





Data quality assurance at research data repositories – results from a survey

IDCC 2022, 2022-06-14

Authors:

Maxi Kindling¹ (maxi.kindling@hu-berlin.de)  [0000-0002-0167-0466](https://orcid.org/0000-0002-0167-0466), Dorothea Strecker¹  [0000-0002-9754-3807](https://orcid.org/0000-0002-9754-3807),
Vivien Petras¹  [0000-0002-8113-1509](https://orcid.org/0000-0002-8113-1509), Yi Wang¹  [0000-0003-1354-3461](https://orcid.org/0000-0003-1354-3461)

re3data COREF team:

Alexandra Axtmann², Heinz Pampel⁴, Kirsten Elger³, Lea Maria Ferguson⁴, Michael Witt⁵, Margarita Trofimenko², Nina Weisweiler⁴, Paul Vierkant⁶, Robert Ulrich², Roland Bertelmann⁴, Rouven Schabinger², Thanh Binh Nguyen²

¹Berlin School of Library and Information Science, Humboldt-Universität zu Berlin

²Karlsruher Institut für Technologie / Karlsruhe Institute of Technology (KIT)

³GFZ German Research Centre for Geosciences / Helmholtz-Zentrum Potsdam – Deutsches GeoForschungsZentrum (GFZ)

⁴Helmholtz-Gemeinschaft / Helmholtz Association, Helmholtz Open Science Office

⁵Purdue University, United States

⁶DataCite - International Data Citation Initiative e.V.



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

 [Search](#)

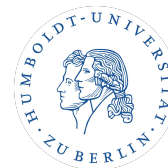
re3data COREF - objective



Increasing re3data's value for all stakeholders, further establishing the service as a tool for finding and referencing research data repositories.

Key Facts

- Starting date - 2020-01-01
- Duration - 36 months
- Funded by the German Research Foundation



Making research data quality assurance visible

making efforts of ensuring data quality at research data repositories visible in re3data

- studies of data quality and data quality assurance practices
- revising the re3data Metadata Schema ([r3d:qualityManagement](#)) based on the findings



Survey

Questionnaire: based on qualitative analyses of CoreTrustSeal self-assessment documents and data journal guidelines

Topics:

- repository characteristics
- formal assessment of data
- data curation
- data review
- quality criteria
- data evaluation by repository users
- indicators of data quality

Survey

Sampling:

Mailing list of 1.897 re3data repository contacts

Field phase:

2021-05-18 - 2021-06-14

Response rate:

332 complete responses (17,5 % response rate)

Data cleaning:

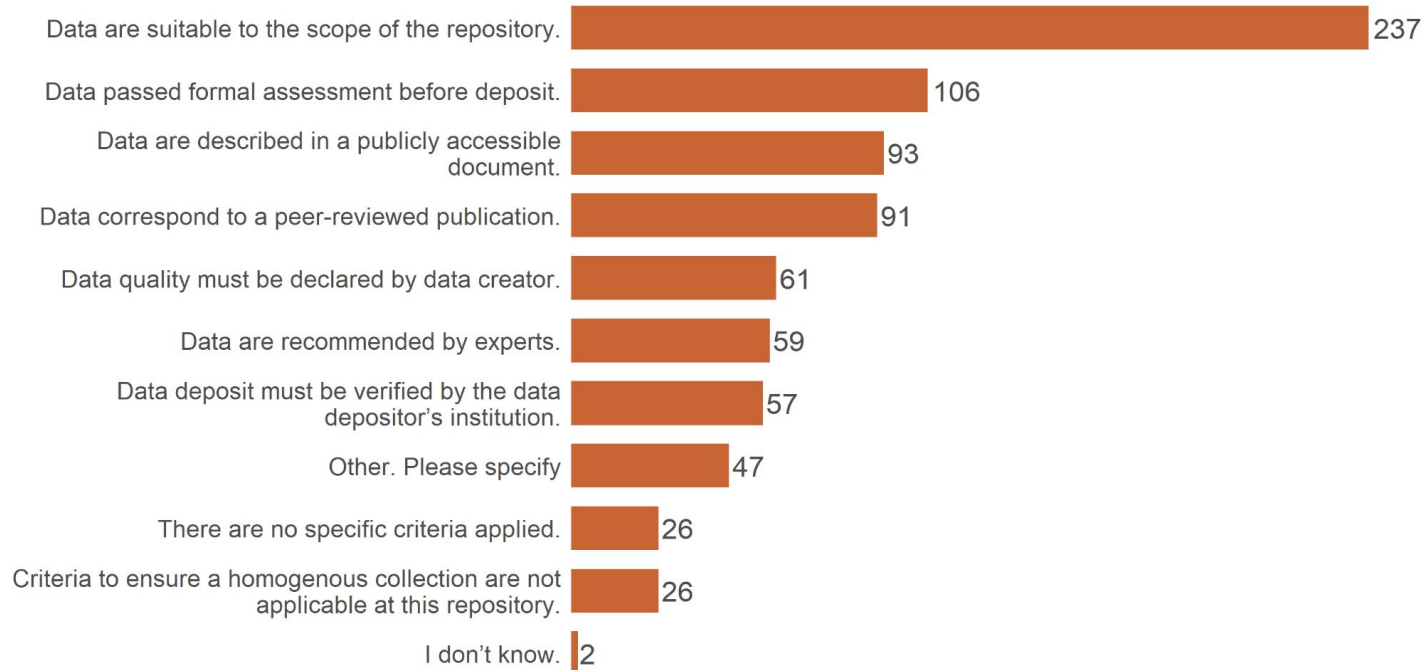
Free-text answers were recoded where necessary

Data availability:

Kindling, Maxi, Strecker, Dorothea, & Wang, Yi. (2022). *Data quality assurance at research data repositories: Survey data* (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.6457849>

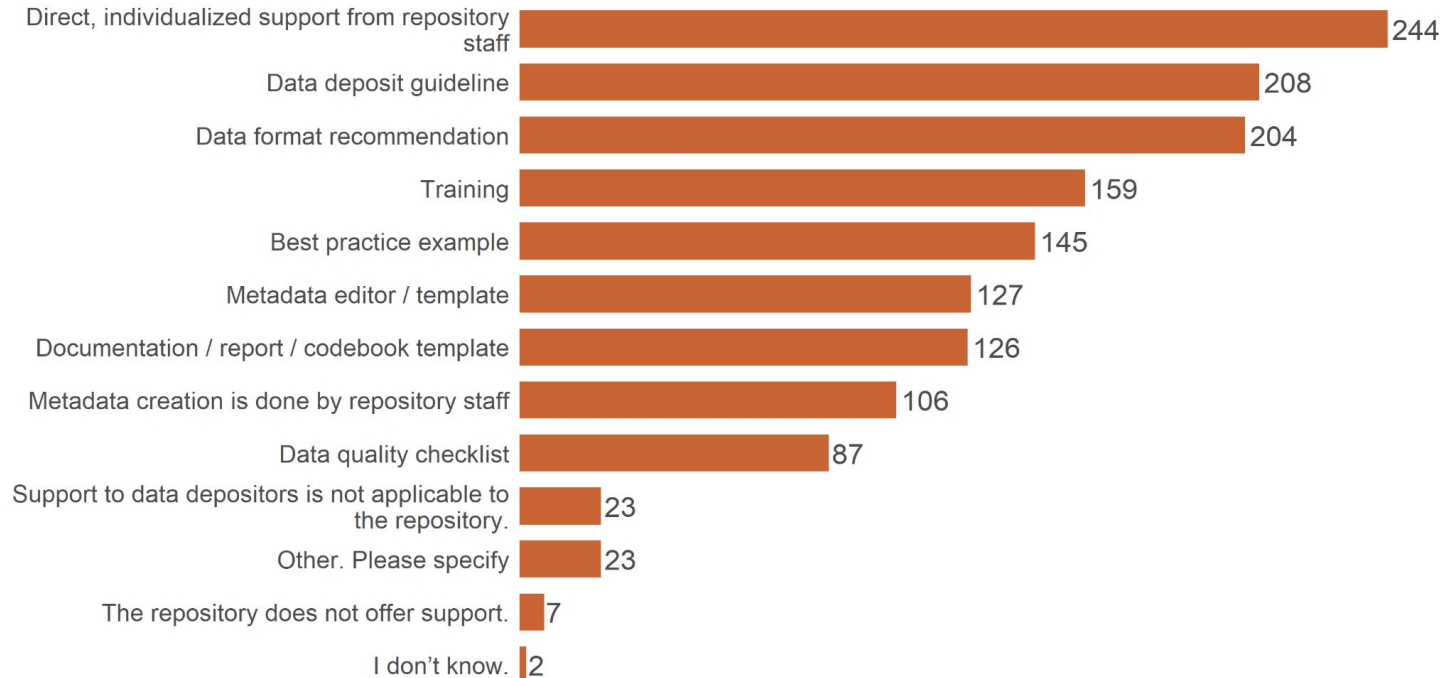
Quality definition

Which of the following criteria are generally applied to the data to ensure a homogenous collection?



Quality development

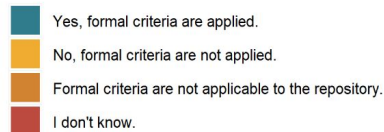
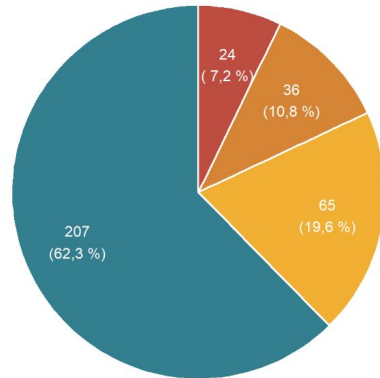
What types of support are offered to data depositors?



Formal assessment and review of data

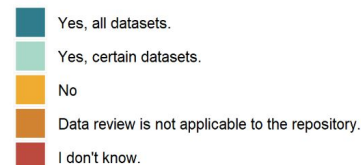
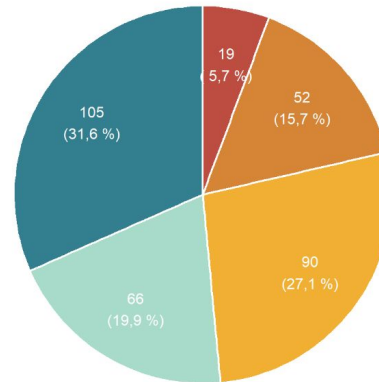
Are formal criteria applied to data before publication?

A



Are data reviewed beyond the application of formal criteria?

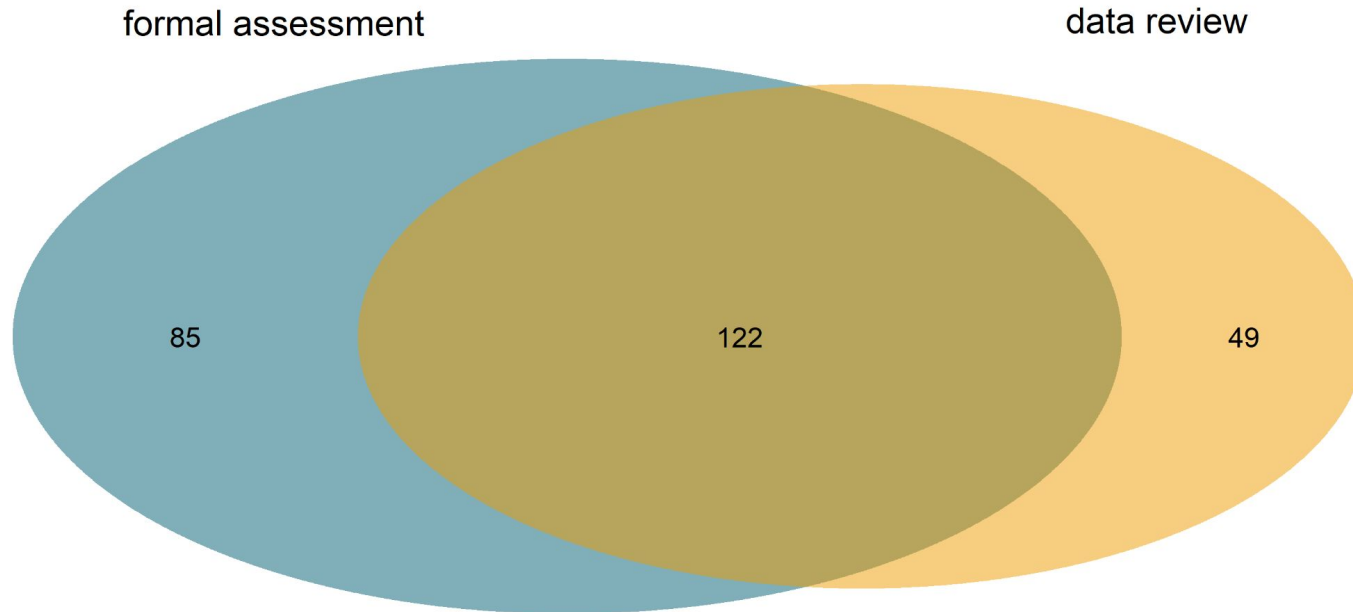
B



N = 332, n = 332

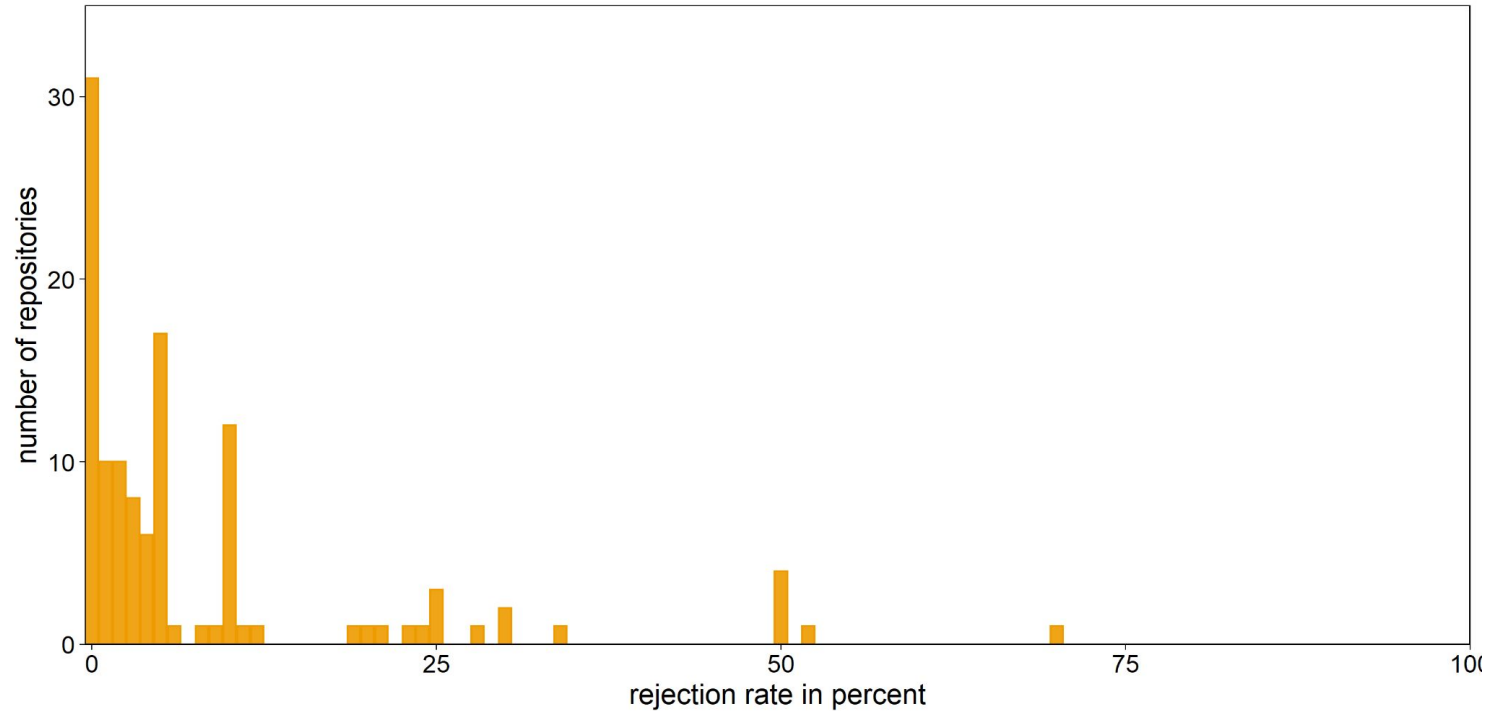
Formal assessment and review of data

22.9 % (n = 76) of the responding repositories perform neither formal assessment nor review of data.



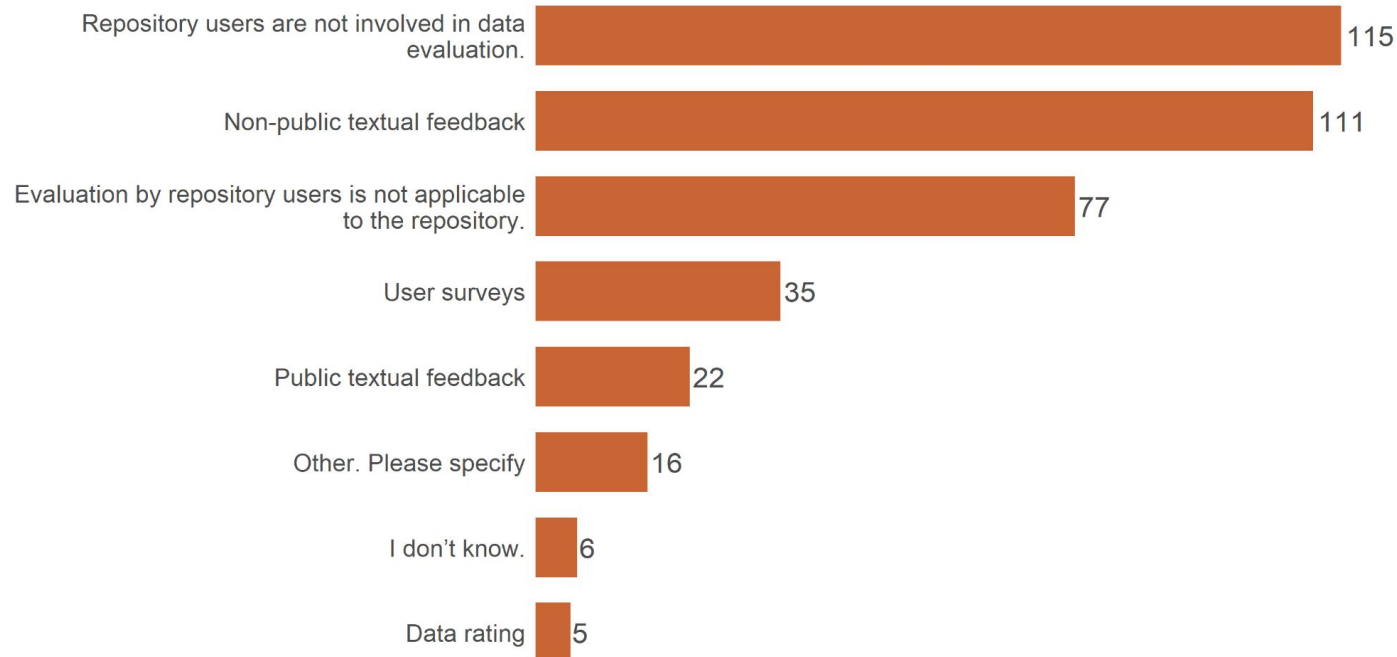
Data rejection

What (estimated) ratio of datasets were rejected by your repository in the last two years?



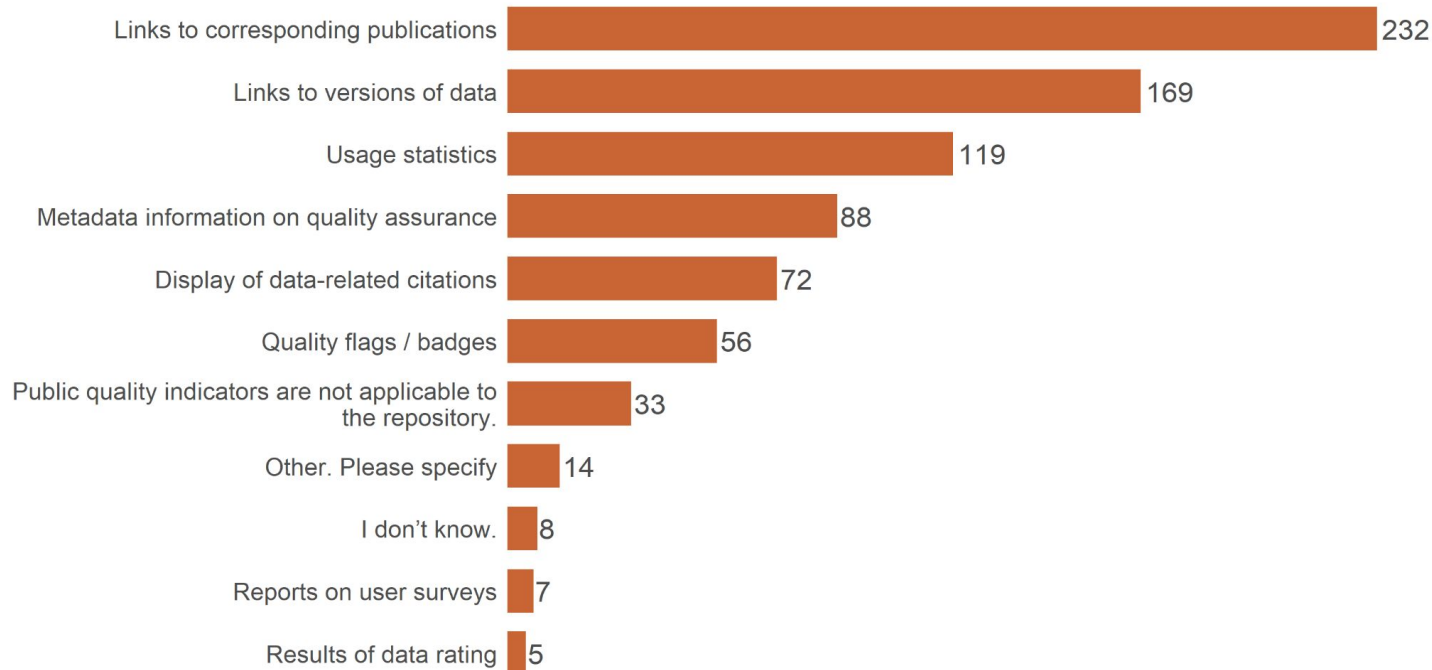
Quality evaluation by repository users

How are repository users involved in data evaluation after data is published?



Quality information

What types of quality indicators are made public to data users?



Conclusion

- Research data repositories significantly contribute to data quality assurance.
- Approaches to ensuring data quality are diverse, and research data repositories perform a variety of measures.
- Contributions to ensuring data quality need to be acknowledged and the visibility of quality assurance measures increased.

Outlook

re3data COREF & CoreTrustSeal workshop:

“Data quality management at repositories” (Online format)

Save the date: 5 October 2022

Survey data:

Kindling, Maxi, Strecker, Dorothea, & Wang, Yi. (2022). *Data quality assurance at research data repositories: Survey data* (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.6457849>

Research article:

Data Quality Assurance at Research Data Repositories
(submitted to the Data Science Journal)

Thank you for your time and attention

 www.re3data.org

 [@re3data](https://twitter.com/re3data)

 doi.org/10.1045/march2017-kindling

