7th Framework Programme
ENV.2010.4.1.2-2
Integrating new data visualisation approaches of earth Systems into GEOSS development

Project Nr: 265178

QUAlity aware VIsualisation for the Global Earth Observation system of systems

Deliverable D6.3
Assessment of standards, protocols and guidelines employed in GeoViQua

Version 1.0.3

Due date of deliverable: 31/07/2013
Actual submission date: 10/12/2013

Document control page				
Title	6.3 Assessment of standards, protocols and guidelines employed in GeoViQua			
Creator	IS_CREAF			
Editor	IS_CREAF and JM_CREAF			
Description	Review of standards, protocols and guidelines used in GeoViQua			
Publisher	GeoViQua Consortium			
Contributors	GeoViQua Partners			
Type	Text			
Format	MS-Word			
Language	EN-GB			
Creation date	4/07/2013			
Version number	0.1			
Version date	10/12/2013			
Last modified by				
Rights	Copyright © 2013, GeoViQua Consortium			
Dissemination level		CO (confidential, only for members of the consortium)		
	X	PU (public)		
		PP (restricted to other programme participants)		
		RE (restricted to a group specified by the consortium)		
	When restricted, access granted to:			
Nature		R (report)		
		P (prototype)		
		D (demonstrator)		
	X	O (other)		
Review status		Draft	*Where applicable:*	
	X	WP leader accepted		Accepted by the PTB
		PMB quality controlled		Accepted by the PTB as public document
	X	Coordinator accepted		
Action requested		to be revised by all GeoViQua partners		
		for approval of the WP leader		
		for approval of the PMB		
	X	for approval of the Project Coordinator		
		for approval of the PTB		
Requested deadline				

Revision history			
Version	Date	Modified by	Comments
0.0	24-07-2012	IS_CREAF	Created the basic content of the deliverable
0.1	30-07-2013	IS_CREAF, JM_CREAF	Second review with comments on standards assessment and GeoViQua's extensions
0.2	10-11-2013	IS_CREAF, JM_CREAF, GC_CREAF	Third review with new content on W3C Prov, QualityML and ISO 19157
0.3	20-10-2013	JM_CREAF	Reorganization of content
0.3	28-10-2013	JB_UREAD	Incorporation of content related to visualization standards
0.3	11-11-2013	LB_ASTON, DC_ASTON	Review of the document and incorporation of new content about several topics
0.3	22-11-2013	LB_CNR	Review of the document and incorporation of new content related to catalogues
0.3	26-11-2013	ST_FRAUN	Review of the document and incorporation of new content related to KML
1.0	08-12-2013	JM_CREAF	Final review

Institution	Contributors
CNR	LB_CNR:Lorenzo Bigagli
CREAF	JM_CREAF: Joan Masó; IS_CREAF: Ivette Serral, GC_CREAF: Guillem Closa
ASTON	LB_ASTON: Lucy Bastin, DC_ASTON: Dan Cornford
UREAD	JB_UREAD: Jon Blower
FRAUN	ST_FRAUN: Simon Thum

Table of Contents

Figures

Tables

1. Introduction

GeoViQua has to provide tools for elicit, search and visualize data based on quality indicators to be integrated in the GCI. The GCI is a collaborative effort where several data provides agrees on using standards and interoperability arrangements to be able to jointly achieve a common platform to discover, share and access geospatial information.

GeoViQua proposed tools are based on the same standards, some times extending them to better include the requirements of the stakeholders.

1.1 GEOSS Standards Interoperability Registry

The intergovernmental Group on Earth Observations (GEO) is leading a worldwide effort to build a Global Earth Observation System of Systems (GEOSS). GEOSS builds upon, and adds value to, functioning Earth observation systems by supporting their interoperability, among other objectives. The added value primarily comes from interoperability facilitating the creation of datasets from disparate observation systems that can be used to obtain vital information for the benefit of society.

Interoperability in GEOSS will be achieved primarily by specifying how GEOSS components exchange data and information at their interfaces. The GEOSS strategy is to realize a system of systems through adoption of selected international standards that enable interoperability. The mechanism that facilitates the interoperability is the GEOSS architecture, which is realized by the components implemented as part of the GEOSS Common Infrastructure (GCI).

Components central to the GCI and the goal of interoperability are the Components and Services Registry and the Standards and Interoperability Registry. These registries maintain the GEOSS resources used for data sharing and the standards, protocols and other specifications enabling interoperability between them. This website hosts the GEOSS Standards and Interoperability Registry. The registry is available at: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157

The group overseeing the advocacy and problem resolution related to interoperability is the Standards and Interoperability Forum (SIF) http://seabass.ieee.org/groups/geoss/. This group is tasked by the Institutions and Development Implementation Board (IDIB)[1] to facilitate the interchange of information and the development of recommendations for standards and interoperability in GEOSS. The purpose of the SIF is to provide advice, expertise, and impartial guidance on issues relating to standards and interoperability. The SIF is also responsible to maintain the SIR.

[1] http://www.earthobservations.org/bo_id.shtml

Next we will review the standards relevant to GeoViQua. Each section provides and overview (including the SIR URN and URL) a standard assessment and a GeoViQua extension if any.

It is worth to mention that GeoViQua has been participating in the regular SIF meeting from the beginning of the project and that the SIF is using the GeoViQua twiki as a collaborative platform for storing meeting minutes and documentation (sif.creaf.cat).

2. Standards used in GeoViQua

GeoViQua has used standards for metadata, and data quality indicators are the foundations for the GeoViQua producer quality model. Mainly these standards come from the ISO 19100 series standards for Geospatial information

Also GeoViQua has used standards for services that come from the Open Geospatial Consortium. These standards are based on the OWS Common model and the ones referring to maps have been used for data visualization. Many of these standards are deeply related to the web standards

Other standards about data formats that can accommodate and spatially distributed quality have also been used.

2.1 Metadata

For geospatial metadata about datasets, the most widely used standards include the Dublin-Core Metadata Initiative (DCMI), the Federal Geographic Data Committee-Content Standard for Digital Geospatial Metadata (FGDC-CSDGM) and, predominantly in the GEOSS GCI, the ISO 19100 series standards for Geospatial metadata. In particular, the ISO 19115:2003 (ISO/TC211, 2003) is related to data products, the ISO 19119:2005 (ISO/TC211, 2005) refers to services, and the ISO 19139:2007 (ISO/TC211, 2007) contains a specific implementation in XML format. This format and some others are supported by the GCI architecture enabling the extraction of metadata records from the Clearinghouse. The standard records include quality elements organized in a hierarchic model. In this sense, ISO 19139:2007 contains the specifications for writing or translating the UML model in ISO 19115:2003 in XML format. In other words, ISO 19139:2007 can be regarded as the realization of ISO 19115:2003 in a XML format. Thus, other formats (e.g., excel files, external documents) must be converted to ISO19139:2007.

Metadata records in the current GEOSS use the ISO 19115 data model and its companion XML encoding (ISO 19139). ISO 19138 defines quality measures, but is expected to be superseded by a new standard, ISO 19157, which also supersedes ISO 19113[2]. On the current ISO standards, it must be remarked that the ISO 19138 (ISO-

[2] From Yang X, Blower JD, Bastin L, Lush V, Zabala A, Masó J, Cornford D, Díaz P, Lumsden J. An integrated view of data quality in Earth observation. Philos Trans A Math Phys Eng Sci. 2012 Dec 10;371(1983):20120072. doi: 10.1098/rsta.2012.0072. Print 2013 Jan 28.

TC211 2006) is being revised. The ISO 19157 (ISO-TC211 2011), currently under development, is prone to provide updated insights in this context.

2.1.1 ISO 15836:2003, Information and documentation -- The Dublin Core metadata element set

Overview

URN: urn:iso:std:iso:15836
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=165

ISO 15836:2003 is applicable to the Dublin Core metadata element set which deals with cross-domain information resource description. It is not exclusive form the GI doman. For Dublin Core applications, a resource will typically be an electronic document.

ISO 15836:2003 is for the element set only, which is generally used in the context of a specific project or application. Local or community based requirements and policies may impose additional restrictions, rules, and interpretations. It is not the purpose of ISO 15836:2003 to define the detailed criteria by which the element set will be used with specific projects and applications.

The Dublin Core allows:

- Creating and maintaining in a simplified way a reduced set of metadata that makes possible an effective search of resources on the Internet.
- Unifying the meaning of its elements among all users worldwide.
- Extending the basic set of metadata elements for the correct definition of newly emerging resources and solving their description needs.

Standard's assessment

Due to the simplicity of the standard and the tools available, must providers prefer the use of geospatial specific standards for metadata. GeoViQua decided to adopt ISO instead. Many concepts in the ISO model can be map back to the DC standard.

GeoViQua standard's extension

This standard has not been extended by GeoViQua.

2.1.2 ISO 19113:2002 Geographic information -- Quality principles; Publisher: ISO; Entry Type: Standard

URN: urn:iso:std:iso:ISO:19113
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=19

Overview

ISO 19113:2002 establishes the principles for describing the quality of geographic data and specifies components for reporting quality information. It also provides an approach to organizing information about data quality.

ISO 19113:2002 is applicable to data producers providing quality information to describe and assess how well a dataset meets its mapping of the universe of discourse as specified in the product specification, formal or implied, and to data users attempting to determine whether or not specific geographic data is of sufficient quality for their particular application. This International Standard should be considered by organizations involved in data acquisition and purchase, in such a way that it makes it possible to fulfil the intentions of the product specification. It can additionally be used for defining application schemas and describing quality requirements.

As well as being applicable to digital geographic data, the principles of ISO 19113:2002 can be extended to identify, collect and report the quality information for a geographic dataset, its principles can be extended and used to identify, collect and report quality information for a dataset series or smaller groupings of data that are a subset of a dataset.

Although ISO 19113:2002 is applicable to digital geographic data, its principles can be extended to many other forms of geographic data such as maps, charts and textual documents.

Standard's assessment

GeoViQua uses Quality elements defined in ISO 19113 as common quality parameters to retrieve and analyse quality embedded in the GEOSS GCI metadata.

Quality parameter	**Quality indicator**	**Definition**
Completeness	Commission	Excess data present in a dataset
	Omission	Data absent from a dataset
Logical consistency	Conceptual consistency	Adherence to rules of the conceptual schema
	Domain consistency	Adherence of values to the value domains
	Format consistency	Degree to which data is stored in accordance with the physical structure of the dataset
	Topological consistency	Correctness of the explicitly encoded topological characteristics of a dataset
Positional accuracy	Absolute or external accuracy	Closeness of reported coordinate values to values accepted as or being true
	Relative or internal accuracy	Closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true
	Gridded data position accuracy	Closeness of gridded data position values to values accepted as or being true
Temporal accuracy	Accuracy of a time measurement	Correctness of the temporal references of an item (reporting of error in time measurement)
	Temporal consistency	Correctness of ordered events or sequences
	Temporal validity	Validity of data with respect to time
Thematic accuracy	Classification correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse (*e.g.*,

		ground truth or reference dataset)
	Non-quantitative attribute correctness	Correctness of non-quantitative attributes
	Quantitative attribute accuracy	Accuracy of quantitative attributes

Table 1 Classification and definition of quality parameters and indicators

GeoViQua has accepted this classification as is. The only concern is about the use of the expression "quality parameter" that can be confused with the UncertML "statistic parameter". To avoid this confusion, GeoViQua prefers the name Quality class that is also used by ISO standards.

The ISO 19113:2002 classification for quality indicators has been widely accepted.

GeoViQua standard's extension

No extensions were needed at this level.

2.1.3 ISO 19114:2003 Geographic information -- Quality evaluation procedures; Publisher: ISO; Entry Type: Standard

Overview

URN: urn:iso:std:iso:ISO:19114
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=20

ISO 19114:2003 provides a framework of procedures for determining and evaluating quality that is applicable to digital geographic datasets, consistent with the data quality principles defined in ISO 19113. It also establishes a framework for evaluating and reporting data quality results, either as part of data quality metadata only, or also as a quality evaluation report.

ISO 19114:2003 is applicable to data producers when providing quality information on how well a dataset conforms to the product specification, and to data users attempting to determine whether or not the dataset contains data of sufficient quality to be fit for use in their particular applications.

Although ISO 19114:2003 is applicable to all types of digital geographic data, its principles can be extended to many other forms of geographic data such as maps, charts and textual documents.

Standard's assessment

This standard was initially developed to reflect classical national map agencies procedures. The QA4EO group has defined quality evaluation procedures that respect the fundamentals of this standard but are better adapted for remote sensing, which is the fundamental source of most information in GEOSS.

GeoViQua standard's extension

Geospatial providers in GEOSS, are well aware and regularly use quality evaluation procedures (also known in the remote sensing arena as CAL/VAL). The results of these evaluations are documented in the GeoViQua quality model but the actual procedure to determine QI is out of scope of the project.

2.1.4 ISO 19115:2003 Geographic information -- Metadata; Publisher: ISO; Entry Type: Standard and ISO 19115:2006 Geographic information -- Metadata (Corrigendum 1); Publisher: ISO; Entry Type: Standard

Overview

URN: urn:iso:std:iso:ISO:19115
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=22

URN: urn:iso:std:iso:ISO 19115:2003/Cor 1:2006
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=23

ISO 19115:2003 defines the data model required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

ISO 19115:2003 is applicable to: the cataloguing of datasets, clearinghouse activities, and the full description of datasets; geographic datasets, dataset series, and individual geographic features and feature properties.

ISO 19115:2003 defines: mandatory and conditional metadata sections, metadata entities, and metadata elements; the minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data); optional metadata elements - to allow for a more extensive standard description of geographic data, if required; a method for extending metadata to fit specialized needs.

Though ISO 19115:2003 is applicable to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.

Though ISO 19115:2003 includes the possibility of describing data quality using concepts form 19113 but also have some important elements associated with quality in a broader sense like lineage, citations etc.

Standard's assessment

This is the most known and implemented standard from the ISO 19xxx series. During its revision period, it received more that 1000 comments. This reflects the interest in the standard, but also reflects its limitations. While some people think it is too complicated

and detailed, others consider that some aspects are poorly solved. Particularly, MD_Usage is a timid effort to approach the user perspective to metadata and needs to be expanded. MD_Citation has to be extended to fully support the range of modern publication citations. Quality Indicators can be specified, as required by QA4EO, but traceability of the QI is not fully supported. The ISO metadata model is perceived as a very complex model. Metadata profiles are intended to simplify it but by doing so they are eliminating important aspects such as quality elements or provenance information.
There are also some difficulties in quantifying how complete a ISO19115 metadata document is. Here, XSLT transformations applied to the XML encoding of the ISO metadata can help. Two examples of this are the Rubric score and the GEOLabel. In fact, a GEOLabel can be automatically generated by providing to the tool in www.geolabel.net the URL locations of the ISO19115-compliant producer XML document and GVQ User feedback document, both in text/xml format.
There is also a problem and a need for clarification on how identifiers (and in particular URIs) for data, metadata have to be used since currently a mix of local and global identifiers are possible in the standard both for data and metadata. This has been addressed by the ISO TC211 in the next revision of 19115-1 still under final steps of approval.

GeoViQua standard's extension

GeoViQua focuses on the results of the three main aspects emphasized in ISO19115:2003 as data quality information: the quality indicator (DQ_Element), the lineage (LI_Lineage) and the usage information (MD_Usage). This element is intended for a brief description of the ways in which the product is being used or has been used.. User requirements in GeoViQua forced us both to migrate the MD_Usage into a GVQ_DiscoveredIssues and to develop a entirely new user feedback model to reflect user comments, ratings, metadata overwrite, report citations and so on.
GVQ_Publication is an specialization of CI_Citation for better paper and report citations. LI_Lineage can be used also to document the provenance of a DQ_Element. UncertML is used in the data quality report to better document the uncertainty of the data and its quality indicators, not just as a number but as a probability distribution.
The BrowseGraphic concept has been extended to allow a reference to a GEO Label representing this dataset.

2.1.5 ISO TS 19138:2006 Geographic information -- Data quality measures; Publisher: ISO; Entry Type: Standard

Overview

URN: urn:iso:std:iso:ISO:ts:19138 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=39

ISO/TS 19138:2006 defines a set of data quality measures. These can be used when reporting data quality for the data quality subelements identified in ISO 19113. Multiple measures are defined for each data quality subelement, and the choice of which to use will depend on the type of data and its intended purpose.

The data quality measures are structured so that they can be maintained in a register established in conformance with ISO 19135.

ISO/TS 19138:2006 does not attempt to describe every possible data quality measure, only a set of commonly used ones.

Standard's assessment

The quality measures presented in ISO1938 are well described but there is a lack of harmonization among them. Some measures represent the same concept with a different mathematical metrics and some quality measures share the same metrics. A good statistical foundation is needed to avoid this confusion. At the beginning of the project the use of UncertML was proposed to fix this but in practice this approach has been insufficient.
Even if the list of classical quality measures used for the cartographical agencies to produce topographical maps is quite complete, during the project it was clear that many quality measures used to parameterize quality in other activities are not equally well addressed. For example, in categorical classifications QIs are missing or poorly addressed.

GeoViQua standard's extension

In fact, this standard has been superseded by the ISO 19157 and GeoViQua has adopted the later. Also, an important effort has been made to formalize URIs for every quality measure included in 19138. QualityML is a language that is dividing every measure description in 19138 into a measured field and the mathematical metric expression (sometimes a statistic) that is used to quantify quality. Other measures for other fields have been included in QualityML and this can be seen as an extension of ISO 19138. For more details see the QualityML section.

2.1.6 ISO/TS 19139:2007 Geographic Metadata (XML Encoding); Publisher: ISO; Entry Type: Standard; Date Submitted: Dec 9th, 2008

Overview

URN: urn:iso:std:iso:ISO:ts:19139 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=174

ISO/TS 19139:2007 defines Geographic MetaData XML encoding, an XML Schema implementation derived from ISO 19115.

Since ISO 19115 does not provide any encoding, the actual implementation of geographic information metadata could vary based on the interpretation of metadata producers. In an attempt to facilitate the standardization of implementations, this comprehensive metadata implementation specification provides a rule-based encoding

for applying ISO 19115. It provides Extensible Markup Language (XML) schemas that are meant to enhance interoperability by providing a common specification for describing, validating and exchanging metadata about geographic datasets. Its applicability for feature or attribute metadata is also possible but not complete defined.

Standard's assessment

XML encoding resulting from this standard is very difficult to fully populate without the proper tools. Good metadata editors are needed. Tools tend to reflect the ISO19115 model classification but this model is not the proper way of presenting the metadata to the user or to build a metadata model on top. Users need a clear flow based around the production workflow of the data that they are documenting but they also need tools that force them to populate essential components such as data temporal extent, quality indicators, provenance...

This standard will be superseded by 19115-3.

GeoViQua standard's extension

For each extension in 19115 or other quality standards, an extension on the XML schema has been done. Also, 19115-2 and 19157 have been adopted and encoded in an schema. The result is a common set of encodings for the data quality representations to be used across the project that can be downloaded from http://schemas.geoviqua.org.

The DAB-Q collects information from the 19139 documents but also from the GeoViQua extended XML documents.

2.1.7 ISO/DIS 19157

Overview

URN: pending URL: pending

The objective of this International Standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a dataset's quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.
This standard supersedes and unifies 19113, 19114 and 19138.

Standard's assessment

This standard draft accumulates 19113, 19114 and 19138. It is mainly following them directly with some small extensions such as the introduction of the metaquality concept and a new type of result called *descriptive results* and some refinements in the documentation of the evaluation method. Some new elements are of metaquality have

been left abstract with no concrete representation (such us confidence, representativity and homogeneity).

GeoViQua standard's extension

GeoViQua has adopted a final draft of ISO157 because some of the small improvements included were recognized as user requirements of the project. Some extensions has been done on top of it like the inclusion on the lineage of a quality indicator as a new metaquality extension and a way to better cite the dataset that was used as reference to assess quality.

At the time of writing this report, the standard is in the FDIS status but no XML schemas has been produced. GeoViQua has produced them and they are available at schemas.geoviqua.org.

2.1.8 USA FGDC CSDGM-Content Standard for Digital Geospatial Metadata; Publisher: OTHER; Entry Type: Standard; Date Submitted: Nov 27th, 2007

Overview

URN: urn:geoss-uuid:422a35856e3754098f03b718098109d3
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=130

The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geospatial data. The standard establishes the names of data elements and compound elements (groups of data elements) to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements.

This standard is intended to support the collection and processing of geospatial metadata. It is intended to be useable by all levels of government and the private sector. The standard is not intended to reflect an implementation design. An implementation design requires adapting the structure and form of the standard to meet application requirements.

The standard was developed from the perspective of defining the information required by a prospective user to determine the availability of a set of geospatial data; to determine the fitness and the set of geospatial data for an intended use; to determine the means of accessing the set of geospatial data; and to successfully transfer the set of geospatial data. As such, the standard establishes the names of data elements and compound elements to be used for these purposes, definitions of these data elements and compound elements, and information about the values that are to be provided for the data elements. The standard does not specify the means by which this information is organized in a computer system or in a data transfer, nor the means by which this information is transmitted, communicated, or presented to the user.

This standard is for the documentation of geospatial data. Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data

Infrastructure," was signed on April 11, 1994, by President William J. Clinton. Section 3, Development of a National Geospatial Data Clearinghouse, paragraph (b) states: "Standardized Documentation of Data. Beginning nine months from the date of this order, each agency shall document all new geospatial data it collects or produces, either directly or indirectly, using the standard under development by the FGDC, and make that standardized documentation electronically accessible to the Clearinghouse network. Within one year of the date of this order, agencies shall adopt a schedule, developed in consultation with the FGDC, for documenting, to the extent practicable, geospatial data previously collected or produced, either directly or indirectly, and making that data documentation electronically accessible to the Clearinghouse network." This standard is the data documentation standard referenced in the executive order.

The elements that are considered under data quality are: attribute accuracy, logical consistency, completeness, positional accuracy, lineage and cloud cover,

Standard's assessment

This standard was used as the basis for ISO19115 so it shares several commonalities. In the quality classes it includes 4 common categories, ignoring temporal quality indicators, and considers cloud cover as a quality element. Lineage is described in a similar way as in the ISO model but lacks the extension that ISO19115-2 introduces.

GeoViQua standard's extension

The DAB-Q accepts and brokers these files into the general GeoViQua metadata model (mainly based in ISO19115).

2.1.9 W3C Prov

This is a recently approved standard for provenance in the web.

The W3C PROV is a generic provenance Data Model (DM) that can accommodate different perspectives of provenance in the web. The core structure of the data model is based on the relations of entities, activities, agents. The entities are digital, physical or conceptual things that are described with provenance. The entities are attributed to (WasAttibutedTo) an Agent, which is something or somebody that have some responsibility over an entity or activity (WasAssociatedWith). An activity is something that occurs over a period of time which use (Used) or produce (WasGeneratedBy) some entities.

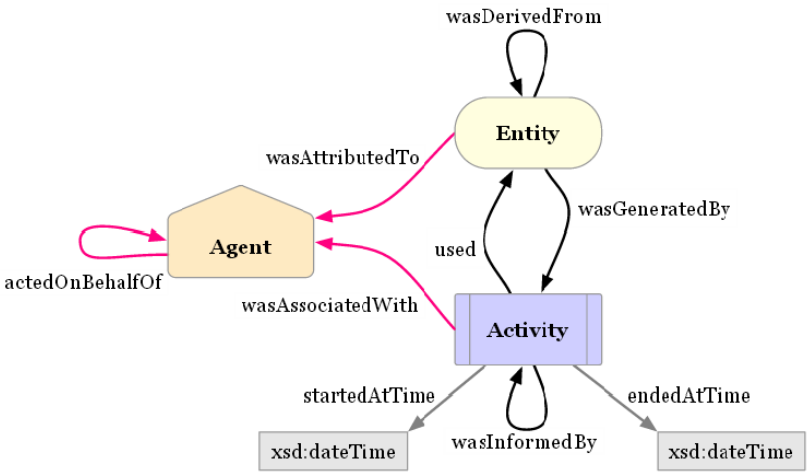


Figure 1 The core structure of the data model is based on the relations of *Entities*, *Agents*, and *Activities (Prov-O)*.

This core structure is expanded with the complete DM to accomplish the representation of provenance in a distributed environment and for different fields. The DM can be encoded in OWL, XML, PROV-N, and JSON.

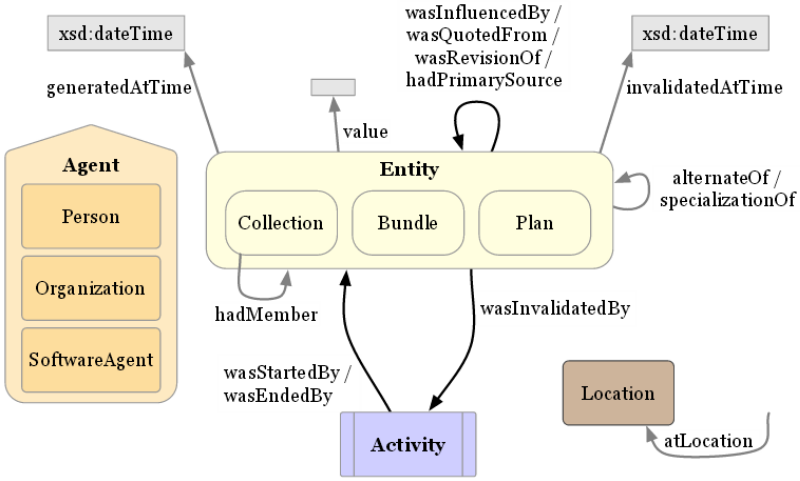


Figure 2 The complete structure of W3C DM (Prov-O).

Even W3C Prov is not a specific data model for describing Geospatial Provenance the model can be customized it for specific needs and for geospatial needs in particular. The way W3C Prov has to be used in the geospatial domain is still under discussion in OGC and will be addressed in OWS10.

Standard's assessment

As the first steep to produce Geospatial Provenance Metadata in W3C specification, we mapped the equivalences between W3C and ISO 19115 (LI_Linage and LE_Linage tags) (Masó, 2011).

W3C name	ISO19115 xpath

Entity	/MD_Identification, LI_Source
Entity Id	/MD_Metadata/MD_Identification/citation/CI_Citation
Activity	/LI_ProcessStep
wasgeneratedBy	/Li_process/process_step
Used	/LI_ProcessStep/source
WasInformedBy	
WasStartedBy	/gmd:LI_ProcessStep/gmd:processor/gmd:CI_ResponsibleParty/
WasEndedBy	/gmd:LI_ProcessStep/gmd:processor/gmd:CI_ResponsibleParty/
WasderivedFrom	/Li_Source/source/
wasQuotedFrom	/gmd:LI_ProcessStep/gmd:source
HadPrimarySource	/Li_Source
Agent	/gmd:LI_ProcessStep/gmd:processor/
WasAtributedTo	/Li_Source /gmd:LI_ProcessStep/gmd:processor/
WasAssociatedWith	/gmd:processor/gmd:CI_ResponsibleParty/
ActedOnBehalf	/gmd:processor/gmd:CI_ResponsibleParty/
Person	/LI_ProcessStep/gmd:processor/gmd:CI_ResponsibleParty/gmd:individualName/
Organization:	/LI_ProcessStep/gmd:processor/gmd:CI_ResponsibleParty/gmd:organisationName/
SoftwareAgent	/LI_ProcessStep/LE_ProcessStep/LE_processing/SoftwareReference
WasinfluencedBy	/Li_Source
Bundle	gmd:fileIdentifier
BundleContenT	/Li_Lineage
Plan	Li_Lineage/gmd:statement

Table 2 Equivalences between W3C and ISO 19115/19115-2

In W3C, provenance can be agent-centered, object-centered and process-centered. We consider geospatial provenance as either process-centered (a succession of activities), or object-centered (a succession of entities).

GeoViQua standard's extension

This standard has not been extended yet but there are plans to work in an spatial extension for W3C Prov by the end of the project.

2.2 Catalogue

Several catalogues are used in GEOSS but in GeoViQua only 3 standards and 1 interoperability arrangement have been actually used.

2.2.1 GeoNetwork Catalog Interface; Publisher: OTHER; Entry Type: Special Arrangement; Date Submitted: Oct 29th, 2009

Overview

URN: urn:fao:geonetwork:2.4
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=281

Geonetwork is an open-source Community portal and catalogue software that supports ISO 19115 (data) and 19119 (service) metadata. It is used to support many catalogues compliant with the ISO 19115 and INSPIRE standards. Schema plugins can be produced for Geonetwork which permit the display and editing of metadata documents that extend the ISO 19139 implementation.

Standard's assessment

Due the complexities of the catalogue standards, this widely used open source software adopted one of the possible interpretations of the standards. Several catalogues used in GEOSS are based on this implementation. The close alignment of the GeoViQua Producer Quality Model with existing ISO standards permitted the addition of our own open-source schema plugin which allows GeoViQua-compliant documents to be served, discovered and retrieved with no change to the interface of the catalogue.

GeoViQua standard's extension

In GeoViQua the metadata model that GeoNetwork accepts was extended to allow storing the GeoViQua producer quality model. The DAB-Q can also read from this GeoNetwork catalog. The plugin which permits this is open-source and is available at: https://github.com/lushc/geoviqua-geonetwork-plugin.

2.2.2 OGC Catalogue Service 2.0.1; Publisher: OGC; Entry Type: Standard

Overview

URN: urn:ogc:serviceType:CatalogueService:2.0.1:HTTP
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=52

The OpenGIS Catalogue Services Specification defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources.

- OGC Cataloguing of ISO Metadata (CIM) using the ebRIM profile of CS-W (0.1.7)
- OGC® Catalogue Services - OWL Application Profile of CSW (0.3.0)

- OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile (1.0.0)
- EO Application Profile for CSW 2.0 (1.4)
- FGDC CSDGM Application Profile for CSW 2.0 (0.0.12)
- OGC® Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products

Standard's assessment

This is the most popular standard for geospatial metadata documents based on ISO 19115. The standard is deeply rooted on the ISO model.

GeoViQua standard's extension

DAB-Q

A new design and developed software components enabling discovery for mining and screening datasets, including quality information by the GeoViQua DAB-Q. These components will be fully compliant with the GEOSS GCI Recommendations since they are conceived to contribute to the GEOSS Infrastructure.

The DAB-Q is an extension of the EuroGEOSS broker and it supports queries based on data quality constraints. It implements the User Quality Model (UQM), the Producer Quality Model (PQM) and the CSW-Q interface.

The DAB-Q enables smart search and discovery functionalities using parameters relevant to GEO products (date, scale/detail, bounding box, SBA, etc), extended to include quality information.

CSW-Q

CSW-Q is an extension of the standard OGC Catalog Service for the Web 2.0.2, ISO Application Profile (CSW ISO AP) to support quality-constrained queries.

Following the standard CSW behaviour, the CSW-Q interface requires explicit quality statements in order to match query constraints. In other words, datasets without quality statements will never be returned as a result of a quality constrained query.

Queryable properties are selected concrete expressions of quality parameters/indicators. Selected means that they are a subset of all the possible ones. Concrete means that they are implementable in an efficient way. We distinguish between PQM and UQM queryables. Furthermore, specific queryables are derived from the proposed Quality Label aspects.

The chosen queryables are a small subset of the possible conceivable representations of data quality.

The PQM introduces indicators to record qualitative and quantitative quality information, and to identify resources (i.e. datasets). The model extends ISO 191151, 19115-22 and 191573, adding means to report publications, discovered issues, reference datasets used for quality evaluation, traceability and statistical summaries of quantified uncertainty. The chosen PQM queryables are:

- Quality info count
- Thematic accuracy result value
- Domain consistency result value
- Lineage process
- Lineage source

The UQM describes the structure and attributes of comments, citations, discovered issues, ratings and reports of usage. It re-uses some ISO quality and metadata elements, and elements of the producer model, but is less strictly bound to existing ISO schemas. The chosen UQM queryables are:

- Rating
- Report aspect
- Main text
- User domain

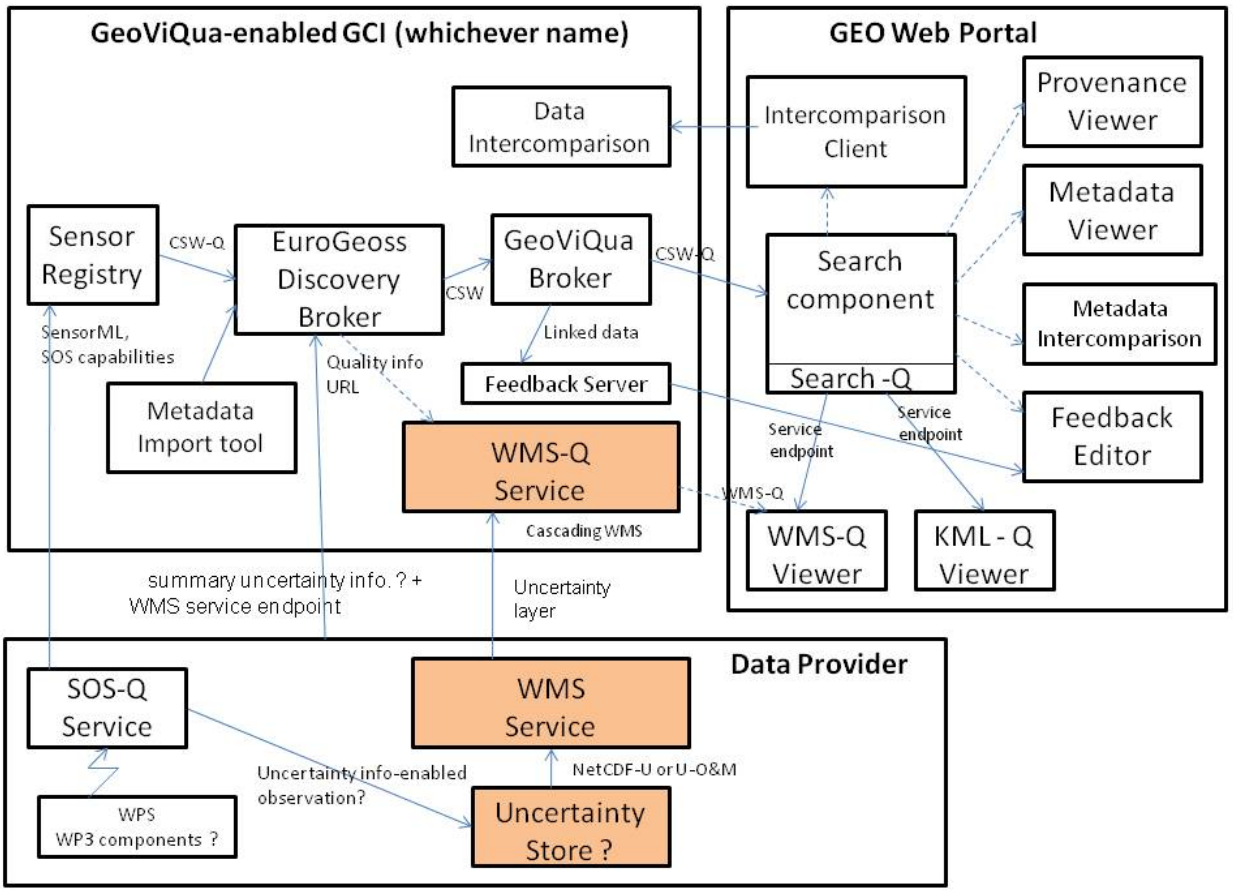


Figure 3 GeoViQua components diagram

2.2.3 Web Accessible Folder (WAF); Publisher: W3C; Entry Type: Special Arrangement; Date Submitted: Feb 4th, 2009

Overview

URN: urn:geoss:waf
URL:
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=183

A Web Accessible Folder (WAF) is an HTTP accessible directory of files, typically metadata files in XML format in which all files and their time-stamps are visible to a web browser or client. Crawlers are able to parse the file listings and date-time stamps and provide a search interface on these documents.

Standard's assessment

This is a simple way to share metadata documents or data by exposing them in a web folder. It is not really a standard but a useful practical solution for a few items,not efficient for the discovery of big datasets.

GeoViQua standard's extension

GeoViQua was been experimenting with metadata files that use the new GeoViQua model in a WAF that CNR is hosting. In the DAB-Q, the WAF points to a URL in an apache http (or ftp) server directory that allows describing a folder of files. Each file is represented by a link that points to a XML file (ISO metadata record for a single resource). The folder acts as a metadata collection (simple catalogue) with no query language.

The GeoViQua DAB connects to several services coming from different partners: UREAD WMS-Q, CREAF WMS-Q, Sensor Registry 52N and also a GeoViQua catalogue (WAF) that is hosted by CNR, the clearinghouse, the extended GeoNetwork catalogue prepared by ASTON (2.2.1) and some community catalogues when needed. Note that the UserFeedback client is linked to the system at a different level also by the DAB-Q.

2.3 Visualization

2.3.1 OpenGIS Web Map Service Implementation Specification (1.1.1, 1.3 and ISO 19128:2005 Geographic information -- Web map server interface); Publisher: OGC; Entry Type: Standard

Overview

OpenGIS Web Coverage Service (WCS) 1.1.0 Implementation Specification
URN: urn:ogc:serviceType:WebCoverageService:1.1.0
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=76

OpenGIS Web Map Service (WMS) 1.3 Implementation Specification
URN: urn:ogc:serviceType:WebMapService:1.3:HTTP
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=67

ISO 19128:2005 Geographic information -- Web map server interface (WMS 1.3)
URN: urn:iso:std:iso:ISO:19128
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=35

The OpenGIS Web Map Service (WMS) Implementation Specification provides three operations (GetCapabilities, GetMap, and GetFeatureInfo) in support of the creation and display of registered and superimposed map-like views of information that come simultaneously from multiple remote and heterogeneous sources.

When client and server software implements WMS, any client can access maps from any server. Any client can combine maps (overlay them like clear acetate sheets) from one or more servers. Any client can query information from a map provided by any server. While programmers need to write code to implement the specifications, end users can take advantage of products that include them to publish and access geospatial information. Software buyers can choose the best solution for their needs

and not worry about if it will work with other solutions; if they all implement the same standard, WMS, they will all work together.

In particular WMS defines:

1. How to request and provide a map as a picture or set of features (GetMap)
2. How to get and provide information about the content of a map such as the value of a feature at a location (GetFeatureInfo)
3. How to get and provide information about what types of maps a server can deliver (GetCapabilities)

The specification 1.1.0 is primarily a correction and clarification of the OpenGIS Web Map Service Interface Implementation (OGC document #01-047r2).

Version 1.3.0 has become also the ISO 19128 standard.

Standard's assessment

WMS is a consolidated standard OGC with 4 main versions. The last one was been also declared as ISO standard. Its popularity makes it an excellent candidate for communicating data quality. Nevertheless, before GeoViQua there was no documented attempt to do so. A WMS layer could show a spatialized quality and the ServiceMetadata document element URL template can help to distribute metadata documents using the GeoViQua models.

The capability to transmit spacialized quality will depend on the styles applied to the data so it will be important to associate the right style to the data. It can be combined with SDL to better encode quality visualization[3].

The standard does not include any recommendation about data quality.

GeoViQua standard's extension

Through task 5.3, GeoViQua visualises quality information through a number of WMS and WMTS clients, including desktop and web-based GIS tools.

The WMS 1.3.0 currently does not describe how to document quality information into WMS. For example, if the dataset-level quality information is presented in an unstructured format (e.g. PDF), or encoded using GeoViQua proposed producer model, or encoded using ISO19157, the current WMS 1.3.0 provides no means to integrate them into the WMS. If the pixel-level uncertainty information is presented using NetCDF-U (which defines NetCDF uncertainty conventions), the current WMS does not describe how data layer can semantically associate with the corresponding uncertainty layers. Hence, we proposed a WMS-Q specification to address these limitations. This WMS-Q

[3] More information in the D5.2 Visualisation of quality information best practice report, http://twiki.geoviqua.org/twiki/pub/GeoViQuaIntranet/D5_2/D5.2_Visualization_of_quality_information_best_proactice_report_v14.pdf

specification was proposed as far as possible within the bounds of the WMS 1.3.0 specification, requiring as few extensions as possible. Discussions of this document have been moved to OGC OWS- 9.

In order to integrate the dataset-level quality information into the WMS, in the WMS-Q, we propose to expand slightly “Type” attribute of “MetadataURL” element to have “unstructured” and “other-structured” options. We also propose to add a “description” element for the “MetadataURL” element. A fragment of capabilities document showing the integration of uncertainty information into WMS is shown in the Figure (where uncertainty is modelled as distribution). The proposed WMS-Q specification can be summarised as follows:

1. The controlled vocabulary for ‘Type' attribute of <MetadataURL> element is expanded. This expansion is compulsory. A ‘<Description>' element for <MetadataURL> element is added, but this extension can be optional.

2. There could be a top-level non-displayable <Layer> that holds all metadata that apply to all layers in the WMS instance. For example, in ncWMS this includes the coordinate reference systems in which all layers can be returned. In WMS-Q this Layer could hold <MetadataURL>s pointing to quality information that pertains to all Layers.

3. If the concept of a “Dataset” (i.e. a collection of Variables) is an appropriate grouping in a given situation (e.g. when including multiple NetCDF datasets in a single WMS), they can be modelled as non-displayable Layers directly underneath the top-level Layer. <MetadataURL>s can be associated to convey quality information at the Dataset level.

4. An uncertain variable can be modelled as a Layer nested inside a relevant Dataset. This Layer may or may not be displayable, depending on the service provider. If it is displayable it needs a Name, and a GetMap request on this name will display some representation of an uncertain variable in a single map image (e.g. a raster field representing the mean, overlain by a contour field representing the variance.) If it is not displayable (i.e. there is no Name) it is simply a container for its children (directly analogous to the NetCDF-U “concept without values”). A Keyword indicates the type of the distribution, from the UncertML vocabulary. Additionally, units of measure could be recorded using a different keyword.

5. Each component of the uncertain variable (e.g. mean, variance) is modelled as a displayable Layer inside its parent. Each component is tagged using a Keyword that defines which component it represents (e.g. mean or variance). A GetMap request on the component will result in a map that displays a raster field of that component only. <MetadataURL>s can be associated with either the individual components or, more likely, with the parent Layer that represents the uncertain variable as a whole.

```
<Layer>
<Title>Some Dataset</Title>
...
<Layer queryable="1">
 <!-- Parent layer, analogous to NetCDF-U "concept without values"-->
 <Title>Sea Surface Temperature</Title>
 <!   The Name is only included if this parent layer is displayable - >
 <Name> sea_surface_temperature</Name>
 <KeywordList>
   <Keyword vocabulary="http://www.uncertml.org/distributions/">
     normal
   </Keyword>
 <KeywordList>

 <!--Now the child layers represent the elements of the distribution -->
 <Layer queryable="1">
  <Name>sea_surface_temperature_mean</Name>
  <Title>SST mean</Title>
  <KeywordList>
    <Keyword vocabulary="http://www.uncertml.org/distributions/">
      normal#mean
    </Keyword>
  <KeywordList>
 </Layer>

 <Layer queryable="1">
   <Name>sea_surface_temperature_variance</Name>
   <Title>SST variance</Title>
   <KeywordList>
     <Keyword vocabulary="http://www.uncertml.org/distributions/">
       normal#variance
     </Keyword>
   <KeywordList>
  </Layer>
 </Layer>
</Layer>
```

Figure 4 “Layer nesting” mechanism: a fragment of capabilities document showing the integration of uncertainty information into WMS.

This extension is documented as a public OGC Engineering report that can be downloaded at: https://portal.opengeospatial.org/files/?artifact_id=52884

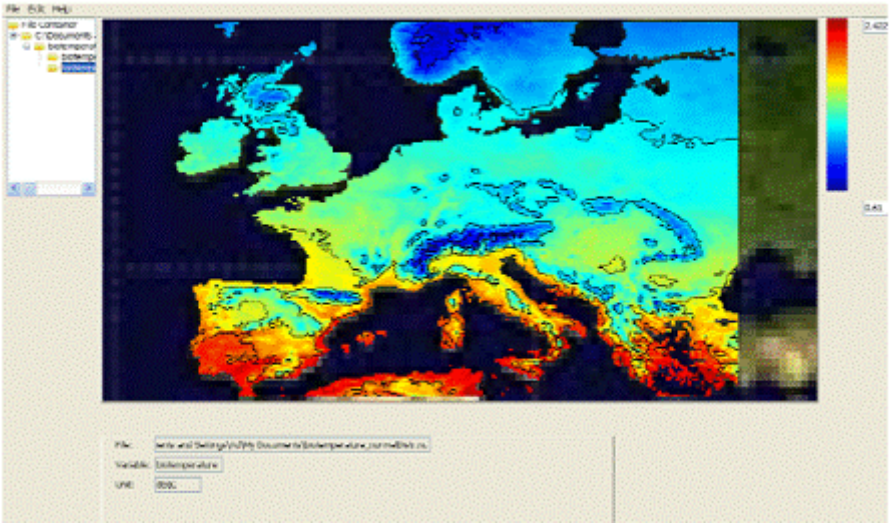


Figure 5 Uncertainty visualisation. ‘variance' is presented as a contour lines overlaid with ‘mean' raster map. Generated with desktop GODIVA-Q WMS-Q Viewer client.

2.3.2 ncWMS

Overview

ncWMS, is an implementation of the Open Geospatial Consortium's Web Map Service (WMS) specification for multidimensional gridded environmental data. ncWMS can read data in a large number of common scientific data formats – notably the NetCDF format with the Climate and Forecast conventions – then efficiently generate map imagery in thousands of different coordinate reference systems. It is designed to require minimal configuration from the system administrator and, when used in conjunction with a suitable client tool, provides end users with an interactive means for visualizing data without the need to download large files or interpret complex metadata. It is also used as a "bridging" tool providing interoperability between the environmental science community and users of geographic information systems. ncWMS implements a number of extensions to the WMS standard in order to fulfil some common scientific requirements, including the ability to generate plots representing timeseries and vertical sections. We discuss these extensions and their impact upon present and future interoperability. We discuss the conceptual mapping between the WMS data model and the data models used by gridded data formats, highlighting areas in which the mapping is incomplete or ambiguous. We discuss the architecture of the system and particular technical innovations of note, including the algorithms used for fast data reading and image generation. ncWMS has been widely adopted within the environmental data community and we discuss some of the ways in which the software is integrated within data infrastructures and portals.

Standard's assessment

ncWMS implements some extensions to the WMS specification in order to add useful functionality whilst remaining fully backward compatible with the WMS 1.1.1 and 1.3.0 specifications. These are discussed further in Blower et al 2013 (http://dx.doi.org/10.1016/j.envsoft.2013.04.002) but are summarized here:

1. The GetMap operation is extended to permit control of the colour map used to style images using parameters including COLORSCALERANGE.
2. A GetMetadata operation is implemented that returns metadata in compact JSON documents, avoiding the problem of parsing very large Capabilities documents in XML, which can be inefficient in a web browser environment.
3. Other operations are defined to generate other plot types of multidimensional data, such as transects, vertical sections and timeseries plots.

This approach was discussed in OGC but since just covers some use cases, did not progress. It could be interesting if it is broad as an official change request.

GeoViQua standard's extension

This approach was used in one of the WMS-Q services in GeoViQua to convey quality information and the Greenland client uses this extension to represent ranges of uncertainty.

2.3.3 OpenGIS® Web Map Tile Service Implementation Standard; Publisher: OGC; Entry Type: Standard; Date Submitted: Nov 28th, 2010

Overview

URN: urn:geoss-uuid:3ad936387044ceea8a4794d2a14c88c1
URL:
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=303

This Web Map Tile Service (WMTS) Implementation Standard provides a standard based solution to serve digital maps using predefined image tiles. The service advertises the tiles it has available through a standardized declaration in the ServiceMetadata document common to all OGC web services. This declaration defines the tiles available in each layer (i.e. each type of content), in each graphical representation style, in each format, in each coordinate reference system, at each scale, and over each geographic fragment of the total covered area. It also provides the optional FeatureInfo documents providing descriptions of the maps at specific locations. This standard specifies several different concrete exchange mechanisms between clients and servers in two different architectural styles. The standard defines the GetCapabilities, GetTile and optional GetFeatureInfo operations for procedure oriented architectural style based approaches using several different message encodings, including messages encoded using Key-Value Pairs (KVP), XML messages, or XML messages embedded in SOAP envelopes. The standard also defines the request mechanisms and endpoint publishing strategy to enable a resource oriented architectural style based on web based URL endpoints allowing clients to simply request the ServiceMetadata, Tile, and FeatureInfo resources as documents.

Standard's assessment

This standard was developed with the WMS in mind with the sole objective that generate faster visualization in highly concurrent requests situation. It can be used to also represent spatialized uncertainty. The main difference with WMS is the lack of layer nesting and properties inheritance that was replaced by a flat layer list and an optional thematic classification of layers. Another important difference is that it is not possible to ask for 2 layers at the same time to combine two parameters on the same visualization on the server side s it relies on clients to do that.

GeoViQua standard's extension

GeoViQua has also extended this standard in the same direction: Using Metadata URLs liked in the Service Metadata document and by using themes structure to express relations between different layers that encode different parameters of a UncertML provability distribution. Since this standard is not intended for dynamic representations of requesting SLD symbolization and it is only limited to expose style names.

More details can be found in the OGC Enginyering Report:: https://portal.opengeospatial.org/files/?artifact_id=52884

2.3.4 OGC-07-063r1 OpenGIS WMS EO Application Profile, 31/10/2008; Publisher: OTHER; Entry Type: Standard; Date Submitted: Mar 24th, 2011

Overview

URN: urn:geoss-uuid:1d450dd7fabdbaee68f30a7aeb91dcf7
URL:
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=311

This profile started as an initiative funded by the British National Space Agency (BNSC) with ongoing support from the European Space Agency (ESA). The specification defines conventions for the Earth Observation (EO) community to use OGC Web Services. The objective is to tightly define how OGC Web Service specifications can be applied by data providers to ensure that the inventory level visualisation of EO data is carried out in a truly interoperable way.

Standard's assessment

This profile of WMS is very relevant to GEOSS due to its relation with EO data. In fact it uses the nesting layer capability and the dimension extension to represent the concept of dataset series, bands and time. It also sets a naming convention that reinforces the relations between layers

GeoViQua standard's extension

GeoViQua analyzed this standard to learn more about WMS profiles before writing the WMS-Q OGC Engineering report. WMS-Q and WMS-EO can be used together in a single instance.

2.3.5 OGC Styled Layer Descriptor and Symbology Encoding Specifications

Overview

The Styled Layer Descriptor (SLD) profile of the OGC Web Map Service enables clients to specify the styling of map layers in much greater detail than is possible with a "plain" WMS. Instead of using a simple named style (e.g. STYLES=rainbow), the SLD specification allows the client to send the server an XML document describing in detail how an image should be rendered. The format of this XML document is defined in the OGC Symbology Encoding (SE) specification.

Standard's assessment

The SLD specification has been found to be adequate for the needs of GeoViQua, but the Symbology Encoding specification does not provide mechanisms to satisfy typical use cases for uncertainty visualization of raster datasets (e.g. using contours or stippling to represent uncertainty).

GeoViQua standard's extension

We have worked on extensions to the SE specification, which allows much greater control of the styling of images based on raster data, including many use cases relevant to uncertainty visualization. These will be described in greater detail in a future deliverable (a revision of D5.1).

2.4 Data formats

Data in GeoViQua has been used in many different formats that will not be described here. Here we are describing the standard formats that had impact on the project development.

2.4.1 NetCDF Network Common Data Form; Publisher: OTHER; Entry Type: Standard; Date Submitted: Nov 23rd, 2007

Overview

URN: urn:geoss-uuid:3a649992bf25bafe1ac8ddc46ec71bef URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=127

NetCDF (network Common Data Form) is a set of interfaces for array-oriented data access and a freely-distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data.

NetCDF data is:

- Self-Describing. A netCDF file includes information about the data it contains.
- Portable. A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- Direct-access. A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- Appendable. Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.
- Sharable. One writer and multiple readers may simultaneously access the same netCDF file.
- Archivable. Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

The netCDF software was developed by Glenn Davis, Russ Rew, Steve Emmerson, John Caron, Harvey Davies, and Ed Hartnett at the Unidata Program Center in Boulder, Colorado, with contributions from many other netCDF users.

Standard's assessment

This standard has become popular because it is very used in the climate change community. The Climate and Forecast community has its own conventions for using netCDF called CF-NetCDF. Its use has been extended to other communities that deal with multidimensional environmental data, including Earth Observation.

GeoViQua standard's extension

Global climate change data is generated from data coming from many sources so it has associate some uncertainty that can vary form one place to another. studied the possibilities of the different conventions and adapted the existing conventions and extended them when necessary.

GeoViQua has continued the work done by other projects in pixel-level uncertainty information can be encoded using NetCDF Uncertainty Conventions (NetCDF-U), which defines some conventions for encoding the uncertainty information following an uncertainty model, namely UncertML 2.0. This is a work initially developed by the FP7 UncertWeb. For example, if the variable "Sea_Surface_Temperature" (SST) follows a normal distribution, it can be characterised using two parameters "mean" and "variance". NetCDF-U models "Sea_Surface_Temperature" as a "concept without values" and uses a variable attribute "ancillary variables" to associate the "mean" and "variance". Discussions are ongoing for including these concepts in the Climate and Forecast conventions.

2.4.2 OGC Geography Markup Language (GML) 3.1.1; Publisher: OGC; Entry Type: Standard

Overview

URN: urn:ogc:dataFormat:GeographyMarkupLanguage:3.1.1 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=68

Also published as ISO 19136. Geography Markup Language is an XML grammar written in XML Schema for the modelling, transport, and storage of geographic information. The key concepts used by Geography Markup Language (GML) to model the world are drawn from the OGC Abstract Specification (available online: http://www.opengeospatial.org/techno/abstract.htm.

GML provides a variety of kinds of objects for describing geography including features, coordinate reference systems, geometry, topology, time, units of measure and generalized values. A geographic feature is "an abstraction of a real world phenomenon; it is a geographic feature if it is associated with a location relative to the Earth". So a digital representation of the real world can be thought of as a set of features. The state of a feature is defined by a set of properties, where each property can be thought of as a {name, type, value} triple. The number of properties a feature may have, together with their names and types, are determined by its type definition. Geographic features with geometry are those with properties that may be geometry-valued. A feature collection is a collection of features that can be regarded as a feature; as a consequence a feature

collection has a feature type and thus may have distinct properties of its own, in addition to the features it contains.

OGC 06-080, GML Application Schema for EO Products, Version 0.9.3, 21/07/2008. It defines an application schema of the Geography Markup Language (GML) version 3.1.1 for describing Earth Observation products (EO products) within the HMA (Heterogeneous EO Missions Accessibility) Application Profile for the OGCTM Catalogue Services Specification v2.0.0 (with Corrigendum) [OGC 04-021r3]

Standard's assessment

This standard is now for its complexities but also for its completeness. The whole idea is that GML can represent almost any geospatial data model but requires an application schema to do it. Objects can be of different types an have associated simple or complex properties.

GeoViQua standard's extension

In OWS-9 the possibility to express object level quality and provenance was explored. Quality properties can be attached to GML elements to do that that are associated with ISO19115 metadata concepts. The results were a considerable increase of the size of the data. In OWS-10 a continuation of this activity is foreseen to look into ways of removing the redundancy but saving the capabilities of this approach.

The OGC Engineering report can be downloaded from: https://portal.opengeospatial.org/files/?artifact_id=51818

2.4.3 OGC KML Version 2.2; Publisher: OGC; Entry Type: Standard; Date Submitted: May 1st, 2009

Overview

URN: urn:doc:is:kml:2.2 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=188

KML is an XML language focused on geographic visualization, including annotation of maps and images. Geographic visualization includes not only the presentation of graphical data on the globe, but also the control of the user's navigation in the sense of where to go and where to look. The standard allows encoding how supported geographic datasets are portrayed, individually as well as in combination.

Standard's assessment

Probably the most important fact for the purpose of GeoViQua is that KML is a mass-market standard. This is at odds with the geospatial expert folk who care enough about quality that they are interested in knowing the spatial distribution thereof. However, KML

is a superior visualisation standard because of its capability to describe how geospatial features are shown and, importantly, refined as appropriate. This especially holds in comparison to other 3D-capable standards that do not have a geospatial capability or exhibit insufficient coverage of geospatial issues.

Another interesting aspect is that most KML elements can have arbitrary, typed attributes without needing to extend KML at the XML level. Such attributes can therefore be leveraged by conforming clients to show e.g. metadata or quality information.

The existence of powerful 3D clients like Google Earth an NASA World Wind allows for a fast experimentation and easy dissemination of the data, given the data is of appropriate form. As data is becoming services this is a waning problem, but it remains a fact that the capabilities of KML are tightly coupled to the infrastructure which is behind it.

GeoViQua standard's extension

In GeoViQua, two approaches to representing quality information in KML were tested under the name "KML-Q". In the first approach the QI is inserted into the appropriate elements as attributes, which enables clients to show this information explicitly as shown the Figure. In the second, the category and the certainty at which the category was detected are combined into a bivariate colour map. This is also shown in the Figure; the two techniques can be combined if desired.

Figure 6 More saturated colours are more certain

The bivariate colour map in the test service is derived using a recent but well-established colour appearance model, CIECAM02. It should therefore be possible to judge quality using intuitively learned colour concepts such as saturation, i.e. with a smaller learning curve compared to colour maps which are not based on proper appearance models. Whether this is the case is subject to further research.

GeoViQua also demonstrated pure 3D techniques and using combination of 3D and colour coding. However, the more important aspect in terms of interoperability is that attributed can be read and, potentially, be made accessible to machines by connecting them using a suitable quality ontology. This will be subject to an engineering report to the KML working group within the OGC when the work is completed. They are not technically extensions, but mere conventions on how to apply KML to achieve a quality representation, visual or as attributes.

2.5 Sensor Web Enablement standards

2.5.1 Observations and Measurements - Observation schema; Publisher: OGC; Entry Type: Standard

Overview

Part 1. Observation schema

URN: urn:ogc:doc:std:om:1.0 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=279

This document specifies the core Observations and Measurements model. it describes a framework and encoding for measurements and observations. This is required specifically for the Sensor Observation Service and related components of an OGC Sensor Web Enablement capability, and also for general support for OGC compliant systems dealing in technical measurements in science and engineering.
The aim is to define a number of terms used for measurements, and the relationships between them. It proposes a conceptual framework for: observation, measurement, result, procedure, feature of interest, observed property, property type, coverage and related terms, presented using UML class diagrams and in equivalent GML conformant XML serialisations. The scope covers observations and measurements whose results may be quantities, categories, temporal and geometry values, coverages, and composites and arrays of any of these.

Part 2. Sampling Features

URN: urn:ogc:doc:std:sam:1.0 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=280

Describes a conceptual model and encoding for sampling features. These feature-types are typically associated with making observations producing estimates of property values that are in some way representative of a domain feature. Sampling features embody a sampling strategy that is suitable for the observation procedure and the observed property. Hence, sampling features are artefacts of the observation process rather than the inherent domain semantics. Similar sampling strategies are used across

a wide range of application domains so a common sampling features schema may be described.
Instances of sampling features may act as the proximate feature-of-interest for observations.

Standard's assessment

Various methods for encoding a quality description are available in the document. Five specific methods ("Description" – a textual description; "Rank" – a value on an ordinal scale; "Error" – a single quantity; "Bounds" – a pair of quantities, "Errors" – a set of quantities).

The original specification recognizes that the mechanisms introduced to allow recording of quality indicators and measures are very rudimentary and need to be developed further.

Connecting this standard with UncertML is required.

GeoViQua standard's extension

This standard was not in the direct scope of the project and has not been extended

2.5.2 OGC Sensor Observation Service; Publisher: OGC; Entry Type: Standard

Overview

URN: urn:geoss-uuid:c621b4fa721b4a4e7413955ad680d6b9
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=71

Sensor Observations Service (SOS) - Standard web service interface for requesting, filtering, and retrieving observations and sensor system information. This is the intermediary between a client and an observation repository or near real-time sensor channel.
A Sensor Observation Service provides an API for managing deployed sensors and retrieving sensor data and specifically "observation" data. Whether from in-situ sensors (e.g., water monitoring) or dynamic sensors (e.g., satellite imaging), measurements made from sensor systems contribute most of the geospatial data by volume used in geospatial systems today.
SOS has three mandatory "core" operations: GetObservation, DescribeSensor, and GetCapabilities. The GetObservation operation provides access to sensor observations and measurement data via a spatio-temporal query that can be filtered by phenomena.

Standard's assessment

By using O&M standard this standard incorporates quality concepts.

GeoViQua standard's extension

No extension had been done to this standard

2.6 Generic standards for services

2.6.1 OpenGIS Web Processing Service (WPS) 1.0.0; Publisher: OGC; Entry Type: Standard; Date Submitted: May 7th, 2009

Overview

URN: urn:ogc:serviceType:WebProcessingService:1.0.0 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=208

Web Processing Service (WPS) defines a standardized interface that facilitates the publishing of geospatial processes, and the discovery of and binding to those processes by clients. —Processes include any algorithm, calculation or model that operates on spatially referenced data. —Publishing means making available machine-readable binding information as well as human-readable metadata that allows service discovery and use.
A WPS can be configured to offer any sort of GIS functionality to clients across a network, including access to pre-programmed calculations and/or computation models that operate on spatially referenced data. A WPS may offer calculations as simple as subtracting one set of spatially referenced numbers from another, or as complicated as a global climate change model. The data required by the WPS can be delivered across a network, or available at the server.
This interface specification provides mechanisms to identify the spatially referenced data required by the calculation, initiate the calculation, and manage the output from the calculation so that the client can access it. This Web Processing Service is targeted at processing both vector and raster data.
The WPS specification is designed to allow a service provider to expose a web accessible process, such as polygon intersection, in a way that allows clients to input data and execute the process with no specialized knowledge of the underlying physical process interface or API. The WPS interface standardizes the way processes and their inputs/outputs are described, how a client can request the execution of a process, and how the output from a process is handled.

Standard's assessment

This standard is different form the others in the OGC family because it allows describing a completely generic process. It improves interoperability in the discovery and execution of processes but not in the definition of the process themselves. This was the definition of a process to correct the same satellite image type with the same algorithm can be done in a completely different way by service implementation, e.g. one exposing a quality report and the other providing just the resulting dataset and a third providing the result and a spacialized quality description layer.

The way to improve interoperability between service implementation of the same algorithm is by the use of WPS profiles (see http://portal.opengeospatial.org/files/?artifact_id=40310)

GeoViQua standard's extension

The flexibility of the WPS is a problem for interoperability but is also an advantage because it allows implementing also any process in a distributed environment. In the GeoViQua case, we encapsulate the GECA intercomparison algorithm into a WPS allowing us to integrate a quality elicitation tool into the GCI. The WPS is located within the data storage network and has very fast access - this is absolutely necessary to use the WPS for EO products that are too big to move them and some time can have direct access restrictions. This way, the WPS can reduce computing time and downloading time for the analyst. Regarding data access, it can be implemented by self-defined identifiers which are resolved by the WPS itself and then data is "just there" locally or downloaded from the FTP.

GeoViQua has not extended WPS but has generated a WPS implementation that demonstrates usefulness of this approach for other quality elicitation tools.

2.6.2 Representational State Transfer (REST); Publisher: OTHER; Entry Type: Special Arrangement; Date Submitted: May 7th, 2009

Overview

URN: urn:geoss-uuid:4527c3aed5dd5b798d3f2870e53cc42f
URL:
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=212

(and also urn:geoss-uuid:e4c9717355fa90cefdd5002d95f5998e
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=318)

REST is a term used to describe an architecture style of networked systems. REST is not a standard, but an architectural style. While REST is not a standard, it does use standards, such as HTTP, URL, resource representations (XML/HTML/GIF/JPEG/etc.), and MIME types (text/xml, text/html, image/gif, image/jpeg, etc).

Standard's assessment

This is a low level concept that emphasises on how the principles that makes the web work and allows to create new services and RESTful APIs on top a the web. It is focus on identifying resources and concentrating in the 4 essential operations of creation, retrieval, update and delete of resources. It is a bit controversial because there are no well defined set of rules but just a fuzzy collection of recommendations.

GeoViQua standard's extension

Since REST is not really a standard but an architectural style, it has been applied with no extensions to the GEO label generator tool and to the user feedback catalogue interface. Both services define a URI template to construct a URL to get resources. In the case of the User Feedback catalogue, also a POST interface is provided.

- GeoLabel API

The GEOLabel service is a web service that transforms a metadata document into an iconic representation, a "label" that summarizes important aspects visually. And at the same time this label is interactive to be able to dig deeper and get detailed information and to explain what it actually represents.

Proposed GEO Label Service Structure:
api.geolabel.net/v1/ - to be the base URL of the GEO label service
dev.geolabel.net/v1/ - to be the base URL for the API development documents

The GEOLabel API is used by the following components in GeoViQua:
- Geoportal: add it to search result pages and dataset detail pages, format: svg, URL is accessed via DAB-Q
- Broker client

Other potential uses:
- ESRI Geoportal Server
- GeoNetwork Catalog client
- Sensor registry search result listings

- User Feedback API

The user feedback model in GeoviQua is intended to connect GEOSS datasets and services to user data. User-centric information such as accounts and authentication and GEOSS concepts are therefore not represented directly, but are linked in at pre-defined junction points. These junction points are the user information and the target. While the model sometimes refers to ISO191xx schema elements, it is not intended to be seen as metadata or directly related to it.

The feedback model's main classes are:
- User information
- Feedback group
- Feedback item
- Feedback focus
- Target

The user information contains self-asserted information about the user which may be used (to assist) in qualifying the feedback he produces. The feedback item is an atomic instance of user feedback, for example, a comment. The primary focus specifies the context of the item, and points to the dataset or resource the feedback is about.

The feedback model is implemented by the feedback catalogue. The feedback catalogue uses the Django framework with a relational database, postgreSQL, as backend. The API framework Tastypie is used to construct a RESTful interface. Via this interface feedback items and collections can be requested and searched and feedback items can be posted. The requests can be done via RESTful URL requests and as response xml files will be returned (formatted according to the GeoViQua xsd schema). The following needs are covered:

- request for feedback items with certain type of feedback or with certain values for feedback
- request for a collection of items, related to a certain (primary) target
- request for (primary) targets with certain type of feedback or with certain values for feedback

An example of the use of the REST idea is the retrieval of feedback collections. A collection of items is something that does not exist directly in the UQM, but only in the combined PQM-UQM. A feedback item consists of user feedback related to one or more targets. A collection however contains all the feedback related to one target. This means that a reverse query needs to be done on the database. The REST API hides this for the user of the user feedback database by offering the search for collections as if it is an existing resource by supporting the following query:
/api/v1/feedback/collections/search?format=xml&target_code=xxx&target_codespace=xxx

2.7 Quality encodings

2.7.1 Guide to the expression of uncertainty in measurement; Publisher: ISO; Entry Type: Standard; Date Submitted: Feb 3rd, 2009

Overview

URN: urn:geoss-uuid:132f7fa4c594d9c1189ee6493c0a4921 URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=177

The Guide to the Expression of Uncertainty in Measurement (GUM) establishes general rules for evaluating and expressing uncertainty in measurement that can be followed at various levels of accuracy and in many fields – from the shop floor to fundamental research. Therefore, the principles of GUM are intended to be applicable to a broad spectrum of measurements, including: maintaining quality control and quality assurance in production; conducting basic research, and applied research and development, in science and engineering; and calibrating instruments and performing tests throughout a national measurement system in order to achieve traceability to national standards.

This Guide is primarily concerned with the expression of uncertainty in the measurement of a well-defined physical quantity that can be characterized by a unique value. If the phenomenon of interest can be represented only as a distribution of values or is

dependent on one or more parameters, such as time, then the measurands required for its description are the set of quantities describing that distribution or that dependence.

Standard's assessment

This standard describes some probability distributions but makes particular emphasis in the normal distribution and its complete description. It was developed in 1993 and the concepts have been include in more modern standards.

GeoViQua standard's extension

The concepts of this standard are included in UncertML language to encode uncertainty in an XML language. UncertML can be seen as an extension of GUM.

2.7.2 Uncertainty Markup Language (UncertML); Publisher: OGC; Entry Type: Standard; Date Submitted: Nov 13th, 2009

Overview

URN: urn:geoss-uuid:9bb66428fe52e2ebd87352e36def03e5
URL:
http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=289

The Uncertainty Markup Language (UncertML) is an XML encoding for the transport and storage of information about uncertain quantities, with emphasis on quantitative representations based on probability theory. The encoding allows an interoperable description of uncertain data, which we define to be random quantities, in a variety of ways including uni- and multi-variate distributions and mixture probability distributions; summary statistics (e.g. means, variances, standard deviations and quantiles; and realisations or sampled data. In addition to the XML encodings, there are JSON and netCDF encodings which permit UncertML to be used in low-bandwidth contexts and on large, multi-dimensional datasets. UncertML is fully documented at http://uncertml.org/. The current version is v2, and this will continue to be supported. However, there is also a v3 in preparation which will:

- Allow the use of named variables, rather than just numerical values, as parameters;
- Include a class of quality metrics, such as skill scores and confusion matrices;

Use more explicit typing which allows validation of JSON-encoded uncertainty information.

Standard's assessment

The quality measures presented in ISO1938 are well described but there is a lack of harmonization among them. Some measures represent the same concept with different mathematical metrics and some quality measures share the same metrics. A good statistical foundation is needed to avoid this confusion. At the beginning of the project the use of UncertML was proposed to address this, and in many cases (such as

encoding a quality assessment in an ISO 19139 quality report) uncertain quantities themselves can be fruitfully described with UncertML. However, the domain-agnostic nature of the language means that many commonly-used and highly-specific geospatial quality indicators fall outside the scope of UncertML. While the content of many quality indicators (probabilities or statistical summaries) can often be expressed in UncertML, some extra description of the semantic nature of the indicator is often required.
Though the list of classical quality measures used by cartographical agencies to produce topographic maps is quite complete, during the project it was clear that many quality measures used to parameterize quality in other domains are not equally well defined. For example, in categorical classifications, QIs are missing or poorly addressed.

GeoViQua standard's extension

Two approaches have been adopted in GeoViQua. To enhance the usability of UncertML across application domains, UncertML 3.0 has been developed. UncertML 3.0 provides two key additions. Firstly it refines the way in which all representations of uncertainty are encoded in XML and JSON to make parsing simpler and allow more freedom in the definition of parameters for the uncertainty representations. This permits the use of UncertML across a number of application domains. Additionally a new Metric type has been added to UncertML, which is designed to provide a new UncertML type that can be used for quality indicators with a sound statistical basis. UncertML also provides the basis of QualityML that has been developed in the GeoViQua context. It is based on the list of measure descriptions that are present in ISO19138 but other measures for other fields have also been included.

QualityML (http://qualityml.geoviqua.org/) (built on top of UncertML (http://www.uncertml.org/)) extends the standard by providing a conceptual model, some quality metrics with XML encoding, and guidelines for including quality measures in metadata. It enumerates quality indicators in the form of metrics or probabilistic uncertainties and statistics. These websites contain the information to quantify and exchange data uncertainties and quality in terms of a detailed description and XML schema, including use case examples. UncertML was developed under the INTAMAP and UncertWeb projects, and extended in the GeoViQua project while QualityML is a follow up development under GeoViQua project in collaboration with some UncertWeb partners.

GeoViQua Quality Indicators Dictionary - QualityML

The QualityML is a dictionary that contains:

- Generic quality element
- Quality measure matrix
- Definition of quality classes
- Definition of quality indicators
- Definition of measurement fields where statistics are applied
- Definition and examples of statistics used in quality measures (includes schemas)

Both *statistics* and *measurement fields* have a URI.

All relative URIs refer to http://qualityml.geoviqua.org/1.0

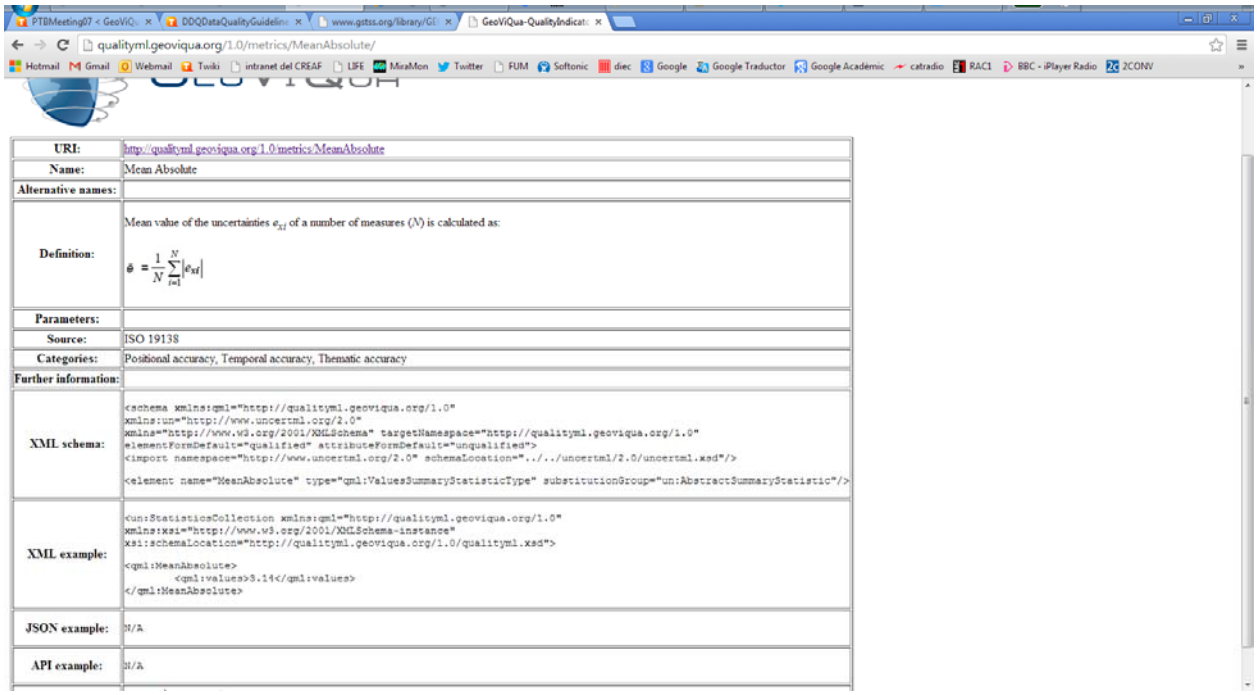
Generic quality element

A Quality element is a combination of a *quality class*, a *quality indicator*, a *quality measurement field*, a *quality statistic* (which include a statistic *name*, statistic *description*, statistical *parameter*, its *values* and *units of measure*). The combination of a quality measurement field and the a quality statistic is commonly knows as *quality measure*. This concepts are maped in the ISO 19139 in the following way:

Concept	ISO 19139 mapping	Example
Quality class and quality indicator	Name of the DQ_Element	gmd:DQ_CompletenessOmission
Quality measurement field	gmd:nameOfMeasure	Missing
Statistic name	gmd:valueType/@xlink:href	http://qualityml.geoviqua.org/metrics/items
Statistic description	gmd:valueType	Indicator of missing items
Statistical parameters	gmd:value/gco:Record/*	qml:rate and qml:max
Statistical values	gmd:value/gco:Record/*	66 and 100
Units of measure	gmd:valueUnit/@href	urn:ogc:def:uom:OGC:1.0:meter

Example using UncertML in an ISO metadata record

Figure 7 QualityML dictionary



URI:	http://qualityml.geoviqua.org/1.0/metrics/MeanAbsolute
Name:	Mean Absolute
Alternative names:	
Definition:	Mean value of the uncertainties e_{xi} of a number of measures (N) is calculated as: $\hat{e} = \frac{1}{N}\sum_{i=1}^{N}\lvert e_{xi}\rvert$
Parameters:	
Source:	ISO 19138
Categories:	Positional accuracy, Temporal accuracy, Thematic accuracy
Further information:	
XML schema:	`<schema xmlns:qml="http://qualityml.geoviqua.org/1.0" xmlns:un="http://www.uncertml.org/2.0" xmlns="http://www.w3.org/2001/XMLSchema" targetNamespace="http://qualityml.geoviqua.org/1.0" elementFormDefault="qualified" attributeFormDefault="unqualified"> <import namespace="http://www.uncertml.org/2.0" schemaLocation="../../uncertml/2.0/uncertml.xsd"/> <element name="MeanAbsolute" type="qml:ValuesSummaryStatisticType" substitutionGroup="un:AbstractSummaryStatistic"/>`
XML example:	`<un:StatisticsCollection xmlns:qml="http://qualityml.geoviqua.org/1.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://qualityml.geoviqua.org/1.0/qualityml.xsd"> <qml:MeanAbsolute> <qml:values>3.14</qml:values> </qml:MeanAbsolute>`
JSON example:	N/A
API example:	N/A

Figure 8 Metrics: Mean Absolute

QualityML defines Quality classes, Quality indicators, Measurement fields and Quality measures in XML encoding providing schemas and examples. It also provides a matrix of combinations of them and its relation with ISO19115 and ISO19157 (formerly ISO19138).

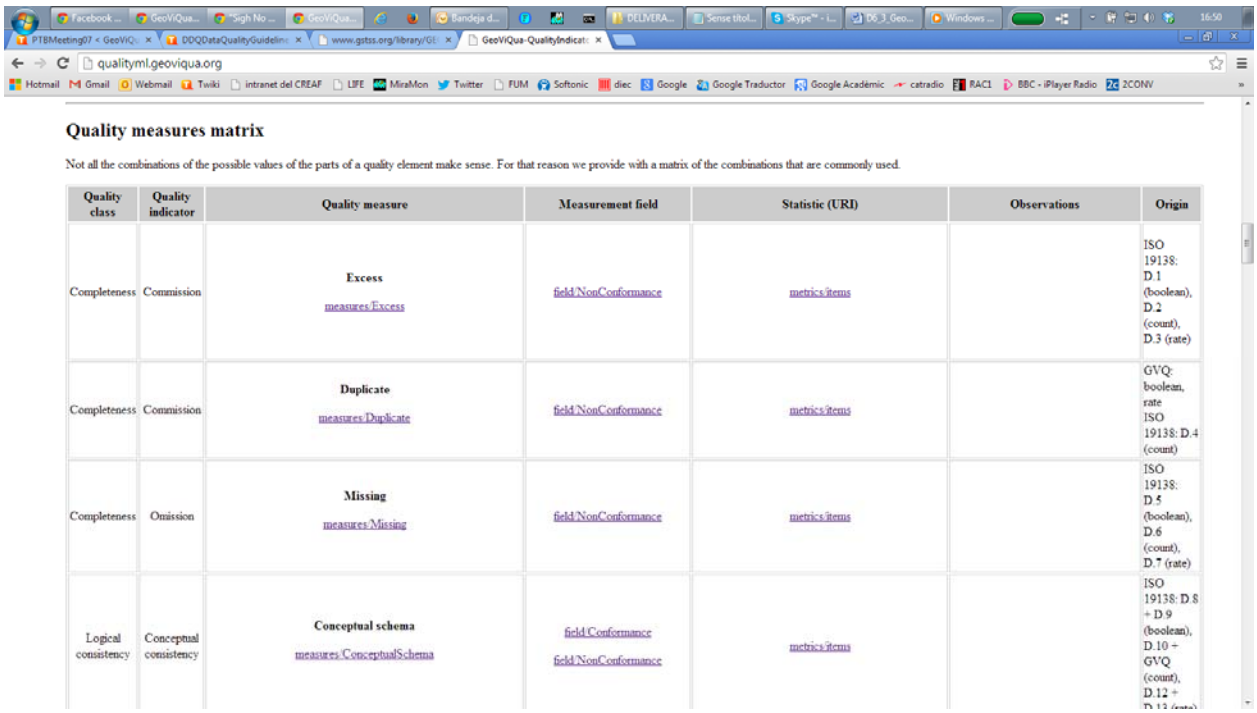


Figure 9 Quality measures matrix

QualityML is also an attempt to introduce quality concepts into the semantic web. For the moment all concepts have a URI but in the future it will be fully integrated in the semantic web.

3. Guidelines, Tutorials and Best practices

3.1 Data Quality Tutorial for GEOSS Providers. GEOSS Best Practices Wiki

GeoVIQua has developed a Data Quality Tutorial for GEOSS Providers available at the Best Practice Wiki. The purpose of this tutorial is to assist the GEOSS providers and GEOSS users in understanding the various ways in which GEOSS can be used to provide and use Earth Observation data and services.

This tutorial can be found directly at the GEOSS Best Practices Wiki (BPW) (http://wiki.ieee-earth.org). It can also be found indirectly via a search at the Geo Web Portal (GWP) (http://www.geoportal.org).

Table of Contents:

1. GEOSS Introduction (SIF)
   1.1. Overview of GEOSS
   1.2. Explanation of tutorials, in general
2. Introduction to the topic
   2.1. Discussion of tutorial use (SIF)

5.12. Finding datasets which meet user-specified quality thresholds

3.2 GeoViQua tutorial for Producer Quality in ISO, User feedback and GeoLabel

This tutorial was shown in the INSPIRE Conference 2013 in Florence explaining how to use the Producer Quality Model, the User Quality Model and the GEO label. http://tutorial.geoviqua.org/. The tutorial is not just a collection of documents to read but contains interactive exercises allowing readers to create metadata documents in a GeoNetwork catalogue, introduce new feedback items in the feedback database and create their own GEO labels based on their own metadata.

The Producer Quality Model extends ISO 19115 and 19157 standards to allow traditional metadata documents to be supplemented with richer information such as citations, discovered issues, information about reference datasets and full statistical representations of uncertainty.

The User Quality Model permits users of datasets to submit ratings, comments, citations and assessments of those datasets to a Feedback server where their comments can be collated and combined with the more traditional metadata to help other users assess the data's fitness for purpose.

The proposed GEO label presents a condensed visual summary of the producer and use metadata, allowing a quick assessment of the availability of data quality information on a dataset, as well as a drill-down feature so that users can query the quality information in more detail.

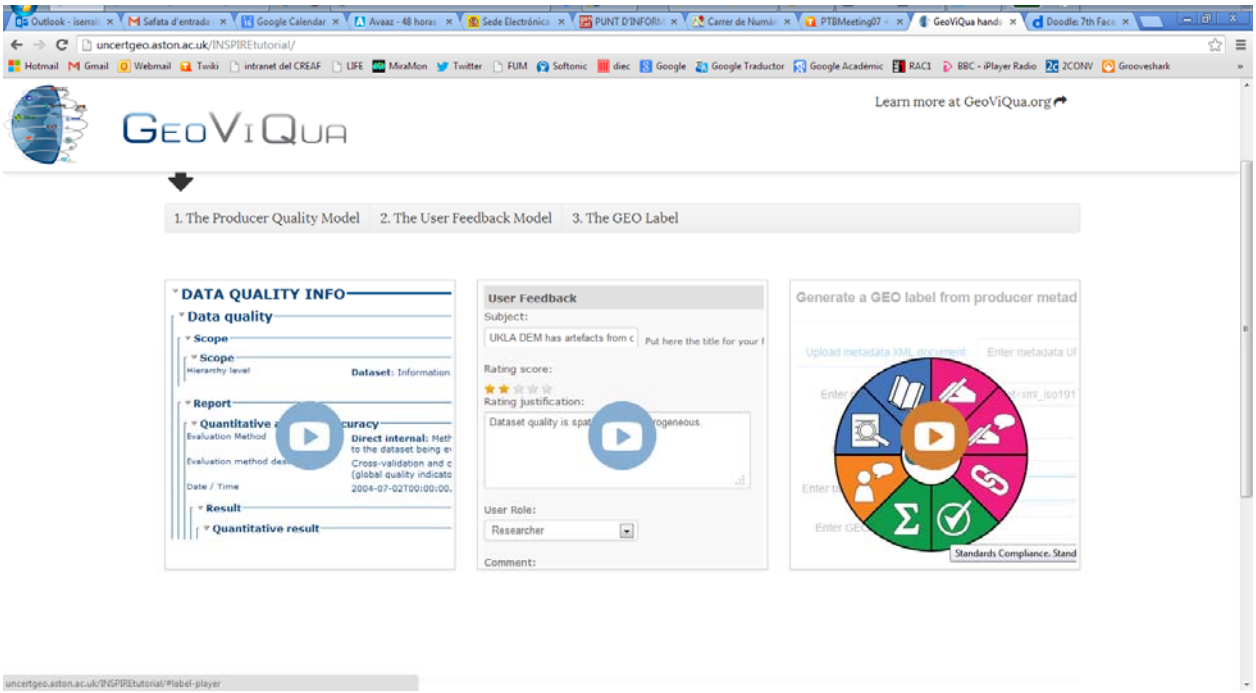


Figure 10 GeoViQua tutorial

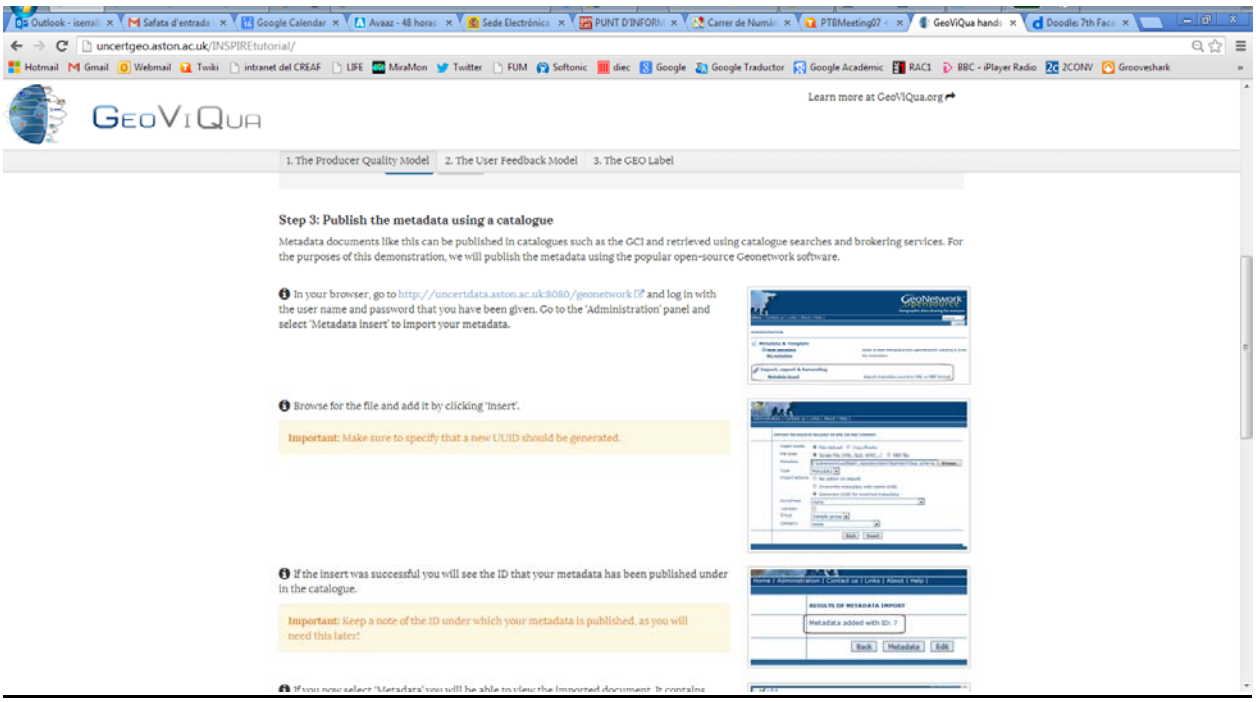


Figure 11 Publication of a metadata record in the modified GEONetwork catalogue

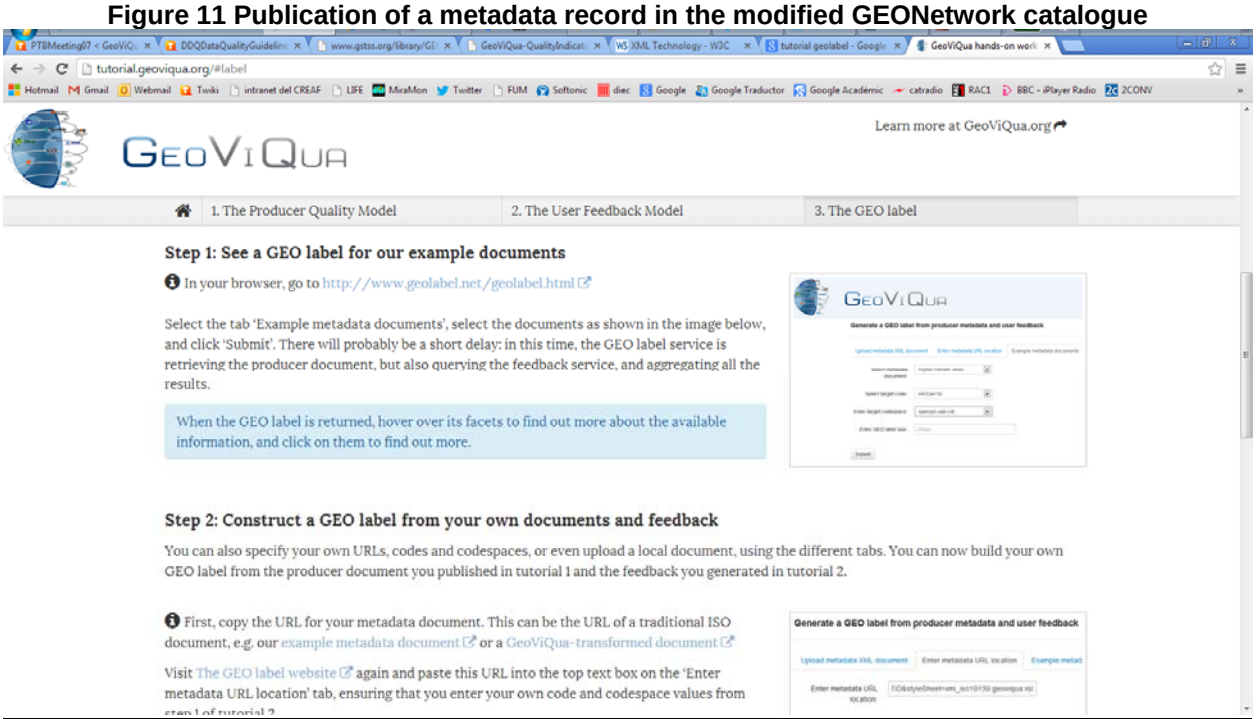


Figure 12 GEO label generation

3.3 Data quality guidelines (Documentation and Data Quality (DDQ) subgroup)

These Guidelines are offered by the GEO Data Sharing Working Group (DSWG) to foster understanding of the importance of data quality information associated with

GEOSS data resources. It is essential for users to understand the quality of data sets and to combine this quality information with other metadata components in order to determine the appropriateness, or fitness, of these data sets for the users' applications and/or purposes. Thorough documentation of data characteristics and data quality will support a wider scope of use than the purpose for which the data resources were originally acquired.

The DDQ Subgroup works towards the establishment of universal principles for optimal data quality management and utilization, including: definition of concepts, tools and systems, and capacity building.

The DDQ Subgroup is committed to expanding, maximizing and providing comprehensive documentation of data quality, including traceability, provenance, uncertainty, and fit for purpose to allow a broader use and understanding of Earth observations.

The DDQ Subgroup will coordinate with GEO entities in related areas of data assurance and data quality such as GEO Infrastructure and Implementation Groups.

A set of GEOSS Data Quality Guidelines were developed with GeoViQua collaboration to foster understanding of the importance of data quality information associated with GEOSS data resources. It is essential for users to understand the quality of data sets and to combine this quality information with other metadata components in order to determine the appropriateness, or fitness, of these data sets for the users' applications and/or purposes. Thorough documentation of data characteristics and data quality will support a wider scope of use than the purpose for which the data resources were originally acquired.

Core metadata and quality tags for fitness-for-purpose assessment include:

- Coverage, including both spatial and temporal dimensions
- Consistency, including long-term consistency
- Uncertainties estimated and documented, including both spatial and temporal dimensions
- Attribution of error sources
- Validation information, i.e. how the data was assessed for uncertainties by comparison with alternative measurements
- Latency from time of observation
- Resolution, including both spatial and temporal dimensions
- Usability, in the sense of being in form and content convenient to use for the specific purpose
- Simplicity, such that the data and the underlying data model are not unnecessarily complicated or difficult to understand or manipulate.

These guidelines can be downloaded from:
http://www.earthobservations.org/documents/dsp/GEOSS_Data_Quality_Guidelines.pdf

3.4 GEOSS Data Citation Guidelines

The GEOSS Data Citation Guidelines were developed by GEO Tasks ST-09-02/ID-03 based on the international discussion and consensus on data citation and give guidance on how to cite a data set.

The GEOSS Data Citation Guidelines might be implemented in the CGI. As a first step, a testbed for data citation will be created. Currently, users of the GEO:Portal are not obliged or encouraged to cite data accessed through GEOSS – if at all, citation requirements come from the individual data providers. The testbed implementation of the GEOSS Data Citation Guidelines will rectify this situation, increase the attractiveness of GEO and GEOSS for scientists by making their contributions visibly acknowledged, and help to identify issues not covered by the Guidelines.

The elements of a data citation referring to quality[4] are:

Publication Date: Whichever is the later of the date the dataset was made available, the date all quality assurance procedures were completed, and the date the embargo period (if applicable) expired.

Publisher: The organisation either hosting the data or performing quality assurance.

Publication Date: Whichever is the later of: the date the dataset was made available, the date all quality assurance.

Publisher Organization: The organization either hosting the data or performing quality assurance. A publisher often has an implied responsibility for stewardship of the data set. This is usually a data centre.

GeoViQua adopted some of these concepts in the Producer Quality model by extending CI_Citation into the GVQ_Publication.

4. Additions to the Standards and Interoperability Registry

The following records were not present in the SIR and have been added to it as a result of this review:

4.1 W3C Prov

URN: urn:geoss-uuid:bc46d5f6cbdacfa06c0deb2e6cf6afab URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=325

[4] GEOSS Data Citation Guidelines. Version 2.0: http://www.gstss.org/library/GEOSS_Data_Citation_Guidelines_V2.0.pdf

Description:
Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.

PROV is a specification that provides a vocabulary to interchange provenance information. Users can do so by marking up their web page using the terms provided or by making available provenance information expressed as linked data.

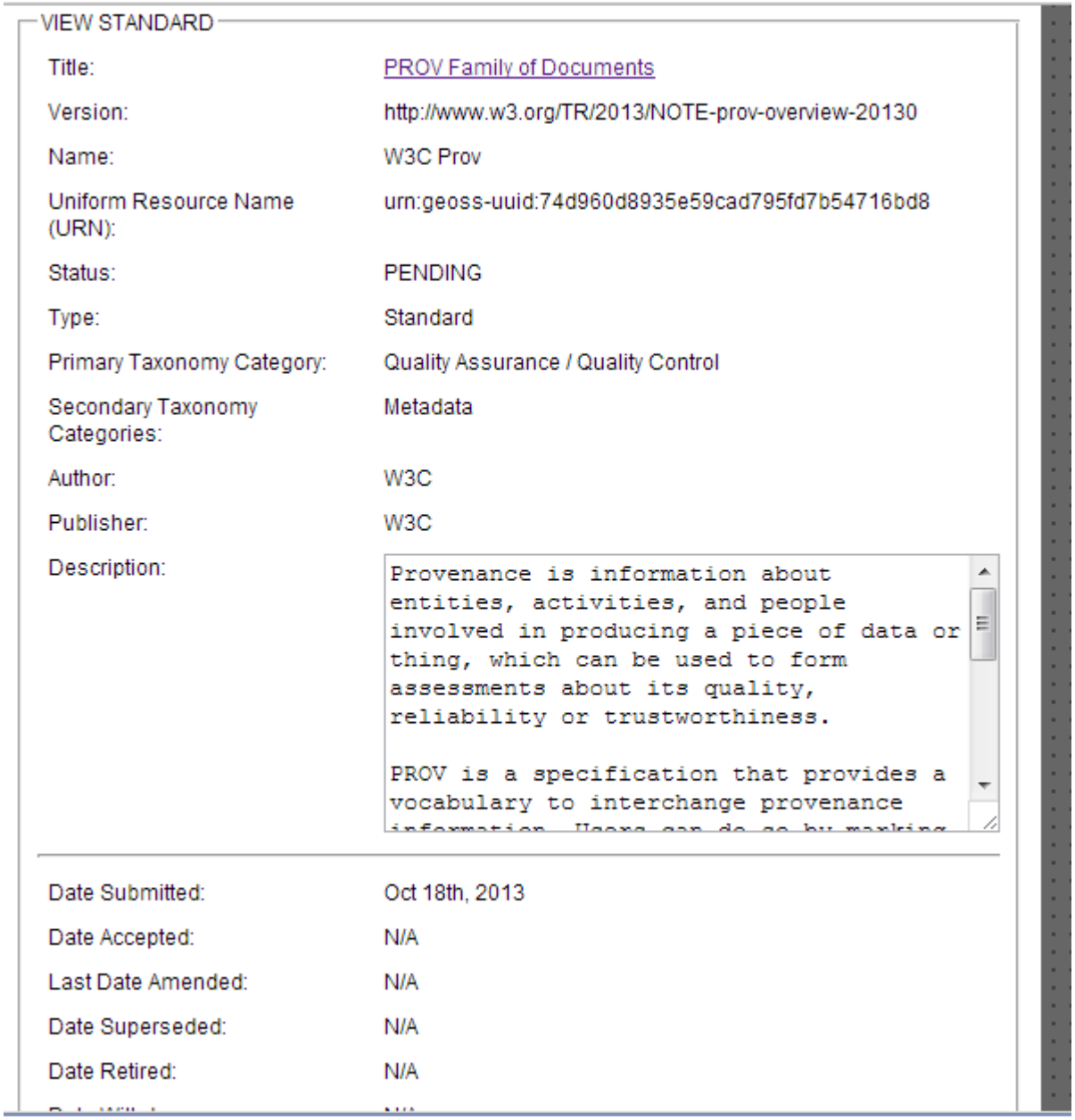


Figure 13 Screenshot of the SIF GEOSS record for W3C Prov

The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance

information in heterogeneous environments such as the Web. This document provides an overview of this family of documents.

4.2 ISO/FDIS 19157 Geographic information — Data quality

URN: urn:geoss-uuid:bc46d5f6cbdacfa06c0deb2e6cf6afab
URL: http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=325

Description:

This International Standard recognizes that a data producer and a data user may view data quality from different perspectives. Conformance quality levels can be set using the data producer's product specification or a data user's data quality requirements. If the data user requires more data quality information than that provided by the data producer, the data user can follow the data producer's data quality evaluation process flow to get the additional information. In this case the data user requirements are treated as a product specification for the purpose of using the data producer process flow.

The objective of this International Standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a dataset's quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.

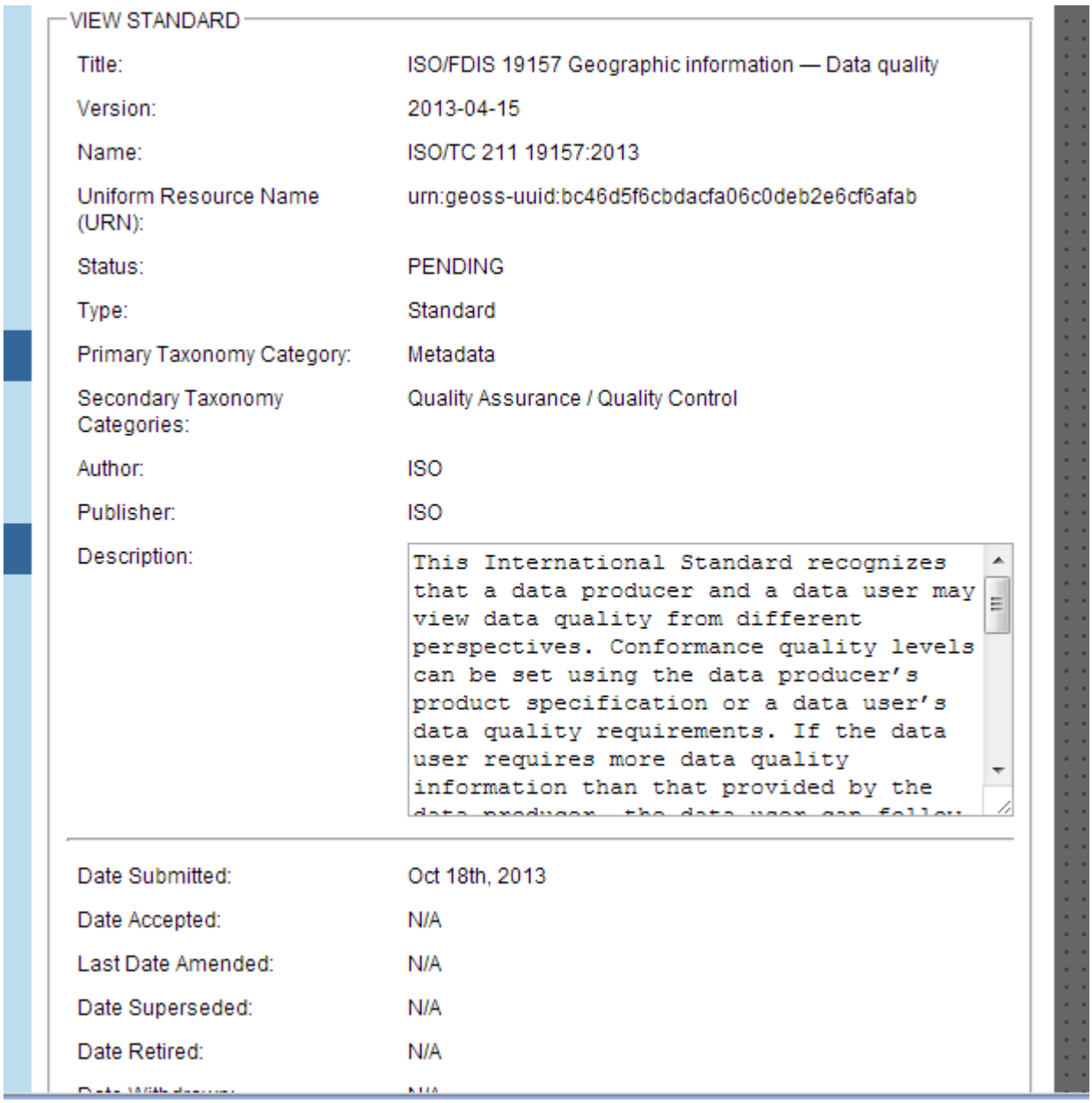


Figure 14 Screenshot of the SIF GEOSS record for ISO 19157