Measuring metadata quality.

Péter Király
Göttingen eResearch Alliance and GWDG (Göttingen, Germany)
peter.kiraly@gwdg.de

Why to measure metadata quality?

The quality of metadata record is an issue which generated many discussions in different contexts, however it is hard to define and hard to measure. Several papers proposed methods to measure structural elements which could be taken as approximations to data quality. In a research project we aimed to create an open source framework which refines and implements these metrics, configurable to work with different metadata schemas. Knowing the issues is the first step in fixing them. Since quality assessment depends on the purpose of data, process should start with functional requirement analysis.

Problem patterns

Consequence of metadata issues

Quality dimensions

The benefits of metadata quality measurement:

- for researchers, data librarians, data stewards: check and improve metadata, propagate good use cases and demonstrate anti-patterns
- for data repository developers, data journal editors: improve data services

Reporting

Completeness is based on the existence, cardinality and importance of fields defined in the metadata schema. Fields support one or more different functions, such as searchability, descriptiveness, identification, browsing, or re-usability. We call the completeness sub-dimensions, and they measure how well the functions are supported.

Multilinguality

Text w/o language annotation ‘Germany’

Text w/ language annotation ‘Germany’@en, ‘Deutschland’@de

Link to multilingual vocabulary <http://www.geonames.org/2921044/>