

A COMPARATIVE STUDY OF SPATIAL DATA HARVEST STANDARDS¹

Yunxuan Yan^a, Guoqing Li^a, Jibo Xie^a, Lixia Guo^a

^aCenter for Earth Observation and Digital Earth, Chinese Academy of Sciences
Beijing, China, 100086
yxyan@ceode.ac.cn

KEY WORDS: Services, Internet/Web, Comparison, Interoperability, Metadata, Standards

ABSTRACT:

So far, thousand kinds of Spatial Data were acquired, produced, published and used for specialized applications. Heterogeneity is obviously one key problem for spatial data in the process of its creating, structuring or managing. The heterogeneity keeps the complete and variety of spatial data, but induces the difficulty of describing, publishing, sharing, interoperability or metadata harvesting. Decentralized, multi-organization and synchronization are the major features of the next generation spatial data service infrastructure. Standard research on geospatial metadata harvesting and data service is becoming a research focus. More and more organizations and consortiums (e.g. ISO/TC 211, OGC) drafted and released their geospatial metadata and service standards. However, these standards are diverse in order to fit to different spatial data format and research fields, which makes spatial data owner hard to choose standard. The paper focuses on analyzing metadata harvesting from multi-organizations archives, including Spatial Data description, publishing, sharing, and interoperability. Several of most widely used Spatial Data standards are introduced, and compared. Advices on how to choose the standards are given for Spatial Data owner.

1. INTRODUCTION

As the central issue of Geographic Information Science, spatial data harvesting has received considerable attention in recent years. Numerous GIS research projects and initiatives have addressed issues of format and representational heterogeneity across spatial datasets, spatial metadata supporting data integration, spatial data interchange standards, etc. Produced by different pre-processing system and to meet varieties of application requirement, spatial data format are diverse and heterogeneous. At the same time, spatial data processing is constituted by data acquisition, production, publishing, sharing, using, etc. The interfaces of the processing web service are distinct. Web Services from different system are tough to work together. It makes spatial data harvest impossible.

To bridge the gaps between different systems, organizations and committees, e.g. OGC, ISO and CEOS, draft a series of standards and specifications for spatial data and web service for interoperation. The interoperation difficulty has been solved to a certain extent.

However, as the spatial data and its processing are complicated, there are spatial data standards that have been release. The relationships among various standards are very complex. And every standard is applicable to specific ranges. Moreover, equipotent standards drafted by different organizations or committees are heterogeneous. Spatial data producer and publisher are confused about to choose which standards for their data and processing web service. The paper aims to solve this confusion. And spatial data harvest discussed in the paper includes metadata acquisition, data description and expressing, and interoperation with other spatial data system.

2. STANDARDS SERIES

2.1 ISO/TC 211

ISO/TC 211⁰ is a standard technical committee formed within ISO, tasked with covering the areas of digital geographic information (such as used by geographic information systems) and geo-informatics. It is responsible for preparation of a series of International Standards and Technical Specifications numbered in the range starting at 19101. Project specification areas in the ISO/TC 211 technical committed include: simple features access, reference models, spatial and temporal schemas, location-based services, metadata, web feature and map services and classification systems. Till now, there are 53 standards on geographic information have been published and 27 on the way. Many of them specify basic feature model or schema, e.g. reference model, spatial and temporal feature schema. Some others specify metadata and services including definition, description, acquiring, processing, analyzing, accessing, presenting and transferring. The rest standards aim at methodology or infrastructure. The paper will discuss the second category.

2.2 OGC

The Open Geospatial Consortium (OGC)⁰ is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services. It composes of 424 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC Standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services and

¹ This work is supported by "Digital Earth Science Platform project" --Informatics study of Digital Earth Systems studies (Y01011101A), The National Basic Research Program of China (973 Program 2009CB723906).

mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications. Ideally, when OGC standards are implemented in products or online services by two different software engineers working independently, the resulting components plug and play, that is, they work together without further debugging⁰. Till now, 52 standards have been published. The OGC Standards Development Process creates Abstract and Implementation specifications⁰. The purpose of the Abstract Specification is to create and document a conceptual model to support the creation of Implementation Specifications. Implementation Specifications are unambiguous technology platform specifications for implementation of industry-standard, software application programming interfaces. Geospatial domain semantics defined in the Abstract Specifications are to be consistent across multiple technology platforms as defined in Implementation Specifications.

2.3 FDGC

The Federal Geographic Data Committee (FGDC) is a United States government committee which promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis⁰. It was created in 1990 and tasked to develop geospatial data standards that would enable sharing of spatial data among producers and users and support the growing National Spatial Data Infrastructure (NSDI). Acting under the Office of Management and Budget (OMB) Circular A-16, and the 1994 Executive Order #12906 creating the US NSDI, FGDC subcommittees and working groups, in consultation and cooperation with state, local, tribal, private, academic, and international communities, develop standards for the content, quality, and transferability of geospatial data. FGDC standards are supported by the vendor community but are independent of specific technologies so they may evolve as technology and institutional requirements change. Most importantly for many stakeholders, FGDC standards are publicly available, typically for free via download from FGDC's Web site⁰⁰.

2.4 CEN

The European Committee for Standardization or Comité Européen de Normalisation (CEN), is a non-profit organisation whose mission is to foster the European economy in global trading, the welfare of European citizens and the environment by providing an efficient infrastructure to interested parties for the development, maintenance and distribution of coherent sets of standards and specifications. The CEN was founded in 1961. Its thirty national members work together to develop European Standards (ENs) in various sectors to build a European internal market for goods and services and to position Europe in the global economy. Some of these standards are voluntary, whereas other standards such as harmonized standards have been made effectively mandatory under EU law. CEN is officially recognised as a European standards body by the European Union; the other official European standards bodies are the European Committee for Electro technical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI). The Vienna Agreement was signed by CEN and the International Organization for Standardization (ISO) in 1991 but came in force in the mid-2000s. Its primary aim is to avoid duplication of (potentially conflicting) standards between CEN and ISO. In the last decade CEN has adopted a number of ISO standards which replaced the corresponding CEN standards.⁰

2.5 Others

OpenSearch⁰: OpenSearch is a collection of technologies that allow publishing of search results in a format suitable for syndication and aggregation. It is a way for websites and search engines to publish search results in a standard and accessible format. OpenSearch was developed by Amazon.com subsidiary A9 and the first version, OpenSearch 1.0, was unveiled by Jeff Bezos at the O'Reilly Emerging Technology Conference in March, 2005. Draft versions of OpenSearch 1.1 were released during September and December 2005. OpenSearch consists of:

- OpenSearch Description files: XML files that identify and describe a search engine;
- OpenSearch Query Syntax: describe where to retrieve the search results;
- OpenSearch RSS (in OpenSearch 1.0) or OpenSearch Response (in OpenSearch 1.1): format for providing open search results;
- OpenSearch Aggregators: Sites that can display OpenSearch results.

OpenSearch "Auto-discovery" to signal the presence of a search plug-in link to the user and the link embedded in the header of HTML pages. OpenSearch is widely used in variety of application. And its Geospatial Extension has been discussed by OGC. A draft version was release in October, 2009⁰.

KML⁰⁰: Keyhole Markup Language (KML) is an XML notation for expressing geographic annotation and visualization within Internet-based, two-dimensional maps and three-dimensional Earth browsers. KML was developed for use with Google Earth, which was originally named Keyhole Earth Viewer. It was created by Keyhole, Inc, which was acquired by Google in 2004. KML is an international standard of the Open Geospatial Consortium. Google Earth was the first program able to view and graphically edit KML files. The KML file specifies a set of features (place marks, images, polygons, 3D models, textual descriptions, etc.) for display in Google Earth, Maps and Mobile, or any other 3D Earth browser (geo-browser) implementing the KML encoding. Each place always has a longitude and latitude. Other data can make the view more specific, such as tilt, heading, altitude, which together define a "camera view". KML shares some of the same structural grammar as GML. Some KML information cannot be viewed in Google Maps or Mobile. KML is complementary to most of the key existing OGC standards including GML (Geography Markup Language), WFS (Web Feature Service) and WMS (Web Map Service). Currently, KML 2.2 utilizes certain geometry elements derived from GML 2.1.2. These elements include point, line string, linear ring, and polygon. The OGC and Google have agreed that there can be additional harmonization of KML with GML (e.g. to use the same geometry representation) in the future.

3. RELATIONSHIPS

The ISO/TC 211 work is closely related to the efforts of the OGC, and the two organizations have a working arrangement that often results in identical or nearly-identical standards often being adopted by both organizations. And many ISO191xx standards are identical standards in OGC.

In Europe the prospective standards (ENVs) and CEN reports (CRs) have been published on many occasions, but then the ISO Geo-Standards came along. This resulted in there being various geographic standards available across Europe for the same topic.

NEN/ENV 12657 and ISO19115 are such pair. At the CEN/TC 287 meeting held in Delft on November 2003, an important decision was made about the existing ENVs and CRs: all of them were taken back so as to secure the implementation and harmonisation of ISO 191xx series Standards for Europe. CEN/TC 287 has a different status to ISO. ISO Standards are voluntary: countries can ignore them. It is also possible to have national Standards that conflict with ISO Standards. CEN Standards have to be published; it is mandatory. A country is not allowed to have any of its own standards conflicting with a CEN Standard. This means that conflicting national standards have to be withdrawn. Many European countries translate all CEN Standards into their own language. There are countries that insist on the use of CEN Standards within their government as well. The European Public Procurement Regulation makes it mandatory to refer to European Standards, if there are any, in the specifications for European Tenders.

4. COMPARE, ANALYSIS AND APPLICATION

Presently, spatial dataset brokers provide two approaches to access datasets. Many of them hold a FTP Server to share their data. Catalogue Server can cache the metadata of all data to harvest this spatial data, and discover the interesting among the cached metadata when user searches data, namely "asynchronous mode". Some others develop some Web Services for accessing their dataset. Catalogue Server can "synchronously" access the Web Services when discovering for user. It is a virtual harvest approach mode. This section tries to analyse the standards for each approach and the relationship between each standard used in the same place. Besides, Sensor Web as a new coming mission data accessing approach is analysed in this section as well.

4.1 Metadata and Data description

4.1.1 Metadata: ISO19115, ISO19139, CEN prENV 12657, FDGC, Dublin Core

Four key standards are considered to describe spatial data metadata in majority spatial data sharing systems or recommended in other standards.

ISO 19115 is fundamental. Majority international projects and standards quote the geographic element and attribute which it defines. ISO 19115 defines how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. The standard defines more than 400 metadata elements, 20 core elements;

ISO 19139 provides the XML implementation schema for ISO 19115 specifying the metadata record format and may be used to describe, validate, and exchange geospatial metadata prepared in XML;

CEN prENV 12657 discusses the same topic as ISO19115. At the CEN/TC 287 meeting held in Delft on November 2003, this standard is taken back for ISO19115 implementation in European;

The American initiative FGDC is a national standard for spatial metadata development for give support to the construction of the United States Spatial Data National Infrastructure. This

standard has been adopted in other countries like South Africa or Canada⁰. In July 2004, the FGDC is tasked to develop metadata content for the U.S. National Profile of ISO 19139⁰;

The Dublin Core set of metadata elements provides a small and fundamental group of text elements through which most resources can be described and catalogued. Using only 15 base text fields, a Dublin Core metadata record can describe physical resources. Metadata records based on Dublin Core are intended to be used for cross-domain information resource description and have become standard in the fields of library science and computer science. Implementations of Dublin Core typically make use of XML and are Resource Description Framework based. Dublin Core is defined by ISO through ISO Standard 15836. And ISO 15836 focuses on common resource description, while ISO 19115 defines metadata of geographic information.

4.1.2 Data description: ISO19118, ISO19136, GML, KML, ISO19131

The discrimination between metadata and data description is that data description focus not only metadata information but also data entity.

ISO 19118 specifies the requirements for defining encoding rules to be used for interchange of geographic data within the ISO 19100 series of International Standards;

ISO 19136 resulted from unification of the OGC definitions and Geography Markup Language (GML) with the ISO-191xx-Normen⁰;

The Geography Markup Language (GML) is an XML encoding in compliance with ISO 19118 for the transport and storage of geographic information modelled according to the conceptual modelling framework used in the ISO 19100 series and including both the spatial and non-spatial properties of geographic features⁰. And GML is widely data description standard used in OGC Web Service specification, e.g. Web Map Service(WMS), Web Feature Service (WFS) and Catalogue Service (CSW);

Keyhole Markup Language (KML), made popular by Google, complements GML. Whereas GML is a language to encode geographic content for any application, by describing a spectrum of application objects and their properties (e.g. bridges, roads, buoys, vehicles etc.), KML is a language for the visualization of geographic information tailored for Google Earth. KML can be used to carry GML content, and GML can be "styled" to KML for the purposes of presentation. KML instances may be transformed lossless to GML, however roughly 90% of GML's structures (such as, to name a few, metadata, coordinate reference systems, horizontal and vertical datum, etc.) cannot be transformed to KML;

ISO 19131 specifies requirements for the specification of geographic data products, based upon the concepts of other ISO 19100 International Standards.

4.2 Data Service

4.2.1 Data Service Model: ISO19119

ISO 19119 provides a framework for developers to create software that enables users to access and process geographic

data from a variety of sources across a generic computing interface within an open information technology environment. It is a basic standard of other data service specifications.

4.2.2 Expressing data: ISO19117, ISO19128, WMS

ISO 19117 concerns portraying geographic information as an image understandable by humans, including the methodology for describing symbols. The portrayal standard will provide applications with a common interface to supported standard symbol sets. Thus this standard does not include standardization of cartographic symbols but provides a standard interface for such standard symbol sets;

OGC Web Map Service (WMS) Implementation Specification standardizes the way in which clients request maps. Clients request maps from a WMS instance in terms of named layers and provide parameters such as the size of the returned map as well as the spatial reference system to be used in drawing the map⁰. And the OGC through its work with ISO announces that its OpenGIS Web Map Service (WMS) Implementation Specification is now available as ISO 19128.

4.2.3 Searching data: CSW, OGC-Filter, ISO19143, OpenSearch

CSW is widely used as resource discovering and harvesting standards. From 2007 to 2009, many spatial data organizations published their CSW for searching their data resource. In CSW, besides query and harvest resource interfaces, the most important query language support is the OGC Filter Specification. Filter, identities as ISO 19143 in ISO 191xx series standards, is very easily to be transformed into SQL WHERE clause in the SQL SELECT statement. So most catalogue services compliant with OGC-CSW are asynchronous searching, basing on resource cache databases. This mechanism is not suitable for huge resource harvest, and difficult to synchronous with data resource. Therefore, many runtime spatial data agencies did not adopt the CSW for sharing their archiving and real time datasets.

OpenSearch was developed by Amazon.com subsidiary A9. OpenSearch consist of four components: XML files to identify and describe the search engine; Query syntax to describe how to retrieve the search results; OpenSearch response in several kinds of format, e.g. RSS, Atom and XML, for providing open search results; Sites that can display OpenSearch results. For its well self-description of search engine and query way, OpenSearch is soon popular as a web auto-discovery implementation. In April 2010, NASA ECHO released their OpenSearch Web Services for users to access the Archived MODIS data. GENESI-DEC adopted OpenSearch to sharing their datasets. Users can access majority of datasets that GENESI-DEC harvested through a recursive query. Then OpenSearch Geospatial Extension was discussed by OGC. And the mechanism of OpenSearch is suitable for a recursive query, which can easily harvest a huge number of spatial data and low-cost to synchronous with datasets. OpenSearch is next generation distributed discovery standards. And OGC is considering absorbing it into CSW implementation Standards.

4.2.4 Retrieving data: ISO19125, OGC-SFA, ISO19142, WFS, WCS, SOS, ISO19125 defines how to access simple feature . It has been separated into two parts. Part 1 establishes

a common architecture for geographic information and defines terms to use within the architecture. Part 2 specifies an SQL schema that supports storage, retrieval, query and update of simple geospatial feature collections via the SQL Call Level Interface (SQL/CLI) and establishes architecture for the implementation of feature tables. ISO 19125 is a joint standard with OGC, namely SFA (Simple Feature Access).

ISO 19142⁰ specifies the behaviour of a web feature service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions. The standard is under development. ISO 19142 standard in OGC architecture is called WFS (Web Feature Service). WFS interface is defined for accessing geographic feature. Majority spatial data owners use it to publish their Vector data type.

WCS (Web Coverage Service) supports the networked interchange of geospatial data as "coverages" containing values or properties of geographic locations. The Web Coverage Service provides access to intact (unrendered) geospatial information, as needed for client-side rendering, multi-valued coverages, and input into scientific models and other clients beyond simple viewers. WCS did not concern about the content of the data. Many Raster data owners use it to publish their data. Moreover, WCS GetCoverage interface provides several parameters, e.g. Format, Bounding Box, and Store, for user to request the server to process data. Therefore, WCS Standard can be applied to build spatial data service on demand. SOS is a specification to access observation data from Remote Sensing platform. And it will be discussed in Section

4.3 Sensor Web

With sensors of all types becoming part of the global information infrastructure, the OGC has approved four Standards and several Best Practices designed to enable sensors to better interoperate with the Web and other information technology assets. The OGC Sensor Web Enablement (SWE) is a set of interfaces and protocols that enable a "Sensor Web" through which applications and services will be able to access sensors of all types over the Web. Foundational components for Sensor Web Enablement have defined, prototyped and tested:

- Observations & Measurements (O&M)⁰: defines an XML implementation of schemas for observations, and for features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities.
- Sensor Model Language (SensorML)⁰: specifies models and XML encoding for the core SensorML, as well as the definition of several SWE Common data components utilized throughout the SWE framework. The primary focus of SensorML is to define processes and processing components associated with the measurement and post-measurement transformation of observations.
- Transducer Markup Language (TML)⁰: describes TML and how it captures necessary information to both understand and process transducer data. TML is intended for communicating transducer data between a transducer node (containing one or more transducers) and a transducer

- processing/control device (application).
- Sensor Observation Service (SOS)⁰: Implementation Specification defines a web service interface for requesting, filtering, and retrieving observations and sensor system information. Observations may be from in-situ sensors (e.g., water monitoring devices) or dynamic sensors (e.g., imagers on Earth-observation satellites).
- Sensor Planning Service (SPS)⁰: Implementation Specification defines an interface to task sensors or models. Using SPS, sensors can be reprogrammed or calibrated, sensor missions can be started or changed, simulation models executed and controlled. The feasibility of a tasking request can be checked and alternatives may be provided.
- Sensor Alert Service (SAS): defines a web service interface for publishing and subscribing to alerts from sensors. Sensor nodes advertise with an SAS. If an event occurs the node will send it to the SAS via the publish operation. A consumer (interested party) may subscribe to events disseminated by the SAS. If an event occurs the SAS will alert all clients subscribed to this event type.
- Web Notification Service (WNS): Standard web service interface for asynchronous delivery of messages or alerts from SAS and SPS web services and other elements of service workflows.

closer and closer to them. The ISO/TC 211 work is closely related to the efforts of the Open Geospatial Consortium, and the two organizations have a working arrangement that often results in identical or nearly-identical standards often being adopted by both organizations. At the same time, OGC keeps absorbing more and more well-used standards, e.g. OpenSearch and KML, into its group. And for developers of Spatial Data, users, owners and publishers of Spatial Data sharing systems, OGC provides majority of standards.

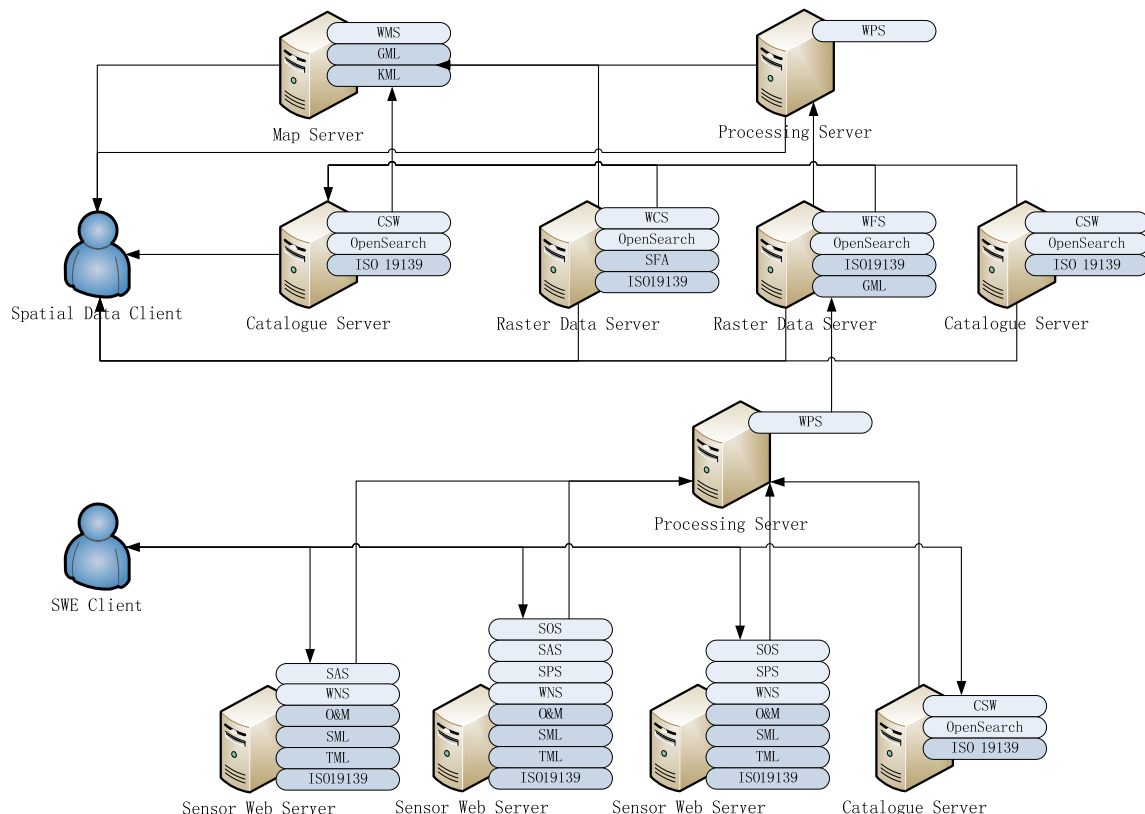
The ISO 191xx/TC211 series standards are exhaustive. Many of them are basic standards, e.g. geographic location (ISO 6709), reference model (ISO 19101), and profiles (ISO 10106). These standards are used by other application standards. Even application standards are suit for different sectors. Table 1 shows whom Standards discussed in the paper are of particular relevance to.

A number of OGC Standards are ISO ones at the same time. The OGC through its work with ISO announces that its OpenGIS Web Map Service (WMS) Implementation Specification is now available as ISO 19128; ISO 19125 is a joint standard with OGC, namely SFA (Simple Feature Access); ISO 19142 standard in OGC architecture is called WFS (Web Feature Service). When user adopts a OGC Standard, he/she maybe choose a ISO standard at the same time.

5. COMCLUSION

The most influence Spatial Data Harvest Standards Committees are ISO/TC211 and OGC. Others are either replaced or move

OGC Standards includes many similar Specifications. But each specification has its particular purpose. The following picture shows the standards that every component in web spatial data platform possibly uses.



Picture 1. OGC Standards used in Spatial Data Sharing Platform

