

## Using metadata quality to filter datasets in data portals

**Susanne Feistel**, Leibniz Institute for Baltic Sea Research Warnemünde (Germany), [susanne.feistel@io-warnemuende.de](mailto:susanne.feistel@io-warnemuende.de)

**Ulrike Kleeberg**, Helmholtz-Zentrum Geesthacht (Germany), [ulrike.kleeberg@hzg.de](mailto:ulrike.kleeberg@hzg.de)

**Jörn Kohlus**, Nationalpark Wattenmeer, (Germany), [joern.kohlus@lkn.landsh.de](mailto:joern.kohlus@lkn.landsh.de)

**Romina Ihde**, Federal Waterways Engineering and Research Institute (Germany), [romina.ihde@baw.de](mailto:romina.ihde@baw.de)

**Carsten Schirnick**, GEOMAR Helmholtz Centre for Ocean Research Kiel (Germany), [cschirnick@geomar.de](mailto:cschirnick@geomar.de)

**Stefanie Schumacher**, Alfred Wegener Institute (Germany), [stefanie.schumacher@awi.de](mailto:stefanie.schumacher@awi.de)

**Susanne Tamm**, Federal Maritime and Hydrographic Agency (Germany), [susanne.tamm@bsh.de](mailto:susanne.tamm@bsh.de)

Sharing geoinformation and data across communities is becoming more important, which is enabled by web services. The Open Data initiative (e.g. data.gov) promoted by public agencies and research institutions enhances this development. However, poor data quality with insufficient quality information may be hindering the acceptance and re-usage of the data by the scientific community in the future. The mandatory metadata within ISO 191\*\* is not sufficient for a comprehensive quality assessment or long-term usability of any dataset. Documentation and quality information is mostly optional and not easily found or filtered. It might be stored within the metadata as information in LI\_Lineage, DQ\_DataQuality, or even within the dataset itself, possibly coded. Ultimately, in most cases, quality information is not directly accessible for interested users.

On the other side, for the data creator, the documentation of quality information can get time-consuming and overwhelming, especially if required all at once in the reporting phase of a project. Furthermore, quality information gets significantly more complex if more than one dataset is considered as in data products such as maps or modeling results.

We present a more practical approach to get a sufficient and standardized quality assessment (quality flag), dynamically generated from the entire quality information of the metadata XML file. While the ISO standard offers a complex array of optional fields to provide quality information, we recommend a manageable number of fields to be filled (Fig. 1). As prepared in (Feistel et al. 2020, in prep.) there is specific information a user needs first and foremost to evaluate the usefulness of a dataset.



Figure 1: Create accessible quality information and record it in recommended fields of the metadata standards of ISO 191\*\*.

Our task group provides a set of tools to assess quality information recorded in the recommended fields of the ISO standard. The first is a web-based form to manually assess a single dataset, the

second is a program for syntactically analyzing multiple dynamically generated XML in ISO 19139. To make the quality assessment compatible and comparable between platforms and quality flag schemes, as well as machine-friendly, we propose a coded summary string in the pattern of the scheme plus a flag, e.g. “SDN::1” (equal to quality flag “1” within the quality flag scheme of “SeaDataNet”) to be put in the ISO field “DQ\_StandaloneQualityInformation” (Fig. 2).

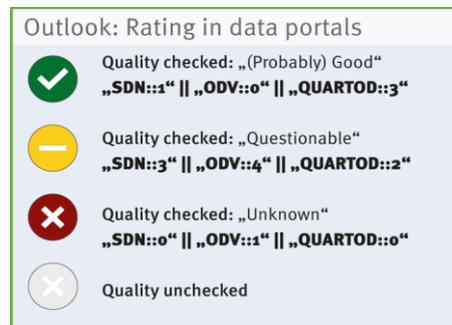


Figure 2: Proposed quality ratings through automatic assessment of metadata quality via the ISO XML checker.

Our toolbox is aimed at data creators to help provide meaningful metadata and at data portals to search for sufficiently documented data.

For data creators, metadata templates can be prepared in advance. This is especially useful for repeating data collection or creation procedures within projects, as well as long-term data series. Known instruments and methods can be inserted in the recommended fields and require little further effort. Thus, the compilation of metadata can be simplified for individual scientists.

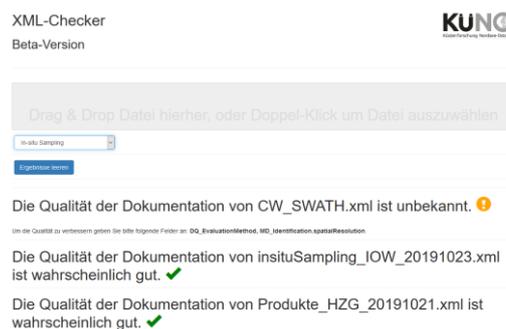


Figure 3: Drag & drop a XML-file in the checker to create a standardized metadata quality assessment.

In data portals the primary advantage of using an automatic assessment of metadata quality via the ISO XML checker is in getting a standardized quality flag based on a common quality flag scheme (Fig. 3). A time consuming individual control is not necessary. The quality control mechanism supports users in finding data relevant for their own work more efficiently. Last but not least it allows for a quality rating and filtering of any search results (Fig. 2).