ISO 19115 for GeoWeb services orchestration

Jan Růžička
Institute of Geoinformatics, VSB-TU of Ostrava
jan.ruzicka@vsb.cz

Keywords: ISO 19115, GeoWeb, Orchestration, BPEL, MIDAS, Dublin Core, INSPIRE

Klíčová slova: ISO 19115, GeoWeb, Orchestrace, BPEL, MIDAS, Dublin Core, INSPIRE

Abstract

The paper describes theoretical and practical possibilities of ISO 19115 standard in a process of generating dynamic GeoWeb services orchestras. There are several ways how to instantiate orchestras according to current state of services and user needs, some of them are briefly described in the paper. The most flexible way is based on metadata that describe geodata used by services. The most common standard used for geodata metadata in the EU is ISO 19115. The paper should describe if the standard is able (without extensions) to hold enough information for orchestration purposes. The paper defines minimal set of metadata items named "ISO 19115 Orchestration Minimal" that must be available for geodata evaluation in a process of orchestration. A second part of the article will be probably less optimistic. It should describe how are (or were, or are planned to be) ISO 19115 possibilities used for metadata creation nowadays in the Czech Republic. This part is based on analyses of ISO 19115 core, MIDAS system, Dublin Core and INSPIRE metadata IR.

Abstrakt

Příspěvek popisuje teoretické a praktické možnosti standardu ISO 19115 v procesu tvorby dynamických orchestrů služeb platformy GeoWeb. V zásadě je možné vytvářet instance orchestrů mnoha způsoby na základě aktuálního stavu služeb a požadavků uživatele. Některé z nich jsou stručně popsány v příspěvku. Nejpružnější způsob tvorby je založen na metadatech, které popisují geodata využívaná službami. V současné době je v rámci EU nejvyužívanějším standardem standard ISO 19115. Příspěvek by měl popsat zda je standard schopen (bez rozšíření) pojmout všechny nezbytné položky pro potřeby orchestrace. V příspěvku je definována minimální sada metadatových položek nazvaná "ISO 19115 Orchestration Minimal", která je nezbytná pro posouzení geodat v procesu orchestrace. Druhá část příspěvku bude zřejmě méně optimistická neboť se bude zabývat jak to vypadá s reálnými možnostmi využití

potenciálu standardu ISO 19115 pro orchestraci v rámci ČR. Tato část je založena na analýze ISO 19115 core, systému MIDAS, Dublin Core a INSPIRE metadata IR.

Orchestras

An orchestration is a process where are modelled processes (real or abstract) in a way of formalized description. A process modelling is a technique that uses several description tools, mainly schemas or diagrams, to describe usually real processes inside enterprise. The processes can lead across several organizations.

A model of a process is transformed from abstract languages (BPMN (Business Process Modelling Notation), UML (Unified Modelling Language)) to a form that can be directly run on a computer. In this area of runnable models of processes is the most known BPEL (Business Process Execution Language). A process run means reading inputs, invoking web services, deciding according to results, repeating some parts of the process and other necessary operations.

A process modelling offers possibilities how to formally describe processes inside an enterprise, to find duplicate processes, to find processes that are not optimised, etc. A process modelling helps with processes optimisation and with sources management optimisation. When it is possible, than the description is available in a form of BPEL-like language and processes can be directly invoked.

GeoWeb services orchestration can be done in many ways. The GA 205/07/0797 team has researched the two ways of possible orchestration.

Simple orchestras

The first way is based on orchestras where the services searched during the building orchestra instance are using the same data sources in a meaning of data sources and algorithms. During the building orchestra instance are searched only services that use the same data source and the same algorithms for data source and input manipulation. Data source content can change only on spatio-temporal extent of the working area. We can speak about services replication (or distribution in a horizontal plane). Current instances of the services that are connected to the orchestra are selected according to current state of the services, such as performance, speed or provider.

These services differ on physical binding. These kind of orchestras is focused on optimisation of orchestras run. For these kind of orchestras is not needed any specific manipulation. There is necessary to identify same services using some key. For our testing purposes we use common identification, based on standardisation organisation identification, standard identification, service identification. Such identification is described on the following example. http://gis.vsb.cz/ogc/wms/1.1.1/ZABAGED/0.1. Items are defined by url. First item is domain of the service type guarantee. Second item is abbreviation of standardisation organisation name. Third item is abbreviation of standard name. Fourth item is a version of the standard. Fifth item is abbreviation of the service. Last item is a version of the service type.

This type of orchestras is simpler to manage than the second one.

Dynamically created orchestras

The second way is based on orchestras where current instances of the services can be just similar to each other in a meaning of data sources and algorithms. For example we can use service that uses railways data source where tracks are just simple lines between stations or we can use service that uses railways data source where tracks are modelled by real headway. We can switch between these sources in many cases, such as routing (finding the best routes) where the main parameter for routing is time. This type of orchestras is more difficult to manage than the first one.

Our research shows that usually the first type of orchestras will be used, but there are still situations when a system for orchestration should be able prepare second type of orchestras. There are two ways how to handle this problem.

The first solution is simple, but difficult to manage in a meaning of long time term, because this solution is rather static than dynamic. There must be simple database (no matter how is organised – relational, XML) where are defined relations between data sources (services). Related services can be named group of similar services.

The second solution is based on data source evaluation based on metadata analyses. This article should describe, why is this way so complicated and probably impossible.

Metadata items useful for data evaluation

In a process of searching available services for dynamic orchestras building we are looking for similar data sources. First of all we have to specify metadata items that can be used for evaluating that the data are similar enough for our orchestra.

There are many different standards in this area that define metadata items, but nowadays probably the most important one is ISO 19115 (ISO 19139). For our research we identify only items from this standard.

We can name this set of items **ISO 19115 Orchestration Full**. Later is described Minimal set of the items that are necessary for running similarity tests.

Administrative metadata

Item	Description of usage and problems
MD_Metadata/ dateStamp	Date that the metadata was created. Useful for evaluation of metadata reliability.
MD_Metadata/ metadataMaitenance	Frequency and scope of metadata updates. Useful for evaluation of metadata reliability.
MD_Identification/ resourceMaitenance	Frequency and scope of data updates. Individual items are described later.

MD_MaintenanceInformation/ **maintenanceAndUpdateFrequency** **userDefinedMaintenanceFrequency** **updateScope** **updateScopeDescription**	Only supplemental information, but useful when information about temporal extent is not available
MD_ReferenceSystem	A reference system is not necessary for analyses, but for using the service. Usually we have enough information in EPSG code, that is included in metadata for a service, but sometimes full description is necessary.

Table 1: Administrative metadata items from ISO 19115 Orchestration Full

Quality metadata

Item	**Description of usage and problems**
MD_DataIdentification/ **spatialResolution** **MD_Resolution/** **equvivalentScale** **distance**	Density of spatial data. Very useful. We can use both options of the resolution, but the distance is better valuable.
MD_Metadata/ **dataQualityInfo**	Quality of a resource. Individual items are described later.
DQ_DataQuality	Very important item. Items (associations are described later).
LI_Lineage/ **statement** **processStep** **source**	Very useful items, but unfortunately only simple table of items and the free text domain is used. Very difficult to handle free text for automatic evaluation. Only items for defining source are not described only by free text, but this is not enough.
DQ_Element/ **nameOfMeasure** **measureIdentification** **measureDescription** **evaluationMethodType** **evaluationMethodDescription** **evaluationProcedure** **dateTime** **result**	This abstract element should be completely included. Of course the main item is result described later.

DQ_Result/DQ_ConformanceResult/ **specification** **explanation** **pass** **DQ_Result/DQ_QuantitativeResult/** **valueType** **valueUnit** **errorStatistic** **value**	This items are quite well defined and useful for evaluation. Even domains are good enough for automatic evaluation.
DQ_Completeness/ **DQ_CompletenessCommission** **DQ_CompletenessOmission**	Described by DQ_Element.
DQ_PositionalAccuracy/ **DQ_AbsoluteExternalPositionalAccuracy** **DQ_GriddedDataPositionalAccuracy** **DQ_RelativeInternalPositionalAccuracy**	Described by DQ_Element.
DQ_TemporalAccuracy/ **DQ_AccuracyOfATimeMeasurement** **DQ_TemporalConsistency** **DQ_TemporalValidity**	Described by DQ_Element.
DQ_ThematicAccuracy/ **DQ_ThematicClassificationCorrectness** **DQ_NonQuantitativeAttributeAccuracy** **DQ_QuantitativeAttributeAccuracy**	Described by DQ_Element.

Table 2: Quality metadata items from ISO 19115 Orchestration Full

Usage metadata

Item	Description of usage and problems
MD_Identification/ **resourceSpecificUsage**	Specific applications for which the resource was used.
MD_Usage/ **specificUsage** **userDeterminedLimitations**	Very useful item, but unfortunately only the free text domain is used. Very difficult to handle free text for automatic evaluation.
MD_Identification/ **resourceConstraints**	Constraints on a resource. Individual items are described later.
MD_Constraints/ **useLimitation**	Very useful item, but unfortunately only the free text domain is used. Very difficult to handle free text for automatic evaluation.

MD_LegalConstraints/ **accessConstraints** **useConstraints** **otherConstraints**	Very useful items, but unfortunately only simple table of items and the free text domain is used. Very difficult to handle free text for automatic evaluation. Information that there is copyright or license is not very useful for evaluation, if the resource can be used in orchestration.
MD_SecurityConstraints/ **classification** **userNote** **classificationSystem** **handlingDescription**	Useful only in some very specific applications. Only simple table of items and the free text domain is used. Very difficult to handle free text for automatic evaluation.

Table 3: Usage metadata items from ISO 19115 Orchestration Full

Extent metadata

Item	**Description of usage and problems**
MD_DataIdentification/ **extent** **EX_Extent/** **description** **geographicElement** **temporalElement** **verticalElement** **EX_GeographicExtent/** **extentTypeCode** **EX_BoundingPolygon/** **polygon** **EX_GeographicBoundingBox** **westBoundLongitude** **eastBoundLongitude** **southBoundLatitude** **northBoundLatitude** **EX_GeographicDescription/** **geographicIdentifier** **EX_TemporalExtent/** **extent** **EX_VerticalExtent/** **minimumValue** **maximumValue** **unitOfMeasure** **verticalDatum**	Spatio-temporal extent. For geographic extent is preferred polygon instead of bounding box.

Table 4: Extent metadata items from ISO 19115 Orchestration Full

Content and structure metadata

Item	Description of usage and problems
MD_DataIdentification/ spatialrepresentationType	Method used for spatial representation. List of available items is very simple. We can use it only for distinguish between raster and vector. The other items described later must be used for better evaluation.
MD_DataIdentification/ language	Language used within the dataset. Necessary for evaluation. We can use dataset with different language usually only when dealing only with geometry or topology.
MD_DataIdentification/ topicCategory	Main theme of the dataset. Not very useful, but can be used for basic evaluation.
MD_Keywords/ keyword Type ThesaurusName	More useful than topicCategory for basic evaluation.
MD_GridSpatialRepresentation/ numberOfDimensions axisDimensionsProperties cellGeometry MD_Dimension/ dimensionName dimensionSize resolution	More precise information about grid. We can include also MD_Georectified and MD_Georeferenceable, but these are not necessary for analyses.
MD_VectorSpatialRepresentation/ topologyLevel geometricObjects MD_GeometricObjects/ geometricObjectType geometricObjectCount	More precise information about vector. Number of object can be significant for analyses of similarity.
MD_FeatureCatalogueDescription/ featureTypes featureCatalogueCitation	Information about used feature catalogue and selected set of features from the catalogue.
MD_CoverageDescription/ attributeDescription contentType dimension	Information about values in grid data cells.

MD_ImageDescription/ **illuminationElevationAngle** **illuminationAzimuthAngle** **imagingCondition** **imageQualityCode** **cloudCoverPercentage** **processingLevelCode** **compressionGenerationQuantity** **triangulationIndicator** **MD_RangeDimension/** **sequenceIdentifier** **descriptor** **MD_Band/** **maxValue** **minValue** **units** **bitsPerValue** **peakResponse** **toneGradation** **scaleFactor** **offset**	Information about digital image record.

Table 5: Content and structure metadata items from ISO 19115 Orchestration Full

Minimal set of Metadata items for automatic data evaluation

Following list shows minimal set of metadata items, that must be available to test similarity of the analysed datasets. We can name this set as **ISO 19115 Orchestration Minimal**. Without these items are not metadata useful for running tests of similarity. This recommendation should be applied to all new created metadata. There are not included items, that are generally useful, but used domain for their specification is not suitable for automatic evaluation. Some of the items are not applicable for all resources (e.g. you can not specify MD_Band for vector data).

MD_DataIdentification/**spatialResolution**

MD_Resolution/equvivalentScale

MD_Resolution/distance

MD_Metadata/dataQualityInfo

DQ_DataQuality

LI_Lineage/**source**

DQ_CompletenessCommission/DQ_Element/DQ_Result

DQ_CompletenessOmission/DQ_Element/DQ_Result

DQ_AbsoluteExternalPositionalAccuracy/DQ_Element/DQ_Result

DQ_GriddedDataPositionalAccuracy/DQ_Element/DQ_Result

DQ_RelativeInternalPositionalAccuracy/DQ_Element/DQ_Result

DQ_AccuracyOfATimeMeasurement/DQ_Element/DQ_Result

DQ_TemporalConsistency/DQ_Element/DQ_Result

DQ_TemporalValidity/DQ_Element/DQ_Result

DQ_ThematicClassificationCorrectness/DQ_Element/DQ_Result

DQ_NonQuantitativeAttributeAccuracy/DQ_Element/DQ_Result

DQ_QuantitativeAttributeAccuracy/DQ_Element/DQ_Result

MD_DataIdentification/extent

EX_Extent/**geographicElement**/EX_BoundingPolygon/**polygon**

EX_Extent/**geographicElement/EX_GeographicBoundingBox**

EX_Extent/**temporalElement**/EX_TemporalExtent/**extent**

EX_Extent/**verticalElement/EX_VerticalExtent**

MD_DataIdentification/spatialrepresentationType

MD_DataIdentification/language

MD_DataIdentification/topicCategory

MD_Keywords

MD_Keywords/**keyword**

MD_Keywords/**Type**

MD_Keywords/**ThesaurusName**

MD_GridSpatialRepresentation

MD_GridSpatialRepresentation/**numberOfDimensions**

MD_GridSpatialRepresentation/**axisDimensionsProperties**

MD_Dimension/**dimensionName**

MD_Dimension/**dimensionSize**

MD_Dimension/**resolution**

MD_GridSpatialRepresentation/**cellGeometry**

MD_VectorSpatialRepresentation

MD_VectorSpatialRepresentation/**topologyLevel**

MD_VectorSpatialRepresentation/**geometricObjects**

MD_GeometricObjects/**geometricObjectType**

MD_GeometricObjects/**geometricObjectCount**

MD_FeatureCatalogueDescription

MD_FeatureCatalogueDescription/**featureTypes**

MD_FeatureCatalogueDescription/**featureCatalogueCitation**

MD_CoverageDescription

MD_CoverageDescription/**attributeDescription**

MD_CoverageDescription/**contentType**

MD_CoverageDescription/**dimension**

MD_RangeDimension/**sequenceIdentifier**

MD_RangeDimension/**descriptor**

MD_Band

MD_Band/**maxValue**

MD_Band/**minValue**

MD_Band/**units**

MD_Band/**bitsPerValue**

MD_Band/**peakResponse**

MD_Band/**toneGradation**

MD_Band/**scaleFactor**

MD_Band/**offset**

MD_ImageDescription

MD_ImageDescription/**illuminationElevationAngle**

MD_ImageDescription/**illuminationAzimuthAngle**

MD_ImageDescription/**imagingCondition**

MD_ImageDescription/**imageQualityCode**

MD_ImageDescription/**cloudCoverPercentage**

MD_ImageDescription/**processingLevelCode**

MD_ImageDescription/**compressionGenerationQuantity**

MD_ImageDescription/**triangulationIndicator**

Expected metadata extent

Previously defined set of items named ISO 19115 Orchestration Minimal will not be probably available generally in the future. We can expect that only a few closed communities e.g. companies can be able have all resources described in this level of detail. In general we can expect that available metadata will not be never so detailed.

We can expect that metadata available in the Czech republic are going to be prepared according to several types of detail. This is necessary to know for geodata evaluation.

These types are:

- metadata according INSPIRE IR (INSPIRE, 2007),
- metadata according to ISO 19115 core (ISO/TC 211, 2003),
- metadata according to Dublin Core basic set (DCMI, 2007),
- metadata according to MIDAS database (CAGI, 2007) completeness.

Other alternatives are not expected.

Metadata according to INSPIRE

The list of items is used from draft implementation rules (INSPIRE, 2007).

Level 1 is a basic level, that will be required always (if the conditional rule does not define different options).

- Resource title.
- Temporal reference – in a case when information is meaningful.
- Geographic extent of the resource.
- Resource language – in a case when text is used.
- Resource topic category.
- Keyword.
- Service type – in a case of a service.
- Resource responsible party.
- Abstract.
- Resource locator – in a case if any reference exists.

The second level is extended level and we can not expect full implementation of this level for all catalogues (datasets or services).

- Constraints.
- Lineage.
- Conformity.

- Service type version – in a case of a service.
- Operation name – in a case of a service.
- Distributed computing platform – e.g. Web Services.
- Resource Identifier – e.g. URI.
- Spatial resolution.

INSPIRE specifies other metadata elements, that can be available, but their usage by data (services) provides is disputable. The same problem is with the second level of metadata, where usage is based on provider decision. We can expect only following items: *resource title, geographic extent of the resource, resource language, resource topic category, keyword, resource responsible party, abstract* and in some cases *temporal reference*. That level of detail is not enough for the orchestration, but it can be used for a basic services selection.

Metadata according to ISO 19115 core

ISO 19115 core is more detailed than INSPIRE requirements and is going to be better applicable for orchestration. But we are still missing for example quality reports. Items in the core are Mandatory (M), Conditional (C) or Optional (O).

- Dataset title (M)
- Dataset reference date (M)
- Dataset responsible party (O)
- Geographic location of the dataset (by four coordinates or by geographic identifier) (C)
- Dataset language (M)
- Dataset character set (C)
- Dataset topic category (M)
- Abstract describing the dataset (M)
- Distribution format (O)
- Additional extent information for the dataset (vertical and temporal) (O)
- Spatial resolution of the dataset (O)
- Spatial representation type (O)
- Reference system (O)
- Lineage (O)
- On-line resource (O)
- Metadata file identifier (O)
- Metadata standard name (O)
- Metadata standard version (O)

- Metadata language (C)
- Metadata character set (C)
- Metadata point of contact (M)
- Metadata date stamp (M)

Metadata according to Dublin Core

Dublin Core is general standard and can be used for definition of own items, but we can not expect that providers will use such capabilities. They will probably use only simple metadata items list.

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Metadata according to MIDAS database completeness

We have analysed MIDAS database and we can probably expect same providers behaviour in the future. The following table categorised metadata items according to completeness in the MIDAS database. MIDAS system contains metadata about 3400 datasets.

Mandatory and conditional items were always filled (was controlled by the system). Optional items were filled in a case, when list of options was available. Very interesting is completeness of alternate title, temporal extent (date from), reference data and dataset usage. Out of interest are quality elements (except lineage).

Completeness	Metadata items
80 – 100 %	Title, abstract, coordinate system for metadata, metadata update, spatial schema, lineage, horizontal spatial accuracy, update frequency, data structure, format, language, classification, direct coordinate system, responsible party.
60 – 80 %	Alternate title, temporal extent (date from), planar extent (by coordinates), reference data.
40 – 60 %	Dataset usage
20 – 40 %	Memo, planar extent (by description)
5 – 20 %	Abbreviated title, version, purpose of production, temporal extent (by description), metadata language, spatial coverage, scale, temporal extent (date to).
< 5 %	English title, English abstract, update date, fees, metadata update plan, vertical spatial accuracy, logical consistency, completeness, homogeneity, resolution, quality, vertical extent, distribution units, medium, indirect reference system, vertical reference system, features description

Table 6: Completeness of the metadata items in the MIDAS database

Comparison to ISO 19115 Orchestration Minimal

ISO 19115 Orchestration Minimal	INSPIRE	ISO 19115 core	Dublin Core	MIDAS*
MD_Resolution	+	–	–	-
LI_Lineage/source	+	+	+	+
DQ_CompletenessCommission	–	–	–	-
DQ_CompletenessOmission	–	–	–	-
DQ_AbsoluteExternalPositionalAccuracy	–	–	–	+**
DQ_GriddedDataPositionalAccuracy	–	–	–	-
DQ_RelativeInternalPositionalAccuracy	–	–	–	-
DQ_AccuracyOfATimeMeasurement	–	–	–	-
DQ_TemporalConsistency	–	–	–	-
DQ_TemporalValidity	–	–	–	-
DQ_ThematicClassificationCorrectness	–	–	–	-
DQ_NonQuantitativeAttributeAccuracy	–	–	–	-
DQ_QuantitativeAttributeAccuracy	–	–	–	+
EX_BoundingPolygon	+	+	+	+
EX_GeographicBoundingBox	+	+	+	+
EX_TemporalExtent	+	+	+	+
EX_VerticalExtent	+	+	+	-
SpatialrepresentationType	–	–	–	+
Language	+	+	+	+
TopicCategory	+	+	+	+
MD_Keywords	+	–	+	-

MD_GridSpatialRepresentation	–	–	–	-
MD_VectorSpatialRepresentation	–	–	–	+**
MD_FeatureCatalogueDescription	–	–	–	+
MD_CoverageDescription	–	–	–	-
MD_ImageDescription	–	–	–	-

Table 7: Comparison to ISO 19115 Orchestration Minimal

** Items completed over 60% has been included*

*** Partly*

The following table shows percent of the items that will be probably included according to selected standard, directive or system.

Standard, directive, system	**Percent of the ISO 19115 Orchestration Minimal items available**
INSPIRE	34
ISO 19115 Core	27
Dublin Core	31
MIDAS	42

Table 8: Percent of the ISO 19115 Orchestration Minimal items available

Conclusion

Results of the research are not so optimistic, because we can not expect in any potential case that metadata are enough detailed for the efficient orchestration. To build orchestras dynamically needs to use alternative ways, how to evaluate served geodata.

According to results of our research, we have decided to use metadata for geodata, but not as only single source for geodata evaluation. We are preparing methodology how to deal with evaluation.

Basic principles of the methodology are summarised in the following points:

- If it is possible use simple orchestras
- Do not base creating groups of similar services on metadata for geodata
- Use experts' evaluation of the orchestras results to create groups of similar services
- Update groups of similar services according to new results evaluation
- Evaluate simple orchestras' results as well

If you are interested in the prepared methodology, please read the arcitle that will be published in the proceedings of the symposium GIS Ostrava 2009.

References

CAGI. (2007). MIDAS. 2001- 2007. at http://gis.vsb.cz/midas/, [accessed 2 July 2007].

DCMI. (2007) *Dublin Core Element Set v. 1.1. – Reference Description*, online[1], [accessed 12 April 2007].

INSPIRE. (2007). *DT Metadata – Draft Implementing Rules for Metadata* at online[2], [accessed 12 April 2007].

ISO/TC 211. (2003). *ISO/FDIS 19115:2003*. ISO/TC 211 Secretariat, Oslo, Norway, 152 p.

Růžička, J., Kaszper, R. Opět o metadatech v geoinformatice. *Proceedings 1. národní kongres v Česku – Geoinformatika pro každého, May 29-31 2007, Mikulov, Czech Republic*, online[3], [accessed 2 July 2007].

Support

The article is supported by Grant agency of the Czech republic GACR as a part of the project **GA 205/07/0797 GeoWeb services orchestration**. The article is supported by **open source** community as well. We have used open source projects GeoNetwork Open Source, WSCO, Apache Tomcat, Jetty, Open Office, GIMP, Dia, PostGIS, PHP, PostgreSQL, Apache HTTP Server, GNU/Linux Ubuntu, GNU/Linux Debian, X11, MySQL, Freefont and others for this article.

[1] `http://dublincore.org/documents/dces/`

[2] `http://www.ec-gis.org/inspire/reports/ImplementingRules/draftINSPIREMetadataIRv2_20070 \ 202.pdf`

[3] `http://mikadapress.com/prednasky/Ruzicka.pdf`