Spatial Data Infrastructure

Henri J. G. L. Aalders
Delft University of Technology
Department of Geodetic Engineering
E-mail: h.j.g.l.aalders@geo.tudelft.nl

Katholieke Universiteit Leuven
Faculty of Engineering

Harold Moellering
Ohio State University
Faculty of Geography
GEOHAL@OSU.EDU

Keywords: Spatial data, Information and Communication Technology, Analytical Cartography, GIS, Standardisation.

Summary
In the analogue era, concern for spatial data and its distribution were tasks for mapping organisations earning a reputation for producing quality products based on their spatial characteristics, including their visualisation in printed map form to be used for many different types of problems.
In contrast, from the second half of the twentieth century onwards, with the developments in computer technology, digital cartography end GIS were also developed to capture, store and analyse spatial data, replacing the tedious analogue map making process.
Nowadays, there is a rapid growth of the availability of digital spatial data and a growing need to use it for all kinds of applications in the field of spatial activities. In addition, with the development of today's communication technology, it becomes possible for every user, surfing the Internet, to collect datasets from a variety of sources and different types of application. This requires translation from the original source data into the user's system, to make the received data user understandable and usable.

SDI, Standards and Metadata
Data transfer from one database to another, requires a spatial data standardisation [Moellering 1991, Moellering and Hogan 1997] in order to allow data providers to store their data into a commonly defined standard way that can be interpreted by the data receiver to display the data meaningful. With the growing number of users, nationally and internationally, the information about what data is available becomes an important knowledge. In many countries, governments, providers and private organisations have been building digital infra-structural networks to enable development of the countrywide spatial data and information flow.
Provisions for the distribution of spatial (or geographic) data can also be viewed as an infrastructure (SDI, Spatial Data Infrastructure). Providers of such data often developed a common site in the Internet to display the type of data they have available, in order to display their products and promote its use. Sometimes, they also make the data available through these channels. In order to understand each other's descriptions also the description of the datasets were standardised: called metadatasets [Mouldering and Crane 2001].
Globally, GSDI aims at the linking of national and international SDI into a global and open process for the co-ordination, organisation management and use of spatial data and related activities. The focus of GSDI is rather on implementation of SDIs rather than on research. GSDI has published a "Cookbook" for implementing SDIs, which will be published on the Internet (http://www.gsdi.org).
Using the forthcoming GSDI Cookbook may help each new SDI implementation to consider its capabilities and legal implications. A comparison of the aspects of the different implementations in different countries – that could be researched by the ICA Commission on Spatial Data Standards in future - will enhance these implementations. Hence, this article will summarise and review these aspects for the ICA Commission on Spatial Data Standards to potentially research the existing SDIs including the technological, legal, administrative, financial and organisational aspects.

Background Information

Both in the governmental as well as in the private sector spatial (geographic) data can be used for decision making and planning. Therefore, in many countries spatial data infrastructures (SDIs) are developed to facilitate the availability and access to spatial data for all levels of government, the commercial and non-profit sector, academia and citizens. It is believed that a SDI will provide spatial information to providers and users with the necessary spatial information background to

evaluate implement, or participate within the growing digital spatial information society for planning and decision making purposes.
The development of national SDI may be divided into two categories [after Masser, 1998]:

- those that are a result of a formal mandate from the government. (e.g. the US National SDI - http://www.fgdc.gov - by Presidential Executive Order in April 1994 and the Portuguese National Geographic Information System - http://www.cnig.pt);
- those that are largely grown out of existing national spatial information co-ordination activities (e.g. the Australia/New Zealand Land Information Council development, relating SDI to other information infra structures (http://www.auslig.gov.au) and the co-ordination amongst participants in the Dutch Ravi to develop a Netherlands Clearinghouse for Geographic Information, http://www.ncgi.nl).

Although the need for spatial data in the society is very large, to make best use of available spatial data the following aspects should be considered:

- use of standards to transfer data from the provider to the user taking in account the existing standards already developed for this purpose. The GSDI Cookbook (http://www.gsdi.org) states that standards should be used and it is advisable to use existing national and international standards rather than developing new ones. This may include several types of standards:
    - **standards for metadata** on order to enable the *discovery* of existing data finding out what datasets are available, and than *explore* these datasets to identify whether or not the data in the available datasets are sufficient for the apparent application. After the correct datasets for the application are found also the *exploitation* of the data should become possible, i.e. what is the process for obtaining and using the required data;
    - **standards for spatial datasets** transfer. The use of spatial datasets is of two types. Firstly after having found the applicable dataset one may create a data transfer from the provider to the user. In this situation the provider has no control over his delivered data. He looses responsibility while the user is stuck with the dataset at the moment of delivery lacking later updates, which is typical also in the "off-line" situation where users will receive data by CD-ROM or tape.
      Secondly, in the other situation, users are allowed to look into the provider's dataset - and maybe also combining this data with data from other providers at the same time - without transferring the dataset but merely transferring the resulting images from the dataset (s). In this way the provider keeps the responsibility over the data while the user will always use the latest available update of the data. This technology is often referred to as Open Data Access, Online-Mapping or Web mapping (whether or not using OpenGIS developments) applying on-line catalog and mapping services and techniques. Integration of datasets from different sources require to convert geometric reference systems towards each other, combine data definition through catalogue services and solve conflicting presentation schemes;
- Internet based communication technology to enable swift transfer of data instead of using the relatively slow postal service by sending a floppy or CD-ROM containing the required datasets. Due to the expected intensive use of the Internet for these purposes, high speed, gigaport connections are required using broad-band and optical fibre communication tools to send large spatial datasets on request;
- supportive organisational strategies and policies with free or low-cast software solutions based on these standards. There are a number of policy conflicts concerning pricing and organisational structures. For example there is the philosophical difference between the USA (with a federal government) and Europe (with separate states) concerning the pricing policy for public access to spatial data. Furthermore the debate about payments for the cost of

developing and upholding a SDI by the providers or the users is not yet finished, apart from the policy of promoting data sharing.
Four types of organisational models exists [Brand, 1998]:

- 'governmental oriented organisations' as CNIG (Comité National de Information Géographique) in France;
- 'business oriented organisations' like the OpenGIS Consortium (http://www.opengis.org);
- 'umbrella type organisations' both on international/regional level like EUROGI (European Umbrella Organisation for Geographic Information) or on national level like the Dutch Ravi (Overlegorgaan voor Vastgoed Informatie);
- 'national professional organisations' like AGI in the UK, ProGIS in Finland or AFIGÉO in France.

History development of SDI

In many countries a spatial information infrastructure is developed to enable users to discover, explore and exploit datasets according to the needs in their applications. These can be seen as first generation information infrastructures; [Masser, 1998] gives an overview of the state of the art. These systems are mostly explicitly nationally organised, and deal with spatial information only. During the UN Conference on Environment and Development in Rio de Jeneiro in 1992 a major resolution was accepted to reverse the impact caused by environmental deterioration e.g.: deforestation, pollution, depletion of fish stocks, monitor and control of toxic waste, etc. Also, the GATT summit in 1999 in Vienna emphasised on the availability of information access over the Internet all over the world. The importance of spatial information to support decision making and management at local, national, regional and global level becomes apparent in many applications and mostly spatial information plays an important role. The capture of spatial data is expensive, forcing the reuse of the same data several times for different applications. And so serving the community through a spatial data network to discover, explore and exploit the available data is developed, firstly at the national level but after the Rio Summit also on a global level.

Scope of a SDI

A SDI facilitates information providers and users to participate in the growing (digital) spatial community at national level. The GSDI links national SDIs to establish connection for all users in the world to share and reuse the available datasets. SDI is often used to denote the relevant basic collection of technologies, policies and institutional arrangements to facilitate the availability of and access to spatial data. The SDI provides a base for spatial data discovery, -conveyance, -evaluation and -application for users. Infrastructure indicates a technology facilitating the conveyance of packages of spatial information virtually unlimited in size.
SDI includes data and attributes, sufficient documentation (metadata), means for data discovery (spatial data query), visualise and evaluate the data using catalogue services and Web mapping and software tools for accessing spatial data.
To make a SDI functional, it must also include the organisational agreements needed to co-ordinate and administer on local, national, regional (and international, in the case of a GSDI) scale. It also should consider the legal aspect to enable or restrict data use.

In many cases the development and research for national SDIs are in an isolated environment, mostly driven by national or federal government. However, often the need for a wide participation and public-private partnerships are encouraged. Regional SDI , being international from scope are encouraged by regional organised bodies like the ESMI project in Europe (http://www.esmi.org) , promoted by the EU in the 1997 GI Meta study (http://europa.eu.int) in order to better understand the requirements for metadata services within the European spatial

information sector. As indicated in the UN conferences on Environment and development as well as during GATT summit the development of a global SDI (GSDI) becomes apparent, linking regional and national
SDIs in order to establish the availability of data all around the world.

Status of SDI development

Studies in the field of SDI list several types of products such as catalogues, indexes and paper directories, CD-ROMs and a growing service for on-line metadata. Nevertheless, there is still a lack of awareness of what a metadata service is and how it can be used. Especially in Europe a multilingual interface is a requirement to use any SDI service, something that maybe of less importance in other regions of the world. Also the metadata service may be:

1. generic, with simple object descriptions of the dataset's content for the general public or
2. specific, with a detailed description of fewer datasets only of interest for professional.

Four types of levels obstruct interoperability:

1. cross-border matching of datasets;
2. cross-sector combinations of datasets from different sectors and different content;
3. cross-type of data e.g. combining raster with vector data;
4. overlap where the same objects coming from different sources contradict both in value and in representation.

Resolving these related issues need a mix of three ingredients:

1. technology;
2. common understanding of data and metadata and the acceptance to mark specific data as 'core data' (requiring specific, generally accepted data definition) to be used as reference in all applications;
3. political support both from the public and private sector.

Metadata standards contents

Meta in the word metadata comes from Greek and means 'change'. Metadata, being always translated by 'data about data', describes the origin of data and tracks its changes. And so, metadata gives the summary of characteristics of a dataset. Metadata is not limited to the description of spatial datasets and has been in use for a long time in libraries and museums to serve as collection management and resource discovery tool. The use of metadata standards in the spatial information scene is useful because experts from different backgrounds develop them in a lengthy discussion process and so they may be seen as a result of a consensus between them, covering many aspects. Using standards allow users to compare and evaluate datasets in a consistent way. Basically, metadata of spatial datasets describe the What, Who, Where, Why, When and How of the data.
Unfortunately, several developments of metadata standards has been undertaken causing debates on metadata and the characteristics that describe datasets at best:

1. Content Standard for Digital Geospatial Metadata, US 1994.
   In the USA the Federal Geographic Data Committee (FGDC) approved this standard. This metadata standard was used for the NSDI. Besides in the USA, this standard has also been adopted and implemented in Canada, the United Kingdom (through the National Geographic Data Framework), the South African Spatial Data Discovery Facility, America countries as Costa Rica, El Salvador, Mexico and Urugay and Asia (Philippines, Japan and Singapore) and elsewhere.
2. CEN (Comité Européen de Normalisation) ENV (Euro Norme Voluntaire) 12657 Geographic Information - Data Description - Metadata was adopted in 1998. The norm is used for several metadata services in Europe as well as projects in GDDD, ESMI and national SDIs as NCGI

in the Netherlands but also in Belgium, Croatia, the Czech Republic, Denmark and Iceland, Germany, Poland, Slovenia an the United Kingdom. Many of them plan to adopt to the international standard:

3. ISO 211 draft standard 19115, (http://www.statkart.no/isotc211/welcome.html) which is not only the metadata standard for the international community but is also adopted by the OpenGIS Consortium.

These standards have different approaches about the characteristics that should be included. Sometimes users need less data, sometimes they want more descriptions about the datasets content.

Metadata may exists at different levels of abstraction:

- collection level, e.g. descriptions of the characteristics of whole topographic or thematic maps series or datasets;
- product level, e.g. describing the characteristics of photographic images or mosaics;
- data unit level, e.g. giving information about the way a dataset is organised as a vector or raster dataset;
- object group level, e.g. describing attributes and class characteristics of similar objects in a dataset for instance all buildings in a dataset;
- instance level, e.g. describing characteristics specific instances of an object appearing in a database as the descriptions of a specific road.

Therefor, metadata should vary according to the purpose and make distinction between core metadata, a set of limited metadata and full descriptions of metadata. In this respect the Dublin Core Set for metadata indicates the basic aspects of a metadata required for any set but specifically for spatial data. Based on this minimal dataset and according to [Moellering et all, 2001], at least the following elements are seen as minimal metadata to describe a dataset:

1) *identification* of the dataset by a *name* given to the dataset by the original producer or publisher, being unique to distinguish the dataset from others. This might also be done by an unambiguous *code* for the dataset;
2) *publisher*, providing the organisation responsible for making the dataset available in its present form, such as a publishing house, topographic service, private company, municipality, etc.;
3) *author* or *original producer* which is the organisation responsible for the original capture of the data in the dataset;
4) *other contributors*. A dataset may consist of several subsets for which the data is gathered by different organisations; their source should also be mentioned;
5) *reference systems* for spatial location reference system, semantic definition internally in the metadata or by an external thesaurus and the definition of time;
6) *extent*, describing the spatial, semantic and temporal coverage of data in the dataset. The spatial coverage may be a description by a set of quadrangles or spatially defined by a bounding polygon. The semantic coverage may be a list of the type of objects and semantic attributes, while the temporal coverage indicates the time period of the validity of data in the dataset;
7) textual *description* of the content of the dataset; the description may include samples in pictures, video or sound;
8) *date*, indicating the date of validity of the data in the dataset. Many ways of writing dates are possible but if used they should be written in a clear and unambiguous manner, e.g. as EN 28601, ISO 10303 – 41, etc.;
9) *language*, of the metadata, especially when the metadata is given in another language as the user understands, it becomes important to introduce a translator;

10) *format* of the dataset transfer, used to identify the software that will be required for reading the data. Standard formats should be used as indicated in standards as listed in [Moellering, 1991 and Moellering and Hogan 1997] or others as CEN PrEN 12658, ISO 10303 - 21;
11) *quality*, describing the spatial, semantic and temporal quality parameters for the dataset. In the quality definition of larger datasets information should be available of the meta-quality indicating the quality indicators e.g. for spatial quality a relative standard error should known as 'average', 'maximum', 'minimum', 'expected', 'required', etc.;
12) *relation* to other datasets that in itself may be datasets too;
13) rights and management indicating the copyrights, constraints on use, way of distribution (e.g. by tiles, by square kilometres, or by polygon as far as the coverage is concerned, but also whether a thematic subset can be delivered or an update since a certain moment in time, etc.).

When creating datasets, users often do not describe their content but merely discuss the capture methods and content and so the data collectors have the metadata in their mind (sometimes it takes years to train novices in the field the metadata descriptions e.g. soil surveyors). With the introduction of the digital era, interpreting metadata by software tools require the creation of digitally available metadata sets. People and organisations often do not find the time to document their metadata or find it too expensive.
Extensible Markup Language (XML) is a capable language for structural definition documentation including metadata. It is used in ISO/TC 211 on Geometics and in the OGC standardisation process. Besides the XML Style Language (XSL) produces style sheet for standardised presentations of dataset content, allowing a structured transfer format of (meta-) data. The metadata standards mentioned above, are content standards and do not prescribe the layout of metadata files. Due to the complexity of these standards almost anything can be conceptually stored in files that conform to these standards, making the conformance to these standards less effective for users. In fact metadata should be comparable with other metadata by software tools to index, search and retrieve documents over the Internet. So metadata must be machine readable and interoperable, meaning to operate with clearinghouse software though openGIS Catalog Services.

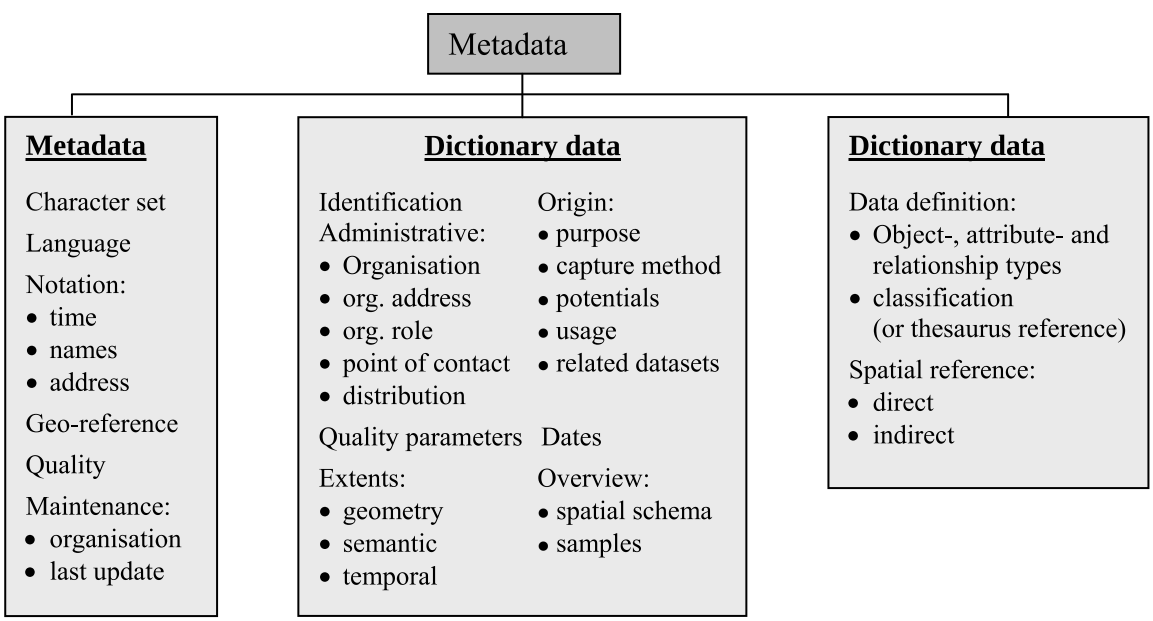

Fig. 1. Contents of Metadata.

SDI framework mechanisms

At national level 'core spatial data' are defined through national agreements by the GI community. Organisations, interested in implementing spatial datasets, must identify and accept these definitions. A framework to build core datasets of spatial data for users has the following aspects:

- specific dataset and their content specifications;
- procedures, technology and guidelines to provide easy access and integration of datasets;
- organisational relationships to encourage the updating and use of the core datasets.

In a networked world, the ability for software to interact with spatial information only exists where public agreements are accepted for data models and definitions. Software can only access objects in remote databases when a common understanding about the nature and composition is achieved – known as semantic compatibility. This can be done by standards for object cataloguing or for common object identifier as some national framework implementations prescribe. Common identifiers allow different representations of the same object (e.g. raster or vector representation of different graphic attributes for different visualisations), while they still can be linked to the same object.

In a national framework various datasets (sometimes also indicated by authentic datasets after they have become a legal status) may be included, such as:

- co-ordinate reference systems;
- object catalogues including object definitions;
- geographic name gazetteers;
- cadastral data ;
- large scale topographic data;
- elevation data;
- administrative subdivisions;
- transportation networks.

A global framework may consist (of combinations) of national core datasets or regional datasets. For these also common reference systems, object catalogue gazetteers for different types of data need to be defined to allow combined use of the different datasets.

Data discovery and access for spatial information is achieved through catalogue gateways and catalogue services. The catalogue gateway contains a set of catalogue servers that have sets of metadata. The metadata files contain instructions of the manner to access the datasets itself. And they all have managers to update the respective information. Therefor, the data access through catalogue servers and metadata has three roles:

1. documenting and locating the data;
2. documenting the content and structures of the datamodel;
3. providing the end-users with detailed information on its appropriate use.

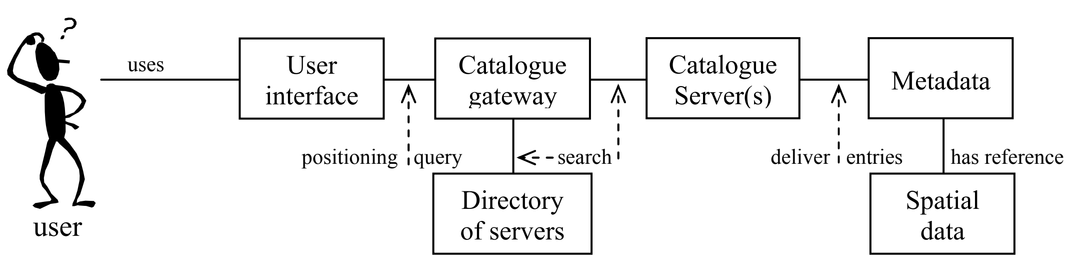

Fig.2. Interaction diagram with basic use of catalogue service and SDI elements.

Framework mechanisms also include:

- conceptual generalisation in order to enable display of spatial data at different scales, by changing the cartographic and spatial characteristics of objects;
- feature extraction to obtain objects from remote sensing and scanned images.
- co-ordinate transformations in order to enable combinations of data using different co-ordinate systems;
- annotation services to add ancillary information augmenting the interpretation of a dataset representation;
- image manipulations to change size, colour and contrast of an image and to conduct mathematical and statistical analyses.
- spatial object manipulations.

Organisational issues

Any SDI can be defined (adopted at the second GSDI Conference in 1998) as a combination of policies, organisational aspects, data technologies, standards delivery mechanisms and financial and human resources necessary to ensure the availability of spatial data to end-users. So, it becomes clear that with the growing demand and availability of spatial data by a network that general issues associated with copyrights, licenses (end-users versus reseller), cost, privacy and use of data formats and standards become larger. A supportive policy and organisational environment is an essential factor in the success of a SDI. Potential participants will only become and stay active if they see the advantages, and do not feel threatened by the infrastructure. The policy /organisational environment varies from country to country and from region to region. All participants, providers and users, both private and public, need to work out a concept, including:

- distributed system to keep the data as close as possible at the source in order to ensure quality of the data;
- being non competitive and collaborative between users and providers, in the public and private sector;
- low level entry barriers both from the supply side as well from the user side. This requires available templates for metadata entry and easy queries for users to access data.

Data access service may offer the following solutions:

- off-line by physical delivery of data by hard or soft copies;
- on-line delivery through email, FTP, etc.;
- brokered, delivering data through request to a third party;
- on-line accessing data by inter-operable datasets and software.

Funding for SDI development

The creation of SDI will depend on partnership both co-operative and financial, among many groups as:

- national mapping agencies play a key role the development and maintenance of spatial data infrastructures. Their datasets are used for many different types of applications and as such, play an important for to facilitate SDI;
- industry providing the technology, data and services. Especially to support information technologies that are consistent with standards as the CEN, ISO, OGC and national standards;
- national and regional SDI initiatives will eventually create to network model for a GSDI.

Funding a SDI does not always have to be in cash but merely in participation. That is to ensure to creation of the infrastructure providing the system with data and advertising the facilities to the users.

Funding can be a major constraint to SDI development and implementation. Therefor, a small project to demonstrate the capabilities of a SDI using Internet available components may result as catalyst to persuade others to join in the initiative before loosing contact with their own market. Also the success stories in other countries may be of use (Australia, Netherlands, UK, US). Apart from private initiatives central governmental funding may accelerate the development of SDIs. Also donor support from other countries, regional bodies or private industries may contribute to the development of SDIs. However one should always consider the local needs and specifications in implementing a SDI by foreign funds. Creation of partnerships by donor and accepting organisations will also lead to the improvement of knowledge about the use of spatial data in general and SDI specifically in the accepting organisation/country resulting in a continuously growing use of spatial data and maintenance of the co-operative developed SDI.

Future developments

With the availability of (G)SDI it becomes possible to use mobile phones and hand-held, palm-top or lap-top computers to access spatial databases solving directions and location problems for any user using WAP technology. The mobile GIS architecture allows usage of satellite positioning systems for the user's apparent location and by entering the destination, querying the appropriate databases through metadata and catalogue services direction indications for route planning may be obtained. Also updating spatial databases becomes possible using mobile GIS systems. And even augmented reality is possible using a real world visualisation on mobile phones or computers showing additional data as real, but invisible objects as cables and pipes, planned objects or add textual information about real world objects.

Therefor, in order to use full advantage of the opportunities of new technology to serve the advancement of the society an (G)SDI is inevitably necessary.

Conclusions

Although a GSDI does not exist yet, many countries are developing their national infrastructure for the dissemination of spatial data. On top the use of spatial data through the Internet is still for experts while many SDIs are developed to enable each citizen to approach, obtain and use the available spatial data.

The development of an infrastructure for the dissemination of spatial data involves many different types of aspects as policies, organisation, data technologies, standards delivery mechanisms and financial and human resources. It would be useful for developments in countries where the infrastructure is not available, as well as where the concern for spatial aspects is missing, to be able to learn from those countries which have undergone the set up solving the difficulties in the different aspects. The ICA Standards Committee, having such experiences in the field of standardisation and metadata, could be of assistance for the making an overview of characteristics of systems to convey digital spatial data. This may be of help for new developments of SDIs as well as being of assistance for the development of a GSDI.

Literature

Brand, M., November 1998
Global Spatial Data Infrastructures: Policy and Organisational issues.
Masser, I., November 1998
The first generation of national geographic information stategies.
Theme papers GSDI Canberra Australia

Bregt, A. en J. Cromvoets 2000
Vlaamse en Nederlandse clearinghouses horen tot de middelmoot (2)
VI Matrix 55 Jaargang 8, nr. 3, mei 2000

Cooper, A.K. and Nielsen, A.S, April 22, 2000
Global Spatial Data Infrastructure White Paper.
Report of the Exploratory Committee for the ICA Commission on Spatial Data Standards.

H. Moellering, editor. 1991
Spatial Database Transfer Standards: Current International Status.
ICA Working Group on Digital Cartographic Transfer Standards.
London, Elsevier Applied Science, ISBN 1-85166-677-X, 235 pp. + index

H. Moellering and R. Hogan, editors. 1997
Spatial Database Transfer Standards 2: Characteristics for Assessing Standards and Full Descriptions of the National and International Standards in the World.
ICA Working Group on Digital Cartographic Transfer Standards.
London, Pergamon/Elsevier Science Ltd., ISBN 0-08-042433-3, 369 pp.+ Table

H. Moellering et all, editors 2001
Spatial Database Transfer Standards 3: Characteristics for assessing metadata standards.
ICA Working Group on Digital Cartographic Transfer Standards. (forthcoming)

Used Internet sites:	*for information about:*
http://www.gsdi.org	Global Spatial Data Infrastructure - Cookbook
http://www.fgdc.gov	USA Spatial Standards and Infrastructure
http://www.cnig.pt	Portuguese Spatial Standards and Infrastructure
http://www.auslig.gov.au	Australian and New Zealand Spatial Standards and Infrastructure
http://www.ncgi.nl	Dutch Spatial Standards and Infrastructure
http://www.opengis.org	Spatial Standards and Infrastructure for the Open GIS Consortium
http://www.statkart.no/isotc211	International Spatial Standards of ISO