An Approach for Automatic Generation of Metadata based on Data Product Specifications

Thorsten Bockmühl[1], Stephan Mäs[1], Wolfgang Reinhardt[1]
[1] *AGIS - Arbeitsgemeinschaft GIS, University of the Bundeswehr Munich, Werner Heisenberg Weg 39, 85577 Neubiberg, Germany*
{Thorsten.Bockmuehl | Stephan.Maes | Wolfgang.Reinhardt @unibw.de}

Introduction

Metadata are an important aspect in Spatial Data Infrastructures (SDI) for discovery, evaluation, access and use of spatial data and services. Thus, the need of metadata is not put into question. The creation of metadata content is mostly manual work, which is often described as boring, tedious and also error-prone [Manso-Callejo, et al. 2009]. Therefore, in many cases only few metadata are captured in practice. But to be able to answer the questions of all potential users, it is vital to capture as much metadata as possible [Pötzsch, et al. 2007].

Some approaches to solve the problems of the manual metadata capturing process have been researched. These approaches can be distinguished regarding to the different sources that are used to generate the metadata content automatically.

For example, [Manso-Callejo, et al. 2004] provide a study, which discusses the possibility to acquire metadata directly from different file formats. Some Metadata can be automatically generated during the updating process of spatial data, as it is for example implemented in ESRI ArcGIS. Therefore the software documents the processing steps which are executed by the user [Kalantari, et al. 2009]. With these methods it is not possible to capture all metadata, which are specified in ISO 19115 – Metadata [ISO 2003].

In this paper we present a new approach to capture metadata automatically, on basis of data product specification (DPS) as defined in [ISO 2007]. This doesn't mean that we solve the whole problem of metadata generation but we introduce an additional source for the metadata capturing process, which has not been considered so far.

Content of ISO 19131 data product specification

While the ISO metadata standard 19115 is well known and often applied the ISO 19131 did not get much attention yet. The data content of the ISO 19131 is structured as displayed in Figure 1.

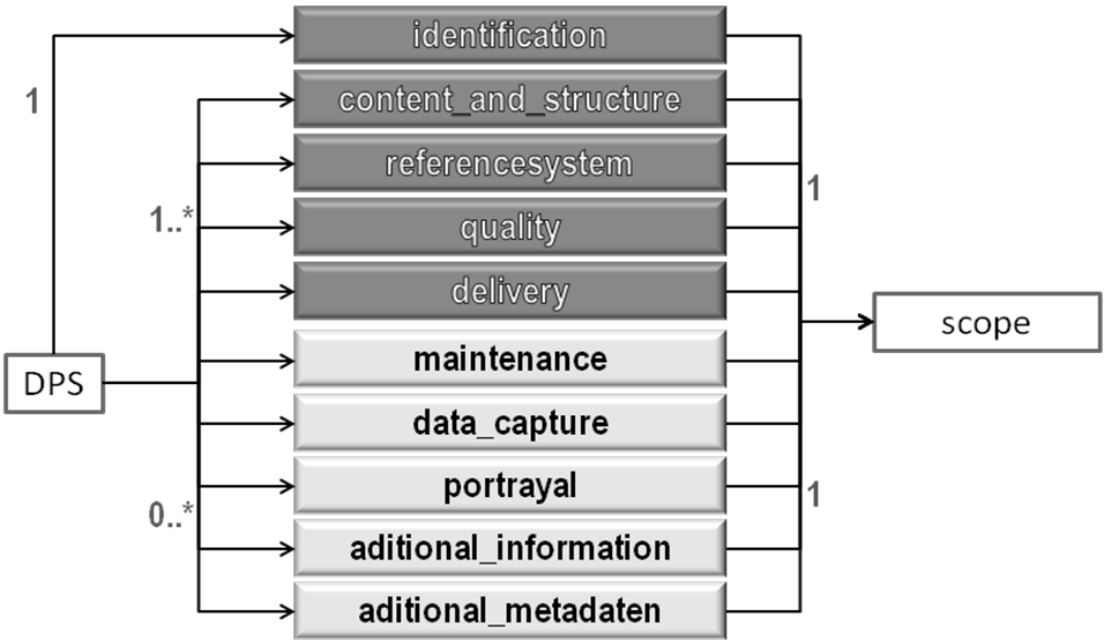

Figure 1: Simplified UML model for DPS

Each DPS has mandatory (dark gray) and optional (light gray) information content groups. They describe the product in general (identification), the methods of quality assurance (quality), the process steps during the spatial data capture and other aspects. All information content groups are linked to a scope which describes the geographical extent for which it is valid. Apart from the identification content group, each group can have more than one entry, such that data of different geographical regions (that is, scopes) can have different entries in the DPS.

For common production processes of spatial data the data content is defined and described (e.g. in a Feature Catalog or Application Schema) before the actual production process. If this information is stored according to ISO 19131 DPS, it can be used during the spatial data production process to support the production workflows in general and to automatically generate metadata of the produced data. Therefore three application possibilities of the DPS have to be distinguished:

- 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

For the use of the DPS it has to be available during the data production workflow. The producer of spatial data must reference to the DPS which describes the production of the data he or she is working on.

Direct mapping of DPS

Especially the information of the content groups “identification”, “content_and_structure”, “maintenance and delivery” can be directly mapped to the two metadata classes of ISO 19115 (Table 1 and Table 2). The ISO 19131 elements are based on ISO 19115 and in many cases the same classes and data types are used. Therefore a direct mapping is technically possible.

Data Production Specification	**Metadata/MD_Identification**
identification/title [1..1]: CharacterString	citation/title [1..1]: CharacterString + individual text
identification/abstract [1..1]: CharacterString	abstract [1..1]: CharacterString
identification/purpose [1..1]: CharacterString	purpose [1..1]: CharacterString
maintenance/ maintenaceAndUpdateFrequency [1..1]: MD_MaintenanceFrequencyCode	resourceMaintenance/ maintenanceAndUpdateFrequency [1..1]: MD_MaintenanceFrequencyCode
maintenance/updateScope [0..1]: MD_ScopeCode	resourceMaintenance/updateScope [0..*]: MD_ScopeCode
maintenance/dataSource [0..1]: CharacterString	resourceMaintenance/maintenanceNote [0..*]: CharacterString
identification/useLimitation [0..1]: CharacterString	MD_Constraints/useLimitation [0..*]: CharacterString
identification/classification [1..1]: MD_ClassificationCode	MD_Constraints/ MD_SecurityConstraints/classification [1..1]: MD_ClassificationCode
identification/ spatialRepresentationType [0..1]: MD_SpatialRepresentationTypeCode	MD_DataIdentification/ spatialRepresentationType [0..1]: MD_SpatialRepresentationTypeCode
identification/spatialResolution [1..*]: MD_SpatialResolution	MD_DataIdentification/spatialResolution [1..*]: MD_SpatialResolution
identification/language [1..*]: MD_CharacterString	MD_DataIdentification/language [1..*]: MD_CharacterString
deliveryFormat/characterSet [0..*]: MD_CharacterSetCode	MD_DataIdentification/characterSet [0..*]: MD_CharacterSetCode
identification/topicCategory [1..*]: MD_TopicCategoryCode	MD_DataIdentification/topicCategory [0..*]: MD_TopicCategoryCode

Table 1: Mapping DPS ISO 19131 to MD_Identification class of ISO 19115

Data Production Specification	**Metadata/MD_ContentInformation**
content/featureCatalogue/ featureCatalogueCitation [1..*]: CI_Citation	MD_FeatureCatalogueDescription/ featureCatalogueCitation [1..*]: CI_Citation
content/specification attributeDescription [1..1]: RecordType	MD_CoverageDescription/ attributeDescription [1..1]: RecordType
content/specification/contentType [1..1]: MD_CoverageContentType	MD_CoverageDescription/contentType [1..1]: MD_CoverageContentType

Table 2: Mapping DPS ISO 19131 to MD_ContentInformation class of ISO 19115

User guidance and workflow support

There are two content groups of DPS which contains information that can be used to support the production process. The content group “data_capture” and “maintenance” contains textual descriptions of the processing steps and data sources, which should be used. This information can be displayed during the production process to support the producer.
At the moment, in DPS the processing steps and data sources for data capturing and maintenance are described in free text elements. For an increased benefit these information must be described formally. Such formalizations of process descriptions is currently also under discussion for the specification of Web Processing Services [Brauner, et al. 2009] and subject of future research.

Support of the manual input of quality metadata

The quality content group of a DPS contains information about the prescribed quality assurance methods, the feature of data which has to be checked and the threshold for a successful quality assurance. Usually producers of spatial data only execute quality checks which are required. And only those have to be documented in the metadata.
Therefore, the information of DPS can be integrated into the metadata entry tool, such that the producer of the metadata receives an adjusted input mask.
For example, a DPS could specify that in the dataset an absolute position accuracy of 5cm must be reached. If this is the only quality requirement, only corresponding quality measure result has to be captured in the metadata.

Practical tests

Currently there is no entry tool for DPS available on the market. Therefore we created an xml schema of the ISO 19131 and developed a standalone entry tool. We used this tool to produce different DPS XML documents which conform to ISO 19131. In the next step an XSLT was developed to execute the direct mapping between the xml schemas of ISO 19115 and ISO 19131 (Table 1 and Table 2). The XSLT also builds the XML structure for the documentation of the quality assurance. The actual quality measure results have to be filled into this structure manually, as described in the paragraph above.
Furthermore an XSLT is developed to automatically create a PDF file, in which the content of “data_capture” and “maintenance” is clearly represented to support the production workflow.

Conclusion

For an automatic generation of metadata different sources can be used. We present a new approach to capture metadata on basis of DPSs. The ISO 19131 DPS is available since three years but it is not really established yet, although there are many application possibilities. The content of the DPS can be used to support the production process of spatial data in many aspects. In this paper we have demonstrated the use for automatic metadata generation.

References

Brauner, J., Foerster, T., Schaeffer, B. und Baranski, B.: (2009) Towards a Research Agenda for Geoprocessing Services. 12th AGILE International Conference on Geographic Information Science, Hannover, Germany, 2009.

International Organization for Standardization (ISO) (2003): 19115:2003, Geographic information -- Metadata.

International Organization for Standardization (ISO) (2007): 19131:2007, Geographic information -- Data product specifications.

Kalantari, M., Rajabifard, A. und Olfat, H.: (2009) Spatial metadata automation : a new approach. Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide, 2009.

Manso-Callejo, M. A., Nogueras-Iso, J., Zarazaga, J. und Bernabe-Poveda, M. A.: (2004) Automatic Metadata Extraction from Geographic Information. AGILE conference on Geographic Information Science, Heraklion, Greece, April 2004.

Manso-Callejo, M. A., Wachowicz, M. und Bernabé-Poveda, M. A.: (2009) Automatic Metadata Creation for Supporting Interoperability Levels of Spatial Data Infrastructures. GSDI 11 World Conference: "Spatial Data Infrastructure Convergence: Building SDI Bridges to Address Global Challenges", Rotterdam, The Netherlands, 15.-19. June 2009.

Pötzsch, M. und Reinhardt, W.: (2007) Metadata management – an analysis from creation till use. GI-Days 2007 - Young Researchers Forum, Münster, Germany,