Using Metadata Standards in a Public Authority

Gerhard NAVRATIL – TU Wien
Georg RAMLER – Office of the Federal Government of Lower Austria
gerhard.navratil@geo.tuwien.ac.at; georg.ramler@gmail.com

Standard in Action – Vienna, May 11 2022

Introduction

- Public authorities must produce metadata descriptions because
  - Public authorities produce data sets
  - Data produced by public authorities should be used by the public
  - Metadata describe data sets
- Data sets produced by domain experts, providing the metadata is quite complex (e.g., quality, fitness for use) – metadata expert?
- Data producers may be faced with multiple standards – how do we avoid duplication of work?

Characteristics of ISO19115:2014

- Needed for internal documentation and data exchange
- Describes one dataset, service or series
- Very few (~5) mandatory attributes
- Additional (~15) attributes recommended for discovery
- Huge amount of optional attributes available
  - -> very versatile
  - -> hard to understand for metadata producers
- Encoded in XML

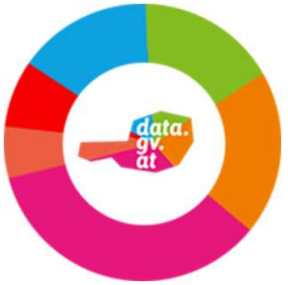
Characteristics of INSPIRE

- Needed to fulfill DIRECTIVE 2007/2/EG (INSPIRE)
- Derived from ISO19115:2003 and ISO19139:2007
- Describes one dataset, service or series
- Lots of (~20) mandatory attributes
- Very restrictive (contents of attributes, order of keywords, …)

  -> hard to understand for metadata producers
- Additional (~5) attributes recommended for interoperability
- Encoded in XML

Characteristics of DCAT-AP/GeoDCAT-AP

- Needed to fulfill DIRECTIVE (EU) 2019/1024 (PSI)
- Describes data catalogs with multiple datasets and services
- Main goal is discovery
- GeoDCAT-AP extends DCAT-AP to map INSPIRE (lossy!)
- Minimal (4) mandatory attributes
- Minimal (4) recommended attributes for discovery
- Encoded in RDF

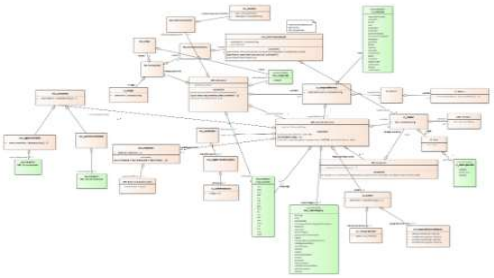


Metadata generation via intermediate model

data producer **metadata form** **relational intermediate model**

fills

maps to



automatic extraction of meta information

dataset or service

ISO19139 XML

INSPIRE XML

generates

DCAT-AP / GeoDCAT-AP RDF

conform and valid metadata

Benefits

- Reduced workload for data producers
- Metadata knowledge encapsulated in mapping to intermediate model
- Manual input reduced through automatic extraction
- Metadata information is stored only once
- Form input support with prefilled information
- Intermediate model can hold more information than published

Challenges

- Intermediate model has to be very strict and is not versatile
- Mapping of existing metadata results in potential information loss
- Standards are in certain aspects not compatible
- Data producers need to fill minimum information for all standards
- Intermediate model is manual work and mappings are hard coded

Conclusions and Future Work

- Harmonized standards avoid costs
- Completely different approaches make harmonization more complicated (e.g., verbal description vs. parameters)

- If we cannot avoid the differences, how can we overcome them?
  - Natural language processing to determine parameters from verbal descriptions?
  - Semantic translation of categories using machine learning?
  - Semi-automatic extraction of an intermediate model?