gov/cise/sci/reports/atkins.pdf> (2023).
- How to encourage the right behaviour. *Nature* **416**, 1–1 (2002).
- Let Data Speak to Data. *Nature* **438**, 531–531 (2005).
- The Royal Society. *Science as an Open Enterprise* [https://royalsociety.org/~media/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf) (2023).
- Data for the masses. *Nature* **457**, 129–129 (2009).
- Data's shameful neglect. *Nature* **461**, 145–145 (2009).
- Science Staff. Challenges and opportunities. *Science* **331**, 692–693 (2011).
- OECD. *OECD Principles and Guidelines for Access to Research Data from Public Funding* (2007).
- European Commission. Commission Recommendation of 17 July 2012 on Access to and Preservation of Scientific Information (2012).
- Wilkinson, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).
- Pampel, H. & Bertelmann, R. in *Handbuch Forschungsdatenmanagement (2011)*. <https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/195> (2023).
- European Commission. *European Cloud Initiative - Building a Competitive Data and Knowledge Economy in Europe* (2016).
- Michener, W. *et al.* DataONE: Data observation network for earth preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine* **17** (2011).
- Parsons, M. A. The Research Data Alliance: Implementing the technology, practice and connections of a data infrastructure. *Bul. Am. Soc. Info. Sci. Tech.* **39**, 33–36 (2013).
- Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World* (The MIT Press, 2016).
- Manghi, P., Manola, N., Horstmann, W. & Peters, D. *An Infrastructure for Managing EC Funded Research Output - The OpenAIRE Project 2010*. <https://publications.goettingen-research-online.de/handle/2/57068> (2023).
- Blanke, T., Bryant, M., Hedges, M., Aschenbrenner, A. & Priddy, M. *Preparing DARIAH in 2011 IEEE Seventh International Conference on eScience*, 158–165 (IEEE, 2011).
- Hey, T. & Trefethen, A. in *Scientific Collaboration on the Internet* (eds Olson, G. M., Zimmerman, A. & Bos, N.) 14–31 (The MIT Press, 2008).
- National Science Board. *Digital Research Data Sharing and Management 2011*. <https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf> (2023).
- Deutsche Forschungsgemeinschaft. *Empfehlungen Zur Gesicherten Aufbewahrung Und Bereitstellung Digitaler Forschungsprimärdaten 2009*. [https://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](https://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) (2023).
- Allianz der deutschen Wissenschaftsorganisationen. *Grundsätze Zum Umgang Mit Forschungsdaten* <https://doi.org/10.2312/ALLIANZOA.019> (2010).
- DINI Working Group Electronic Publishing. *Positionspapier Forschungsdaten*. <https://doi.org/10.18452/1489> (2009).
- OpenDOAR. <https://beta.jisc.ac.uk/opendoar> (2023).
- Deutsche Forschungsgemeinschaft. *Re3data.Org - Registry of Research Data Repositories. Community Building, Net working and Research Data Management Services GEPRIS*. <https://gepris.dfg.de/gepris/projekt/209240528?context=projekt&task=showDetail&id=209240528&> (2023).
- Pampel, H. *et al.* Making research data repositories visible: the re3data.org registry. *PLoS ONE* **8** (ed Suleman, H.) e78080 (2013).
- Witt, M. *Databib: Cataloging the World's Data Repositories 2013*. <https://ir.inflibnet.ac.in:8443/ir/handle/1944/1778> (2023).
- Witt, M. & Giarlo, M. *Databib: IMLS LG-46-11-0091-11 Final Report (White Paper) 2012*. <https://docs.lib.purdue.edu/libreports/2> (2023).
- Buys, M. *Strategic Collaboration 2022*. [https://datacite.org/assets/re3data%20and%20DataCite\\_openHours.pdf](https://datacite.org/assets/re3data%20and%20DataCite_openHours.pdf) (2023).
- Vierkant, P. *et al.* *Metadata Schema for the Description of Research Data Repositories: version 2.2*. <https://doi.org/10.2312/RE3.006> (2014).
- Brase, J. *DataCite - A Global Registration Agency for Research Data in 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, 257–261 (IEEE, 2009).
- Witt, M. *DataCite's Re3data Wins Oberly Award from the American Libraries Association* [https://doi.org/10.5438/0001-0HN\\*](https://doi.org/10.5438/0001-0HN*).
- Deutsche Forschungsgemeinschaft. *Re3data - Offene Und Nutzerorientierte Referenz Fur Forschungsdatenrepositorien (Re3data COREF) GEPRIS*. <https://gepris.dfg.de/gepris/projekt/422587133?context=projekt&task=showDetail&id=422587133&> (2023).
- re3data COREF. *Re3data COREF Project* <https://coref.project.re3data.org/project> (2023).
- FAIRsFAIR. *Repository Discovery in DataCite Commons* <https://www.fairsfair.eu/repository-discovery-datacite-commons> (2023).
- Strecker, D. *et al.* *Metadata Schema for the Description of Research Data Repositories: version 3.1*. <https://doi.org/10.48440/RE3.010> (2021).
- re3data. *Examples for using the re3data API GitHub*. [https://github.com/re3data/using\\_the\\_re3data\\_API](https://github.com/re3data/using_the_re3data_API) (2023).
- Schabinger, R., Strecker, D., Wang, Y. & Weisweiler, N. L. *Introducing Re3data - the Registry of Research Data Repositories*. <https://doi.org/10.5281/ZENODO.5592123> (2021).
- re3data COREF. *How Open Are Repositories in Re3data?* <https://coref.project.re3data.org/blog/how-open-are-repositories-in-re3data> (2023).
- re3data. *Suggest* <https://www.re3data.org/suggest> (2023).
- Kindling, M. & Strecker, D. Data quality assurance at research data repositories. *Data Science Journal* **21**, 18 (2022).
- Kindling, M. *et al.* Report on re3data COREF/CoreTrustSeal workshop on quality management at research data repositories. *Informationspraxis* **8** (2022).
- Kindling, M., Strecker, D. & Wang, Y. *Data Quality Assurance at Research Data Repositories: Survey Data* (Zenodo, 2022).
- L'Hours, H., Kleemola, M. & De Leeuw, L. CoreTrustSeal: From academic collaboration to sustainable services. *IASSIST Quarterly* **43**, 1–17 (2019).
- Vierkant, P. *et al.* *Re3data Conceptual Model for User Stories*. <https://doi.org/10.48440/RE3.012> (2021).
- Weisweiler, N. L. *et al.* *Re3data Stakeholder Survey and Workshop Report*. <https://doi.org/10.48440/RE3.013> (2021).
- Webster, P. *Integrating Discovery and Access to Canadian Data Sources. Contributing to Academic Library Data Services by Sharing Data Source Knowledge Nation Wide* in. In collab. with Haigh, S. (2017). <https://library.ifa.org/id/eprint/2514/> (2023).
- Dearborn, D. *et al.* *Summary Report: Canadian Research Data Repositories and the Re3data Repository Registry* in collab. with Labrador, A. & Purcell, F. (2023).
- Helmholtz Open Science Office. *Community Building for Research Data Repositories in Helmholtz* <https://os.helmholtz.de/en/open-science-in-helmholtz/networking/community-building-research-data-repositories/> (2023).
- Research Data Alliance. *Libraries for Research Data IG* <https://www.rd-alliance.org/groups/libraries-research-data.html> (2023).
- Research Data Alliance. *Data Repository Attributes WG* <https://www.rd-alliance.org/groups/data-repository-attributes-wg> (2023).
- Research Data Alliance. *Data Granularity WG* <https://www.rd-alliance.org/groups/data-granularity-wg> (2023).

53. re3data. *API* <https://www.re3data.org/api/doc> (2023).
54. HATEOAS <https://en.wikipedia.org/w/index.php?title=HATEOAS&oldid=1141349344> (2023).
55. Ninkov, A. B. *et al.* *Mapping Metadata - Improving Dataset Discipline Classification*. <https://doi.org/10.5281/ZENODO.6948238> (2022).
56. Pampel, H. *Re3data.Org Reaches a Milestone and Begins Offering Badges* <https://doi.org/10.5438/KTR7-ZJJH>.
57. DataCite. *DataCite Commons* <https://commons.datacite.org/repositories> (2023).
58. Wimalaratne, S. *et al.* *D4.7 Tools for Finding and Selecting Certified Repositories for Researchers and Other Stakeholders*. <https://doi.org/10.5281/ZENODO.6090418> (2022).
59. Brandon University. *Research Data Management Strategy* <https://www.brandonu.ca/research/files/Research-Data-Strategy.pdf> (2023).
60. Technical University Berlin. *Research Data Policy of TU Berlin* <https://www.tu.berlin/en/working-at-tu-berlin/important-documents/guidelinesdirectives/research-data-policy> (2023).
61. The University of Edinburgh. *Research Data Management Policy* <https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy> (2023).
62. University of Eastern Finland. *Data management at the end of research* <https://www.uef.fi/en/datasupport/data-management-at-the-end-of-research> (n. d.).
63. Western Norway University of Applied Sciences. *Research Data* <https://www.hvl.no/en/library/research-and-publish/publishing/research-data/> (2023).
64. Gates Open Access Policy. *Data Sharing Requirements* <https://openaccess.gatesfoundation.org/how-to-comply/data-sharing-requirements/> (2023).
65. European Commission. *Horizon Europe (HORIZON) - Programme Guide 2022*. [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide\\_horizon\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf) (2023).
66. European Research Council. *Open Research Data and Data Management Plans - Information for ERC Grantee 2022*. [https://erc.europa.eu/sites/default/files/document/file/ERC\\_info\\_document-Open\\_Research\\_Data\\_and\\_Data\\_Management\\_Plans.pdf](https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf) (2023).
67. National Science Foundation. *Dear Colleague Letter: Effective Practices for Making Research Data Discoverable and Citable (Data Sharing)* <https://www.nsf.gov/pubs/2022/nsf22055/nsf22055.jsp> (2023).
68. National Institutes of Health. *Repositories for Sharing Scientific Data* <https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/repositories-for-sharing-scientific-data> (2023).
69. Taylor and Francis. *Understanding and Using Data Repositories* <https://authorservices.taylorandfrancis.com/data-sharing/share-your-data/repositories/> (2023).
70. Scientific Data. *Data Repository Guidance* <https://www.nature.com/sdata/policies/repositories> (2023).
71. SAGE. *Research Data Sharing FAQs* <https://us.sagepub.com/en-us/nam/research-data-sharing-faqs> (2023).
72. Wiley. *Data Sharing Policy* <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html> (2023).
73. re3data. *Publications* <https://www.re3data.org/publications> (2023).
74. Buddenbohm, S., de Jong, M., Minel, J.-L. & Moranville, Y. Find research data repositories for the humanities - the data deposit recommendation service. *Int. J. Digit. Hum.* **1**, 343–362 (2021).
75. DARIAH. *DDRS* <https://ddrs-dev.dariah.eu/ddrs/> (2023).
76. Witt, M. *et al.* in *Digital Libraries: Supporting Open Science* (eds Manghi, P., Candela, L. & Silvello, G.) 86–96 (Springer, 2019).
77. DataCite. *DataCite Repository Selector* <https://repositoryfinder.datacite.org/> (2023).
78. OpenAIRE. *Open Science Observatory* <https://osobservatory.openaire.eu/home> (2023).
79. The Lisbon Council. *Open Science Monitor Methodological Note 2019*. [https://research-and-innovation.ec.europa.eu/system/files/2020-01/open\\_science\\_monitor\\_methodological\\_note\\_april\\_2019.pdf](https://research-and-innovation.ec.europa.eu/system/files/2020-01/open_science_monitor_methodological_note_april_2019.pdf) (2023).
80. European Commission. *Facts and Figures for Open Research Data* [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en) (2023).
81. Devaraju, A. & Huber, R. *F-UJI - An Automated FAIR Data Assessment Tool Zenodo*. <https://doi.org/10.5281/ZENODO.4063720> (2023).
82. Berlin Institute of Health. *ChariteMetrics Dashboard* <https://quest-dashboard.charite.de/#tabMethods> (2023).
83. Boyd, C. Understanding research data repositories as infrastructures. *P. J. Asso. for Info. Science & Tech.* **58**, 25–35 (2021).
84. Khan, N. A. & Ahangar, H. *Emerging Trends in Open Research Data in 2017 9th International Conference on Information and Knowledge Technology*, 141–146 (2017).
85. Hansson, K. & Dahlgren, A. Open research data repositories: Practices, norms, and metadata for sharing images. *J. Asso. for Info. Science & Tech.* **73**, 303–316 (2022).
86. Bauer, B. & Ferus, A. Österreichische Repositorien in OpenDOAR und re3data.org: Entwicklung und Status von Infrastrukturen für Green Open Access und Forschungsdaten. *Mitteilungen der VOB* **71**, 70–86 (2018).
87. Cho, J. Study of Asian RDR based on re3data. *EL* **37**, 302–313 (2019).
88. Milzow, K., von Arx, M., Sommer, C., Cahenzli, J. & Perini, L. *Open Research Data: SNSF Monitoring Report 2017–2018*. <https://doi.org/10.5281/ZENODO.3618123> (2020).
89. Schopf, J. in *Schopf, J. & Rebouillat, V. Research Data Sharing and Valorization: Developments, Tendencies, Models* (Wiley, 2022).
90. Gomez, N.-D., Mendez, E. & Hernandez-Perez, T. Data and metadata research in the social sciences and humanities: An approach from data repositories in these disciplines. *EPI* **25**, 545 (2016).
91. Li, Z. & Liu, W. *Characteristics Analysis of Research Data Repositories in Humanities and Social Science - Based on Re3data.Org in 4th International Symposium on Social Science* (Atlantis Press, 2018).
92. Prashar, P. & Chander, H. *Research Data Management through Research Data Repositories in the Field of Computer Sciences* <https://ir.inflibnet.ac.in:8443/ir/bitstream/1944/2400/1/43.pdf> (2023).
93. Kindling, M. *et al.* The landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine* **23** (2017).
94. re3data. *Statistics* <https://www.re3data.org/metrics> (2023).
95. re3data COREF. *Re3data COREF Blog* <https://coref.project.re3data.org/> (2023).
96. Witt, M., Weisweiler, N. L. & Ulrich, R. *Happy 10th Anniversary, Re3data! DataCite*. <https://doi.org/10.5438/MQW0-YT07>.
97. Felden, J. *et al.* PANGAEA - Data Publisher for Earth & Environmental Science. *Scientific Data* **10**, 347 (2023).

## Acknowledgements

This work has been supported by the German Research Foundation (DFG) under the projects *re3data.org - Registry of Research Data Repositories. Community Building, Networking and Research Data Management Services* (Grant ID: 209240528) and *re3data - Offene und nutzerorientierte Referenz für Forschungsdatenrepositorien (re3data COREF)* (Grant ID: 422587133). The article processing charge was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491192747 and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

### Author contributions

H.P., N.L.W., D.S. and M.W. wrote the first draft. P.V., E.K., R.B., M.B., L.M.F., M.K., R.K. and V.P. provided critical feedback and helped shape the manuscript. H.P., N.L.W., D.S., M.W., P.V., E.K., R.B., M.B., L.M.F., M.K., R.K. and V.P. contributed to the final writing and revision of the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to H.P. or N.L.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023