Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data Infrastructure

background paper for NIEeS Workshop on Activating Metadata, 6-7 July 2005, Cambridge, UK

by Roger A Longhorn, Director, Info-Dynamics Research Associates Ltd, UK
EC Projects Office, 1A Potters Cross, Wootton, Bedfordshire MK43 9JG
e-mail: ral@alum.mit.edu or ral@i-dra.org

Introduction.

The main standards now being implemented in the geospatial world emanate from the International Organization for Standardization (ISO) Technical Committee 211 - Geomatics/Geographic Information (generally referred to as 'TC 211'), in their 19xxx series of standards. Because some EU Member States have legal requirements that 'European standards' be used for various types of work, if they exist, the European standards organisation CEN is creating 'European profiles' of ISO standards in the 19xxx series. Prior to this, CEN had developed a series of geomatics related standards in their Technical Committee TC 287, but these were not progressed beyond pre-standards (preENs) before TC 287 stopped work, as by that time the ISO TC 211 work has begun, and much of the TC 287 work was subsumed into the ISO work. In the USA, the Federal Geographic Data Committee (FGDC) created its geospatial standards in the mid-1990s which became the basis for the metadata creation work and clearinghouse development mandated by Presidential Executive Order in April 1994 creating the USA NSDI[1]. Much of the FGDC standardisation work was also subsumed into the ISO TC 211 work from 1998 onwards.

Note: The annexes to the main report contains hyperlinks to most of the standards mentioned in this document, although not all may be readily downloadable without payment of a fee, e.g. most ISO standards are not available for free, nor are some CEN standards.

CEN (Europe).

The European standards organisation CEN has a policy that no new regional European standards will be produced if an international standard (ISO) already exists. Thus the CEN TC 287 work focuses mainly on creating European "profiles" for existing ISO standards from ISO TC 211. See Annex A for an update on how far this profiling work has progressed as of May 2005. Having ceased operation in 1999, CEN/TC287 was rejuvenated in late 2003 to bring ISO work into the European arena, as CEN standards are compulsory in many European countries, whereas ISO standards are not. Topics discussed at the TC287 October 2004 meeting included:

- Procedural issues on how to make the ISO work and outputs compatible with CEN's work, i.e. how to overcome administrative complexities, as the ISO work moves on and CEN work struggles to catch up (because it started much later). Many of the published ISO standards became EN standards by the end of 2004 (see Annex B), and ISO work in its earlier stages will become joint with CEN in 2005 and into 2006.
- The work of CEN TC287 Working Group 1 on standards for spatial data infrastructures is examining the standards required for SDIs (with a focus on INSPIRE). WG1 is considering whether full standards are required or simply profiles of ISO standards. Where the former is the case, CEN will request that ISO picks up the necessary work; otherwise CEN will publish and mandate the profiles as ENs. Practical implementation and use guides are also being considered.
- Whether a European conceptual schema language is needed (or desirable). This was left for further reflection – it is an issue of whether GML provides a sufficiently specific language or not.

The latest CEN TC287 meeting took place in Sweden in conjunction with the ISO meetings in June 2005 (reports from that meeting are still awaited) and the next meeting will be in early 2006, in Dublin.

[1] The core USA standard was the Content Standard for Digital Geospatial Metadata - CSDGM - (FGDC-STD-001-1998).

Federal Geographic Data Committee (FGDC) (USA).

The Federal Geographic Data Committee 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 site.

FGDC is covered in this report mainly because of the thematic profiling work that has been undertaken by the committee (see Annex A), much of which remains to be carried out in regard to ISO (and thus CEN) geospatial data standards. Profiling is necessary to fully engage different user communities to 'buy in' to what are otherwise typically generic geospatial data standards. The second reason for looking at FGDC's work is that these standards were adopted by many geospatial initiatives across the globe and are still in use today, for example in the UK's NGDF (until it is converted to the ISO 19115 metadata standard later in 2005) and in the geospatial standards regimes of Canada, Australia/NZ and several other countries.

ISO TC 211 (International Organization for Standardization).

At the end of 1999, there were 12,524 ISO standards in print, totalling 356,427 pages of text! Standards relating to geospatial information are thus only a very small part of the ISO regime.
ISO Technical Committee TC 211 (Geographic Information / Geomatics) began work in earnest in 1995 and various working groups began to publish a variety of standards and/or technical specifications relating to different aspects of geospatial data, tools and services as early as 1996, often in close cooperation with other international bodies, such as FIG, the International Federation of Surveyors[2], national bodies such as FGDC and regional bodies such as CEN TC 287. As can be seen from the list of topics shown in Annex C for which TC 211 has been or is developing international geomatics standards (the 19xxx series), the committee has been exceptionally busy ever since, and new work items appear annually. As mentioned earlier, much of the work of the FGDC in the USA and CEN TC 287 was subsumed into the work of ISO TC 211, within which working groups are often led by the same persons who supervised development of the FGDC and CEN TC 287 standards.

It is important to remember that ISO standards - and most other 'formal' standards - are normative specifications which, by their very nature, are technical reference documents and are not written to introduce non-experts to the complex field of standards implementation. Because of this, there is an on-going need to develop user-oriented educational material to make geospatial standards more accessible to users, ideally targeted to thematic user groups where the standard can be presented in a format applicable to the needs of such groups.

A third issue relating to ISO's geospatial standards is that, as they become the established global norm, modified perhaps at regional (European) level mainly by profiling, the need arises to provide 'cross-walks' between existing (national, thematic) geospatial standards and the international standards. This is not always straightforward as there is never a convenient one-to-one correspondence between an existing standard and a new one. Establishing cross-walks or links between standards is considered in a later sub-section of the report.

Not all ISO standards of importance to the geospatial community arise from work of TC 211. As well as generic standards for country names, time representations, etc., other standards in effect for a number of years are also embedded in the new standards - and some of these have now been found

[2] ISDI is fortunate in having a member of the Advisory Board who is closely associated with geospatial standards in the person of Iain Greenway of OSi, also chairman of the FIG Standards Network.

wanting. Both ISO 6709 (Standard representation of latitude, longitude and altitude for geographic point locations) and ISO/IEC 11179 (Specification and Standardization of Data Elements) are focus of new work items for ISO TC 211 in 2005 as updates or extensions to these standards are needed. Also, as mentioned below in relation to the eGovernment standards arena, ISO 15836:2003 (Dublin Core Metadata Elements) is a key discovery metadata standard for which a mapping between its elements and the 'core' metadata elements of ISO 19115 has been published.

Open Geospatial Consortium, Inc. (OGC).

The Open Geospatial Consortium, Inc. (OGC), formed in 1994, with affiliates in Europe and AustralAsia, is a member-driven, not-for-profit, international geospatial industry trade association which is developing computing standards for the geospatial and location based markets. Using an open, consensus-focused process, the OGC works with government, private industry, and academia to deliver publicly available, royalty-free interface and geospatial encoding standards for use in GIS, location services, sensor webs, portals, and other related mainstream technologies and applications. OGC's mission is "to promote the open development and use of advanced open systems standards and techniques in the area of geo-processing and related information technologies". This work is conducted mainly by volunteers from member organizations, with staff and support infrastructure for Consortium activities paid for from membership fees.

OGC produces "specifications", i.e. documents that detail engineering aspects and rules for implementing interfaces or encoding that solves a specific geospatial interoperability problem. See Annex D for a list of the main specifications which have already been adopted or are under consideration by OGC working groups. The documents are defined, discussed, tested, and approved by OGC members using a formal process incorporating SIGs (special interest groups), working groups, requests for proposals, requests for comment, Interoperability Experiments, etc. Information about currently approved Open GIS Specifications can be found at the OGC web site[3]. The specifications are freely and publicly available. OGC makes every effort to ensure that no specification has IPR restrictions.

It is not (usually) necessary for end users of geospatial products or services be familiar with the work of OGC *per se*, as the technical specifications developed by the organization, once tested, verified and certified, are then embedded into the software tools provided by the vendor members, e.g. for creating Web Feature Services, Web Map Services, Catalogue Services, etc. Such toolsets then typically earn and display the "OGC compliant" tag. Of course, it is useful that users, and especially geospatial data publishing units, are fully aware of what specifications are being developed, how they might impact on their geospatial use and publishing remits, when the new functionality/ies is/are likely to become available via vendor toolsets, etc. Also, infrastructure oversight or advisory committees, such as the Irish SDI Work Group or a future ISDI Coordination Committee, also need to be kept fully up to date on work of OGC, as these specifications are seen as crucial tools for implementing interoperable geospatial data infrastructures.

There is a more direct link between OGC and ISO TC 211 (and thus CEN TC 287) in relation specifically to metadata. One of the major specifications to come from OGC deals with the XML (eXtensible Markup Language) implementation of GML (Geography Markup Language). OGC enjoys a Class "A" liaison with ISO TC 211 under which certain OGC specifications eventually become ISO standards. Such is the case for GML which is in the process of becoming ISO 19136 and ISO 19128, the proposed ISO standard for Web Map services, based on OGC's Web Map Server specification (see Annexes C and D).

Role of ISO TC 211, CEN TC 287 and OGC in INSPIRE.

The EU draft Directive INSPIRE - Infrastructure for Spatial Information in Europe - envisions a pan-European SDI (ESDI) comprising interlinking national and/or regional SDIs within and between EU Member States. This interlinking will be achieved only if the individual SDIs which will comprise the

[3] see www.opengis.org

ESDI can communicate via a common global infrastructure. For purposes of geospatial data discovery, (eventual) access and use (exploitation), standards become of paramount importance. The INSPIRE draft Directive is a legislative document stating that "international standards" will be used to record and publish metadata about geospatial data holdings at all levels of government, from local to national. The actual standard to be used is ISO 19115, which is being confirmed by the newly formed INSPIRE Drafting Team on metadata which is expected to begin work in September 2005. The other ISO standard of most relevance to INSPIRE metadata is ISO 19139 (Geographic information - Metadata - Implementation specifications), which has yet to be adopted by ISO and which covers XML schema standards for representing 19115 metadata. It is through such XML schema that interoperability tools compliant with OGC technical specifications will permit ever easier linking of disparate geospatial datasets from multiple sources in innovative ways and with considerably less effort (and cost) than is required today in 'hard coding' such dataset linkages.

CEN TC 287 is involved in the INSPIRE standards-related activities, since it is completing the European 'profiling' of the ISO standards, including ISO 19115.

Other e-Government Standards Relevant to Spatial Data Infrastructure.

The Dublin Core Metadata Initiative (DCMI) produced the Dublin Core (DC), a consensus derived, discovery level metadata standard now issued by ISO as ISO 15836 and already adopted by many national and regional (transnational) eGovernment initiatives. Cross-walks have been prepared between Dublin Core and ISO 19115, with FGDC, and with other metadata standards. In CEN, this was accomplished via a CEN Workshop Agreement (CWA), a form of pre-standard or publicly available specification (PAS) issued under CEN's ISSS (Information Society Standardisation System) by expert working groups, in this case the CEN MMI-DC (Multimedia Information - Dublin Core) W.G., as CWA 14857 - Mapping between Dublin Core and ISO 19115[4].

CWAs are official CEN publications. Copies can be obtained from CEN member organizations (European national standards bodies). In addition, the CEN Members have agreed that certain CWAs may be provided on this web-site for electronic downloading free of charge. The CWAs concerned are produced by Workshops whose volunteer expert participants have concluded a special arrangement, or they relate to the eEurope standardization initiative of the European Commission. CWAs are CEN copyright. Typically they can be freely downloaded from the Web on the condition that they may not be modified, re-distributed, sold or repackaged in any form without the prior consent of CEN, and are only for the use of the person downloading them. For additional copyright information, please refer to the statements on the cover pages of the CWAs concerned.

Cross-walks and Links between Standards.

Accommodating legacy databases in new technology applications is a massive headache. Literally dozens of data formats may have been used in recording the data across a number of database platforms/applications over the years. Metadata may range from non-existent to quite detailed - but not produced to any current metadata specification standard. Interlinking such databases under new applications has historically been accomplished by 'brute force' programming. Numerous cross-walks between different types of data standards have been developed, as presented elsewhere in this report, as well as below.

To give some idea of what is involved in defining and implementing a 'cross-walk' (link) between existing GI metadata standards, FGDC produced a mapping between the FGDC Content Standards for Digital Geospatial Metadata (CSDGM) (FGDC-STD-001-1998) and the ISO-19115 Metadata Standard (ISO 19115:2003(E)) in the form of a set of spreadsheet tables, which propose equivalency between spatial metadata content elements defined in the two standards. This is available for download from the Web[5] and shows the amount of work that goes into such a cross-walk mapping exercise. The Digital Geographic Information Working Group (DGIWG) produced a simple table

[4] Download CWA 14857: ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14857-00-2003-Nov.pdf
[5] see http://www.fgdc.gov/CrossWalk/ISO-FGDC-METADATA-CROSSWALK-V4.xls

version of the crosswalk[6], which highlights the elements for which there is no direct link between FGDC and ISO 19115. It also indicates how much work is required in implementing such linking or conversion from one standard to another - a resource requirement that should not be forgotten nor underestimated when introducing new standards via an initiative such as the Irish SDI. A visit to the FGDC site[7] where the history of development of the cross-walk is explained is a useful reminder of the amount of work involved - work that spanned the period April 1997 to September 2000 and involved hundreds of hours of standards and geospatial experts' time.

Before metadata became a ubiquitous buzzword, a descriptive and standardized format for exchanging information about scientific data sets was conceived and implemented. This was the Directory Interchange Format, DIF, a product of an Earth Science and Applications Data Systems Workshop (ESADS) held February 24-26, 1987 on catalogue interoperability. The workshop recommended making a "...first step towards data system interoperability, Catalog Interoperability (CI), the ability to find information about data held at other sites."

From 1987 through 1991, this standard underwent formal acceptance by a number of national and international groups who needed such a standard, including the Committee on Earth Observation Satellites (CEOS), the Interagency Working Group on Data Management for Global Change (IWGDMGC) which later became the Global Change Master Directory (GCMD) still used today. Elements of the FGDC's Content Standard for Digital Geospatial Metadata (CSDGM) were incorporated in the DIF in 1994. Required elements and appropriate modifications to DIF were approved by the CEOS Interoperability group and incorporated into DIF to achieve full ISO 19115compatibility in 2004.

Because of its wide use, cross-walks have been created to a number of other, newer or national standards, as shown in Table 1. Required elements and appropriate modifications to DIF were approved by the Committee on Earth Observation Satellites (CEOS) Interoperability group and incorporated into DIF to achieve full ISO 19115compatibility in 2004. DIF has been successful for more than 17 years because its structure has been flexible enough to evolve with growing metadata requirements, especially for the geospatial disciplines.

Table 1 - Historical / Thematic Metadata Standards and mapping between standards (crosswalks)

Directory Interchange Format (NASA DIF Guide)	http://gcmd.nasa.gov/User/difguide/difman.html
DIF to ISO 19115 mapping	http://gcmd.nasa.gov/Aboutus/standards/difiso.html
Dublin Core to DIF mapping	http://gcmd.nasa.gov/Aboutus/standards/dublin_to_dif.html
FGDC standard to DIF mapping	http://gcmd.nasa.gov/Aboutus/standards/fgdc_to_dif.html
ANZLIC to DIF mapping	http://gcmd.nasa.gov/Aboutus/standards/anzlic_to_dif.html

To conclude the standards section, note that Annex E contains a list of on-line links to national geospatial standards initiatives in other countries which are pertinent to the work being undertaken in Ireland. Of potential importance is the UK's GEMINI geospatial metadata standard, because of cross-border geospatial data sharing initiatives between Ireland and Northern Ireland.

Interoperability Introduction.

Figure 1 shows the relationships between the different kinds of components in a distributed system that lead to the need for interoperability standards. This illustrates well the potential complexity that arises when trying to interconnect users (user applications) with geospatial data (content repositories) via service providers (geoprocessing services), with both relevant services and content located via

[6] see http://metadata.dgiwg.org/dgiwgcrosswalks/FGDC_to_ISO_03-06-02.doc
[7] http://www.fgdc.gov/standards/standards.html

catalogs, e.g. directories of datasets and services available from widely distributed data custodians and service providers. Without some degree of standardisation, at global level, for data and metadata to describe both data and services, this complex combination of actors would be nearly impossible to interact efficiently and at least cost.

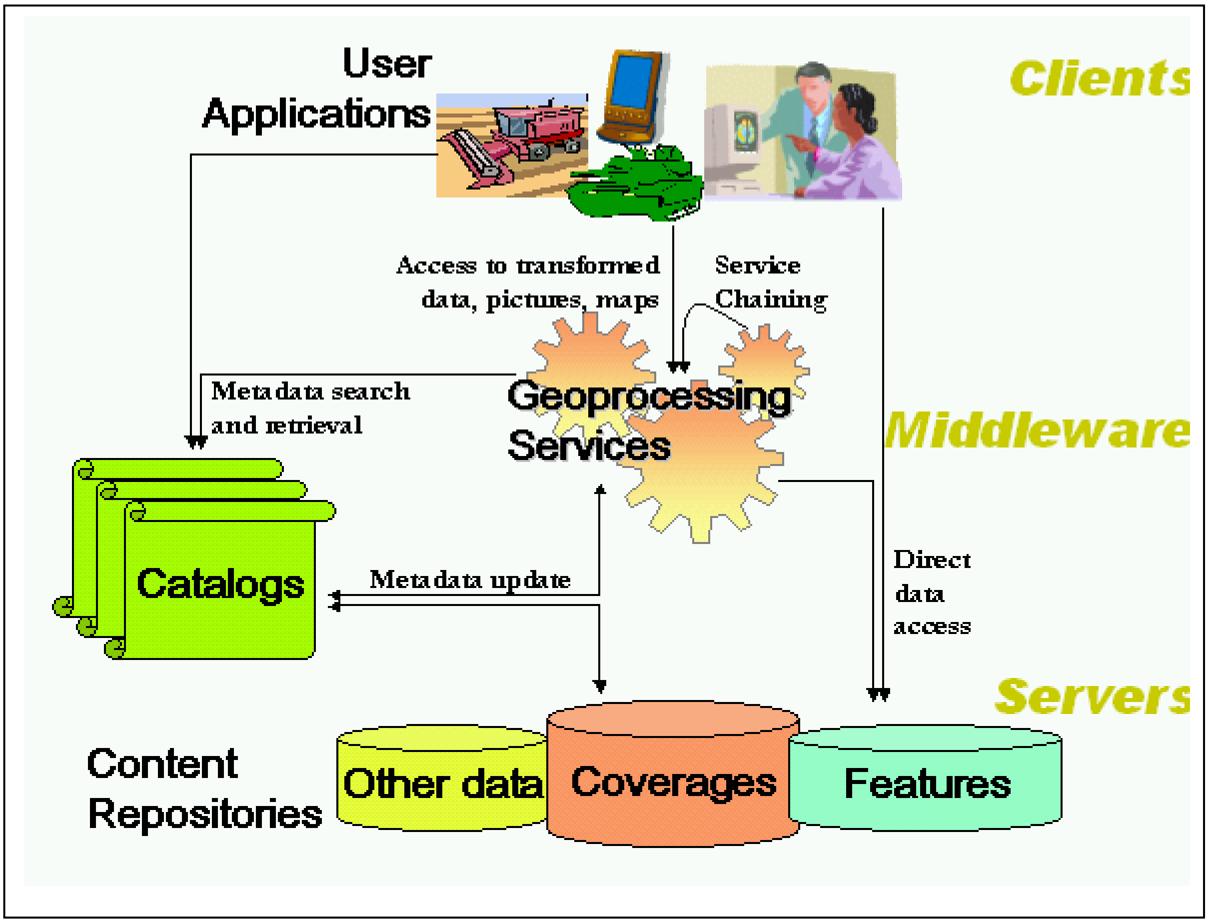

Figure 1 - Relationships between the different kinds of components in a distributed system

(from "A Geospatial Interoperability Reference Model (G.I.R.M.)" version 1.1, December 2003, by FGDC Geospatial Applications and Interoperability (GAI) Working Group, edited by John D. Evans)

Terminology - Harmonisation, Integration and Interoperability.

Many creators and users of information use the terms 'harmonisation', 'integration' and 'interoperability', sometimes incorrectly and frequently inconsistently. We define the term 'harmonisation' to mean the steps one takes in relation to data *per se* to prepare a dataset for possible (or more efficient) exploitation in conjunction with other datasets. For example, one dataset can be 'harmonised' in relation to another by using consistent metadata tags to describe the data fields of both or by developing data models and schemas which describe the datasets in a way that can be interpreted by human or programme agents to fully describe the datasets unambiguously.

Two or more datasets are 'integrated' only by applications - whether these are software programmes or knowledgeable users physically viewing contents of the datasets in such a way as to intelligently use the contents of both to examine a specific problem. Harmonisation does ease integration, but is not mandatory and many disparate datasets are 'integrated' today using fairly 'brute force' methods for extracting and merging data elements, typically under programme control.

Datasets are made "interoperable" typically by computer software which uses information about the datasets and the data elements, contained in metadata, to extract and process the data in numerous, often new and innovative ways, many of which were not envisaged at the time the datasets were created. Thus, data can be harmonised, but only applications provide interoperability and interoperability should aide in smoother and less costly data flow between applications, data owners/custodians and users in achieving integration of multiple datasets.

One goal for data management best practice should be to enhance the likelihood that a dataset can be used "interoperably" in the future, thus increasing its potential value to users - and for uses - other than those for which it was originally created. How interoperability is achieved in practice varies widely based on the type of datasets being created, by whom, for what initial purpose, and for what potential future use. Interoperability should also reduce the cost of sharing data and applications across departmental, sectoral and national boundaries.

> *"Overall the project that adopted and implemented geospatial interoperability standards saved 26.2% compared to the project that relied upon a proprietary standard."*
>
> *"Standards lower transaction costs for sharing geospatial information when interfaces are standardized and can facilitate machine-to-machine exchange."*
>
> from "Geospatial Interoperability Return on Investment Study" by Booz Allen Hamilton for NASA Geospatial Interoperability Office, USA, April 2005. [see http://gio.gsfc.nasa.gov/docs/ROI%20Study.pdf]

An important consideration is the ability for others (or you, some years from now!) to understand the data so that it can be interpreted and used correctly. Future users need to know why the dataset was created, how was the data collected, when, by whom, using what methodologies, was there a formal quality control programme in effect, did it involve conversions or transformations, were these parameters the same for all items in the dataset, and a host of other questions. If, as a new user, you don't know the answers to these questions, you probably should not be using the dataset, or at the very least you should be careful in how you interpret any results when combining that dataset with others. Lack of such descriptive information about a dataset, encoded in standardised (and thus interpretable) metadata, could lead to serious misuse of datasets, even if unintentional. This is the reason that the importance of geospatial data and metadata standards was stressed earlier in the report and why we also need to look at interoperability standards in the following section.

Interoperability Standards.

Most OGC interoperability specifications are of little relevance to end-users, other than to know that the added functionality exists in 'OGC-compliant' toolsets available to their information systems developers. Users and developers of Web-based geospatial portals need to know what the different functionalities are, and these continue to be updated over time, as new specifications are released and as previous specifications are updated or extended. The continued evolution of extensions to existing interoperability capabilities is something that IT departments need to monitor on an on-going basis. This is one of the tasks that could be assigned to the proposed Technology and Standards Advisory Group within the recommended ISDI Unit.

OGC interoperability specifications appear in two forms - abstract specifications and implementation specifications (see the full list in Annex D). The implementation specifications of most direct relevance to spatial data Web portal developers are shown in Table 2. Note also that some of the OGC implementation specifications have already become, or are proposed to become, ISO 19xxx series international standards. GML (Geography Markup Language) is of special importance as use of this metadata markup specification is a prerequisite for use of some of the other OGC specifications. Also, all the specifications are freely available from the Web[8].

[8] see http://www.opengis.org/specs/?page=baseline

Table 2 - OGC Implementation Specifications

OGC Implementation Specification	Brief Description
Geography Markup Language (GML 3.0)	GML is an XML grammar written in XML Schema for the modelling, transport, and storage of geographic information.
Web Feature Service (WFS)	WFS allows a client to retrieve geospatial data encoded in XML using Geography Markup Language (GML) from multiple Web Feature Services.
Web Map Service (WMS 1.1.1)	Web Map Service allows a client to overlay map images for display, served from multiple Web Map Services on the Internet
Catalog Interface (CAT)	CAT provides discovery, access (direct or brokered) and management services permitting clients to locate metadata, access services relating to the data and to change metadata in a catalogue.
Web Coverage Service (WCS)	WCS supports electronic interchange of geospatial data as 'coverages', i.e. digital geospatial data representing space-varying phenomena, providing access in forms useful for client-side rendering, multi-valued coverages, and input into scientific models and other clients.
Filter Encoding (Filter)	Defines an XML encoding for filter expressions based on the BNF definition of the OpenGIS Common Catalog Query Language.
OpenGIS Location Services (OpenLS): Core Services [Parts 1-5] (OLS Core)	OpenLS defines access to the Core Services and Abstract Data Types (ADT) that comprise the GeoMobility Server, an open location services platform.
Web Map Context Documents (WMC)	A Context specification accompanying WMS 1.1.1, that states how a specific grouping of one or more maps from one or more map servers can be described in a portable, platform-independent format for storage in a repository or for transmission between clients.
Styled Layer Descriptor (SLD)	Specifies the format of a map-styling language for producing georeferenced maps with user-defined styling, transforming geographic data into an explanatory, decision-support tool (e.g. a map).
Grid Coverages (GC)	GC is designed to promote interoperability between software implementations by data vendors and software vendors providing grid analysis and processing capabilities.
Coordinate Transformation Services (CT)	Provides interfaces for general positioning, coordinate systems, and coordinate transformations wherein coordinates can have any number of dimensions, i.e. can handle 2D, 3D coordinates plus 4D, 5D, etc.
Simple Features CORBA (SFC)	Provide interfaces to allow GIS programmers to develop applications that expose functionality required to access and manipulate geospatial information comprising features with 'simple' geometry using OMG's CORBA (Common Object Request Broker Architecture) technology.
Simple Features OLE/COM (SFO)	Specifies an architecture comprising three components - **OLE DB** for implementing data providers - **ADO** for presenting a simplified data access model on top of OLE DB - **Geometry and Spatial References** for detailed geometry and reference operations - all implemented using the Microsoft Component Object Model (COM)
Simple Features SQL (SFS)	Defines a standard SQL schema that supports storage, retrieval, query and update of simple geospatial feature collections via the ODBC API

Sharing Geospatial Data with Others.

The main trend in serving up spatial data to interested users is via the Web, which in the future will be ever more dependent upon Web Services. The OGC Web Services (OWS) type infrastructure supports 'always-on' data serving, e.g. serving of soils map from a government soils database to all OGC-compliant data readers, whether on a desktop or hand-held, mobile device, or to GIS tools and databases which further process the original data

The XML/GML Data Serving format supports service requests, but also core dataset exchange, either dynamic data serving, e.g. on-line road traffic data, or non-dynamic data downloads, e.g. topographic mapping in GML, PSI Information Asset Register metadata in XML, etc.

All these mechanisms have a role in supporting a federated approach to data management and exchange, as opposed to centralized warehousing. Because of the complexity of systems that already exist across the wide range of stakeholders concerned with a Spatial Data Infrastructure, it is difficult (if not impossible) to impose any specific data transport model or mechanism(s) on such stakeholders today.

Data Description Models.

Data models are created in order to express a vision of the real world in a formalised way, which can then be used by software applications to achieve defined goals, in other words “a data representation of reality” (Kennedy, 2001). Models are created using a modelling language or methodology, of which there are many. Some models describe the contents of a database created for a specific purpose or application. Other models describe how applications use one or more databases to achieve a required output, for example a report or some form of transaction processing. Modelling starts with the capture of user requirements in the form of descriptions of data and processes needed to achieve a stated objective, for example satisfying a user’s request for information about location of buried utility cables.

While it is beyond the scope of this report to provide a full introduction to data and process modelling, in this section we briefly introduce the concept of data models, which today tend to be relational data models or object-oriented data models. The relational model expresses specific, usually fixed, relationships between the different data elements in a data record. Relational databases are typically implemented via tables in which columns represent different data attributes (name, age, sex, etc.) and rows exist for each instance of that record (John, Jane, Bill, etc.). Relational data models are implemented via relational database management systems (RDBMS) and until recently were more common than object-oriented database implementations.

The object-oriented approach defines objects as collections of data attributes that are determined by the object data model. Thus, ‘John’ in the above relational data model could become an object in the object-oriented data model and would exist independently, with a unique identifier, to be acted upon by different software applications in different ways. Object-oriented data models are implemented using object database management systems (ODMS) which are designed to work well with object-oriented programming languages such Java and C++.

Different GIS toolsets exist that are built around both relational and object-oriented data models, although any one toolset typically supports only one model type due to the inherent difference in how data is modelled between the two methodologies.

Creating a data model is recommended (and typical) practice for any new geospatial information application. The process of creating a formal data model helps ensure that all aspects of the data requirement have been considered. The data model will include identification and specification of all the types of data needed to satisfy the user or application requirement, including relationships (dependencies) between the data items, their format, validity tests, etc. In a more complex application, the data model will also indicate how data items held in one database may be related to items held in other databases. A separate process model (also called a data flow model represented by a data flow diagram) illustrates how software applications (programs) interact with the data items to deliver system outputs. The two models are typically developed in parallel, since they depend upon each other.

Implementing Data Models.

Data models can be expressed in a number of ways, including data dictionaries (detailed descriptions of database contents, often in formal, standardized ways), DDT (Data Definition Tables) or graphical representations, such as that used in the Unified Modelling Language (UML). Logical data models describe data contents and relationships, in ways and using terminology often specific to the methodology used to create the logical model. XML (eXtensible Markup Language) plays an important role in how data is handled by various applications, how it is transported (logically) between computers (over the Internet or via intranets) and how applications are able to interpret data contents that have been ‘tagged’ using XML conventions.

Especially for geospatial information, OGC’s GML 3.0 specification, Geography Markup Language, is an eXtensible Markup Language (XML) grammar written in XML Schema to enable modelling, transport, and storage of geographic information in a standardised way. GML 3.0 is also now the focus of an ISO standard working group, as ISO 19136. The key concepts used by Geography Markup Language (GML) to model the world are taken from the OGC Abstract Specification.[9]

GML specifies different kinds of 'objects' for describing geography generally, including features, coordinate reference systems, geometry, topology, time, units of measure and generalized values. GML treats geographic features as abstractions of real world phenomenon, i.e. a geographic feature is associated with a location relative to the Earth. A digital representation of the real world thus becomes a set of features, and a feature is defined by one or more properties, where each property can specified by three parameters, i.e. name, type and value.

GML 3.0 conforms with other ISO standards, including:

- ISO DIS 19107 Geographic Information – Spatial Schema
- ISO DIS 19108 Geographic Information – Temporal Schema
- ISO DIS 19118 Geographic Information – Encoding
- ISO DIS 19123 Geographic Information – Coverages

GML has already been adopted and implemented by many major actors in the SDI arena, including Ordnance Survey GB, for its totally digitised MasterMap ® database. It is being adopted by other European NMAs (national mapping agencies), including OS NI and OS Ireland.

Data Quality.

Data quality is a multifaceted issue. First, there is the issue of creating 'quality' data in the first place. Then different issues arise in assessing the quality of a dataset after it has been created, depending upon the purpose for which the data is now being accessed, which may not be the same as that for which it was created. Finally, there are different issues on metadata that can be used to describe or infer the quality of a geospatial dataset. Yet ensuring 'data quality' is a complex and increasingly vital issue, with no easy or obvious solutions.

Data Quality Definitions and Standards.

Much of the initial work on spatial data quality was conducted during formation of early data standards, either for data capture or transfer, beginning in the early 1980s. Interestingly, some GIS texts offer no comprehensive definition or discussion of 'data quality', except to include it with other 'problem' issues waiting to be resolved. This is as true for several texts published in the early 1990s as in 2004, so we can assume that data quality issues are not going to disappear anytime soon.

A US Office of Management and Budget circular from 2001 defined quality as:

> "Quality" is an encompassing term comprising utility, objectivity, and integrity. Therefore, the guidelines sometimes refer to these four statutory terms, collectively, as "quality." (OMB 2001)

The six principal areas of data quality identified by the FGDC (1998) and Guptill and Morrison (1995) include:

- positional accuracy
- attribute accuracy
- completeness
- logical consistency
- semantic accuracy, and
- currency.[10]

[9] http://www.opengis.org/techno/abstract.htm
[10] after Fisher, P. in "Approaches to Uncertainty in Spatial Data", book chapter in otherwise unspecified book.

The FGDC Content Standards for Digital Geospatial Metadata (CSDGM) contains a detailed specification for "Spatial Data Quality"[11] as set out in chapter 3 of part 1 in the US Spatial Data Transfer Standard (SDTS)[12], which reflects the six categories above. While whole volumes have been written on spatial data accuracy (Goodchild and Gopal, 1989), less attention is given to the other elements of data quality listed above. The most complete treatment we found is over a decade old, yet still a valuable contribution to the spatial data quality discussion (Guptill and Morrison, 1995).

A draft ISO standard on "Quality Principles" (ISO 19113) was published in 2002, along with an accompanying standard (ISO 19114) on "Quality Evaluation Procedures", both of which have also been endorsed by CEN in 2005, as European Standards (ENs). However, there is little evidence that these have been widely adopted, or even promulgated, yet, as is the case with many of the more recent standards that emanated from ISO TC 211. ISO 19113 states that:

> "Quantitative quality information shall be reported as metadata in conformance with the requirements of ISO 19115" and "Quantitative quality information shall additionally be reported using a quality evaluation report in conformance with the requirements of ISO 19114".

The quality principles are the focus of a new (requested) ISO TC 211 work item, stating that there are "obscurities and imperfections" and "differences" between the quality specification ISO 19113 and ISO 19115 in regard to metadata expression. The work item request also states that "A fundamental problem is that implementation of 19113 through the UML schemas of 19115 is difficult bordering on the impossible".[13] This may (partly) explain why ISO 19113 (and 19114) have yet to make any great impact on the geospatial community.

Alternatives in Determining Data Quality.

The term 'quality' is used in reference to data, metadata, data services, value added products, etc. Across the information industry generally, not just in the geospatial data sector, there has been little (to no) progress in agreeing on the parameters that would suffice to underpin issuing of quality 'kite marks' to different types of data, products or services. Therefore, other types of quality indications (as opposed to indicators) have been attempted. One example of this was the system by which users of STM (scientific, technical, medical) databases could record their personal experiences and comments with the Centre for Information Quality Management[14] concerning the suitability of specific databases from named industry suppliers. These comments were then made available publicly on a Web site maintained by CIQM, initially under an EU-funded RTD programme. Over a number of years, the STM information community felt that existence of the CIQM, which was established in 1993, had encouraged some STM data suppliers to upgrade the overall quality of their databases (mainly abstracting services for STM-related papers, patents, etc.), mainly by a "name and shame" policy. An inherent problem, of course, is that data of sufficient 'quality' for one user's application may be totally unsuitable for that of another user or application.

Metadata doesn't resolve data quality issues, either. All that metadata can do is describe data items in a dataset or document or offer some text about the provenance of the dataset. Metadata cannot state that "this is an excellent data element that suits your purpose". 'Quality' as a term used in regard to metadata usually refers to whether or not the metadata was created at all, was created by someone knowledgeable about the dataset to which the metadata is attached, is complete (i.e. metadata exists for all required items in the dataset), is created to a known standard (preferably international standard), or similar considerations. Yet all of these conditions can be met and a potential user may still not be able to use a dataset with complete confidence, especially if the user has little experience in the field of expertise of the creator of the dataset and thus is not able to interpret the metadata appropriately, or if the dataset is to be used for a totally different application.

---

[11] see http://www.fgdc.gov/metadata/csdgm/02.html
[12] Department of Commerce, Federal Information Processing Standard 173: Washington, Department of Commerce, National Institute of Standards and Technology (see http://mcmcweb.er.usgs.gov/sdts/index.html)
[13] ISO/TC 211 N1770 of 2005-02-10, see www.isotc211.org
[14] see www.i-a-l.co.uk/ciqm_index.html

Another possibility in determining if a specific geospatial data resource is suitable for any specific application may be to examine the metadata for specific parameters (if reported) relating to the data, i.e. date created, creator (do they have expertise in their field?), stated purpose why the dataset was created, for what application, stated resolution, etc. The problem with this approach is that it cannot be easily automated, requiring human interpretation of the metadata found. The appropriate metadata may not (always) exist for the data resource.

It would be worth investigating the possibility of creating an interactive front-end to spatial databases, which would permit the user to specify which types of quality elements are of most interest while interrogating the dataset in real time. Thresholds could be set for many parameters below which data items or instances (whole records) could be automatically rejected.

To conclude, we repeat that data quality is a complex issue that has proven resistant to easy solution for over three decades now. We cannot expect any simple solutions to suddenly appear in the next decade, since spatial data and the myriad applications to which it is put seem to get ever more complex. Convergence of spatial data with other data types, especially multimedia, and onto non-traditional devices, especially mobile devices, will simply add to the complexity.

Ontologies.

Why an Ontology?

An ontology is defined as a "specification of a conceptualization" of the real world[15]. It is a formal representation of our knowledge about an area of interest. Think of ontology as a written description of one aspect of the world in which you have special interest and expert knowledge, using an agreed vocabulary (standard terminology) to describe what you see in that world and how these things interact with each other - and with you. The 'formal' aspect is key, in that the expression of the ontology must be in a format that is understandable not only by humans, but also by machines, two types of 'agents' that could use this description of the world in different ways. An ontology is a "description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents" (Gruber 2004).

An ontology is written (described) as a set of definitions of formal vocabulary for practical reasons, since it permits our conceptualization of reality, our "knowledge", to be shared by and with clever software applications. Ontologies are created by groups of experts who attempt to codify their knowledge of a particular field of interest, establishing agreed terminology between different members (agents) of that community of interest, e.g. marine researchers or forestry experts or urban planners.

Returning to the 'formal' aspect again, the 'conceptualization' must be expressed in a representation language (similar in concept to a computer programming language), of which several exist. For applications on the Web, it is important to have a language with a standardized syntax.. Because XML is the current accepted standard for exchanging information via the Web, XML syntax is used to exchange ontologies, which greatly simplifies the job of writing parsers (programmes that can interpret an ontology when it is expressed in XML). An intermediate step between describing the ontology in an informal way (which is how the process starts) and arriving at formal (agreed) XMl schemas to represent the data types required to capture the knowledge represented by the ontology, is to use the Unified Modelling Language (UML) - see the next section.

Ontology & semantics (for a thematic community) ➔ **UML model** ➔ **XML Schema**(s) to implement

Of course, different communities will create different ontologies for their fields of interest, and some communities may even create different ontologies to express different aspects of their world, e.g. the marine community developed one ontology to specify their knowledge of marine navigation and a different one for marine science. Once there are a number of ontologies accessible via the Web, the

[15] Gruber, T. see http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

next step is to make them interoperable, which is accomplished (less easily) by defining mappings between different ontologies.

OWL - Web Ontology Language.

For those wishing to create and publish their ontologies on the Web, W3C (the World Wide Web Consortium) recommends OWL, the Web Ontology Language. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans (for example via HTML tags). OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF (Resource Descripton Framework), and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.[16]

Unified Modelling Language (UML)

The Unified Modelling Language combines graphical and textual notations to describe systems, using several diagram types and associated descriptive sections which provide systems (and ontology) developers with a large set of views for representing system requirements, architecture, behaviour - and knowledge. Using UML, systems requirements are defined by Use Case Diagrams, which describe interactions between 'actors' and the system, defining the functional capabilities of the system, and then by Sequence Diagrams, which specify individual processes performed by or within the system. Class diagrams, collaboration diagrams, state diagrams and activity diagrams are also developed during system design (or knowledge capture). Component and deployment diagrams are used to present system deployment information.

It is beyond the scope of this report to provide a course in UML, but a brief overview is readily available in Alan Moore's "A Unified Modeling Language Primer"[17] and there are many more detailed primers and full UML specifications available from the Web[18]. A wide range of products exists to help create UML models graphically, as well as to map between XML, databases and flat files. Much shareware and 'freeware' also exists. It is necessary to use such software tools when developing UML models, as the specification for UML runs to more than 700 pages and the graphic nature of the modeling methodology is crucial to the process.

> "… UML provides a visual foundation for using an object-oriented approach. The UML is primarily a set of notations and does not prescribe a specific development process. Based on OO principles, the UML lends itself well to a pragmatic, iterative, and elaborative approach to systems development. As such, the representations a systems developer constructs will evolve over time as he captures more detail about the system." (Moore, 2004)

Note that references to 'systems' above are equally interchangeable with the term 'knowledge base' or 'conceptualisation of the real word", which is why UML can be used to capture the formal representation of an ontology. Note that UML is not a 'model' but a modelling language. The reason we introduce it in this report is that UML is used extensively in the standards world, by both ISO and OGC in formally specifying their respective standards and interoperability specifications. UML is especially useful in the latter, as it can be used to express process-related information, not only data-related information.

Semantics and Metadata.

In the previous two sections we have referred to 'standard terminology' or 'agreed vocabulary' more than once. This is because it is impossible to formally codify descriptions of the world, of a database, of a user requirement unless everyone is speaking the same language, i.e. using commonly understood terminology. When a meteorologist speaks of 'waves' he may or may not be speaking of the same

[16] see http://www.w3.org/TR/owl-features/ for specifications, primers, and guides to using OWL.
[17] see http://www.techonline.com/community/related_content/14413
[18] see http://www.uml.org maintained by OMG the Object Management Group (www.omg.org)

‘waves’ as a marine scientist or coastal engineer and neither is probably speaking of the same ‘waves’ of the electronic engineer. This is why the properties of ‘wave’ must be explicitly defined in a formal way when developing an ontology or building some other form of system description using UML or another modelling language.

Yet another level to this problem is only now receiving more formal investigation (see Comber et al, 2005a). This relates to how even the same school of experts may view terms such as “field” or “forest” quite differently. For example, the agricultural/forestry community uses more than 25 official definitions for ‘forest’ globally, depending very much on where they are located in the world or why and how the definition is being used, e.g. for environmental management reasons versus perhaps payment of financial subsidies.

According to Comber et al (2005):

> “Current metadata standards (e.g. ISO, OGC) are inadequate. They don’t describe the meaning of the concepts embedded in the data that would allow users to make some assessment of the applicability of the data to their problem. … in many cases environmental information supply is either a monopoly or an oligopoly. For instance, consider the availability of land cover, soils, geology, climate, species distribution, agricultural census and population census: often the user has the choice to use product “x” or not, rather than a choice between “a”, “b”, “c”… etc. Even so, current metadata does not give any information to the user about how to best exploit the data.”

The main problem today with standards relating to metadata is that the standards relate only to format or forms of expression of metadata, which, while important, is only part of the problem. The Booz Allen Hamilton study (2005) found that their case study project that fully implemented geospatial interoperability standards had a significantly higher return on investment (ROI) and saved 26% overall in comparison to the case study project that used proprietary standards, due to a variety of benefits and efficiencies presented in the full report. However, initial implementation cost was also higher for the project that fully adopted interoperability technology from the outset, due partly to the higher cost of “achieving semantic agreement” between all parties who interacted with the datasets in use by that project. Yet the long term costs relating to lower transaction costs were significant, contributing to the better overall performance of that project.

A much better understanding is required of the concepts of metadata ‘from first principles’. Also, immediate attention should be given to an investigation of semantics, especially any formal terminologies or controlled vocabularies that are now in use by the many stakeholders in ISDI. Once the degree (or not) to which such standardisation occurs is known, an initiative could be undertaken, perhaps as an ISDI research project, to see if such lists can be harmonised and the ramifications of such harmonisation, for example on legacy database systems.

Summary.

This short report has attempted to look at a broad range of data related issues, some in more detail than others, depending upon which issues may have more direct impact on immediate SDI implementation needs versus longer-term considerations. Of most immediate importance are:

- creation of metadata to international standards (ISO 19115 and its CEN equivalent),
- understanding XML metadata schemas and creating such schemas for both key legacy datasets and all new datasets,
- understanding data modelling and representational issues, including ontology development and semantics (controlled vocabularies or terminologies, data dictionaries, etc.),
- maintaining awareness (observatory? formal watching brief?) of developments in the standards world (evolution of ISO standards and OGC interoperability specifications) and policy relating to SDI (INSPIRE).

References and Bibliography.

Booz Allen Hamilton. 2005. Geospatial Interoperability Return on Investment Study, National Aeronautics and Space Administration, Geospatial Interoperability Office, USA. [http://gio.gsfc.nasa.gov/docs/ROI%20Study.pdf]

Coomber, A., Fisher, P. and Wadsworth, R. 2005. What is land cover? Proceedings of GISPlanet 2005 and to be published in Environment and Planning B.

Flater, David. 2001. Impact of Model-Driven Standards, OMG Model Driven Architecture.

Goodchild, M. and Gopal, S. 1989. Accuracy of Spatial Databases. Taylor & Francis Ltd, London & Bristol.

Gruber, T. R. 2003. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220. (see http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html)

Gruber, T. R. 1993. Toward principles for the design of ontologies used for knowledge sharing. Presented at the Padua workshop on Formal Ontology, March 1993 (see http://ksl-web.stanford.edu/KSL_Abstracts/KSL-93-04.html)

Guptil, S.C. and Morrison, J.L.. 1995. Elements of Spatial Data Quality. ICA Commission on Spatial Data Quality, Pergamon/Elsevier Science.

Joffe, Bruce and Todd Bacastow. 2005. Geodata access while managing digital rights: The Open Data Consortium's Reference Model. *URISA Journal*, 10 January 2005 (under review). See http://www.urisa.org/Journal/UnderReview/joffe/geodata_access_rights

Joffe, Bruce. 2003. Data Distribution Policy Issues - a summary review, prepared for the Open Data Consortium project. See http://www.opendataconsortium.org/documents/Data_Distribution_Issues.pdf

Kennedy, Heather. 2001. Dictionary of GIS Terminology, ESRI Press: Redlands, CA, USA.

Nativi, Stefano. 2001. THREDDS - A Proof of Concept Systems. UNICATA-UCAR, Boulder, CO, USA. December 2001.

OMB. 2001. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, Office of Management and Budget, Executive Office of the President, USA, 1 October 2001. (http://www.whitehouse.gov/omb/fedreg/text/final_information_quality_guidelines.html)

Reichardt, Mark. 2003. The Standards Edge: Dynamic Tension, The Bolin Group

Annex A. FGDC Metadata Standards Summary

Content Standard for Digital Geospatial Metadata (CSDGM)(FGDC-STD-001-1998)

This FGDC standard is that referred to in Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," signed on 11 April, 1994, by then President William Clinton. Section 3 of the order, Development of a National Geospatial Data Clearinghouse, paragraph (b) states: "Standardized Documentation of Data, ... 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." Thus, the FGDC standards were legally mandated and became the backbone for the US NSDI's now extensive geospatial information clearinghouse (portal) network.

This standard provides a common set of terminology and definitions for the documentation of digital geospatial data. 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 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, the 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 for data transfer, nor the means by which the information is transmitted, communicated, or presented to the user.

Biological Data Profile of the Content Standard for Digital Geospatial Metadata (FGDC-STD-001.1-1999) endorsed on 26/10/99.

This standard provides a user-defined, theme-specific profile of the FGDC Content Standard (CSDGM) to increase its usefulness in documenting biological resources data and information. The standard supports increased access to, and use of, biological data among users on a national (and international) basis and broadens the understanding and implementation of the FGDC metadata content standard within the biological resources community. It is also the metadata content standard for the National Biological Information Infrastructure (NBII)[19]. The standard is used to specify metadata for all biological resources data and information, including biological data which is explicitly geospatial in nature, and non-geospatial data (e.g. data resulting from laboratory-based research), including research reports, field notes or specimen collections.

Metadata Profile for Shoreline Data (FGDC-STD-001.2-2001)

This metadata profile is used as an extension or profile to the existing CSDGM with a glossary and bibliography as informative annexes providing a basis for understanding the shoreline and related issues. CSDGM only permits documentation of generic geospatial data, so the FGDC Bathymetric Subcommittee developed a metadata profile addressing shoreline data and data that intersects with the shoreline to capture the critical processes and conditions revolving around shoreline data. The metadata produced using this standard is important for clearinghouse/portal activities to locate potential data sets and to indicate the fitness for use and accuracy of a given data set. Shoreline data are important for coastal zone management, environmental monitoring, resource developments, legal land jurisdictional issues, ocean and meteorological modelling, engineering, construction, planning, and many other uses. In the US NSDI, shoreline is an integral component of the geospatial data framework.

[19] see http://www.nbii.gov/datainfo/metadata

CSDGM Extensions for Remote Sensing Metadata (FGDC-STD-012-2002) endorsed 9/10/ 2002

The purpose of this standard is to provide extensions to the FGDC Content Standard for Digital Geospatial Metadata for metadata describing geospatial data obtained from remote sensing. Efforts are being made to make these extensions compatible with the framework and content of the ISO metadata standard now undergoing the approval process, in order that the FGDC standard can be converted to ISO form for use as remote sensing extensions to the ISO standard. The extensions define content standards for additional metadata, including metadata describing the sensor, the platform, the method and process of deriving geospatial information from the raw telemetry, and the information needed to determine the geographical location of the remotely sensed data. In addition, metadata to support aggregation, both the components of an aggregate data set and the larger collection of which a data item may be a member, will be supported.

Thematic Supplement for Geospatially Referenced Cultural and Demographic Data Metadata

This supplement aims to maintain consistency in creating FGDC compliant metadata and encourage use of the CSDGM, while meeting the unique needs of data producers and users interested specifically in geospatially referenced cultural and demographic data sets. The Supplement will become an FGDC standard, defining the metadata core for geospatially referenced cultural and demographic data, based on existing metadata elements of the CSGDM, as well as additional "user-defined" metadata elements. For each additionally defined metadata element, the Supplement will include (1) a reference to other metadata collections, if applicable, (2) the metadata element's definition, type, and domain, (3) the source of the metadata element, and (4) the CSGDM metadata element the metadata element leads/follows.

USA - FGDC Metadata standards - 1994 onwards - links	
Content Standards for Digital Geospatial Metadata (CSDGM) - Federal Geographic Data Committee (FGDC) standard for the USA NSDI (FGDC-STD-001-1998)	http://www.fgdc.gov/metadata/contstan.html
CSDGM Biological Data Profile (FGDC-STD-001.1-1999) for the National Biological Information Infrastructure	http://www.fgdc.gov/standards/status/sub5_2.html http://www.nbii.gov/datainfo/metadata
CSDGM Shoreline Metadata Profile (FGDC-STD-001.2-2001)	http://www.fgdc.gov/standards/status/sub5_6.html
CSDGM Extensions for Remote Sensing Data	http://www.fgdc.gov/standards/status/csdgm_rs_ex.html
CSDGM Geospatially Referenced Cultural and Demographic Data	http://www.fgdc.gov/standards/status/sub2_5.html
Z39.50 “GEO” Profile	http://www.geoconnections.org/architecture/technical/specifications/geodata_registry/geo22.htm

Annex B. CEN TC 287 European 'profiles' of ISO 19xxx series Standards

CEN TC 287 Published Standards Relating to ISO 19xxx Standards

Standard reference	Title
EN ISO 19101:2005	Geographic information - Reference model (ISO 19101:2002)
EN ISO 19105:2005	Geographic information - Conformance and testing (ISO 19105:2000)
EN ISO 19107:2005	Geographic information - Spatial schema (ISO 19107:2003)
EN ISO 19108:2005	Geographic information - Temporal schema (ISO 19108:2002)
EN ISO 19111:2005	Geographic information - Spatial referencing by coordinates (ISO 19111:2003)
EN ISO 19112:2005	Geographic information - Spatial referencing by geographic identifiers (ISO 19112:2003)
EN ISO 19113:2005	Geographic information - Quality principles (ISO 19113:2002)
EN ISO 19114:2005	Geographic information - Quality evaluation procedures (ISO 19114:2003)
EN ISO 19115:2005	Geographic information - Metadata (ISO 19115:2003)

CEN/TC 287- ISO 19xxx-compatible Standards under Development

TC 287 Project Reference Code	Title	Current Status	Date Available
prEN ISO 19106	Geographic information - Profiles (ISO 19106:2004)	Under Approval	2006-03
prEN ISO 19116	Geographic information - Positioning services (ISO 19116:2004)	Under Approval	2006-03
prEN ISO 19125-1	Geographic information - Simple feature access - Part 1: Common architecture (ISO 19125-1:2004)	Under Approval	2006-03
prEN ISO 19125-2	Geographic information - Simple feature access - Part 2: SQL option (ISO 19125-2:2004)	Under Approval	2006-03
	Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructure	Under Development	2006-07
prEN ISO 19111 rev	Geographic information - Spatial referencing by coordinates	Under Development	2007-12

CEN TC 287 & ISO 15836 Dublin Core	
CEN-TC287 for geographic data - main committee site	http://www.cenorm.be
CEN ISSS **CWA 14857** - Dublin Core to ISO 19115 crosswalk	ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14857-00-2003-Nov.pdf
CEN ISSS **CWA 14856** - Guidance on mapping DC-ISO	ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14856-00-2003-Nov.pdf
CEN ISSS **CWA 15858** - DC Spatial Application Profile	ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14858-00-2003-Nov.pdf
Dublin Core at 2005-06-15 (now ISO 15836)	http://dublincore.org/documents/dcmi-terms/
Dublin Core - **Abstract model**	http://dublincore.org/documents/abstract-model/
ANSI-NISO Z39.85-2001	http://www.niso.org/standards/resources/Z39-85.pdf
ISO 15836:2003 - Dublin Core ISO Standard	http://www.iso.org/iso/en/ISOOnline.frontpage (€30.00)

Annex C. ISO Published standards/reports and documents under publication (May 2005)

ISO 6709:1983	Standard representation of latitude, longitude and altitude for geographic point locations
IS0 19101:2002	Geographic information – Reference model
ISO 19103:2005 (PRF)	Geographic information – Conceptual Schema Language
ISO 19105:2000	Geographic information – Conformance and testing
ISO 19106:2004	Geographic information – Profiles
ISO 19107:2003	Geographic information – Spatial schema
ISO 19108:2002	Geographic information – Temporal schema
ISO 19109:2005 (PRF)	Geographic information – Rules for application schema
ISO 19110:2005	Geographic information – Methodology for feature cataloguing
ISO 19111:2003	Geographic information – Spatial referencing by coordinates
ISO 19112:2003	Geographic information – Spatial referencing by geographic identifiers
ISO 19113:2002	Geographic information – Quality principles
ISO 19114:2003	Geographic information – Quality evaluation procedures
ISO 19115:2003	Geographic information – Metadata
ISO 19116:2004	Geographic information – Positioning services
ISO 19117:2005	Geographic information – Portrayal
ISO 19118:2005 (PRF)	Geographic information – Encoding
ISO 19119:2005	Geographic information – Services
ISO/TR 19120:2001	Geographic information – Functional standards
ISO/TR 19121:2000	Geographic information – Imagery and gridded data
ISO/TC 19122:204	Geographic information – Qualifications and Certification of personnel
ISO 19125-1:2004	Geographic information – Simple feature access – Part 1: Common architecture
ISO 19125-2:2004	Geographic information – Simple feature access – Part 2: SQL option
ISO/TS 19127:2005	Geographic information – Geodetic codes and parameters
ISO 19133:2005 (PRF)	Geographic information – Location based services tracking and navigation

Work Items Still Underway (July 2005)

19104 (15046-4)	Geographic information – Terminology
19120/Amedmend 1	Geographic information - Functional standards - Amendment 1
19123 (17753)	Geographic information - Schema for coverage geometry and functions
19124 (17754)	Geographic information - Imagery and gridded data components
19125-3:	Geographic information - Simple feature access - Part 3:COM/OLE option
19126	Geographic information - Profile - FACC Data Dictionary
19128	Geographic information - Web Map server interface
19129	Geographic information - Imagery, gridded and coverage data framework
19130	Geographic information - Sensor and data models for imagery and gridded data
19131	Geographic information - Data product specifications
19132	Geographic information - Location based services possible standards
19134	Geographic information - Multimodal location based services for routing & navigation
19135	Geographic information - Procedures for registration of geographical information items
19136	Geographic information - Geography Markup Language (GML)
19137	Geographic information - profiles of spatial schema and similar important other schemas
19138	Geographic information - Data quality measures
19139	Geographic information - Metadata - Implementation specifications

Annex C (continued)

ISO/IEC 11179 - Specification and Standardization of Data Elements

ISO/IEC 11179 specifies basic aspects of data element composition, including metadata. It applies to formulation of data element representations and meaning as shared among people and machines; it does not apply to the physical representation of data as bits and bytes at the machine level. It describes the standardizing and registering of data elements to make data understandable and shareable. Data element standardization and registration as described in ISO/IEC 11179 allow the creation of a shared data environment in much less time and with much less effort than it takes for conventional data management methodologies. The purpose of ISO/IEC 11179 is to give concrete guidance on the formulation and maintenance of discrete data element descriptions and semantic content (metadata) that shall be used to formulate data elements in a consistent, standard manner. It also provides guidance for establishing a data element registry. ISO/IEC 11179:

- facilitates acquisition and registration of data,
- expedites access and use of data,
- simplifies data manipulation by intelligent software by enabling manipulation of data based on characteristics described by metadata,
- enables the development of a data representation metamodel for CASE tools and repositories, and
- facilitates electronic data interchange and data sharing.

In ISO/IEC 11179, metadata about data elements is stored in a data element registry, which supports data sharing with descriptions of data. Registration is the process of documenting metadata to support data shareability and should be carried out at the level of data elements to promote and maximize semantic value. ISO/IEC 11179 enables the end user to interpret the intended meaning confidently, correctly, and unambiguously.

ISO 15836:2003 (Dublin Core Metadata Elements)

ISO 15836:2003 is applicable to the Dublin Core metadata element set which deals with cross-domain information resource description. 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.

ISO 19115:2003 Geographic information - Metadata

ISO 19115:2003 defines the schema 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. [Note: Certain mandatory metadata elements may not apply to these other forms of data].

ISO - International Organization for Standardization	
ISO 6709 - Standard representation of latitude, longitude and altitude for geographic point locations	new ISO TC 211 work item on a long established ISO standard
ISO/IEC 11179 - Specification and standardization of data elements	http://www.iso.org Search on “15836” and “Standards” (€30.00)
ISO 19101 - Geographic Information - Reference model	
ISO 19103 - Geographic information – Conceptual schema language	
ISO 19107 - Geographic information - Spatial schema	
ISO 19108 - Geographic information - Temporal schema	
ISO 19109 - Geographic information – Rules for application schema	
ISO 19115 - Geographic information - Metadata	http://www.iso.org or http://www.isotc211.org Search on “19115” and “Standards” (€135.00)
ISO 19118 - Geographic information – Encoding	
ISO 19136 - Geographic information - Geography Markup Language	
ISO 19139 - Metadata implementation specification	http://www.iso.org or http://www.isotc211.org/ http://www.isotc211.org/protdoc/211n1535/
ISO TC211 N1859 - Core Cadastral Domain Model discussion paper	http://www.isotc211.org/protdoc/211n1859/

Annex D. - Relevant OGC Abstract and Implementation Specifications

Legend: ATB = Approved Technical Baseline
IS = Implementation Specification
AS = Abstract Specification

ATB	OpenGIS® Reference Model (ORM)
IS	OpenGIS Location Services (OpenLS): Core Services [Parts 1-5] (OLS Core)
IS	Web Coverage Service (WCS)
IS	Web Map Context Documents (WMC)
IS	Geography Markup Language (GML 3.0)
IS	Catalog Interface (CAT)
IS	Styled Layer Descriptor (SLD)
IS	Web Feature Service (WFS)
IS	Web Map Service (WMS 1.1.1)
IS	Filter Encoding (Filter)
IS	Grid Coverages (GC)
IS	Coordinate Transformation Services (CT)
IS	Simple Features - CORBA (SFC)
IS	Simple Features – OLE/COM (SFO)
IS	Simple Features - SQL (SFS)
AS	Topic 0 - Overview
AS	Topic 1 - Feature Geometry
AS	Topic 2 - Spatial Referencing by Coordinates
AS	Topic 3 - Locational Geometry Structures
AS	Topic 4 - Stored Functions and Interpolation
AS	Topic 5 - Features
AS	Topic 6 - The Coverage Type
AS	Topic 7 - Earth Imagery
AS	Topic 8 - Relationships Between Features
AS	Topic 10 - Feature Collections
AS	Topic 11 - Metadata
AS	Topic 12 - The OpenGIS Service Architecture
AS	Topic 13 - Catalog Services
AS	Topic 14 - Semantics and Information Communities
AS	Topic 15 - Image Exploitation Services
AS	Topic 16 - Image Coordinate Transformation Services
AS	Topic 17 - Location Based Mobile Services
AS	Topic D1 - Telecommunications Domain

Annex E - Example National Geospatial Standards

United Kingdom

UK **GEMINI** Standard version 1.0 (2004-10-12) ISO 19115 compatible, eGMS compatible, superseded older NGDF standard (FGDC-based)	http://www.gigateway.org.uk/metadata/pdf/UK_GEMINI_v1.pdf
eGMS eGovernment Metadata Standard - (based on Dublin Core/ISO 15836 with UK eGov extensions)	http://www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=872

Australia - New Zealand Spatial Metadata

ANZLIC - Australia New Zealand Land Information Council - Core Metadata Elements for Land and Geographic Directories ANZLIC **Metadata Guide**	http://www.anzlic.org.au/infrastructure_metadata.html http://www.anzlic.org.au/download.html?oid=2358011755

Canada

Canadian (CGDI) Metadata Standards uses FGDC CSGDM; converting to ISO 19115 standard	http://www.geoconnections.org/CGDI.cfm/fuseaction/technical.indextospecs/gcs.cfm http://www.geoconnections.org/CGDI.cfm/fuseaction/technical.metadata_for_geodata/gcs.cfm