e-infrastructures austria · Deliverable

> *„What makes it possible for the web to be composed of small pieces loosely joined? Metadata“*
> Jeffrey Pomerantz

Metatada Quality¶

e-infrastructures austria · Deliverable

Metatada Quality¶

Work-Package-Cluster:	**Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht**
Leitung des Clusters:	Susanne Blumesberger Universitätsbibliothek Wien susanne.blumesberger@univie.ac.at
Datum:	12.05.2016
Version	1.1
Autor:	José Luis Preza Zentraler Informatikdienst, Universität Wien jose.luis.preza@univie.ac.at

e-infrastructures austria · Deliverable

Kurzbeschreibung (Deutsch):	Metadaten Qualität ist entscheidend. Daten müssen gefunden werden. Metadaten erleichtern die Zugang zu Daten und Ressourcen. Je besser und genauer ein digitales Objekt beschrieben wird, desto leichter wird es sein, es zu finden.
Description (English):	Metadata quality is critical. Data needs to be found. Metadata is good for providing access to data and resources. The better and more accurate a digital object is described, the easier it will be to find it.
Schlagwörter (Deutsch):	Metadata Quality – Österreich - Austria
Keywords (English):	Metadata – Quality –Austria

office@e-infrastructures.at
www.e-infrastructures.at

Table of Contents

office@e-infrastructures.at
www.e-infrastructures.at

e-Infrastructures Austria

e-Infrastructures Austria is a project for the coordinated establishment and development of Repository infrastructures for digital resources in research and science throughout Austria to securely archive and publish digital publications, multimedia objects and other digital data resulting from research and education.

Key facts

Begin: 1. January 2014
Ends: 31. December 2016
Project coordination: University of Vienna
Partners: 25 Partner institutions
Funding agency: Federal Ministry of Science, Research and Economics BMWFW

Website: http://www.e-infrastructures.at/

Coordination Office: Library of the University of Vienna
E-Mail: office@e-infrastructures.at
Tel: +43 1 4277 DW 15172

Executive Summary Deutsch

Metadaten Qualität ist entscheidend. Daten müssen gefunden werden. Metadaten erleichtern die Zugang zu Daten und Ressourcen. Je besser und genauer ein digitales Objekt beschrieben wird, desto leichter wird es sein, es zu finden.

Metatada quality

Metadata quality is critical. Data needs to be found. Metadata is good for providing access to data and resources. The better and more accurate a digital object is described, the easier it will be to find it.

Defining and Ensuring Metadata Quality

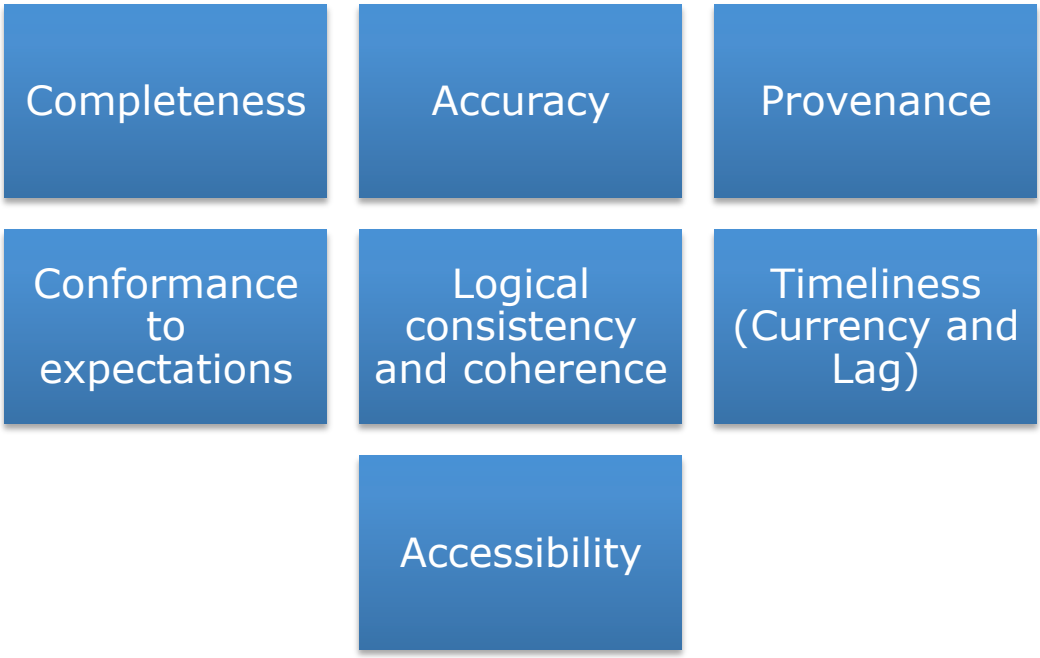
Defining "quality" can be very challenging.

There are techniques for evaluating and enforcing consistency and predictability in metadata that will assist with improving metadata quality.

One of them is to clearly set the Criteria for measuring the quality of the metadata.

Criteria for Quality Measurement of Metadata

Criteria to evaluate and measure metadata quality includes:



Completeness

Metadata should describe the target objects as completely as economically possible

Accuracy

Information should be correct

- Eliminate typos
- Ensure conforming name expressions
- Ensure standard abbreviations

Provenance

Who prepared the metadata?
What do we know about the preparer?
What methods were used to create the metadata?
Is it human created or created by machine?
What transformations have been applied since creation?
Where has it been before?

Conformance to expectations

Does the metadata contain elements a community would expect to find?

Controlled vocabularies are well-chosen and explicitly exposed to downstream users
Metadata is reflective of community thinking about necessary compromises

Logical consistency and coherence

Similar structures and appearance are enabled for search results.

Timeliness (Currency and Lag)

Currency: metadata should be up to date

No Lag: metadata should be updated whenever the object changes. It is an iterative process

Accessibility

Barriers to accessibility may be economic, technical or organizational:

- Metadata as “premium” or proprietary information
- Unreadable for technical reasons (file formats, etc.)
- Metadata may not be properly linked to relevant object(s)

Best practices for Data Quality

Best practices can be applied before, during and after metadata has been collected.

Before data collection

- *Design an efficient storage system for the data.*
- *Establish metadata standards*
- *Assign responsibilities: Quality*

During data entry/collection

Minimize the number of times the data need to be entered. Instead, use reference mechanisms such as a relational database.

Use consistent terminology.

Reduce data to one piece of information per cell.

Always document any modifications to the dataset. This avoids duplicate error checking.

After data collection

Look for missing or irregular data entries.

Evaluating metadata

There are several strategies to evaluate metadata quality, including:

Metadata analysis

Find inconsistencies, missing values

Validate schema

These tasks can be done with software tools or even Excel

Metadata Analysis: Missing Elements (Scatter Plot)

This is an example of how a set of metadata elements can be visually analyzed and evaluated for missing elements.

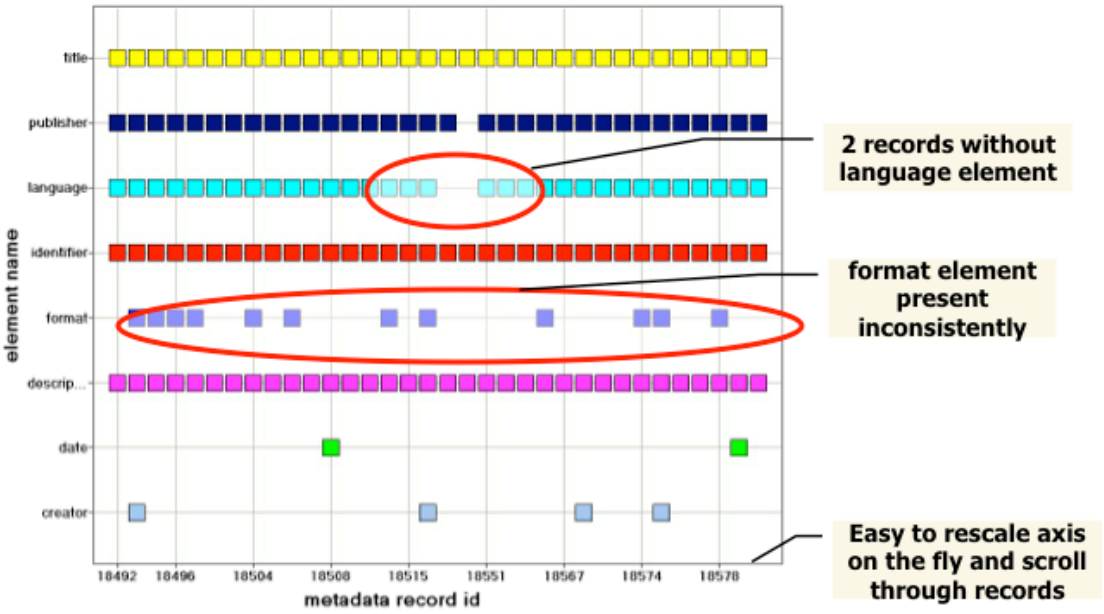


Figure 1 - Metadata Analysis: Missing Elements (Scatter Plot)

Metadata Analysis: Element Names vs. Record Ids (Scatter Plot)

This is an example of how a set of metadata elements can be visually analyzed and evaluated for missing elements against record ids.

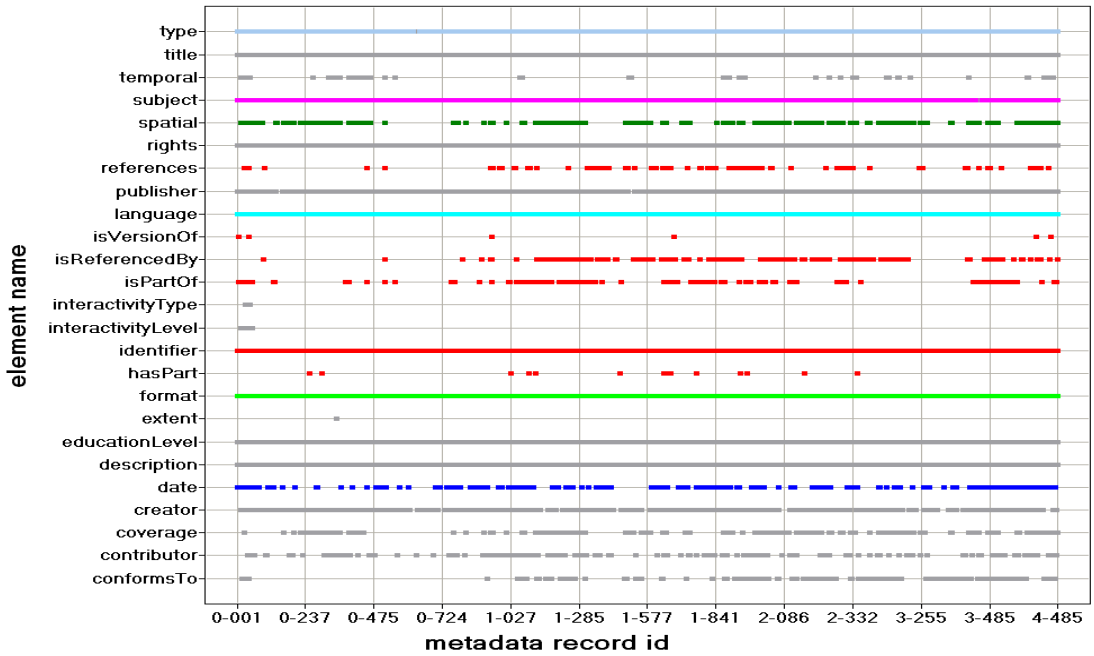


Figure 2 - Metadata Analysis: Element Names vs. Record Ids (Scatter Plot)

Recommendations for improving Metadata Quality

Documentation: metadata should be well documented

Use of standards, best practice guidelines (DC, METS, etc.)

Exposure and maintenance of local and community vocabularies

Use of Application Profiles: set of metadata elements, policies, and guidelines defined for a particular application

Provide Training materials, tools, methodologies

Refereces

Nichols, Paynter, Chan, Bainbridge, McKay, Twidale
and Ann Blandford 2008. METADATA TOOLS FOR INSTITUTIONAL REPOSITORIES. The U

Rettig, Liu, Hunter, Level. 2008. Developing a Metadata Best Practices Model: The Experience of the Colorado State University Libraries. Journal of Library Metadata, Vol. 8(4) 2008.

Table of Figures

e-infrastructures austria

University of Vienna Library and Archive Services, 1010 Vienna, Universitätsring 1
T: +43-1-4277-15172, office@e-infrastructures.at