

Quality Assurance for Spatial Research Data

A summary - Original by: M. Wagner and C. Henzen - Published in: ISPRS International Journal of Geo-Information, vol. 11, no. 6, p. 334, 2022 [1]

1st Niklas Lübcke

Faculty of Information, Media & Design
SRH Heidelberg
Heidelberg, Germany
niklas.lubcke@gmail.com

2nd Michael Wagner

Center for Information Services and High Performance Computing (ZIH)
Technische Universität Dresden
Dresden, Germany
michael.wagner@tu-dresden.de

3rd Christin Henzen

Working Group Geo UX, German UPA
Technische Universität Dresden
Dresden, Germany
christin.henzen@tu-dresden.de

Abstract—The authors of [2] propose an Assertion Framework (II-A) and a Quality Assurance Workflow (II-B) for spatial data sources in Earth System Science (ESS). The framework is comprised of an evaluation of the 5-star rating system for open data by Tim Berners-Lee, a revised maturity matrix including FAIR (Findable, Accessible, Interoperable and Reusable) criteria and a spatial data quality matrix where maturity levels are related to quality metrics. The metrics of openness, maturity and quality are then mapped to phases within the research data lifecycle to produce the proposed QA workflow.

Index Terms—quality assurance; data maturity; maturity matrix; spatial data quality; FAIR

I. INTRODUCTION

M. Wagner and C. Henzen recognised the level of quality of spatial data as an indication of the relevance of the extracted results of scientific work in the field. They suggest that there is a vacuum for tools to assess the quality of spatial data sources. In addition, despite the high quality of scientific data creation and manipulation workflows, the authors identified the monitoring and reporting of adequate information as a challenge in research data management (RDM). This leads to information being output only at the end or after the completion of the data lifecycle.

II. RESULTS

The primary results of [2] are the following, with the mapping of openness metrics to maturity and quality metrics being summarised in the Assertion Framework (II-A)

- 1) Assertion Framework (II-A)
- 2) Quality Assurance Workflow (II-B)
- 3) Proof of Concept (II-C)

A. Assertion Framework

The aim of the Assertion Framework is to provide a tool for assessing the quality of spatial data sources according to the following categories:

- 1) Openness
- 2) Data maturity
- 3) Data quality

The authors of [2] do not calculate an average of all the metrics or categories included. Instead, the output of the framework is 3 independent ratings between 1 and 5 stars. The ratings are created by following a ladder principle, where a higher step adds another aspect to the metrics of the data source.

It is possible for categories to have limited rating ranges. As shown in [2], e.g. the openness category can only produce a rating between 2 and 4 stars.

B. Quality Assurance Workflow

To create the best possible data sources as part of the scientific data lifecycle [2], the metrics of II-A are mapped to each phase of the lifecycle and its appropriate roles to create a workflow that acts as a QA for scientific teams.

Furthermore, the workflow was implemented as a plugin in the interactive questionnaire tool RDMO (Research Data Management Organiser), a Python software for self-hosting on the World Wide Web.

C. Proof Of Concept

To apply their proposed framework, the authors rate and discuss SPAM2010 (A land use dataset) 3.5 stars in the openness category, 4 to 5 stars in the maturity category and 5 stars in the quality category.

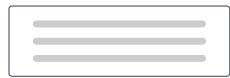
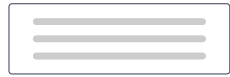
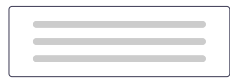
REFERENCES

- [1] S. SCImago (n.d.), “Isprs international journal of geo-information,” 2023. [Online]. Available: <https://www.scimagojr.com/journalsearch.php?q=21100427639&tip=sid&clean=0>
- [2] M. Wagner and C. Henzen, “Quality assurance for spatial research data,” *ISPRS International Journal of Geo-Information*, vol. 11, no. 6, p. 334, 2022.

APPENDIX

- [3] Figure of the proposed Assertion Framework and the QA Workflow

Quality Assurance for Spatial Research Data

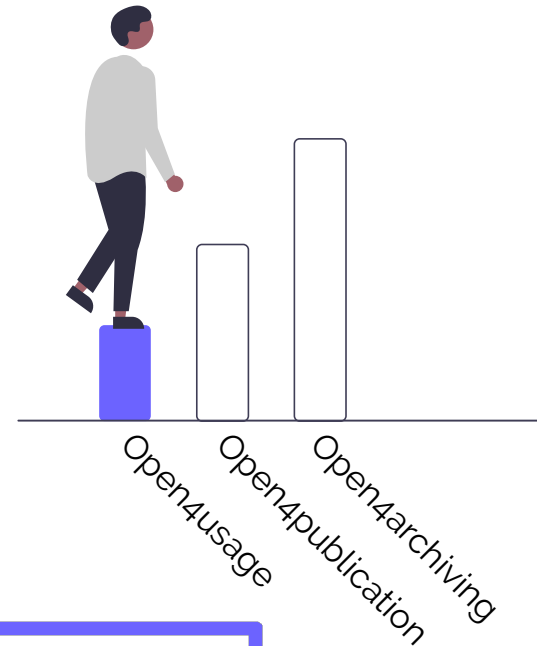


The Authors Michael Wagner and Christin Henzen propose a framework to help researchers rate **spatial data** sources in **three categories** from 1 to 5 stars

Categories:

1. Openness
2. Maturity
3. Quality

- Ratings are not averages
- Ratings follow a ladder scheme
- A higher step adds another aspect to the metric mix
- A higher rating automatically includes the aspects from the lower ratings
- There can also just be ratings from 2 to 4 like in the **openness category**



Research Data Lifecycle Mapping

The metrics from the categories are mapped to appropriate phases in the research lifecycle and fall into the responsibility of one of the three roles shown in the process figure. This mapping serves as a QA process to help scientists create amazing sources for spatial data.



As a proof of concept the authors rated the SPAM2010 dataset a **3.5** in openness, a **4 to 5** in maturity and a **5** in quality