Metadata of spatial data
Brno, June 3rd 2010

Dipl.- Umweltwiss. Thorsten Bockmühl
University of the Bundeswehr Munich
thorsten.bockmuehl@unibw.de
www.agis.unibw.de
Overview

- Introduction of AGIS and University of the Bundeswehr Munich
- Military standardization in DGIWG
- Metadata
  - General questions about metadata
  - Typical user requests
  - Difficulties between produced metadata and user requirements
- Data Product Specification ISO 19131
  - Background ISO and DGIWG Profile
  - Use for an automatic generation of metadata
UniBw München

Around 3500 Students

- Private university within the range of the ministry of defense
- Approved by the state of Bavaria, courses and degrees are equivalent to other universities
- Officers have to study!
- Most students live on campus
- Most students are soldiers for 13 years
- Staff is civil
- Trimester system!
AGIS: GI-Lab at UniBw M, Chair of Geoinformatics
- Around 10 scientists, mostly financed from 3. Party funds
MILITARY STANDARDIZATION IN DGIWG
DGIWG is the multi-national body responsible for geospatial standardization for the defence organizations of member nations.

Aim is to create required standards and procedures to enable the provision, exchange and use of standardized geospatial information.

18 Member + 5 Observer nations.
requirements

Nations

NATO

MGCP

DGI WG

ISO/ TC 211

Open Geospatial Consortium

Standards
METADATA
Metadata overview
# Metadata overview

**Resource title**: Image2000 Product 1 (uk7) Panchromatic

**Resource abstract**: IMAGE2000 product 1 individual orthorectified scenes. IMAGE2000 was produced from ETM+ Landsat 7 satellite data and provides a consistent European coverage of individual orthorectified scenes in national map projection systems. The year 2000 was targeted as reference year, but a deviation of maximum 1-year was allowed to obtain a full coverage of Europe, which involves approximately 450 Landsat TM Frames. Where Landsat 7 data were not available, Landsat 5 data have been used instead. The spatial resolution is 25 metres for multispectral and 12.5 metres for panchromatic imagery.

**Resource type**: dataset

**Resource locator**: Code space: null

**Unique resource identifier**: Code: image2000pr1uk7panc

**Topic category**: boundaries, economy

**Couple Resource**: Not applicable to dataset
What is Metadata?

• Definitions in standards of ISO TC211
  • data about data
  • information about data

• INSPIRE Directive
  • information describing spatial data sets and spatial data services and making it possible to discover, inventory and use them

• Extended definition
  • all **needed information** to describe spatial data sets and services **without a direct access** to them.
How is metadata generated?

1. Manual collecting in a Metadata Entry Tool
   - The common way to collect metadata
   - Great effort and additional work
   - The huge amount of metadata enable a wide description of a resource

2. Automatic generation by accessing spatial data
   - The quantity of metadata depends on the data format

3. Automated generation by monitoring the production process of spatial data
   - A controlled environment is needed
Which Metadata should be generated?

• The purpose of metadata is to inform a user
  → metadata is user oriented
• Nobody knows during the production of spatial data and metadata, which metadata users need
  → all metadata should be generated
• To decrease “all” in a practical quantity
  → all metadata of ISO standards should be generated

But the standardized metadata are only an approach to fulfill the user requirements
Typical user request

Background

• An ongoing process is the development of a NATO metadata profile
• The first draft version of the profile was created as a subset of ISO 19115
• Typical user request (freetext) were defined by NATO
• There was an analysis if the typical user request can be answer with the metadata elements of the first draft of NATO metadata profile
Additional information to answer a request

Typical user request:
I need datasets which describe main supply routes of ISAF RC East.

To answer this request with common metadata some additional information is needed.

Questions:
- Which features contain main supply routes (e.g. roads, bridges, tunnels)?
- Is an attribute named supply classification needed?
- Which area contains ISAF RC East?

Missing information:
- The semantic relation between "main supply routes" and the feature types or keywords must be defined
- The area of ISAF RC East must be identified
Transformation to answer a request

Typical user request:
I need all obstacles higher than 30 m in the surrounding of 10 km from Kandahar air field.

The request has to be transformed in a “metadata” request
e.g.: Give me all datasets with:
• Features = rocks, trees and buildings
• Attribute = height
• Attribute accuracy of height = 5 m
• Geographical extend of Kandahar air field + Buffer of 10 km
Typical user request (examples)

In what area is no data on rivers available? (different to the question in which area are no rivers)

I need a 3D urban model of Kabul with textures of all buildings.

• There is no element in ISO 19115 for textures of buildings
Typical user request (examples of analysis)

• If the system tells me that snipers cannot see me, can I trust that information?

• Is the route really the shortest route?

• I need to transport oversized equipment. Can I rely on the results of the routing analysis?

→ Open question: How can I describe the reliability of an analysis, because it depends not only on the spatial data?
Results

- Metadata should be user oriented but often collected in a production view
- There is a discrepancy between the production view and the user requirements
- Users are different in:
  - Knowledge (expert – non expert)
  - Domain (civilian – military)
  - Tasks (customer – producer)

It is an ongoing research to solve the problem to create user oriented metadata.
AN APPROACH FOR AUTOMATIC GENERATION OF METADATA BASED ON DATA PRODUCT SPECIFICATIONS

Thorsten Bockmühl, Stephan Mäs, Wolfgang Reinhardt

Paper is published on INSPIRE conference 2010 Krakau
Background

The creation of metadata content is mostly manual work, which is:

- boring,
- tedious and also
- error-prone.

It is vital to capture automatically as much metadata as possible.
Overview of data product specification (DPS)

Not so well known as Metadata standard ISO 19115

• Google Hits: “ISO 19131” (7,840) vs “ISO 19115” (298,000)

Data product specification (DPS) is defined in ISO 19131

Metadata about a product and the production process

A product is a dataset or dataset series

Refers to the elements of the ISO metadata standard (ISO 19115)
Content of ISO 19131 data product specification

- identification
- content_and_structure
- referencesystem
- quality
- delivery
- maintenance
- data_capture
- portrayal
- additional_information
- additional_metadata

1..*

0..*

1

DPS

scope

Brno Workshop 2010
Additional Content in DGIWG Profile

1. **identification**
2. **content_and_structure**
3. **referencesystem**
4. **quality**
5. **delivery**
6. **maintenance**
7. **data_capture**
8. **portrayal**
9. **additional_information**
10. **additional_metadata**
11. **overview**
12. **terms_abbreviation**

DPS

Brno Workshop 2010
Application of DPS

Assumption for common production processes of spatial data

- Data content is defined and described (e.g. in a Feature Catalog or Application Schema).
- The information is stored according to ISO 19131 DPS

A DPS is applicable for:

- Direct mapping of DPS contents to the metadata of the produced data
- User guidance and workflow support during the production process
- Support of the manual input of quality metadata
# Direct mapping of DPS

<table>
<thead>
<tr>
<th>Data Production Specification</th>
<th>Metadata/MD_Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>identification/title</td>
<td>citation/title</td>
</tr>
<tr>
<td>[1..1]: CharacterString</td>
<td>[1..1]: CharacterString + individual text</td>
</tr>
<tr>
<td>identification/abstract</td>
<td>abstract</td>
</tr>
<tr>
<td>[1..1]: CharacterString</td>
<td>[1..1]: CharacterString</td>
</tr>
<tr>
<td>identification/purpose</td>
<td>purpose</td>
</tr>
<tr>
<td>[1..1]: CharacterString</td>
<td>[1..1]: CharacterString</td>
</tr>
<tr>
<td>maintenance/maintenanceAndUpdateFrequency</td>
<td>resourceMaintenance/maintenanceAndUpdateFrequency</td>
</tr>
<tr>
<td>[1..1]: MD_MaintenanceFrequencyCode</td>
<td>[1..1]: MD_MaintenanceFrequencyCode</td>
</tr>
</tbody>
</table>
User guidance and workflow support during the production process

textual descriptions of the processing steps and data sources can be displayed during the production process to support the producer
Support of the manual input of quality metadata

DPS specifies that the dataset should have an absolute position accuracy of 5cm. How many percent of the data pass the level?

Only the required quality measure result has to be captured in the metadata.
Conclusion

• Metadata should be user oriented but often collected in a production view

• There is a discrepancy between the production view and the user requirements

• Different Approaches are needed for an automatic generation of metadata

• The ISO 19131 data product spezification can be used as a source for an automatic generation of metadata
Thank you for your attention