Departamento de Ingeniería Topográfica y Cartografía
Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía

The use of metadata for the development of a model of interoperability for Spatial Data Infrastructures

Author: Miguel-Ángel Manso-Callejo (Telecommunications Engineer)

Tutors:

PhD. Mónica Wachowicz

PhD. Miguel Ángel Bernabé Poveda

2009

ACKNOWLEDGEMENTS

This piece of work would not have been possible without the help of many people.

I wish to express here my gratitude to all my colleagues of the Department of Surveying Engineering and Cartography of the Polytechnic University of Madrid, to all my colleagues of the MERCATOR Work Group – Geoinformation Technologies and the Geographic Information Technology Laboratory (LatinGEO) for their support and encouragement.

In addition, I express my gratitude to my supervisors, Dr. Monica Wachowicz for her guidance, support and useful discussions and to Dr. Míguel Ángel Bernabé Poveda for his advice, patience and understanding.

My recognition to the Polytechnic University of Madrid for its Program of Appointments for Visiting Professors which has enabled Dr. Wachowicz to act as supervisor of this thesis, and to the National Geographic Institute for its collaboration.

Finally I thank my parents for their incessant education efforts and their continuous incitement to progress. My gratefulness to Fabiola for her immeasurable love, support and help with grammar and drafting, and to my son who has been of great help with his joy and smiles in critical times.

ABSTRACT

An interoperability model is proposed based on the Spatial Data Infrastructure (SDI) metadata, as well as a method for automatic metadata creation and a methodology allowing analysis of the interoperability provided by them. Metadata constitute an essential piece for SDI's; they catalogue geographic information (GI), describe its characteristics, quality, conditions, etc, and their roles are: discovery, evaluation, access and exploitation of GI. Interoperability is an essential aim for GI to be shared, cooperated, communicated and exchanged in SDI's. The formulation of interoperability models allows analyzing system behavior from different approaches or levels. The lack of interoperability models applied to SDI's, and the lack of studies analyzing the interoperability provided by metadata and of methods of automatic metadata creation constitute the research aims of this thesis.

The proposed interoperability model for SDI's considers the levels defined in the models applied to the systems of systems since they are considered a specific case of those, and an additional level to deal with legal and organizational aspects. In the context of the metadata-based interoperability models, it seems necessary to possess an original method advancing the automatic metadata creation as well as a methodology allowing analysis of the interoperability provided by them. The proposed method to automatically create metadata organize the process of information compilation and handling, it composes and stores metadata in a standard fashion and may be integrated into GI workflows. The analysis of the interoperability provided by the ISO-19115 metadata has allowed their interpretation from an alternative viewpoint different than the traditional function-based approach. The validation of the model with the help of an expert survey has dispelled the uncertainty around the subjectivity of the interoperability identification provided by the metadata. The analysis of the potential interoperability of metadata has resulted in the identification of favored levels (semantic, dynamic and organizational) as well as the deficiencies. Finally, the study of the automatically created metadata with the proposed method has enabled knowledge of their interoperability potential and clearing up whether automatic creation fulfils the requirements of institutions and organizations.

INDEX

FIGURE LIST

TABLE LIST

1 INTRODUCTION

The aim of this research work is to increase the knowledge about the models and levels of interoperability in Spatial Data Infrastructures (SDI) and to substantiate with evidence how manually or automatically generated metadata describing Geographic Information (GI) may help enhance that interoperability. The specific objectives are: (a) to define an interoperability model for SDIs based on interoperability levels provided by metadata and to assess the contribution of every element of a metadata register to each model level; (b) to define a procedure enabling automation of GI metadata creation, and (c) to define a methodology enabling evaluation of the interoperability provided by metadata. By using that methodology and the interoperability model, it will be possible to get information about the intensity and variability of the interoperability levels provided by metadata. Regardless of the way those had been created, different metadata profiles will be able to be analyzed in order to assess their merits and deficiencies at the level of interoperability, so that this will be used in future revisions of the norms to include new elements of a metadata register ('items' from now on) maximizing SDI interoperability.

Throughout this research work an original model of interoperability has been defined based on the general system models, adapted to the particular characteristics of SDIs; an original methodology has been developed for interoperability evaluation based on the content of the metadata items for the SDI context; a new procedure has been developed that enables the automatic creation of metadata; the interoperability model has been validated and finally, the interoperability that the created metadata may provide has been analyzed with the new methodology according to the new model.

The aim of this introductory chapter is to provide an overview on the use of metadata for the development of an interoperability model for SDIs. The first section explains the grounds for the research project and its context. The second section is an in-depth view of the thesis with the description of the research project and its context. In the third section the main objectives are described. The first section explains the motives of the research project and its context. The second section is an in-depth description of the research project. In the third section the main objectives of this thesis are presented, and finally in the fourth section the structure of the thesis is exposed.

1.1 Metadata in the SDI interoperability

The location and access to Web distributed data through the use of specialized search engines implies their cataloguing by means of homogeneous descriptors, i.e. their metadata. For this reason metadata are an essential element of the SDIs and they are regarded both as a necessary tool and a formality to be able to access spatial data either manually or automatically (Najar and Giger 2006), in the latter case, in such a way as to allow interoperability between the systems handling spatial information. Metadata are "data about data" (ANZLIC, 1996; Kildow, 1996; ANZLIC, 1997) that are being widely applied in every type of electronic information resource (Milstead and Feldman, 1999), and in the special case of GI they are used to describe the content, quality, conditions and other characteristics of data so that GI users may access and exploit them.

Historically metadata have played a secondary role; they have been generally created after data production or acquisition since by and large organizations have regarded metadata creation as an additional cost in many cases (Najar, 2006). This view is countered by the following statement: "if you think the cost of metadata creation is too high, it is because you have not calculated the cost of their non-existence: loss of information due to personnel changes, data replication and conflicts, responsibility, bad use and poorly reasoned decision taking" (CGIAR-CSI 2004).

A consequence of metadata being manually generated after the passage of time of creation of the spatial data they describe is the proclivity to errors and the lack of information, leading sometimes to a great difficulty or even inability to carry out the task (Beard, 1996). The search for tools of automatic metadata creation should be emphasized, which by shortening the time and reducing the efforts of GI cataloguing, should prevent typographic transcription errors as well as the errors of interpretation and treatment of geographic data properties, such as spatial reference system or coordinates.

Interoperability, understood as the ability to exchange information between two systems, may be analyzed from different viewpoints: data, services, applications organizations or others (Williams, 2002, Gordon, 2003). It may be also analyzed at different levels: technological, syntactic, semantic (Tolk, 2003; ISO 19101; Turnitsa, 2006). Within the context of the Geographic Information Systems (GIS), interoperability has become a research objective in this past decade, the aim being the implementation of computer applications oriented to data exchange. Eight interoperability levels have been identified: 1) users and institutions; 2) corporations; 3) application; 4) tools and utilities; 5) intermediaries, 6) data stores; 7) distributed computation, and 8) Web (Goodchild et al. 1997).

The formulation of interoperability models (archetypes in which to structure the objectives or classes of interoperability), enabling to appraisingly measure the connectivity of different interoperability levels, is an essential task for abstraction development. In 2004 NATO defined its first interoperability model to define system architecture infrastructures (NC3SAF) based on the directive for their development. Later the Software Engineering Institute (SEI: Carnegie Mellon University) proposed several interoperability models: Levels of Information System Interoperability (LISI), Levels of Conceptual Interoperability Model (LCIM), System-of-Systems Integration (SOSI) and levels of conceptual interoperability (LCI) (Tolk, 2003). Although there is abundant literature about interoperability models and their respective levels, research on the development of interoperability models for implementation of the SDIs has not yet been carried out (Groot and McLaughlin, 2000; Bernard *et al.*, 2005); no references have been found about the use of the items stored in GI metadata in relation to interoperability models either. Metadata-based interoperability models (MBIM) with application to SDIs are unknown.

1.2 Description of the research project

The main purpose of this research project is to enquire into the use of an MBIM between systems to apply it to the automatic metadata creation for the SDIs.

At present two concepts and forms of GI presentation may be distinguished in the literature:

- As a set of rules included in the computer applications handling GI to describe the internal structure and schemes of the data (Codd, 1990; Korth and Silberschatz, 1991; Wilson, 1998) and,
- As independent products associated to the geographic datasets providing information to data and service catalogues and SDIs, in such a way as to enable sharing of data, in addition organizing and keeping a data inventory or providing information making its transfer and use possible (Phillips, 1998; FGDC, 2000; Najar, 2006).

The working hypothesis is that metadata is the support that allows defining the different interoperability levels for SDIs with efficiency, completeness and precision.

The innovation of the research work lies in: the definition of an interoperability model for SDIs; the analysis of the interoperability provided by the ISO 19115 International Metadata Standard; and the definition of a new methodology of GI metadata automatic creation.

The research contributions made are:

- The design of a new model of interoperability between systems applicable to SDIs. The common aspects of the systems and the aspects related to organizations have both been considered in the design.
- The design of a new procedure for automated GI metadata creation suitable to be integrated into the metadata creation workflow.
- The study of GI metadata from the point of view of interoperability, contributing a methodology of interoperability analysis favored by the items of the ISO 19115 International Metadata Standard, applicable to the different profiles made from it and extensible to new standards.
- The application of the metadata automatic creation procedure to different GI types (raster, vector, databases) and different storage formats (the scope of the procedure at the level of metadata items that may be theoretically generated and added for the different GI typologies, is objectively described).

- The assessment of the procedure for automated metadata creation from the interoperability viewpoint (the created items are analysed taking into account the interoperability levels of the model, and the results are compared with the maximum possible for the standard and the items of its core).

1.3 Issues intended to be addressed by the research work

This thesis answers and extends the original motivation of the research work, the main purpose being to define an interoperability model appropriate for SDIs, analysing the ability of metadata thereon to provide the different levels of the model and studying how the automatically generated metadata provide interoperability, also obviating the monotony of its creation and the unavoidable errors.

The applicability of the general interoperability models defined for systems within the SDI context, the usefulness of metadata within the context of interoperability, the existing bias against GI metadata and the difficulty of its manual creation have led me to pose several research questions.

The five research questions around this piece of work are as follows:

- Is it possible to frame a system interoperability model for SDIs?
- What is the contribution of the information contained in metadata in terms of interoperability?
- Is it possible to create useful GI metadata automatically and effectively?
- What proposal is most appropriate to validate a system interoperability model within the context of SDIs?
- What are the strong and weak points of manually and automatically generated metadata from the perspective of the systems that will exploit them (SDI)?

From this list of research questions it may be inferred that the thesis focuses on methodological issues and that it is an applied piece of research work. In short, the main purpose of this thesis is the design, development and assessment of an interoperability model based on automatically created metadata within the SDI context. This objective implies: (a) designing an assessment methodology of interoperability provided by metadata; (b) designing a procedure for automation of GI metadata creation.

1.4 Structure of the thesis

To finish this introduction (Chapter 1), the structure of this thesis is described. Chapter 2 deals with the state-of-the art and knowledge about metadata. Chapter 3 deals with interoperability. Next, in Chapter 4, answering the three first research questions, the procedure of automatic metadata creation is dealt with; the interoperability model for SDIs is defined and the classification of the items of ISO 19115 International Metadata Standard is carried out according to the interoperability model. Chapter 5, answering the fourth and fifth research questions, presents the implementation and the results of applying the procedure of automated metadata creation, the analysis of the interoperability provided by them and the validation of the model. Chapter 6 concludes the thesis with the answers to the research questions and a proposal for future work.

2 METADATA

2.1 Introduction

Metadata are an essential part for SDIs; they describe resources (data, services and other objects) and enable users and applications to search for them in catalogs. Metadata may be considered a formality and a necessary tool enabling access to data and services automatically (Najar and Giger, 2006), thus providing the interoperability of the systems that use spatial information. Metadata are defined as "data about data" (ANZLIC, 1996; Kildow, 1996; ANZLIC, 1997) and they are used widely in all types of electronic information resources (Milstead and Feldman, 1999); in addition, they are used in the GI domain to describe the content, quality, conditions and other characteristics of data. In this chapter a bibliographic revision about the term 'metadata' is made, the roles they play and the different forms of being stored, perceived and dealt with. The procedures of metadata creation have also been reviewed, with special emphasis on the automatic methods.

2.2 Definitions of metadata

Next different definitions of the term 'metadata' are presented. These definitions express the different viewpoints attributed to the term, including those searching for its meaning in its roots, the definitions coined in the context of Biblioteconomy, the current definitions related to the Internet and the semantic Web and finally those directly related to the GI.

Metadata (*metadata*, from the Greek μετα, 'beyond' and from 'data', plural form of the Latin 'datum-i', 'what is given', 'data'), which literally means 'beyond the data', describe other data, with the general understanding that a set of metadata describes a dataset or a group of resources.

Lack E. Myers (1969) coined the term '*metadata*' to describe sets of data or products. The term appeared in print for the first time in a booklet in 1973. From then on both 'metadata' and 'meta-data' have been adopted by the knowledge domains of Computer Science, Statistics, Databases and Biblioteconomy with the meaning of ‘information describing data’ (the expression ‘data about data’ has also become very popular). From this approach, the term ‘metadata’ encompasses all the attributes of the data describing them, they provide a context, they indicate quality or they just document the characteristics of an object or data.

From a computational viewpoint, metadata are regarded as a set of rules included in the applications of GI handling that describe the internal structure of the data schemes (Codd, 1990; Korth and Siberschatz, 1991; Wilson, 1998).

Some authors (Milstead and Feldman, 1999; Caplan, 1995) distinguish between the action of creating metadata and cataloging resources. They identify the main difference in the metadata and they indicate that they only refer to electronic information.

For Caplan (1995) the term ‘metadata’ is used in a neutral way to set aside the possible bias those people kept away from the library world may have, placing all professional sectors involved in its development on equal footing.

In Biblink’s report (Heery, 1996) metadata is defined as information about a publication supplementing its content. It includes a bibliographic description and it also contains relevant information such as topics, price, usage, etc.

Ercegovac (1999) states that metadata describes the attributes of a resource, taking into account that the resource may consist of a bibliographic object, archival registers and inventories, geospatial objects, visual or museum resources or software implementations. Although metadata may present different levels of specificity or structure, the main objective is always the same: to describe, identify and define a resource in order to retrieve, filter and inform about conditions of use, authentication and evaluation, preservation and interoperability.

Some authors such as Sheldon (2001) propose the definition ‘information about data’, Steinacker (2001) ‘data about information or Swick (2002) ‘information about information’

Other authors such as Durval *et al.* (2002) and Woodley *et al.* (2003) who contribute specific definitions for the term '*metadata*' for certain types of information resources in physical or digital form (archives, museums, libraries), emphasize the roles played by metadata and suggest the definition 'structured data about data' for that term.

From the GI perspective, the SDI Cookbook v2 indicates that the concept 'metadata' is also familiar to most people handling spatial subjects, e.g. it is underlined that a map legend contains information about the editor, publishing date, type of map, description, spatial references, scale and accuracy among others, all of them being metadata applied to the documentation of geospatial information. It is indicated that information in its many aspects is described in such a way as to be able to answer to questions of the type 'what', 'who', 'where', 'why', 'when' and 'how' about the described data and that the sole important difference is the emphasis on the spatial component – the 'where' aspect.

Finally, for another group of authors metadata may be considered independent products associated to geographic datasets that provide information to data and service catalogs and to the SDIs, so that it may be possible to share data in addition to organize and maintain a data inventory or to provide information allowing the transfer and use of the data (Phillips, 1998; FGDC, 2000; Najar, 2006)

In this thesis the following definition of metadata has been adopted: "Structured dataset describing other data, their internal structure and their services, whose purpose is to increase knowledge about them and answer questions of the type 'what', 'who', 'where', 'when', 'how much' and 'how'. They may also be considered independent products associated to data that allow keeping an inventory of the data, facilitate their publication and query through the catalogs in the SDIs and favor the reutilization of data and the exploitation of the services".

2.3 Functions performed by metadata

GI metadata may be classified according to the role they play or the function they perform. The review of functions as proposed by different authors is coincident in a high percentage. Table 2.1 shows the functions identified for metadata by chronological order and by authors.

	Año	Buscar / localizar / descubrir	Recuperar / extraer / acceder	Transferir	Evaluar / explorar	Administrar	Describir / interpretar	Preservar	Usar / utilizar / explotar	Técnicos	Gestionar	Identificar
Beard	1996	X	X	X	X							
Gilliland-Swetland	2000					X	X	X	X	X		
Danko	2002	X	X		X				X			
Senso y Rosa Piñero	2003					X	X	X	X	X		
Jones y Taylor	2003	X			X				X			
ECNBII	2003	X	X						X			
ISO 19115	2003		X		X				X		X	X
Nebert	2004	X			X				X			
Oosteron	2004	X	X		X				X			
Johnston	2005	X	X		X		X				X	
Ostensen *et al.*	2005	X	X		X				X			
Moellering and Brodeur	2006	X	X		X				X			
Gayatri y Ramachandran	2007	X	X								X	X
Díaz *et al.*	2008	X			X				X			
		11	9	1	10	2	3	2	11	2	3	2

Table 2.1: Functions performed by metadata according to different authors and dates

After having identified the different functions, the definitions contributed by the mentioned authors to describe metadata functions in more detail are presented. The aim is to identify the highest number of common functions present in the literature.

Functions identified by Kate Beard (1996):

- Search: metadata should provide sufficient information either to discover whether there are data of interest within the available data collection or just to know they are there.
- Retrieval: users should be able to acquire information of their interest as provided by metadata. The analogy with a library would be in the procedure to follow in order to get a book. The component retrieving the data from the metadata may be as simple as providing a URL identifying the location of a

digital dataset and as complex as involving security issues or having to carry out a financial transaction to be able to access the information. Within the ‘retrieval function’ must be included the information describing how to locate the contact person, the format of data distribution or any access constraint, as well as information about costs

- Transfer: metadata should provide the necessary information for the users to utilize the retrieved files in their machines. This component would include information about the size of the dataset (and the corresponding metadata), and the logical and physical structure of data and metadata.
- Assessment: this function is perhaps the most complex. The metadata supporting assessment may consist of any information helping the users to determine if data are going to be useful for a certain application.

For Gilliland-Swetland (2000) five metadata types may be distinguished: administrative, descriptive, preservative, for use and technical. The metadata for use include the intellectual property; the last three terms are usually included within the administrative category and a structural type is added. The descriptive metadata are usually identified as those providing location (Caplan, 2003). This author states that metadata describe the characteristics and the functionalities of objects along their life cycle, from creation, the different versions, organization (register, catalog and indexations), search and retrieval, use (reproduction, modification) and preservation (refreshing, migration, integrity check), finishing with their possible elimination.

Senso and Rosa Piñero (2003), in addition to making a literature review related to the definition of the term and the functions performed by metadata, among which the taxonomy of Iannela and Waugh (1997) stand out, they propose the following functions: use, technique, preservation, descriptive and administrative. The definitions proposed for these functions are:

- Use: Level and type of use made of the computer resources.
- Technique: Relative to the operation of the systems or the behavior of the metadata.
- Preservation: To safeguard the information resources.
- Descriptive: Used to represent the information resources.

- Administrative: Used in management and administration of the information resources.

Jones and Taylor (2003) propose the following definitions for metadata functions:

- Discovery in the Web: These metadata provide enough information to discern the content, format and scope of a dataset. In general the information answers the questions ‘what’, ‘who’, ‘where’, ‘how’ and ‘when’; this allows deciding whether the dataset is potentially useful. In order to be able to exploit this functionality of metadata, a site is necessary in which to carry out the searches and locate the existing data.
- Exploration: Once a dataset has been located, its suitability should be assessed to meet the requirements. This *adequacy to the objective* is a function of data quality. The Association for Geographic Information has worked out a set of guidelines to describe the content and quality of GI (Parker et al, 1996); it identifies five aspects of geographic information: completeness, thematic accuracy, temporal accuracy, positional accuracy and logical consistency. Exploration metadata should contain sufficiently detailed descriptions of the five aspects of data quality to allow assessment of the suitability of a dataset.
- Exploitation: The third level of metadata is related to the process of acquiring and using a dataset. These metadata may contain information about data source and use restrictions (Parker et al, 1996). Technical details such as data format will guide the user in the selection of the datasets compatible with his system.

ECNBII (Environmental Canadian National Biological Information Infrastructure Canada, 2003) proposes these functions:

- Search in the Web: Most of the environmental information items will be located at this elementary level; the information may be a collection of simple objects or a database.
- Access: By using the full geospatial profile and/or the biological profile, this level will ensure the detailed description and the location of data.
- Use: This level will allow the use of biological or geospatial metadata at the workstation for their visualization and extraction through Web services.

Nebert (2004) approaches metadata functions as answers to questions by the users:

- Location metadata: What datasets contain the type of data I am interested in? This allows organizations knowing and publicizing the portfolio of available data.
- Exploration metadata: Do the identified data contain sufficient information to carry out a reasonable analysis according to objectives? That is the documentation that should be provided to ensure that others use data correctly and wisely.
- Exploitation metadata: What is the process of acquiring and using the required data? This helps the end users and the providing organizations to efficiently store, reuse, maintain and file their data collections.

Danko (2004), Oosteron (2004), Ostensen y Danko (2005) address metadata functions from the viewpoint of the items added to metadata to allow development of their functions:

- Locate: These are metadata items enabling users to locate the GI they are searching for and enabling producers to publicize their data. They enable organizations to locate external data and to search for partners with whom to share information capture and maintenance costs. In addition, these items favor management, storage, retrieval and reuse of data.
- Evaluation: After data location, other metadata items are needed in order to determine whether the data fit the intended use. Among the items that enable data evaluation are quality and accuracy, spatial and temporal schemas, content and definition of geographic features and the spatial reference systems used.
- Extract: In many instances users need to access data after these have been located and their adequacy has been evaluated. The metadata items for data extraction enable knowledge of location of a dataset, its size, price and use restrictions.
- Manage: Once data are downloaded, users need to know how to manage or handle them. For this reason some additional metadata items are described, among others to know how to fuse and combine data with the users' own data,

how to apply them correctly and fully understand their properties and limitations.

Johnston (2005) proposes the following targets for metadata functions:

- Locate: to find the resources of interest using their descriptions published by their developers, distributors or third parties.
- Assess: allows the user assessing the appropriateness of the resource.
- Access: should provide information about the mechanisms needed to access resources after they have been assessed.
- Interpret: to help users interpret data.
- Manage: may help people in charge of data in other management tasks; they allow them preserving all the information concerning intellectual property and rights.

Moellering and Brodeur (2006):

- Locate: to find the location of a geographic dataset in reference to a specific set of characteristics, e.g. topography of an area; in many cases this process of location takes place in the Web in the setting of an SDI network.
- Assess: to make sure the geographic data of the spatial database have the characteristics desired by the user: accuracy, validity, etc.
- Extract: to transfer the spatial database from its location, usually through the Web, to an appropriate location for the user.
- Apply: to use metadata to successfully process the geographic database, to analyze, perhaps solve an ongoing issue or problem.

Gayatri and Ramachandran (2007) propose:

- To find/locate and access resources: Metadata help locate or discover relevant information according to given criteria. They may also help screen similar types of resources and to separate different resources; they help with specific searches. They may discover information about resource location. In some instances they provide a preview of data as a sample.
- Digital identification: The digital identifiers such as the name of the URL file and the Digital Object Identifier (DOI) which form part of metadata items help in resource identification.

- Management and organization of resources: Metadata help with the organization of several links associated to the resources based on the client's request. These resources are dynamically created from the metadata databases. This helps to easily browse through the acquired information.
- Interoperability: Metadata support interoperability since metadata standards has been defined and there are sharing protocols; the discovery of information resources has been seamlessly integrated. Protocols such as Z39.50 have helped in simultaneous searches. The Open Archives Initiative for Metadata Harvesting (OAI-PMH) has also been of great help.
- Archiving and preservation: Digital objects corrupt and they may easily get altered, hence the need for their conservation. Metadata are a key element to ensure survival of resources and their accessibility in the future (National Information Standards Organization - NISO, 2004).

To finish this review, Díaz *et al.* (2008) identify 3 basic roles for metadata: location, evaluation and support of resource use.

In this thesis the metadata functions adopted have been; *location* of resources, *evaluation* of their suitability for specific purposes, *access* or acquiring of data and finally, *use* of resources. These functions have been selected because they meet the needs and aims of both users and final applications that wish to use data and services; in addition, they seem to be the most important functions according to the literature review carried out.

2.4 Taxonomy of metadata

Different authors classify metadata following various criteria. Next the taxonomies of metadata found in the literature are mentioned, then synthesized and structured.

Jokela (2001) calls *implicit* metadata those strongly attached to data. They may be essential, i.e. those necessary to use the data, e.g. number of rows, columns and bands of an image or type of data compression, or non-essential for the use of data. In the realm of databases, Morgenstern (1998) regards as implicit metadata those that are not declared and depend on the interpretation given to the context of use.

Balfanz (2002) suggests another interpretation of the implicit and explicit concepts. As an example of *explicit* he proposes the name of the data storage format and he regards as implicit those that form part of the geodata (e.g. types of stored geometries, count of rows and columns) or those that can be calculated, e.g. geographic extent.

Díaz *et al.* (2008a) regard as implicit metadata those that can be inferred (obtained from other metadata or from the actual data), linking the methods proposed by Beard and Goodchild. For Beard (1996) inferred metadata are those obtained by defining logical rules that allow deducting values from known data, e.g. if you are in a certain region with a given type of climate and the temperatures are below the typical values, it may be inferred that it is winter. For Goodchild (2007) inferred metadata may be those obtained by data mining or automatic retrieval techniques.

For Wilson (1998), implicit metadata may also be the set of underlying rules that point to the way the data should be handled by the applications. This may be the case of the information and the set of rules implemented in the GIS to manage the spatial reference systems by coordinates.

Codd (1990), Korth and Silberschatz (1991) indicate that for a long time the Database Management Systems (DBMS) have been making use of metadata to describe the internal structure of the data schemas; these metadata could also be regarded as implicit.

As to the life cycle of metadata, Jokela (2001) proposes to class them as *static* and *dynamic*. Static metadata never change their content; dynamic metadata change with the passage of time and they should be refreshed or recalculated. This author also mentions the *temporal* metadata that are created with a certain purpose and after a period of time are disposed of. An example would be the status and programming of the information workflow.

From the point of view of the roles played by metadata, according to the classification of Boll *et al.* (1998) mentioned by Jokela (2001), three categories are proposed: *structural, control* and *descriptive* metadata. For NISO (2004) the previous categories are called *structural, control* and *administrative* metadata. *Structural* metadata are strongly related with the essential metadata necessary to use the information. The *control* metadata are used to manage the content flow; they are data describing whether the content is ready for the next phase, and in some cases they may be regarded as temporal metadata (machine control, service quality and management of errors). According to Jokela (2001) the *descriptive* metadata are in turn classified into *contextual* and *semantic*. These metadata have to do with the aspects concerned with intellectual property and access rights and privileges. *Contextual* metadata concern temporal information and the system used to manage information while *semantic* metadata describe the semantic quality of the contents; they answer questions relative to the meaning of things (e.g. subject, location, names, and keywords). For NISO (2004) the *administrative* metadata are subdivided into intellectual property management metadata and metadata for preservation of the information.

For Durval *et al.* (2002) from the point of view of their storage or form of access, metadata may be:

- Stored in the *resource proper*: by using marks they add value and visibility to the data.
- Stored in *archives coupled* to the resources: They are advantageous in the independent creation of data and they are inconvenient in the simultaneous management of data and the archives storing the data.
- Stored in an *independent repository*, generally in a database. This makes direct queries difficult. Maintenance may also become difficult if it is carried out by an organization not having the control of the data.

Durval *et al.* (2002) also distinguish between *objective* and *subjective* metadata. The *objective* metadata are related to author, date, and all the information stored by some applications, such as text processors, in the form of properties associated to the archives. The *subjective* metadata are those who may be interpreted from different viewpoints (keywords, abstracts) and when they depend on the domain, context or culture.

From the viewpoint of metadata creation, the data, that once structured, form metadata, may be obtained in different ways that define a classification or a taxonomy. In this thesis the following types of metadata have been considered: *implicit* metadata, those that are strongly linked to data and their use; *explicit* metadata, those linked to type of data and storage; *calculated* metadata, those that may be obtained through some type of calculation or treatment; *inferred*, those that may be obtained through logical rules allowing to deduct values from other values; and *contextual*, those that may be obtained or imposed by the context in which they are created: date, application, machine, etc. The remainder of the types has not been considered since they are not related to the automatic creation of metadata.

2.5 Metadata creation procedures

Since

- The manual metadata creation is a burdensome, error prone process (Batcheller, 2008, Wyoming, West and Hess, 2002, Guptill, 1999);
- The automatic procedures of metadata creation cannot provide the information that both data producers and metadata compilers may contribute (Campbell, 2008, Guy *et al.*, 2004, JORUM, 2004);
- The manual metadata creation on the part of the data author or their automatic creation cannot provide the cataloging experience of the experts in information management (Currier *et al.*, 2004, Guy *et al.*, 2004, JORUM 2004);

it is therefore timely to carry out a review of the existing procedures and methodologies of metadata creation (manual, automatic and mixed) to propose then a procedure to automate, so far as possible, metadata creation in the GI domain (geodata).

From the viewpoint of cataloging, Colleman (2002) distinguishes two types: descriptive and subject-oriented. There are several procedures and methodologies of metadata creation proposed in the literature, implemented in the workflows and metadata management and there are also various procedures used in the different disciplines that need them.

Beard (1996) in a workshop on environmental metadata proposed five methods to create or compile metadata:

a) Manually, by keyboard (traditional method);

b) Enlarging the stored information with values obtained through a search in a reference table (e.g. given the geographic extent, searching of a geographic identifier by using a gazetteer);

c) Through automated measurements and observations;

d) Extracting and calculating;

e) Inferring new metadata from other elements.

The five proposed methods may be combined as best as possible to meet needs, for example when dealing with metadata about moving objects that collect environmental information, in addition to position the methods b) and c) may be combined to obtain both static and dynamic complementary information derived from movement. On the other hand, if one wants to create metadata about an existing geographic dataset, the combination of the methods d) and e) may end up being the best way of creating metadata, as Beard (1996) indicates.

Bailer and Schallauer (1998) examine the functions of metadata in the process of production of audiovisual media: pre-production (conception of the product), production (content creation), data and metadata storage, post-production (editing, modification), exploitation (delivery of resources) and visualization. When the authors analyze metadata in the process of creation, they identify as possible information sources: a) the capture (by means of devices capturing data, cameras, GPS, etc); b) the information inherited from the old materials; c) the manual annotations; d) the information extracted from the context and e) the semantic analysis of texts.

Next the information sources are described:

- In the process of capture, the metadata describing the configuration of the device used to capture the photo or scene are important, typically remote sensing and photogrammetry. Other metadata that can be captured are date, time and place (with global or relative coordinates). When an analogical material is being digitized (e.g. a map), the information relative to the quality and defects of the material may be captured.

- The information inherited from old materials, such as abstracts, comments, press articles, books or documentaries may be important to preserve and maintain in the metadata.
- Manual annotations are also very important since they provide semantic abstraction of the meaning; however, it is a very costly process. On some occasions a global description of the content is made.
- The analysis of contents allows creating a large amount of metadata at a low cost; yet the results achieved from extracting that information from the basic and middle levels have not been good enough. This fact is known as "semantic gap". The results may improve if information of the context is available.
- Semantic analysis of texts is the commonly used technique to extract semantics. In this type references to entities with name (persons, organizations, and places), the link with other entities belonging to ontologies, the detection of themes together with the classification of their contents in segments and the link of the elements with inherited or related information are all included.

Balfanz (2002) backs into the automatic creation and the visualization of metadata in the geographic context. For this author the creation of metadata follow certain objectives and the degree of its success should be measured. The type of metadata that can be extracted and calculated automatically is the implicit type. The means to obtain these metadata may be varied; every method should be characterized by an estimate of the required time and an indicator of the quality of results. Thus the operator of the metadata creation and visualization system may select the method best fitting in each case and for each type of need.

Jane Greenberg (2004) identifies two methods of automatic creation of metadata: extraction and harvesting. The extraction uses data mining techniques and indexation to retrieve items or labeled contents. The harvesting uses techniques of collecting already existing labeled contents. Later Greenberg et al (2006) review the tools and the applications developed for the automatic creation of metadata about electronic resources with a certain librarian bias; they state that the use of these automatic methods enables orientation of the human resources effort to more intellectual aspects. Depending on how automatic is the process of creation of metadata or its human requirements, the distinction is made between generators and editors of metadata, which integrate the automatic and human processes.

The Automatic Metadata Generation Applications (*AmeGA*) Project carried out a review of the limitations of the applications generating metadata to propose the desirable functionalities for the applications. The project made an inquiry to learn the opinions of the personnel involved in metadata. It probed automatic creation and the desirable functionalities that enable its creation within the context of the Dublin Core Metadata Initiative (DCMI). The questions of the enquiry concerned the most appropriate option to complete the metadata items (DCMI: 15 items): automatic, semi-automatic and manual. The conclusions of this study point to methods that would make the creation of metadata automatic although most of the respondents conclude that the tools should help the human operator and not replace him. The functionalities most asked for by the respondents were:

- Possibility of executing the automatic algorithms so that a person may be able to evaluate the results, and
- Integration of content standards such as the thesauri.

Concerning the algorithms of automatic metadata creation, the research on automatic indexation based on the use of thesauri in specific domains such as the medical field (Nadkarni *et al.*, 2001), the research on automatic creation of text abstracts (Johnson, 1995) and the research on automatic classification (Losee, 2003) are mentioned.

The EDINA and MIMAS teams, which have participated in the JORUM Project (2004), set on the production of metadata on resources for educational electronic settings (e-learning), identify several models for metadata creation:

- Metadata are created by the author.
- Metadata are created by an information expert.
- Use of a collaborative tool so that the author includes part of the information and the information expert includes another part.
- Metadata are created by another person and somebody belonging to the project validates them before publishing.

Every model has its advantages and inconveniences; for this reasons those teams propose the use of hybrid models.

Guy *et al.* (2004), within the frame of the ARIADNE Project propose a workflow for metadata production in which the fully automatic, semi-automatic and manual creation is anticipated by the author of the data first and then by an expert in information management, as shown in Figure 2.1. That workflow anticipates all possibilities: a) automated creation; b) automated creation enhanced by the author of the data; c) automated creation enhanced by the author and an information expert; d) manual creation by the author and enhanced by an information expert and e) creation by an expert in information management.

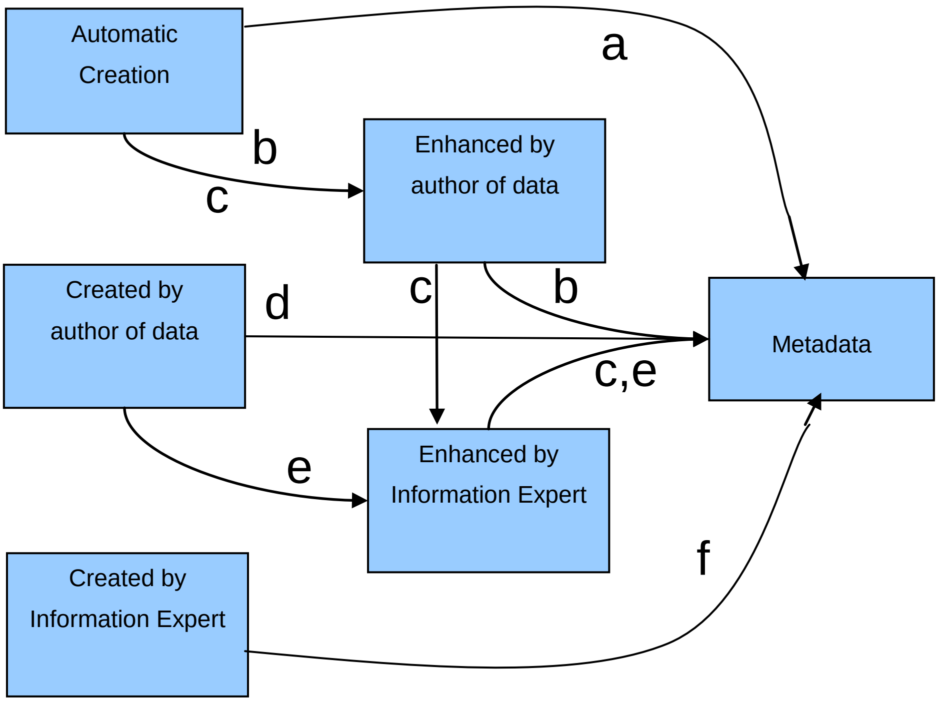


Figure 2.1: Possible methods of metadata creation (Adapted ARIADNE Project)

Baird (2006), in the review of the methods for automation of the existing or potential metadata workflow, identifies three states: contribution, catalog and revision. The JORUM Project, in the review of the workflow of August 2006, supplements those states with a fourth one: publishing. Initially it is about a manual method of metadata creation in which most of the metadata are provided by the creator of the data, then passed on to the flow that manages them. The metadata may be published, though with reservations, while waiting for the cataloger to revise and validate them. This task finished, the metadata will be valid until the time they are reviewed or until the data they describe change.

A similar process of metadata collection in the context of educational electronic contents is proposed by Currier *et al.* (2004): whoever creates the contents also creates the metadata and the information management expert reviews them, supplements them and catalogs them. Baird (2006) proposes certain catalogers' operations and even some operations of the content creation to be automated. It may be stated that the proposal of metadata workflow is a collaborative process.

Morris *et al.* (2007) have proposed a framework for metadata creation (Figure 2.2) in the North Carolina Geospatial Data Archiving Project (NCGDAP). First a template or profile is defined for the organization or institution (a); this template is personalized for a specific geographic data collection (b); metadata are processed to adapt them to the template (c); if there are no metadata, they will be created (d); next the lineage information is added (e) and finally, a process of synchronization with a commercial metadata extraction tool is applied (f). Other authors, e.g. Hedorfer and Bianchin (1999) also suggest the use of templates as a mechanism supporting metadata creation.

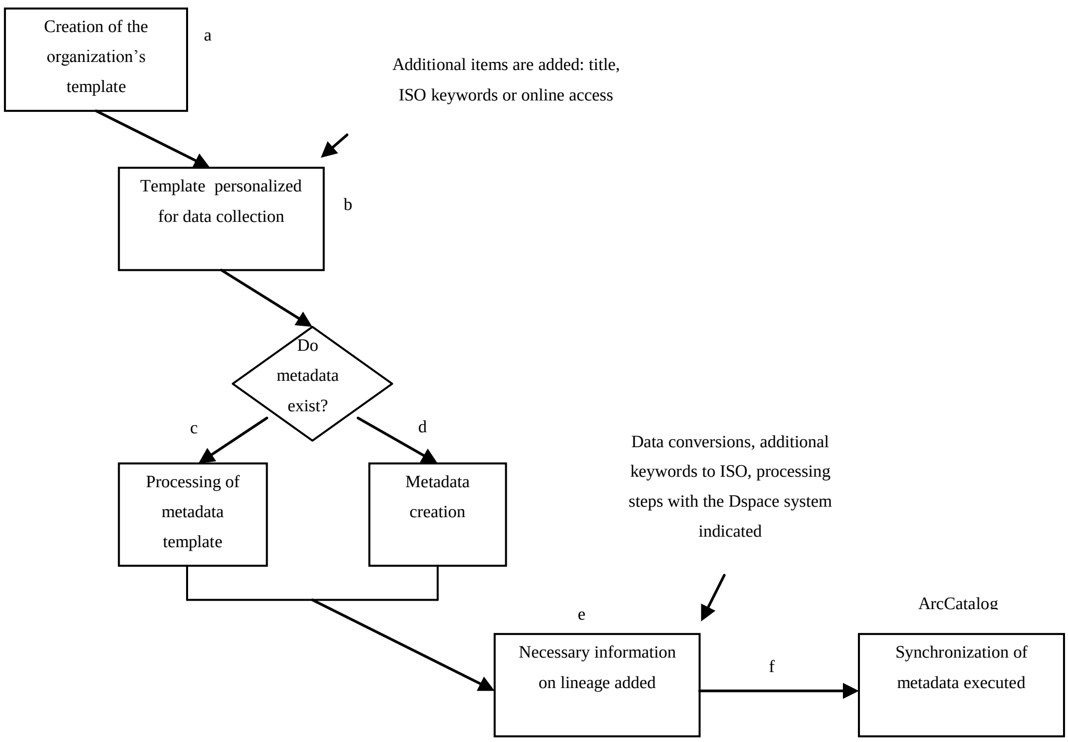


Figure 2.2: Metadata creation workflow (Source: NCGDAP Project)

Mustacoglu (2007), in a research proposal, makes a brief review of the collaborative settings in which it is necessary to describe resources through metadata. *YouTube*, *LibraryThing*, *Flickr* and *43things*, all using tagging profusely, are mentioned as well as *CiteULike*, *Connotea*, *Bibsonomy* or the *Delicious* systems (*del.icio.us*), which he deals with as tools supporting the sharing of annotations about resources and their links (URL). Concerning metadata creation, the author identifies two types or mechanisms: professional (specialized) metadata, working with complex cataloging schemas, and metadata created by the content authors, even if they are inexperienced. In the former case, the lack of scalability of cataloging systems is mentioned as a problem whenever metadata have to be applied to a large number of data on the Web. The metadata created by the authors are vulnerable since the proposed descriptions are often inadequate. These collaborative proposals have been called *folksonomy*; they create metadata through tagging and the actual users are responsible for the metadata. Some advantages of this type of techniques are: a) ease of access and use of the tools; b) ease of discovery of new contents, and c) support of creation of new community niches. Drawbacks are: a) the lack of standards to use keywords; b) difficulty in finding misprints and in detecting synonyms and acronyms, and c) difficulty in using inference hierarchical relationships between the labels since there are no taxonomies.

Batcheller (2008) proposes the customization of commercial tools to generate metadata more efficiently that would have the following functions: a) to harvest the existing metadata items; b) to extract the items of implicit metadata stored together with the data; c) to harvest the metadata templates prepared by the person in charge; d) to integrate the previous data enabling their editing and visualization, and e) to provide import and export tools in standardized formats.

Campbell (2008) proposes a methodology of metadata creation, based on a commercial tool, with three stages: a) extraction; b) investigation, and c) compilation and export.

- The extraction requires the existence of metadata, their subsequent retrieval and placement in files with the appropriate formats, such as Word documents or spreadsheets.

- The investigation requires obtaining the non-existing information; to this end personal interviews should take place with the authors of the data or with persons who were in the organization at the time of creation of metadata.
- The compilation and export take the information retrieved by the previous stages and they mold it appropriately on the document; In addition, the consistency and completeness of the metadata for all datasets should be verified.

The automatic extraction of certain items is carried out by means of the commercial tool ArcCatalog.

2.6 Summary

As mentioned above, metadata are used to describe the context, condition and characteristics of data so that the users can locate them and understand them. It was evident in the definition that metadata have different functions. After having identified the main functions according to different authors, they are synthesized:

- Discovery/location: this is an essential aspect shared by the authors, although with different names. The objective is to find the data of interest. It realizes in the metadata items enabling GI users to locate them and producers to publicize the data.
- Evaluation: in addition to helping to make searches, metadata should also help users to determine whether the described data and their use are apt to meet their expectations.
- Access data: after having found and assessed data applicability, the third step is to transfer them for exploitation together with the own data or with others obtained by this procedure; the data may be accessed directly or a connection may be established to exploit them online. Users need information items in metadata describing the form of access or transfer data for their use.
- Use: the final aim of the search process is the use of the data. Metadata should contain information items enabling the most appropriate use of the data, their adequate fusion and combination; for this reason a full knowledge of their properties, limitations and use restrictions is required.

- Preservation: a smaller number of authors identify the functions of management, storage and preservation. The objective of the metadata items meeting these functions enable organizing and maintaining a data inventory, so as to preserve knowledge; they support organizations both in arbitrating conflicts derived from the inappropriate use of the data and in safeguarding the information resources.

It has been shown both in the definitions of the term metadata and in their functions that part of this information that describe data may be stored or it may be obtained in different ways. As a synthesis of the analysis carried out about metadata taxonomies, it may be concluded that metadata may be classified according to the following criteria:

- From the point of view of their existence, as *implicit* and *explicit*;
- From the point of view of their life cycle, as *static*, *dynamic* and *temporal*;
- Depending on the role they play, as *structural*, *of control* and *descriptive*;
- According to their interpretation, as *objective* and *subjective*;
- According to their storage (or access), as *embedded*, *associated archives* and *external repositories*.

3. INTEROPERABILITY

3.1 Introduction

In order to have a comprehensive idea about the meaning of the term, the different definitions of interoperability proposed by a number of authors and organizations will be reviewed. Since in many cases interoperability is handled in a structured way, several organizations and initiatives have defined different models. These interoperability models and their definitions classify interoperability in levels. Finally the interoperability levels are reviewed and analyzed individually. The purpose of the definition of the different levels is to break up the subject in order to be able to solve it.

To study the interoperability degree achieved, interoperability measurements may be defined. These measurements would allow detection of weaknesses and strengths; metrics are still to be defined.

The objectives to be achieved with these reviews are as follows:

- To define the term interoperability.
- To analyze the existing interoperability models.
- To analyze the interoperability levels identified.
- To carry out a concise review of the measures of interoperability.

This study will be useful to propose a valid and applicable interoperability model in the SDI context answering the first question of the planned research work: Is it possible to establish a system interoperability model for SDIs?

3.2 Value of interoperability for SDIs

SDIs are institutional initiatives supporting access to GI through the Internet; they comprise organizational, cultural, political and technological aspects. These infrastructures may be defined as the set of rules, standards, procedures, guidelines, instructions, policies and technology that enable creating, collecting, processing, storing, maintaining, accessing and using spatial data (Crompvoets and Bregt, 2003). Some GI authors and users regard SDIs as the technological advancement introduced in the traditional GIS that allow them accessing to GI through distributed Web applications (Najar, 2006). At the present time SDIs may be regarded as a process comprising the above-mentioned aspects that may take place in different ways, different speeds, different costs and variable impact (Longhorn, 2008).

The GIS have evolved technologically (Dangermond, 1991; Longley *et al.*, 2000) and an increasing number of corporations are found developing or building computer applications in these settings through the Internet. A consequence of this growth is the increasing number of systems and formats of data in use, although users may not be aware of that multiplicity and heterogeneity. The organizations of standardization involved in the SDIs (ISO, OGC) have the interoperability in its diverse connotations as a main objective.

The term interoperability has many connotations, including the objectives of communication, information exchange, and cooperation and sharing of resources between different types of systems. In fact, the essence of interoperability is to ensure the relations between systems, every relation being a way of sharing, communicating, exchanging and cooperating (Carney *et al.*, 2005). Our attention is focused on the SDIs in which technologies, systems, networks, standards, data, persons, policies, agreements, organizational aspects and the mechanisms of data delivery to final users (GSDI, 2004, p. 8; Georgiadou *et al.*, 2005; Williamson, 2004) should facilitate localization, evaluation, access and use of GI in a transparent way for the users, whether these are human agents or computer applications; in the latter case interoperability is a more restrictive requirement.

In the literature different definitions of the term interoperability may be found, which by and large differ in the description of the relations and system components. On the other hand Georgiadou *et al.* (2005), in agreement with Bernard *et al.* (2005), state SDIs are a special case of specifically GI-oriented Information Infrastructures (II). Béjar *et al.* (2009), in agreement with Maier (1996), propose other frameworks of support of SDIs which they regard as Systems of Systems (SoS) components. The main conclusion of the latter proposal is that the analyzed concepts are similar although they are examined from different perspectives; even if the argument goes that the reference framework of

SoS is a broader concept than that of the II, it is in this reference framework that interoperability for SDIs will be analyzed. Finally, William (2002) and Gordon (2003) propose that interoperability may be studied from different viewpoints: data, services, applications and organizations. For others, such as ISO 19191 (2002) or Tolk (2003) and Turnitsa (2006), interoperability may be analyzed at different levels: technological, syntactic and semantic.

3.3 Definitions of interoperability

The Institute of Electrical & Electronics Engineers (IEEE, 1990) defines interoperability as the ability of two or more systems or components to exchange information and use it.

The security agencies and institutions, Department of Defense, NATO, Alliance Defense Fund and Command and Control Information System Plan (1995/1996) describe interoperability as *"the capability of systems, units or forces to provide services and accept services from other systems, units or forces, interchanged in such a way as to be able to operate jointly and efficiently without altering or degrading information"* (Glossary of the Command and Control Subordinate Systems Study Phase 1 Report) (C2SS WG 1996). The same study provides another two simpler definitions: the ability of an entity to serve another entity and the need of a group to interact in some way with another one.

The Open Geospatial Consortium (OGC 1998) in its document "OpenGIS Guide" defines interoperability in the context of the OGC Specifications as components of computer applications working reciprocally with one another to avoid burdensome, systematic conversion tasks, the hurdles of data imports and exports and the access barriers to the distributed resources imposed by processing settings and the data heterogeneity.

Miller (2000) regards interoperability as a process which organizations are engaged in to ensure that the systems, processes and the actual culture of the organization are optimally managed, so that the opportunities of exchange and reuse of information, both internally and externally, are maximized.

Flater (2002), in the document “Impact of Model-Driven Standards” proposes a definition of interoperability based on two concepts: “compatibility” and “adaptability” (integratability). Interoperability is the process pursuing the integration of systems so that they become compatible. “Adaptability” is the capability for adapting incompatible systems or the data they exchange, so that they can cooperate.

ISO TC211 in the ISO 19101 Standard (2002) defines the Reference Model for the entire group of standards ISO 19100, and contributes the definition of interoperability as the capability of systems or components for exchanging information and being able to ensure a cooperative process between applications.

The interoperability refers to the following capabilities:

- To discover information and processing tools whenever needed regardless of its physical location;
- To understand and use the discovered information and tools without limits due to the platforms to be used either in local or remote contexts.
- To develop processing settings for commercial use without imposing any market limitations by trusts;
- To rely on the information and processing offered by third party infrastructures allowing for the needs of the different market niches without fear of failure when the support infrastructure will mature and evolve.
- To participate in a free, transparent market where goods and services meet consumers’ needs and channels open as the market grows enough to support them.

ISO 19101 describes the aspects of the interoperability between systems and lists the following interoperability levels: network communication protocols, file systems, remote procedure calls and database search and access. Finally the Standard mentions two classes of interoperability: syntactic and semantic.

For Gordon (2003) interoperability is the result of applications, data and solutions lying in different places being capable of sharing information and functionalities correctly, thereby providing added value to a sole product within which all of them will be integrated.

Rawat (2003) proposes the concept of interoperability in the GIS domain as the capability for exchanging GI and data coming from different organizations so that society could apply them to any type of application through the networks.

The Association of Library Collections and Technical Services (ALCTS, 2004) defines interoperability as the capability of two or more systems or components to exchange information and use it without that being a special effort for each of them.

For the National Information Standards Organization (NISO, 2004) interoperability is the capability for exchanging and sharing data between different systems that use different architectures of teams and computer programs, different interfaces and data structures.

Taylor (2004) states that two systems are compatible when both can exchange information and use it without the need for any special treatment, coining the concept "compatibility".

The Police Information Technology Organization (PITO) defines interoperability as the capability of two systems or components to exchange and use the information, and their capability to provide to or get service from other systems in such a way as to use the exchanged services to jointly operate effectively (ALCTS, 2004).

The Dublin Core Metadata Initiative (DCMI) Glossary (Woodley *et al.*, 2005) defines interoperability as the ability of different types of computers, networks, operative systems and computer applications to work jointly and effectively without requiring previous communications, so as to exchange information usefully and in a valid manner. Three aspects of interoperability are identified: syntactic, structural and semantic.

ISO 19119 (2005) proposes a definition of interoperability that is applied to every type of information concerning space and geographic data: *"Geographic interoperability is the capability of the information systems to freely interchange every type of spatial information concerning the Earth, the objects and phenomena taking place above, below and on the Earth surface, and to execute* *programs capable of handling that information cooperatively on communication networks".*

Carney *et al.* (2005) extend the previous definitions focusing on the aim (the objective for interoperability) and the context (the setting in which the system exists).

For this thesis the following definition of interoperability has been adopted: capability of a collection of system components, or a system in an SoS, to exchange and share specific information and treat that information in accordance with a shared semantics with the purpose of achieving a specific objective within a given context. In the SDI context, interoperability should enable users and systems to share data and services by exchanging information according to a set of shared semantic rules depending on different contexts and purposes.

This definition leads to the notion of interoperability models to ensure the realization of interoperability between systems according to the different purposes and contexts.

3.4 Interoperability models

The past and current efforts made to create interoperable computer systems have produced models that help solve the syntactic, semantic and structural heterogeneity of data, service interfaces and metamodels that describe them (Lemmens, 2006).

The origin of the formulation of interoperability models lies in the information systems and the need to integrate them. NATO (2004) in its document NATO C3 System Architecture Framework (NC3SAF) defines a first interoperability model in which the general guidelines for the development of system architectures are proposed. Later the Software Engineering Institute, Carnegie Mellon University (SEI-CMU) defined several models of interoperability between systems among which are Levels of Information System Interoperability (LISI), Levels of Conceptual Interoperability Model (LCIM), System-of-Systems Integration (SOSI) and Levels of Coalition Interoperability (LCI) among others (Tolk, 2003).

From the point of view of interoperability models, GIS building may be approached in different ways. Every approach presents advantages and drawbacks depending on the setting (Lewis *et al.*, 2005). The main advantages of the interoperability models are the ability to define a common vocabulary allowing the analysis and discussion about meaning, providing suggestions concerning the structure of the solutions and support for assessment of new ideas and different options (SEI-CMU).

Every interoperability model proposed in the literature deals with different use purposes. However, most of these models do not take full advantage of some types of systems such as the platform-independent characteristic or the tool interoperability. Besides, the models also have some limitations, especially when they want to be interchanged or more information about the context wants to be expressed, such as occurs with distributed architectures.

At the present time every interoperability model defines a particular taxonomy with the aim of supporting different use purposes or achieving interoperability in different contexts. The concepts or classifications commonly used to define the model taxonomies are the layers, the dimensions, the levels and the areas. Depending on the type of context, one interoperability model or another will be used to build a system. Next some examples of interoperability models are presented that have been successfully applied in contexts foreign to GIS or SDI:

- The most common, traditional perspective on interoperability models has been represented by the Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) Architecture Framework with the LISI Model (1998). The five levels of the LISI Model are: isolated, connected, functional (distributed), domain (integrated) and enterprise (global).
- The Enterprise Interoperability Maturity Model (EIMM) (Athena, 2005) helps detect collaborative processes within the organization. This model defines six areas of interest in the evaluation: business strategy and processes, organization and competences, products and services, systems and technology, legal environment, security and trust, and enterprise modeling, with a maturity scale of five levels: performed, modeled, integrated, interoperable and optimizing.

- In the Organizational Interoperability Maturity Model (OIMM) of Clark and Jones (1999) the stages of progression or evolution of systems, processes or organizations are described depending on how they are defined, built or optimized. The concept of level is intrinsic in this model; it is used to characterize the status of a system or organization. Its levels are: independent, cooperative, collaborative, combined and unified.
- The Organizational Interoperability Agility Model (OIAM) proposed by Kingston *et al.* (2005) focuses on the dynamic aspect of coalition work. In this model, a scalable organizational interoperability has been developed going from least to most agile; this enables to relate a set of levels with a set of attributes and factors. The five levels of the model are static, docile, complacent, open and dynamic.
- The Levels of Conceptual Interoperability Model (LCIM) defined by Tolk (2003) and refined by Tolk and Turnitsa (2006) defines seven interoperability levels: no connectivity, technical, syntactic, semantic, pragmatic, dynamic and conceptual. Subsequently Turnitsa *et al.* (2007) proposed the use of metadata in the design of systems as an integrating or enabling mechanism of interoperability.

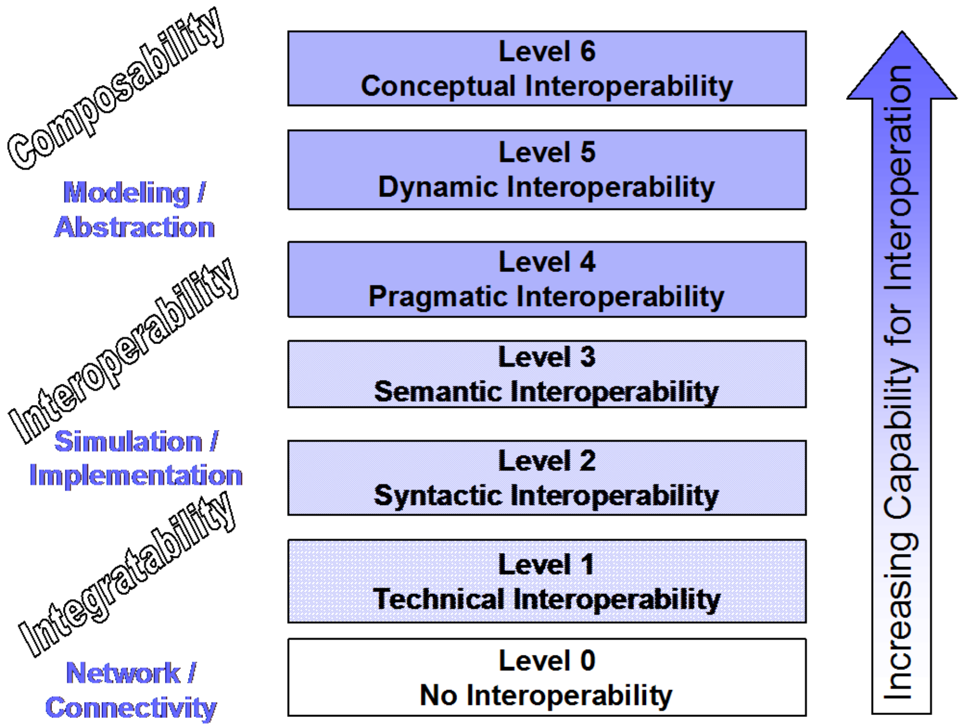


Figure 3.1: LCIM of Turnitsa and Tolk (2006)

In the GIS context, interoperability started to be dealt with a decade ago in the Project Panel "Computational Implementations of Geographic Concepts" (Goodchild *et al.*, 1997). In this panel a schema of eight interoperability levels was adopted: information community and institutions, enterprise, application, tools, middleware, data store, distributed computing environment and network.

In the SDI context, OGC and ISO TC211 are the main players promoting data and system interoperability. In addition to defining data and metadata standards, they are building object models and XML schemas for information storage and transfer, and open processing service interfaces.

Goodchild *et al.* (1997) defined an integrated model of interoperability for GIS. They stated that the prospect of reaching interoperability depends on many factors such as will or predisposition, economic considerations and legal or organizational issues. They described five levels in their model: engineering and networks, technology and platforms, computational architecture and computer applications, data and information conceptual models, and initiative or enterprise. The same discussion group defined an interoperability schema based on eight levels where different organization dimensions are combined, including scale and abstraction, as indicated above and in table 3.1.

For Yasher Bishr (1998), the interoperability model in the GIS context, that tries to avoid semantic barriers in particular, is made up of six levels: protocols, hardware and operative systems, spatial data formats, database management systems, data models and application semantics.

A	Interchanges with	B
Information Communities, Institutions	Policies, values, culture	Information Communities, Institutions
Enterprise	Agreements, consensuses	Enterprise
Application	Cooperation and coordination	Application
Tools	Services	Tools
Middleware	Distributed objects	Middleware
Data store	Data	Data store
Distributed computing environment		Distributed computing environment
Networks		Networks

Table 3.1: Schema of interoperability with eight levels (Goodchild *et al.* 1997)

Intermodel5, proposed by Shanzhen *et al.* (1999) identifies the interoperability problems in the connection and interchange of data between databases and systems. For these authors interoperability is reached by defining the interfaces through specifications and standards. Shanzhen *et al.* (1999) state that research on interoperability is usually made about the procedures for distribution and integration of systems: equality, reciprocity, interchange, diversity, independence and belonging to a domain.

Intermodel5 defines the following interoperability levels: resource location, resource transformation, application service, semantics, and institutional. The authors establish relations between the model and the data or the spatial information industry (SII), in which the system architecture defines four levels: data, technical, operational and institutional (see Table 3.2).

System A	Interoperability model	System B
Institution	Policies, culture, values	Institution
Semantic	Semantics, translators, metadata, GI, Formalization of systems	Semantic
Application or service	Distributed object agents, CORBA, OpenGIS	Application or service
Processing resources	Virtual databases, MultiDatabase, OGC, SDTS data, Warehouse Framework	Processing resources
Location resources	Metadata, digital libraries, catalogues, clearinghouse	Location resources

Table 3.2: Intermodel5, interoperability model of 5 levels (Shanzhen, 1999)

The Project InterOP (IST508011) of an interoperability model for the Information Society defines an interoperability framework distinguishing different viewpoints regarding aspects related to data, services, processes and business; the possible barriers may be classified into conceptual, technological and organizational and finally, regarding the type of approach of the interoperability model (in accordance with ISO 14258), we can talk about unified, integrated and federated.

After having reviewed the definitions of interoperability and identified the models that try to analyze their relations and different aspects, an in-depth review of the interoperability levels proposed or mentioned in the literature will be undertaken, so as to have a comprehensive overview of the issue and of the proposed solutions so far.

As indicated in Section 3.2, the SDIs may be regarded as a particular case of the II and the II in turn are also a particular case of the SoS. In this section about interoperability models, a review has been made of the existing models in the geographic domain (Goodchild *et al.*, 1997, Bishr and Intermodel5), in the SoS domain (LISI and LCIM) and in the organizational context (OIAM and OIMM).

Every model tries to respond to a set of interoperability needs, yet without studying more deeply the interoperability levels proposed in the literature, selecting a model applicable to the SDIs would be subjective and unjustified. Once the interoperability model to be used in the SDI context has been identified and its levels have been justified, analysis of how metadata may favor interoperability is intended, such as Tolk, Diallo and Turnitsa (2007) had proposed.

3.5 Interoperability levels

The interoperability levels of a model are the concepts, abstractions or categories distinguishing a taxonomy related to interoperability. These levels are used to define the capabilities systems must comply with in order to reach a certain degree of interoperability.

The literature review shows quite a number of levels: semantic, technical, legal, organizational and others, as can be seen in Table 3.3. In spite of the many levels described, it appears difficult to get an understanding of them individually, since the proposed classifications may be similar or related, and the definitions of different levels may have common characteristics.

The fifteen interoperability levels identified are: semantic, syntactic, technical, pragmatic, organizational, schematic or structural, dynamic, legal, conceptual, social, intra-communities, political/human, international, empirical and physical.

Author	Year	Interoperability levels														
		Technical	Schematic or structural	Semantic	Organizational	Physical	Empirical	Syntactic	Pragmatic	Social	Intra- Communities	Political / Human	Legal	International	Dynamic	Conceptual
ISO	1990	x		x				x								
Goh	1997		x													
Goodchild *et al*	1997	x	x		x			x	x				x		x	x
Bishr	1998		x	x				x								
Vckovski	1998			x				x								
Harvey, et al.	1999	x		x						x		x				
Shanzhen *et al*	1999			x	x			x	x						x	
Ouksel & Sheth	1999	x	x	x				x	x	x					x	
Miller	2000	x		x							x	x	x	x		
Nedovic-Budic & Pinto	2001				x											
Tolk	2003	x	x		x							x				x
Tolk & Muguira	2003	x		x				x	x						x	x
Bermudez	2004			x												
Shekhar	2004		x	x				x								
Schekkerman	2004	x	x	x												
Stroetmann	2005			x				x								
Ding	2005			x				x								
Kuhn	2005	x		x	x			x		x						
Nowak *et al*	2005		x	x				x								
Mohammadi	2006	x			x					x		x	x			
Kalantari	2006	x		x							x		x			
Vas Assche	2006					x	x	x	x	x						
Turnitsa &Tolk	2006	x		x				x	x						x	x
Whitman *et al*	2006	x		x				x	x							
Dekkers	2007	x		x	x											
Chen, D.	2007	x		x	x			x								
Zeigler & Hammonds	2008			x				x	x							

Table 3.3: Interoperability levels identified in the literature

Figure 3.2 shows the quotation frequency of each level. As can be seen, the physical, empirical and international levels are only quoted by only one author while the semantic level is the most quoted followed by the syntactic, technical, pragmatic and organizational levels. For the remainder, the quotation frequency is low.

As mentioned above, in order to understand the issue comprehensively, the definitions of the interoperability levels as interpreted by the respective authors are presented below. This review has been extended to all the levels identified whether they belong to an interoperability model or not.

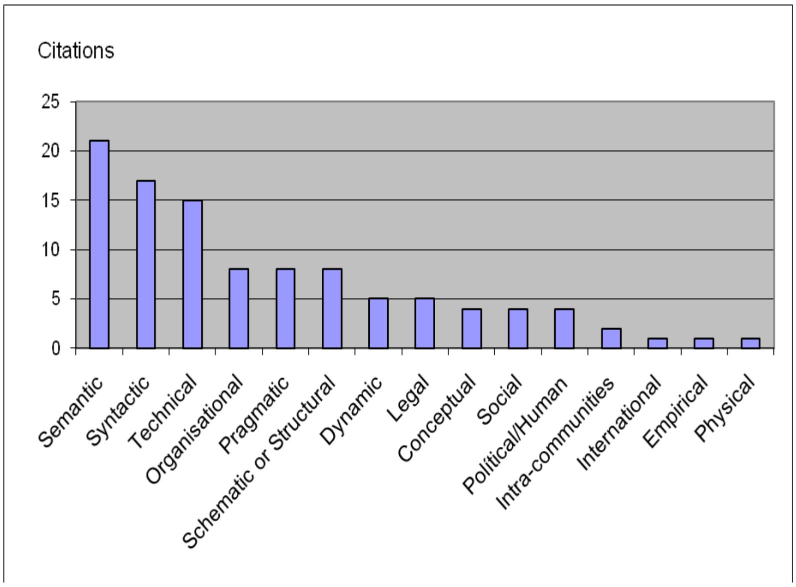


Figure 3.2: Histogram of interoperability levels

3.5.1 Semantic interoperability

SEI, in the LCIM, identifies semantics as the element allowing data and context transfer, so that the meaning of the data may be shared unambiguously.

ISO in the family of 19000 standard series includes in its ISO 19101 Geographic Information – Reference Model and ISO 19119 Geographic Information – Services, the interoperability types, and they also deal with semantics.

Goodchild *et al.* (1997) indicate that any semantic interoperability level can manage the difficulties that may come up when sharing meanings. The OpenGIS Guide (1998) identifies semantic problems in the geographic or spatial data. With the purpose of reaching semantic integrity, the use of common languages and conceptual models is proposed, and the use of common concepts is suggested.

Bishr (1998) sees semantic problems in the way the different disciplines or user communities describe the objects of the real world in the databases, thus encouraging semantic heterogeneity. As an example, a highway network has for a pavement manager different semantic description than the transportation infrastructure data in a GIS database designed for topographic cartography applications at small scales.

Vckovski (1998) defines the concept of semantic diversity which usually occurs when information communities cooperate; it happens because concepts, objects, etc have not been defined semantically. Each information community generally understands data and its implicit assumptions differently. However, these assumptions are very dangerous. According to the author, the only way of solving this problem is reasoning. The author suggests using concepts in systems of support to decision-taking to solve semantic heterogeneity. Another highlighted aspect is the lack of foresight about the future use that could be given to geographic data.

Harvey *et al.* (1999) indicate that the semantic interoperability levels are not achieved through standardization, since the meaning of the concepts is different in different domains. Interoperability problems come up when merging geographic data, either because they come from different sources or because there is a cultural frontier. The authors identify three aspects of this type of interoperability on which there is ongoing research: cognitive, computational and linguistic. They propose the concept of semantic similarity as a metric to measure the degree of interoperability achieved.

Ouksel (1999) identifies six elements in the infrastructure of open systems for social interaction related to semantic interoperability: meaning, propositions, validity, veracity, relevance and denotations.

Shanzhen (1999) identifies as objective of this interoperability level the semantic exchange between knowledge domains or user communities, using special standards of information and correspondence rules between communities. To reach this objective it is indispensable to have a common, fundamental knowledge corpus about GI and its theory.

Bermúdez (2004) and Goh (1997), in their respective PhD theses related to ontologies in the metadata context, have identified semantic problems of interoperability in synonyms and homonyms, in scales and measurement units, comprising aspects concerning the erroneous interpretation of concepts.

The Project IDABC (Interoperable Delivery of pan-European eGovernment Services to Public Administrations, Businesses and Citizens) (2004) identifies the need for multilingual interfaces in the provision of services to users as a key aspect of semantic interoperability.

Schekkerman (2004) focuses on the meaning of the exchanged information. In this case semantic interoperability is related to the aim of assuring that any other application developed for other purposes understands the meaning of the exchanged data accurately.

Shekhar (2004) defines semantic interoperability as a necessary requirement for the information systems destined to spatial data mining. This author emphasizes the need of specifying, through a document, the content of spatial data, the models and the relations. Ideally he proposes sharing a vocabulary and an ontology of concepts. Failing this, he proposes using a well defined translation system that enables relating the concepts used by the systems that send and receive data.

Hao Ding (2005) proposes storing the relations between ontologies and the use of fuzzy cross relations as a solution to the problems of semantic interoperability.

Kuhn (2005) suggests that the semantic interoperability is the only useful type, although later he mentions other interoperability aspects as relevant. His work implies that the semantic Web does not deal with all the semantic aspects, since it does not go into the meaning of the operations offered by the services, giving more relevance to semantics of parameters and results. The author distinguishes two types of semantic heterogeneity when matching data: syntactic and structural. He had previously proposed the concept of semantic reference systems (Kuhn, 2003) in parallel with the spatial reference systems; the rules for transformation from a system into another would have to be defined.

Nowak (2005) defined as aspects of semantic interoperability those related to the difference in the meaning of the same global entity in different databases (difference in the information context).

Pokraev (2005) states that problems of semantic interoperability occur when the 'objects domain' model at the level of user is different than at the level of system. The author refers, as 'objects domain' model, to the part of the world the message is all about (Wieringa, 2003).

Assche (2006) identifies as semantic interoperability problems those related to the meaning and validity of whatever is being expressed.

Antonovic and Novak (2006) propose that the players must be able to understand the information they are exchanging, i.e. to share the meaning of the information elements with the least error and ambiguity possible.

Kalantari *et al.* (2006) stress the harmonization of terminology and the interpretation of concepts.

Probst (2006) identifies semantic interoperability problems by analyzing the OpenGIS Observations and Measurements Encoding Standard from an ontological viewpoint. A conceptual model is proposed and it is pointed out that there are aspects in the document that depend on human interpretation when contributing free text descriptions. As a consequence of this circumstance, problems are identified in the process of location and evaluation of geographic object sources.

Turnitsa and Tolk (2006) highlight the need for a common reference model to exchange information and they emphasize the need of sharing the meaning of the data.

Chen *et al.* (2007), in the InterOP Project, identifies conceptual type barriers and they regard semantic problems as an important barrier for interoperability. They relate these difficulties to the fact that semantics to be used to represent information and knowledge has not been clearly defined in the models or in the computer applications.

Dekkers (2007) defined this type of interoperability as “the understanding of the meaning of information, keeping in mind the data of other persons”. He groups together the issues related to type of information proper, the responsibilities for the data and its maintenance as well as the aspects related to use conditions and restrictions; he also relates all those issues with the structure (metadata standards), vocabularies, values and classifications.

Zeigler and Hammonds (2007), in their information interchange model Net-Centric, indicate that the XML, XML-Schema and UML languages provide syntactic and semantic interoperability, being able to express semantics from an ontological viewpoint. They identify the lack of metadata items allowing enhancement of semantics on the dynamics in a practical way.

3.5.2 Syntactic interoperability

Bishr (1998) understands by syntactic interoperability problems those related to the implementation of databases in different paradigms (relational, objects) or the geometric representation of objects (raster, vector).

Vckovski (1998) refers to syntactic diversity as the source of problems for geodata exchange which is due to the diversity of storage and transfer formats in spite of standardization efforts. The author points out that these problems are especially important in the context of data quality management. Many problems of syntactic interoperability imply the manual conversion of formats which affects in many cases their quality; as an example of those problems, the capability of expressiveness of the formats is mentioned. In some cases the formats contain information the tools cannot handle or are unaware of. This implies the need for especially trained personnel to carry out those conversions. The author also points to some cases where semantic problems could be interpreted as syntactic, and as an example he suggests the homonyms. It may also occur that syntactic heterogeneity is not detected (e.g. text encodings, byte order). The author concludes stating that the source of these problems is usually the lack of explicit format specifications. In many cases implicit assumptions are not made in a concise manner.

Ouksel (1999), in the infrastructure of open systems for social interaction, mentions some syntactic aspects for the formalization of structures, languages, logic, data, registers, deductions, programs and files.

ISO 19101 (2002) and ISO 19119 (2005) Standards also regard the syntactic interoperability level as one of the objectives to achieve.

Shekhar (2004) defines syntactic interoperability as the specification of common formats for messages (e.g. use of tags and labels) to exchange spatial data, models and relations.

Nowak *et al.* (2005) regard as syntactic aspects causing lack of interoperability the hardware and software platform heterogeneity, the geometric and thematic representations and the relations of spatial objects (coordinate systems, geometric resolutions, geometric representation, quality).

Turnitsa and Tolk (2006) mention that the syntactic interoperability level should be supported by a common structure for information exchange where, for instance, common data formats are required.

Assche (2006) attributes to this interoperability level aspects concerning used language, structure and logic.

SEI has introduced this level in LCIM, indicating that it is the one level enabling data exchange using standardized formats.

Chen *et al.* (2007) identify conceptual type barriers in the InterOP Project; syntactic differences may be found when different persons or systems use different structures to represent information and knowledge.

Zeigler and Hammonds (2007) take a backward glance at the lack of syntactic aspects in the old communication models where information is arranged serially and a specific information structure is expected. At the present time these aspects have been resolved by the service-oriented architecture technologies and the information encoding in XML formats.

3.5.3 Technical interoperability

Goodchild *et al.* (1997) consider as technical aspects of interoperability the distributed computing, the communication networks, the technologies in themselves and the distributed calculation platforms. He extends the need to enhance the technical interoperability to data and middleware.

Ouksel (1999) proposes as a solution to the system heterogeneity the Internet General Inter-ORB Protocol (GIOP), for interaction between object request brokers.

Miller (2000) identifies as technical aspects of interoperability the communication standards, transportation, storage and representation; as an example he mentions the communication protocol Z39.50 (ISO 23950) destined to support search and retrieval of information in different systems.

In ISO 19101 (2002) Geographic Information – Reference Model the technical interoperability level is described, ascribing to it communications, information transfer and running of programs in functional units.

The IDABC Project (Interoperable Delivery of pan-European eGovernment Services to Public Administrations, Businesses and Citizens) (2004) identifies as technical aspects the ones related to service connection and computer systems at different levels: open interfaces, interconnection services, integration of data, exchange and presentation of data, accessibility and security services.

Schekkerman (2004) states that the technical interoperability is basically related to the connection of the computer systems and the services.

Turnitsa and Tolk (2006) also identify the technical aspects of interoperability with data exchange. They suggest the use of common communication protocols with the aim of systems exchanging bits and bytes.

Kalantari *et al.* (2006) regard technical interoperability as the development of standards of communication, exchange, modeling and storage of data as well as the portals and Web services that interoperate equipped with friendly interfaces.

Antonovic and Novak (2006) introduce key aspects to reach technical interoperability: open interfaces, service interconnection, integration of data, middleware, presentation and exchange of data, accessibility and security services.

Mohammadi *et al.* (2006) identify and catalog the following problems as aspects of the technical interoperability: disparity in the computing environment, lack of standards, 3D (vertical) topology, spatial reference systems, scales, quality, data models, metadata, formats and semantics.

The SEI-CMU proposes taking into account the technical aspects of interoperability when creating models, coinciding with Tolk and Turnitsa (2006) by highlighting the information and the exchange of bits at the level of communication protocols.

Chen *et al.* (2007) identify problems of technical interoperability in the InterOP Project relating them mainly to the incompatibility of the information technologies (architectures, platforms, infrastructure, etc).

Dekkers (2007), DCMI Management Director, defines these levels of interoperability relating it to the interconnection, presentation and interchange of data, the accessibility and the security. He groups in this level the aspects related to communication protocols and service interfaces, formats, encodings, measures of accessibility and security solutions.

3.5.4 Pragmatic interoperability

In the document "Infrastructure of Open Systems for Social Interaction", Ouksel (1999) identifies as aspects of the pragmatic interoperability the intentions, communications, conversations and negotiations.

Shanzhen (1999), in Intermodel5 for distributed computing, proposes the application service level relating it to the distributed computation models that enable the geoprocessing service interchange and the analysis between different communities and departments.

For Pokraev *et al.* (2005) a problem of pragmatic interoperability comes up when the effect of a message differs from the expected effect.

Assche (2006) attributes to this interoperability level the objectives, responsibilities and hidden consequences in the information or the messages. The pragmatic interoperability reaches every user of interoperable services having compatible objectives, roles and consequences relative to the own services and the exchanged information. In this context the pragmatic interoperability is especially related to the fact that all parties involved, users of a service, have their roles defined.

For Turnitsa and Tolk (2006) this level of interoperability is achieved when the systems know and are able to use the methods and procedures provided and implemented by the systems, in other words, when the use of the data or the context of their applications are understood by the participating systems.

SEI, in the LCI Model, groups in one level the pragmatic and dynamic interoperabilities. It identifies them as those allowing identification of the data context and the possible forms of applying them without possible ambiguities.

Zeigler and Hammonds (2007) in their information interchange model define the pragmatic level relating it to transmission, error detection, their correction or the negotiation of retransmission.

3.5.5 Organizational interoperability

SEI proposes considering the organizational and cultural aspects of interoperability to create models that enable its evaluation, but not as a model of interoperability between systems.

Goodchild *et al.* (1997) identify the institutional aspects that may be the most problematic since they depend on many factors: behavioral, economic, legal and organizational.

Ouksel (1999), in the document “Infrastructure of Open Systems for Social Interaction”, points to aspects such as behaviors, expectations, contracts, laws and commitments.

Shanzhen (1999) states that the problems of institutional interoperability come up when there are no policies of data exchange or they are not known; these exchanges may be conditioned by the differing policies, cultures, values and aspects related to privacy. He states that it is necessary to coordinate the political and cultural aspects as well as the values in the relations between different communities.

Nedovic-Budic and Pinto (2001) deal monographically with the organizational interoperability within the GIS context. They identify four working areas related to interactions, implementation, coordination and results. The responsibility aspects, the policies of access to data and the agreements for resource management are treated in the first area. Regarding implementation, seven aspects are mentioned to be taken into account: resources, experience, stability, coordination, data access, capability for implementation, leadership aspects and management. Regarding coordination, the agreements, rules and roles are mentioned. Finally, concerning results they indicate that the objective is to improve relations between organizations.

Schekkerman (2004) identifies as aspects of organizational interoperability the ones related to the business aims, the business process models and the information exchange between organizations. As a solution to these aspects, he proposes to consider the requirements of the user community, trying to make services available, accessible, easily identifiable and user-oriented.

Mohammadi *et al.* (2006) emphasize at the institutional level, considered equivalent to the organizational level, the aspects related to the models of collaboration and funding and the links between the data management units as well as knowledge of their existence.

For the SAGA Project (Standards and Architectures for eGovernment Applications (2006), the main objective of organizational interoperability is to determine when and why data are exchanged. Therefore the approach to this interoperability type is to define the legal reference framework for data exchange.

Dekkers (2007) includes in this level of cooperation between organizations known as BPI (Business Process Integration), the business objectives and process modeling. Information exchange between partners (scientific domain, clients and providers, administration agencies) is included as well as the whole coordination of the business tasks such as the user needs of communication or the results of costs and benefits analysis.

The IDABC Project (2004) reasserts the aspects described by Dekkers (2007) taking into account the user requirements related to identification, access and availability.

Chen *et al.* (2007), in the InterOP Project mention organizational interoperability problems by identifying the authority and its responsibilities within the organization. They indicate that human factors and technologies are related to human and organizational behaviors, so that interoperability might be impossible.

3.5.6 Schematic or structural interoperability

Goh (1997) mentions the following structural aspects of interoperability: types of data (different data primitives in different systems), label conflicts (synonyms and homonyms in different schemas), aggregation discrepancies (different forms of design or attribute assignation to entities) and conflicts of generalization (ways of relating entities to each other).

Bishr (1998) identifies as schematic heterogeneity circumstances such as the objects of a database being considered as properties in another database, or the fact that object classes may have different hierarchical levels of aggregation and generalization, even though they describe the same objects of reality.

Ouksel and Sheth (1999) propose the use of the RDF (Resource Description Framework), developed with the general purpose of describing information sources or object models for Web information exchange (Manola, 1998). The MPEG-4 (Motion Picture Experts Group Layer-4) format is another example for description of structure or level of video objects, MHEF-5 for multimedia and hypermedia, KIF (Knowledge Interchange Format) for knowledge representation or OKBC (Open Knowledge Base Connectivity) as a base for distribution of knowledge.

Shanzhen (1999) states that data coming from different sources have different structures and schemas and he proposes in Intermodel5 the level of resource transformation as a solution to solve this type of data heterogeneity. In this sense the FGDC (Federal Geographic Data Committee) has specified the SDTS (Spatial Data Transfer Standard). The OGC has proposed the Simple Features Specification, a positive contribution to this interoperability level to solve structural heterogeneity issues.

Shekhar (2004) suggests an intermediate level of structural interoperability so as to provide the means to specify the semantics of data schemas (metadata) and to be able to share them.

Nowak *et al.* (2005) define the heterogeneity of data model and database schemas (class hierarchy, attribute structure, etc) as aspects related to structural interoperability.

3.5.7 Dynamic interoperability

Shanzhen (1999), in Intermodel5 emphasizes a resource location level which is based on the existence of metadata standards and in the capability of locating resources for their exploitation; first interoperable objects that would solve our needs are searched for; at the level of resource location, the objects and the search method are defined. There are many methods such as the one developed by the FGDC, developing metadata standards for the NSDI Project. Among the descriptions contained in the metadata are the spatial information and the potential application of the data.

Turnitsa an Tolk (2006) state that this interoperability level is achieved when the systems are capable of reacting to the state change of the others or to conditions affecting data exchange and are also capable of taking advantage of such change.

3.5.8 Legal interoperability

Miller (2000) identifies as aspects of legal interoperability those related to intellectual property rights and with the laws or standards facilitating dissemination of information in the public sector among others.

Kalantari *et al.* (2006) propose the creation of guidelines, rules, parameters and instructions to manage the workflow, considering the information and incorporating the communications in the management of the territory. This may be extended to the SDI context.

Mohammadi *et al.* (2006) identify aspects such as intellectual property rights, restrictions and responsibilities, licensing and access constraints or data privacy as aspects of the legal interoperability.

3.5.9 Conceptual interoperability

SEI in the LCIM states that this level should allow establishing a common view of the world e.g. based on epistemology. This level should contain the relations between elements in addition to the implemented knowledge.

Goodchild *et al.* (1997) state that business interoperability will be achieved when conceptualization will be evident at the individual level.

Turnitsa and Tolk (2006) define this type of interoperability as the one to be achieved when a conceptual model is documented through methods used in engineering, so that it can be interpreted and evaluated by a third party. In this case, if the conceptual model is aligned (e.g. with the assumptions and the constraints of abstraction about a reality), it means that the highest level of interoperability has been achieved. Additionally, the method used to document the conceptual model should not influence the implementation and it should not depend on the model.

3.5.10 Social interoperability

Ouksel (1999), in the “Infrastructure of Open Systems for Social Interaction” identifies some social aspects such as the interests, expectations, contracts, laws, culture and commitments.

Harvey *et al.* (1999) just mention the social aspects of interoperability.

Assche (2006) attributes to this level the aspects related to interests, beliefs and shared conclusions as results.

Mohammadi *et al.* (2006) add to the previous proposal the cultural aspects, the education and the identification or knowledge of the responsibilities.

3.5.11 Intra-community interoperability

Miller (2000) defines as aspects of the intra-community interoperability those related to the common solutions for different levels of detail in the geographic description and in different domains of knowledge, science, etc.

Kalantari *et al.* (2006) identify this type of interoperability with the coordination and alignment of the business processes and the information architectures that comprise both persons and partners (private sector) and public sector.

3.5.12 Political/human interoperability

Harvey et al (1999) mention, without describing, the political issues as aspects of interoperability.

Miller (2000) identifies the policies or guidelines of organizations, oriented to dissemination and maintenance of information, as political and human aspects of interoperability. In order to ensure interoperability, the information must be maintained updated and disseminated, or failing that it might disappear together with the persons in charge of its maintenance.

Mohammadi *et al.* (2006) identify the deficiency in legislation matters, pricing policies or priority identification and assignment.

Although we have come across another document referring to political or human interoperability in the health context, it has not been mentioned since it deals with the same aspects already described for the previous authors.

3.5.13 International (linguistic) interoperability

For Miller (2000) the language in which the data are provided or described defines international interoperability. The aim is to avoid problems concerning the understanding of certain languages on the part of users. This category also includes aspects related to the practices of use, culture, expectations and needs of the users. We have not come across any other authors dealing with the international interoperability level.

3.5.14 Empirical interoperability

Assche (2006) attributes to this interoperability level the aspects concerning entropy, confusion and diversity. We have not come across any other authors dealing with the empirical interoperability level.

3.5.15 Physical interoperability

Assche (2006) understands by physical interoperability the aspects related to physical appearance, the environment and the degree of contact or interaction between systems. We have not come across any other authors dealing with the physical interoperability level.

In the reviewed literature no methods have been identified that would validate the interoperability models. Some authors, e.g. Whitman *et al.* (2006), propose to carry out the validation through matching the obtained results with the expected ones. For this reason, in this thesis the existing literature is reviewed in the context of interoperability measurement as a first step to propose a method to validate interoperability, either on the model or on the different levels.

3.6 Interoperability measurement

Measuring the degree of interoperability between two systems allows learning the strengths and weaknesses of both when they operate jointly. The measurement may help increase interoperability and solve deficiencies. Some progress has already taken place in this direction (Pridmore and Rumens 1989; Hamilton et al, 2004; Kasunic and Anderson, 2004; Janowicz *et al.*, 2008), yet the main stumbling block comes up when trying to define the metrics. Metrics, in the context of software engineering is defined as any measure or set of measures that allow characterizing software or information systems. In our case it is not about characterizing an information system but the interoperability between information systems. Daclin *et al.* (2006) emphasize how arduous it is to identify the parameters characterizing interoperability to be able to apply a measurement.

Pridmore and Rumens (1989), in the document “Interoperability - how do we know when we have achieved it?” indicate that first of all it is necessary to define the interoperability requirements and then to carry out the measurements to verify if it has been achieved. In order to measure interoperability, metrics is needed which defines the common reference framework. In addition to defining metrics, weighting of the indicators relative to the achievement of interoperability within the system must be established depending on how critical or important is the part of the system being observed. The measurement process does not finish there; weighting must be established for every measurement in order to have an estimate of confidence on the results.

Application of metrics means to compare the real system with a model or measurement scale (defined on the basis of requirements), and to express results by means of simple terms. The essential steps are:

- Identification of each measurement with a system requirement;
- Comparison of result of the measurement with the objective;
- Normalization of the result and definition of the degree achieved for the requirement;
- Estimate the confidence of the result for each comparison
- Weighting of the results with relative values of the whole system;
- Combination of the results to shape the interoperability measurement and its confidence.

The authors conclude their work with a critical reflection: it would be possible to interpret that a metric does not provide a normalized mechanism to measure interoperability even though it may provide a reference framework and a method to express and evaluate interoperability problems. They also indicate that metrics may help answer questions such as: What is the general status of interoperability? Does the system comply with all the objectives and will it work? Will the result of making any change be better or worse? Which are the problem areas and how serious are the problems? Is confidence well defined? Have all the due verifications been made? Or what confidence should be assigned to the result?

Hamilton *et al.* (2004) have worked on the development of interoperability metrics and state that interoperability is indeed very hard to measure, so that generally the use of simple models is proposed. The models to carry out the measurements must have metrics available enabling the measurement of the degree of interoperability achieved. The metrics may evaluate interoperability from a quantitative and qualitative viewpoint.

Kasunic and Anderson (2004), in the document "Measuring Systems Interoperability: Challenges and Opportunities", state that the development and use of accurate measurements in such a complex and multidimensional environment as interoperability is hard to achieve; they deal with its multiple facets associated to a domain and they propose four sets of measurements to address the following aspects of this difficult problem: (a) technical conformity relative to the accomplishment of the norms; (b) measurements of systems interoperability focused on information flow; (c) exploitation interoperability, focused on measurements that check whether the specific requirements of the node-to-node information flow are met, and (d) organizational and cultural measurements.

Daclin *et al.* (2006), in the document "Enterprise interoperability measurement – Basic concepts", also state that the interoperability measurement has the purpose of defining metrics to determine the degree of interoperability. The application of metrics with the aim of measuring the degree of interoperability is related to two principles: (1) the identification of parameters relative to interoperability, and (2) the characterization of these parameters through metrics. The degree of interoperability of a given system may be defined by a vector made up of three types of measurements: (a) interoperability potential measurement, (2) interoperability compatibility measurement, and (3) interoperability performance measurement.

The interoperability potential measurement is related to the identification of a set of system properties having an impact on the development of interoperability. The problem comes up when the systems involved evolve dynamically and the measurement must suit the new system. The compatibility measurement is carried out at the engineering stage, i.e. when the system is re-engineered with the purpose of establishing the interoperability. The performance measurement shall be carried out during the operative phase, i.e. during the performance, to evaluate the joint working of the two systems.

Chen and Daclin (2007), in the document "Barriers driven methodology for enterprise interoperability" have proposed a method similar to the development cycle of computer applications made up of (a) definition of objectives and needs, (b) analysis of the current system, (c) selection and combination of solutions, and finally (d) application and testing. In parallel they proposed the participation of four groups of players as defined in the engineering methodology "Results and Interrelated Activities Graphic": project team, synthesis group, expert group and interviewer group.

Recently Daclin *et al.* (2008), in the document "Enterprise interoperability methodology", provide some solutions or improvements for interoperability and a method to measure and evaluate the capability of interoperation (degree of interoperability and performance). This methodology takes into account only two participants and the operative results are limited to the technical aspects (communication performance and information exchange), and finally another limitation of this methodology is the type of interoperability it is related to, namely the information exchange between the persons belonging to different enterprises.

Janowicz *et al.* (2008) indicate in the document “Semantic similarity measurement” that those measurements are typical of the cognitive sciences and that historically modeling similarity has been tried and reasoning models have been developed. The recent computer science research has been applied to the theories of computational similarity as a support of reasoning for the processes of information retrieval and organization. The authors state that the characteristics of the measurements in the geospatial domain are: (a) similarity is an asymmetrical relations according to the developed theories; (b) contexts such as age, knowledge and cultural experiences, motivation and the actual application are key aspects to measurement similarity; (c) the theories about similarity are tied to the language used to represent it; (d) for a theory to be usable, the computational representation and the descriptions used by the participants must be comparable. Finally the authors indicate the future research areas in similarity measurement: explanation and approximation of similarity values, study of the context influence, the similarity between objects that do not last indefinitely, the differences between similarity and semantic analogy, and the semantic similarity within the semantic reference systems.

In this thesis a new way of measuring interoperability based on the use of metadata is proposed. Metadata within the SDI context are indispensable components to describe datasets and services. As indicated in Chapter 2, metadata perform location, evaluation, extraction and use functions in this context due to standards defining their items, their semantics and in some cases restricting the value range. From our perspective, metadata are elements enabling interoperability: they describe syntax and semantics, they define how to access data and interfaces, they catalog resources, describe legal and organizational aspects concerning data and services and their exploitation, and for these reasons, aligned to the proposal of Tolk, Diallo and Turnitsa (2007), the use of metadata as a tool that allows measuring interoperability in the SDI context is proposed.

3.7 Summary

In this chapter the significance of interoperability in the GI context, in particular SDI, has been described. The definitions of the term interoperability have been reviewed and one has been selected as a framework. The interoperability models in the GI, SoS and organizations have been reviewed as well as the definitions or scope that the different authors grant to the interoperability levels, finishing with the measurement of interoperability as a mechanism of validation.

Interoperability as standardization. One of the aims of SDIs is to provide access, use and reuse of data. From the services viewpoint, standardization is needed for users and applications to exploit them. From the technological viewpoint, an SDI is considered both a set of systems and an II, and at the same time it may be regarded as a system within which many subsystems coexist (SoS). For SDIs, II and SoS, interoperability is an objective per se.

Review of the interoperability concept. The interoperability concept has been analyzed and the different definitions interpreted. As a consequence it may be stated that the interoperability may be perceived from different perspectives. As an example, from the point of view of the standard organizations, special attention is paid to the technical aspects (resource exchange between systems). However, for political organizations and institutions, the definitions proposed for the term interoperability must emphasize information reuse.

Review of models. The interoperability models have been reviewed within the GI and SoS contexts. The research challenge about interoperability models consists of examining a unified model based on the strengths of the different models or on gateways or transformations among models allowing their integration.
Interoperability is a complex issue that should be analyzed with a piecemeal approach and both the actual definitions and the interoperability models call their parts “levels”. An extensive literature review on interoperability levels in the context of systems and GI has been undertaken. Fifteen different interoperability levels have been detected with different quotation frequency and relevance with respect to the geographic and SDI contexts.

Interoperability measurements. Since interoperability is an objective for SDIs, the degree achieved between the components and applications operating on them is an objective as well. For this reason the literature concerning forms of measurement has been reviewed. In some cases the issue is detecting barriers, in others ensuring a certain level. Finally, the main efforts focus on semantic interoperability and the metrics that enable measurement of the degree achieved.

After having reviewed the literature on the concept of interoperability and the taxonomy of levels, some types of interoperability in the SDI context are analyzed, then an interoperability model is proposed consisting of seven levels: technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational. This has been done after surveying the definitions proposed by the authors and reflecting on how to simplify the model, getting rid of those levels of no application in this context and through reclassification of others because of semantic affinity.

Of the interoperability types shown in this chapter, the following considerations must be taken into account:

- The *physical* and *empirical* interoperabilities proposed by Assche (2006) relate to aspects of the man-machine interaction or to the amount of information provided to the student within the e-learning context. Since these interoperability types have little application in the SDI context and in the case of the SoS, no consideration is given to them.

- The *international, political/human, social and/or cultural, intra-communities, legal and policy* interoperability levels may be lumped together as aspects affecting users, institutions and/or organizations. The legal aspects often slip beyond the competences of the institutions, the regulation coming from higher levels, i.e. national for the autonomous communities/states or international laws for the nations. Collaboration and cooperation are aspects affecting both institutions and organizations that allow them defining interrelations at the regional, national, intra-communities and international levels. Some of the mentioned authors (Harvey *et al.*, 1999; Kuhn, 2005; Assche, 2006) write about the lack of interoperability at the social level or they relate it to political, cultural or value aspects. Likewise Goodchild et al, 1997 regard organizational interoperability as the most difficult to achieve.

- The *schematic* or *structural* aspects mentioned by Nowak and Nogueras (2005), Goh (1997) and Shekhar (2004) can be seen included in other interoperability types. As an example, Nowak and Nogueras (2005) regard the differences in data models as lack of schematic or structural interoperability. This lack of interoperability could also be considered as *conceptual* lack, since in many cases the definition of the actual data models is not described by means of modeling languages giving independence to the model of their implementation. Goh (1997) and Shekhar (2004) also deal with these aspects even though they may be considered as syntactic and conceptual aspects by regarding the data models and the metadata as mechanisms that enable sharing of the data schemas.

Finally, it should be noted that the following interoperability models have been considered for the definition of the seven level model: *Integrated Interoperability in GIS* (Shanzhen Yi *et al.*, 1999; Goodchild et al, 1997), *Coalition Model* (Tolk, 2003) and *Integrated Interoperability Model* (Turnitsa and Tolk 2006).

4 METHODOLOGY

4.1 Introduction

This chapter presents the proposed methodology for automatic metadata creation, the integrated interoperability model designed for SDIs is described, and finally the methodology adopted for analysis of the interoperability provided by ISO 19115 metadata standard items is presented.

In the first place we should answer the third research question put forward in the PhD project: is it possible to create useful GI metadata automatically and efficiently? The methodology to extract the metadata implicitly or explicitly stored in the data is described. We also expose the different analysis and treatment phases carried out with the information and the inference phases that enable proposing another set of items that catalog the metadata or describe their data models.

In the second place the integrated interoperability model proposed for its application within the SDI context is presented. The choice of levels of the model is explained and the aims of those levels is described at large.

In the third place we should answer the second research question put forward in the PhD thesis: what is the contribution in terms of interoperability of the information contained in metadata? In order to answer this question a thorough analysis of the items that make up the standard core on the one hand and of the whole items on the other hand has been carried out. In addition to aggregate-level analysis of how the items provide interoperability, it has also been studied how they simultaneously favor several levels.

Finally the conclusions of the chapter are presented.

4.2 Methodology of automatic metadata creation

As has already been mentioned, the automatic metadata creation (Greenberg, 2004) is necessary and useful; it should be integrated in the data workflow and may require important computing developments. Next a methodology developed in the form of a procedure will be presented, which is susceptible of being automated through a computer system for metadata creation on spatial data repositories or stores. The methodology stages and the results that can be achieved are described. Figure 4.1 shows a block diagram with the different stages and their relations.

From a global perspective the method consists of extracting the largest amount of information possible from the GI repository (3) to complete metadata with the greatest detail possible.

The methodology foresees the case where the GI repository does not store any information identifying the coordinate type used and the case where there is a metadata template available for completion. In addition, it is anticipated that the user or application exploiting this methodology can select the type of desired result: an XML file (17) with the metadata or a ZIP file with MEF (18)1 containing both the metadata and the additional files generated: preview (12), UML data model (13) and XSD application schema (15).

The figure starts with the input (1) where a grey-shaded text shows the optional elements, i.e. those that may or may not be provided. The only necessary, actually mandatory element is the URI describing where and how to access the GI repository. It has been designed this way in order to be able to connect with the different transfer protocols (http, ftp, file) and with different databases and stores (oracle, postgres, mysql, db2, informix, sqlite, etc). The latter may require additional data relative to the identification system based on user and password in addition to the database identifier at the administrator.

[1] MEF (Metadata Exchange Format) defined in the GeoNetwork Project. http://geonetwork3.fao.org/workshops/workshop2007/GeoNetwork%20MEF%206-11-2007.pdf

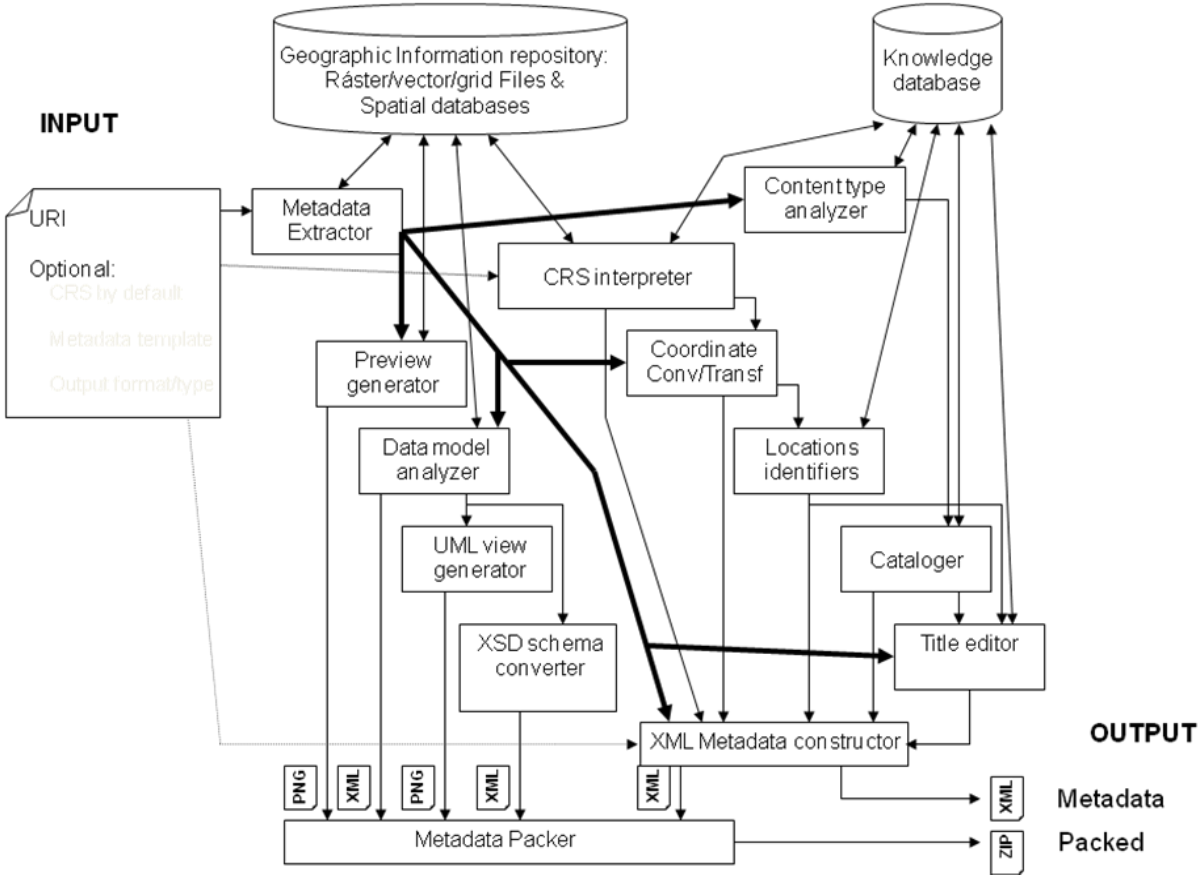


Figure 4.1. Method of automatic metadata creation: stages and relations

The first stage of the methodology consists of metadata extraction; the *Metadata Extractor* (2) uses the information contained in the *URI* (1) to access the *GI Repository* (3), thus getting the largest amount of implicit information thereof. The information retrieved by the *Metadata Extractor* (2) will depend on the storage format used by the GI Repository (Manso et al, 2004). It has been possible to identify more than 90 storage formats for images and rasters, more than 25 storage formats for vector or drawing data and more than 10 DBMS with capability to store spatial information. Next the extraction procedure for information stored in the file or database for every format is specified if useful to build metadata:

- The general procedure for the databases consists of accessing recursively all their tables to read the names and types of columns, the column containing the geometries, the type of geometry/ies stored, the number of table registers and the coordinates max and min values.
- The general procedure for vector files is similar to the procedure for databases, searching for the same properties, although the access is more complex since different data types (equivalent to a table) with different geometries are mixed in one layer. A particular case of the vector formats are the CAD formats, even more complex due to the different solutions adopted to encode the elements which makes difficult obtaining reliable information about the type of geographic features contained in them.
- The general procedure for images consists of retrieving the information relative to the number of bands, data type or number of bits used for each pixel, the image dimensions (width and height), the max and min coordinates, the pixel resolution and the statistics: average and typical deviation of each band.
- The general procedure for grids is similar to that of images since they could be regarded as a special case of image; there is only one band and it usually contains numerical values representing attributes such as altimetry or assigned class identifier, and they represent a rectangular surface. For grids it is interesting to obtain the max and min values of the band in addition to the already mentioned values for the images.

This information will be used by the content type analyzer (5) to try to determine what type of content is stored in the images and consequently to use it for the purpose of GI cataloging.

The complexity of this extraction stage lies in the high number of available geographic data repository formats and in the lack of uniformity relative to the type, amount and manner of storing the implicit information (that may be handled with a procedure) and the explicit information (that may be handled individually) describing the data. It may be stated that there is a significant number of formats for which it is not possible to apply a common procedure to recover the additional explicit information they contain.

The second stage in importance is the *CRS Interpreter* (4). By extracting the information identifying the coordinate system used from the *GI Repository* (3), it is possible to find out that there are different encoding types to represent that information (Manso and Bernabé, 2005)

- As text information, with a pre-defined structure (e.g. "*European_Datum_1950_UTM_Zone_30N*");
- As WKT structures (e.g. *PROJCS[ED50/UTM zone 30N", GEOGCS["ED50", DATUM ["European_Datum_1950", SPHEROID ["International 1924",6378388,297, AUTHORITY["EPSG","7022"]], AUTHORITY ["EPSG","6230"]], PRIMEM["Greenwich",0,AUTHORITY["EPSG","8901"]], UNIT ["degree", 0.017453292519943 28,AUTHORITY[(..)"*);
- As mnemonics (e.g. Ermapper datum *ED50* and *NUTM30* Projection);
- As particular numerical codes (e.g. Intergraph datum: *4*, projection: *7*);
- It is also usual for some people to use attribute-value pair sequences (e.g. Proj4 *+proj=tmerc +lat_0=0 +lon_0=93 +k=1.000000 +x_0=16500000 +y_0=0 +a=6378140 +b=6356755.288157528 +units=m +no_defs*);
- Finally, coinciding with the present trend, several repositories use standard numerical codes (e.g. EPSG: 23030).

The objective of this stage consists of identifying the encoding type in the format used by the spatial information repository, obtaining the stored values and relating those values to the numerical codes standardized by the European Petroleum Surveyor Group (EPSG), presently integrated in the International Association of Oil & Gas Producers (OGP). The last operation of this process requires the exploitation of the information stored in the knowledge database (6) where the relations between the particular encodings of each format and the EPSG codes have been stored. The result will be incorporated into the metadata and used to carry out coordinate conversion/transformation calculations as required.

The complexity of the CRS interpreter lies in the correct identification of the coordinate type and the datum used in the repository to assign the EPSG code to it.

The *Preview Generator* (12) has the purpose of generating an image in PNG format visualizing all the information of a layer or set of bands contained in the GI repository. This image will form part of the additional information accompanying the metadata and it will help a user decide if the located dataset meets the anticipated needs, same as with the metadata extractor where the complexity of this block lies in the great variety and heterogeneity of the existing GI storage formats.

The stage of *Coordinate Conversion/Transformation* (7) has the purpose of obtaining the geographic coordinates using the WGS84 as the reference datum, associated to the geographic BoundingBox? (North, South, East, West) of the layer or band of information, from the coordinates expressed in the CRS (4) stored in the data repository.

As can be seen in Figure 4.1, it is necessary to know the stored coordinates and the SRS used to carry out the appropriate calculation: transformation (with change in the datum) or conversion (keeping the datum). The remarkable characteristic of this block is that it should be able to interpret many CRS types defined by their EPSG code and to know how to carry out the necessary transformations and conversions to switch types. The coordinates obtained after conversion or transformation, besides forming part of metadata, will be used as input to the *Geographic Locator* (8).

The *Data Model Analyzer* (13) interprets the type of GI repository to determine the typology of data. When the GI has associated alphanumerical information in table form, in order to suggest a data model based on entities and relations, the possible relations between the table attributes will have to be determined. This is made easier when the GI repository is a database since most of the DBMS store metadata corresponding to the tables, attributes and relations (primary keys, foreign keys, constraints, etc). In the case of files with GIS structure and tables or alphanumerical databases with tables, this stage of the process becomes more complex; relations have to be inferred since they are nor explicitly defined. The resulting data model is represented in UML and stored in an XML format of metadata interchange (XMI) so that it could be treated automatically by other applications. Besides, the inferred data model will be the information source that the stages generating a graphic view of the UML classes diagram will use (13), and also the source that generates the converter to the XSD application schema format (15).

The *Geographic Locator* (8) will use the latitudes and longitudes marking off the geographic BoundingBox in order to determine the most relevant geographic identifiers of the zone to include them in the metadata. This identification will be carried out on the basis of a coverage containing the polygons with the geographic bounds of the most important countries and regions of the world stored in the *Knowledge Database* (6). It is a spatial query whose results are arranged by relevance of the region or town, thus acting as an inverse Geocoder that retrieves a toponym from the coordinates of the geographic extent. The complexity of the geographic locator lies in the need of accessing or exploiting thesauri specialized in placenames. The calculated geographic identifier will be included in the metadata and will be used by the *Title Editor* (10).

The *UML View Generator* (14) builds an image representing the diagram of classes contained in the data model and it stores the image in graphic PNG format. This graphic facilitates the interpretation of the data model by users since the interpretation of a UML model in graphic form is easier than in XML language.

The *Cataloger* (9) uses the information retrieved through classification methods, data mining and inference, together with lists of terms belonging to multilingual thesauri, to propose keywords cataloging the GI repository. Besides suggesting a set of keywords, this stage of the procedure selects one or several terms of the controlled list that identifies the *topicCategory* of the resource. The results of this stage are included in the metadata and also used later by the *Title Editor* (10).

The *XSD Application Schema Converter* (15) carries out the necessary transformations to represent the UML data model in the form of a GML application schema, storing it in the form of an XML Schema (XSD). This application schema models the repository or information layer in the language standardized by the OGC and ISO for describing geographic data models in markup languages (GML). The result of this stage is the generation of an XML file that will accompany the metadata.

The *Title Editor* (10) composes a phrase that intends to synthesize the description of the data as best as possible. The title should contain sufficient information to answer a good number of questions: what? where? when? of whom? what scale? etc. In order to compose the title, this stage will use the information obtained by the geographic locator (8) (where?), the classification made by the *Cataloger* (9) (what?) and the information stored in the *Knowledge Database* (6).

The *XML Metadata Constructor* (11) assembles the entire information provided, extracted, calculated, inferred and worked out, described in the previous stages. If, as input of the procedure there is metadata or a metadata template, this stage will contribute that information over the metadata or the template. Otherwise new metadata will be built and information will be inserted over it. The result of this stage is an XML file conformant with ISO 191125-19139.

The *Metadata Packer* (16) takes the XML file with the metadata assembled by the metadata constructor, the XML files containing the UML diagram of classes, the GML application schema and the graphic files containing the preview of the data and the diagram of classes, and it builds a structure of files and directories in which the entire information is integrated.

Finally the whole structure of files and directories is compressed in a document with *Zip* format (18). The internal structure contained in the file is conformant with the metadata exchange format (MEF) proposed by the GeoNetwork Project.

4.3 Interoperability model applied to SDIs

Our interoperability model consists of seven levels: technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational.

As far as relations between the different interoperability levels are concerned, we propose that they should not be hierarchical, as advanced in the reviewed literature. Our model points to the existence of different types of relations between the different interoperability levels (Figure 4.2). This assumption is based on some evidence and reflections, e.g. the fact that in order to achieve conceptual interoperability, based on data models and application schemas, the syntactic and semantic interoperabilities are important and the pragmatic and dynamic interoperabilities are not. Concepts need syntax and semantics to be expressed but they do not need the pragmatic and dynamic levels, except in the case in which those interoperabilities are necessary to access an existing repository of concepts or data models. Another similar example is about the legal aspects such as the intellectual property or access restrictions and/or use of resources. It is reasonable to think of the need of syntactic and semantic interoperabilities to express or specify these legal aspects while the conceptual, dynamic or pragmatic interoperabilities are not relevant. These two examples reveal that the dependence relations between interoperability levels are not hierarchical

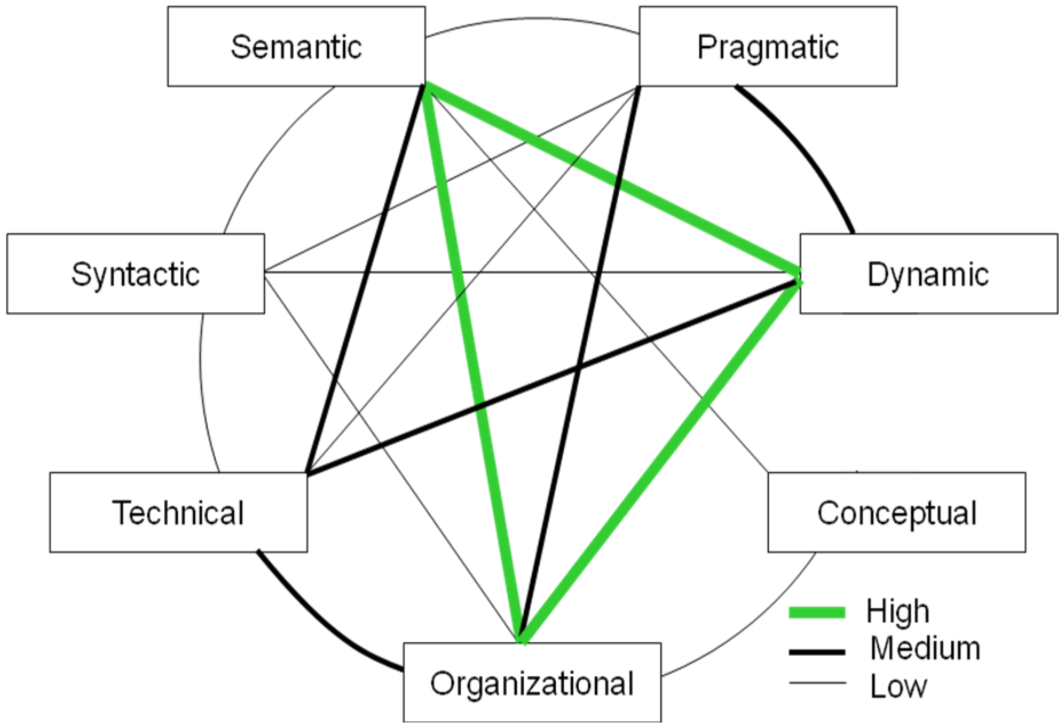


Figure 4.2. Integrated interoperability model proposal

Based on the integrated interoperability model proposal, we intend to analyze the interoperability levels provided by the metadata items describing a resource, as Shanzhen (1999) proposed in Intermodel5 with metadata functions. For this reason we assume that the metadata items may be used to build and describe the relations between the interoperability levels in the integrated model; that is our hypothesis.

Some authors such as Tolk, Diallo and Turnitsa (2007) emphasize the usefulness of metadata to achieve interoperability between systems in the LCIM context. As an example they mention the usefulness of metadata in the communications between intelligent software agents to 'communicate about situations', 'allow them selecting different components and evaluate their composition', 'support in decision taking' and finally 'to support the composition and arrangement of components nimbly, at least up to the dynamic level'.

Both the reference of Tolk, Diallo and Turnitsa (2007) and the integrated interoperability levels have prompted us to classify metadata according to the levels their items provide, so that the type of existing relations between the levels of the interoperability model for the SDIs, based on metadata, may be inferred.

Next the aims and scopes of the model levels are defined.

4.3.1 Technical interoperability

Like Turnitsa and Tolk (2006) and by similarity of models we define as technical the interoperability that enables interconnections of systems at the level of protocols and the exchange of information at its most basic level: bits.

Some examples related to SDIs have been identified: character sets, character encoding used in the data, file identifiers, description of the processing environment, file names, service types and versions, storage means, protocols and points of access to services.

4.3.2 Syntactic interoperability

It enables information exchange in a common format. By aspects of the syntactic interoperability we mean the standard formats of information exchange. It is the case of the XML format and the rules defining the data structure in the form of schemas (XSD) for every type of alphanumerical information. It is also the case of the image graphic formats (JPEG, PNG, and GTIFF). Within the SDI context, aspects of syntactic interoperability are all the XML schemas defined by the OGC for applications and Web services (WMS, WFS, WCS, CS-W, WPS, SOS), in addition to the formats for data encoding (GML, O&M, SensorML, TML) and the definition of presentation styles of objects (SLD and SE) or the syntax to define the filters (FE). All of them have been identified as initiatives increasing syntactic interoperability.

4.3.3 Semantic interoperability

It enables information exchange by using a commonly shared vocabulary avoiding inaccuracies in the interpretation of the meaning of terms.

Aspects of the semantic interoperability are the standards and/or specifications defining the schemas of information exchange and the meaning of every item unambiguously.

Examples are the Web Service Description Language (WSDL) and Simple Object Access Protocol (SOAP) for service interconnection, Geographic Markup Language (GML) for GI vector transfer, Symbology Encoding (SE) defining how to encode object symbology, Style Layer Descriptor (SLD) defining how to use symbology together with the WMS service, Common Query Language or Filter Encoding (CQL – FE) for queries and filters, etc. Likewise the ISO 19100 family of standards includes sections with the definition and controlled lists of terms that help share a common meaning. A particular case is ISO 19115 (Metadata Standard – 2003) containing at least 24 controlled lists of enumerated terms. Another case is the INSPIRE European Directive which is also making controlled lists of terms to classify data themes described in the Annexes in the categories defined by ISO 19115 (2003) for classification of resources.

4.3.4 Pragmatic interoperability

It enables the systems to know and exploit the methods and procedures provided by the other systems.

Aspects of pragmatic interoperability are the standards and specifications defining services taxonomy and their exploitation interfaces. Examples are ISO 19128 Standard, OGC Service Specifications (WFS - Web Feature Service, WCS - Web Coverage Service, CS-W - Catalog Service Web, SOS - Sensor Observation Service, WNS - Web Notification Service, WAS - Web Alert Service, LBS - Location Based Services, etc.), for which the standards or specifications define the interfaces allowing the exploitation and the parameters they are capable of handling. A characteristic aspect of the SDI and OGC contexts relative to the services is the mandatory operation getCapabilities, provided by all of them. This operation allows querying any service about its capabilities, thus obtaining a description of its capabilities and the implemented operations, contributing the point of access for its exploitation. This also occurs, in a similar but less enriching manner, in the Web Services and the WSDL.

4.3.5 Dynamic interoperability

It enables the systems to correct their own working when there are changes in information transfer, obtaining benefit from it.

Aspects of dynamic interoperability are the capability of dynamically replacing one system for another if inaccessible or if the service quality does not cover the needs. In this sense, the systems should have mechanisms to dynamically discover the existence of services complying with the requested requirements. A priori this interoperability level requires an exploitation of the semantic component making possible the discovery of the services based on the information describing them (metadata).

The definition of service taxonomies, as ISO 19119 does with the concept Service Organizer Folder (SOF) or the metadata implementation rules promoted by the INSPIRE Directive (2008), establishes the use of a set of identifiers for the services which are able to promote this interoperability type. Thus the INSPIRE Directive in its implementation rules promotes the classification of services according to their function into discovery, visualization, downloading and transformation. ISO 19119 classifies services into human interaction, and management of the model, tasks, processing, communications and systems. Additionally the INSPIRE Directive (2008) establishes a classification of the data types in which the categories defined by ISO 19115 (2003) are classified (topicCategory), facilitating service exchange dynamically, at least from a theoretical viewpoint.

4.3.6 Conceptual interoperability

It enables to know and reproduce the functioning of a system based on the documentation expressed in a format used in Engineering.

Aspects of conceptual interoperability are those describing the data and system models in the form of standardized and interchangeable documentation from an engineering viewpoint, without depending on the model used to describe it. The description through UML of the data model, either from a data store or provided by a Web service, makes this interoperability type possible. Some of the OGC service specifications such as WFS (ISO19142) provide a conceptual description of the features as a response to the describeFeatureType operations. In these cases a conceptual description of the feature in the form of GML application schema is being recovered. At the present time, in the implementation rules of the INSPIRE Directive for data types, models are being defined using CASE tools that allow exchanging those models of classes and UML language restrictions by using standard exchange formats such as XMI, e.g. *Addresses.*

4.3.7 Organizational interoperability

It enables to know the business objectives, the process models, the laws and access policies and the use of data and services.

Aspects of organizational interoperability are those enabling knowledge and understanding of access policies and use of data and services, personal and institutional responsibilities and the objectives and goals defined by the organization when creating a certain type of data or providing a type of service. An important part of the information concerning the policies of use and access is considered identification or description of constraints; it is useful information to evaluate the use of the resources described by metadata.

4.4 Classification of metadata items according to the interoperability model

In order to analyze the role metadata may play in the different interoperability levels of the proposed model, a classification of the different items GI metadata are made up of has been proposed (ISO 19115/19139), focusing on the interoperability levels they may provide.

Since the ISO 19115:2003 Metadata Standard may potentially contain over 400 items (ANZLIC, 2005) depending on the type of GI being described, we have opted for making the analysis in two stages. In the first stage the analysis is applied only to the core of the standard and in the second stage the analysis is extended to the entire standard.

For the first case the items making up the core have been brought together in a spreadsheet and the interoperability provided by each item has been studied. For the second case, the items of the metadata standard have been dealt with in accordance with the packages/classes defined by the standard: identification, quality, restrictions, maintenance, distribution, reference system, application schema, spatial representation and content information.

Next the results of the classification have been analyzed at the descriptive level with two scopes: a) incidence of interoperability levels assigned to items, and b) incidence of relations between pairs of interoperability levels of the model.

The first analysis has consisted of measuring the incidence of interoperability provided by the levels of the model within the scope of the study (items of the core and total items of the standard). The second analysis has consisted of measuring the incidence of relations between levels of the model or the incidence of metadata items providing simultaneously two interoperability levels, also within the scope of the study.

Next (1) the interoperability levels of the metadata items are identified, (2) the levels provided by the metadata items belonging to the core and to the entire standard are analyzed and (3) the results are compared. Now, the incidence of relations between interoperability levels is analyzed and the process is repeated; first the core items and then the total items, then the results are compared. Finally the relations between the levels of the integrated interoperability model are interpreted within the SDI context.

4.4.1 Identification of the interoperability levels of the metadata items

The identification of the interoperability levels provided by the ISO 19115 metadata items has been carried out on a spreadsheet for each package/class of items in the metadata standard, so as to make its treatment and subsequent analysis easy and thorough. The names of the 52 items, a brief description and seven columns for the levels of the model have been incorporated into the spreadsheet, as can be seen in Figure 4.3. The first column contains the names of the items, the second column is a description of the items, and the following seven columns identify the interoperability levels provided by the item. In a column not shown in Figure 4.3, a description explaining the choice of interoperability levels is presented. The red background of a cell indicates the item is mandatory and the orange background indicates conditionality (mandatory item in certain conditions).

Packet	**MD_Metadata**							
Element	**Description**	**Technical**	**Syntactic**	**Semantic**	**Pragmatic**	**Dynamic**	**Conceptual**	**Organisational**
fileIdentifier	unique identifier for this metadata file	x			x			
language	language used for documenting metadata			x				x
characterSet	full name of the character coding standard used for the metadata set	x		x				
parentIdentifier	file identifier of the metadata to which this metadata is a subset (child)	x			x			
hierarchyLevel	scope to which the metadata applies			x			x	x
hierarchyLevelName	name of the hierarchy levels for which the metadata is provided							x
contact (CI_ResposableParty)	*party responsible for the metadata information*							
dateStamp	date that the metadata was created			x		x		x
metadataStandardName	name of the metadata standard used		x	x				x
metadataStandardVersion	version (profile) of the metadata standard used		x	x				x
dataSetURI	Uniformed Resource Identifier (URI) of the dataset to wich the metadata applies	x		x	x			
locale (PT_Locale) (ISO19139)	*Information about linguistic alternative*							
spatialRepresentationInfo (MD_SpatialRepresentation)	*digital representation of spatial information in the dataset*							
referenceSystemInfo (MD_ReferenceSystemInfo)	*description of the spatial and temporal reference systems used in the dataset*							
metadataExtensionInf (EX_Extent)	*information describing metadata extensions*							
identificationInfo (MD_identification)	*basic information about the recource(s) to which the metadata applies*							
contentInfo (MD_ContentInformation)	*provides information about the feature catalogue and describes the coverage and image data characteristics*							
distributionInfo (MD_Distribution)	*provides information about the distributor od and options for obtaining the resource(s)*							
dataQualityInfo (DQ_DataQuality)	*provides overall assessment of quality of a resource(s)*							
portrayalCatalogueInf (MD_PortrayalCatalogueReference)	*provides information about the catalogue of rules defined for the portrayal of a resource(s)*							
metadataConstrains (MD_Constraints)	*provides restrictions on the access and use of metadata*							
applicationSchemaInf (MD_ApplicationSchemaInformation)	*provides information about the conceptual schema of a dataset*							
metadataMaintenance (MD_MaintenanceInformation)	*provides information about the frequency of metadata updates, and the scope of those updates.*							

PT_Locale	Defines the locale in which the value (sequence of characters) of the localised character string is expressed							
language	Designation of the locale language			x				x
country	Designation of the specific country of the locale language			x				x
characterEncoding	Designation of the character set to be used to encode the textual value of the locale	x		x				

Figure 4.3. Data schema of the interoperability analysis document

Table 4.1 shows the incidence of interoperability levels in the core items and Figure 4.4 shows those values as a histogram.

Interoperability level	Total items 52	%
Technical	8	15,38
Syntactic	3	5,7
Semantic	40	77
Pragmatic	3	5,7
Dynamic	31	59,6
Conceptual	1	1,9
Organizational	43	82,7

Table 4.1. Incidence of the interoperability levels in the core items

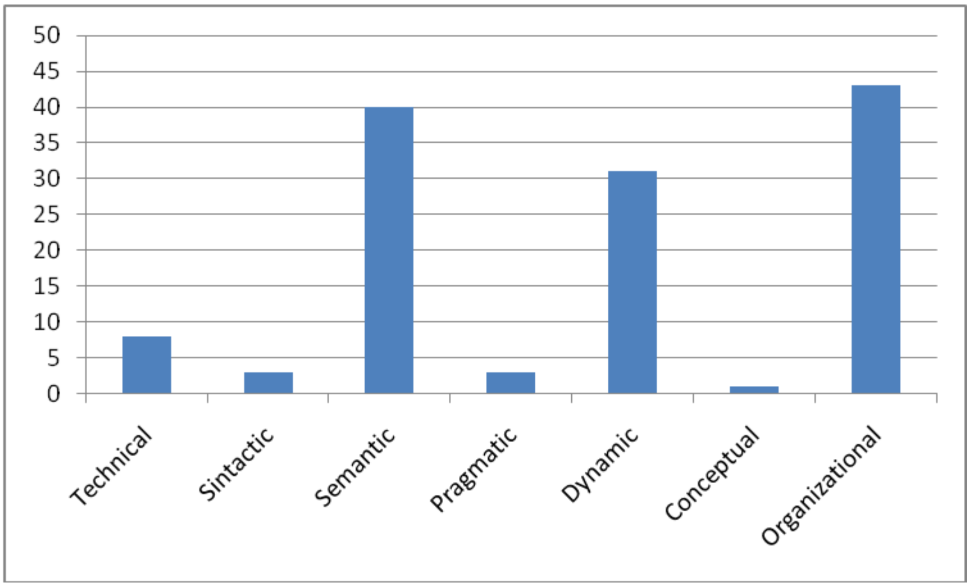


Figure 4.3. Histogram of the interoperability levels in the ISO 19115 core items

Figure 4.4 shows that the organizational, semantic and dynamic levels are most favored by the core items of the standard. Approximately 80% of the items provide with organizational and semantic interoperability, 60% of them with dynamic interoperability and the remainder of the levels are scarcely favored, with 15% and 2%. In view of these results uncertainty arises about the analysis of the total items of the standard. Will the results be similar? Table 4.2 and Figure 4.5 show the results.

Interoperability level	Total items 235	%
Technical	37	16
Syntactic	6	2,6
Semantic	128	54
Pragmatic	15	8,7
Dynamic	99	42
Conceptual	7	3
Organizational	217	92

Table 4.2. Incidence of interoperability levels in the total items of the standard

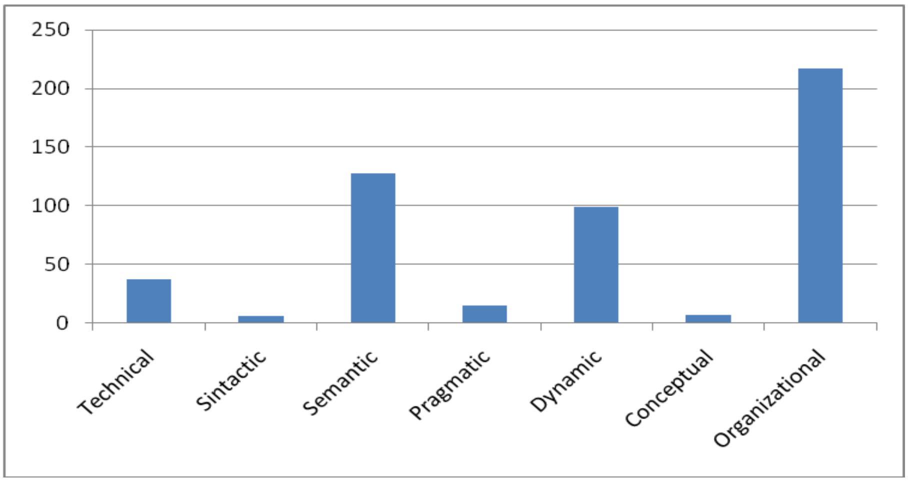


Figure 4.4. Histogram of the interoperability levels in the items of the ISO 19115 standard

The main conclusions that can be drawn are:

- There is a remarkable resemblance between both analyses. The most favored levels are the organizational, dynamic and semantic levels in both cases. Least favored by the metadata items are the conceptual and pragmatic levels.
- The percentage of items provided by the organizational, pragmatic and conceptual levels is maintained in the two scopes of the study. The remainder of interoperability levels lessens their representation when studying the total items of the standard versus the items of the core.
- The lack of items provided by the pragmatic and conceptual interoperability levels persists in both scopes since the percentages of items favoring those levels are very low: 6% and 2% respectively.

- Therefore it may be stated that the ISO 19115/19139 metadata standard provides chiefly the organizational interoperability, contributing much information describing the policies of access and use of data, responsibilities, objectives and goals; in brief, useful information to evaluate the use of data and services.
- It may also be stated that the standard does not favor the conceptual interoperability of data. The metadata standard includes only seven useful elements to describe the data model from an engineering viewpoint.

4.4.2 Relations between the interoperability levels of the model

First it is necessary to analyze how metadata provide interoperability levels and then to describe the relations between levels and their incidence. The method used for the analysis has been to count the number of metadata items of ISO19115 providing simultaneously two or more levels of the model, defining two scopes for the study: (1) the core items, and (2) all the items.

Concerning the core items, Table 4.3 shows the number of items simultaneously favoring two interoperability levels of the model, so that in the intersection of a row and a column the absolute values (simultaneous number of items) for the two levels, and the relative values of the level identified in the column are observed. As an example, the values 28/70%, showing at the intersection of the Dynamic row and the Semantic column, indicate that there are 28 items that simultaneously provide the two levels, representing 70% of the 40 items providing semantic interoperability.

By analyzing the values of Table 4.3 it is observed that the absolute values of the main diagonal are the frequencies studied above. Likewise it may be seen that those numbers appear symmetrically in Table 4.3 and the results are consistent with the method used to obtain the data, which consisted in counting the number of items providing two interoperability levels, dealing with them independently.

	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Technical	**8**	0	4	2	3	0	3
	100%	0%	10%	67%	10%	0%	7%
Syntactic	0	**3**	3	0	1	0	2
	0%	100%	8%	0%	3%	0%	5%
Semantic	4	3	**40**	0	28	1	29
	50%	100%	100%	0%	90%	100%	67%
Pragmatic	2	0	0	**3**	2	0	1
	25%	0%	0%	100%	6%	0%	2%
Dynamic	3	1	28	2	**31**	0	23
	38%	33%	70%	67%	100%	0%	53%
Conceptual	0	0	1	0	0	**1**	1
	0%	0%	3%	0%	0%	100%	2%
Organizational	3	2	29	1	23	1	**43**
	38%	66%	73%	33%	74%	100%	100%

Table 4.3. Relations between the interoperability levels according to count of core items

It may be observed in Table 4.3 that the number of metadata items providing the organizational, semantic and dynamic interoperability levels, taken in pairs, is high, and an average of 70% of the items favoring one of these levels favors the other two. Table 4.3 only shows the number of items providing two levels of the model simultaneously. Table 4.4 counts the number of items providing only one, two, three… levels of the model simultaneously and Figure 4.6 shows the distribution graphically.

# simultaneous levels	1	2	3	4	5	6	7
# items favoring them	11	17	27	1	0	0	0

Table 4.4: Number of core items favoring one, two, three… levels of the model

It should be noted that the 11 items, appearing in the second column of Table 4.4 indicating they only provide one interoperability level, do it at the organizational level.

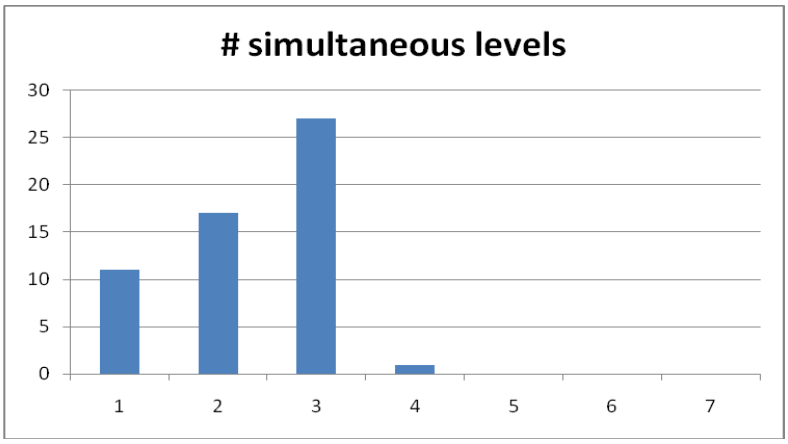


Figure 4.5: Distribution of the number of core items providing only one number of levels of the model

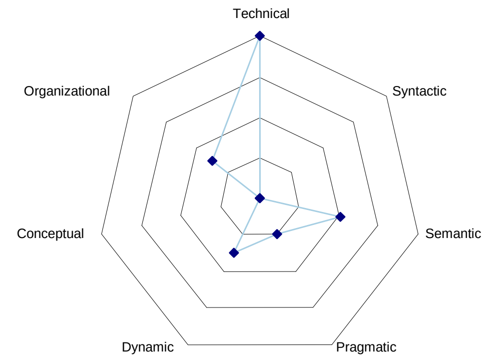
Table 4.3 shows the number of items providing two interoperability levels simultaneously. For this reason Table 4.5 has been drawn with the purpose of showing the remainder interoperability level combinations enabled at once by the core items of the metadata standard. It can be observed that there is an important number of items simultaneously providing the semantic, dynamic and organizational levels.

Interoperability levels provided by the items	# items
Technical, Semantic and Dynamic	1
Technical, Semantic and Organizational	1
Technical, Dynamic and Organizational	3
Syntactic, Semantic and Dynamic	2
Syntactic, Semantic and Organizational	2
Semantic, Dynamic and Organizational	20
Semantic, Conceptual and Organizational	1
Pragmatic, Dynamic and Organizational	1
Technical, Semantic, Dynamic and Organizational	1

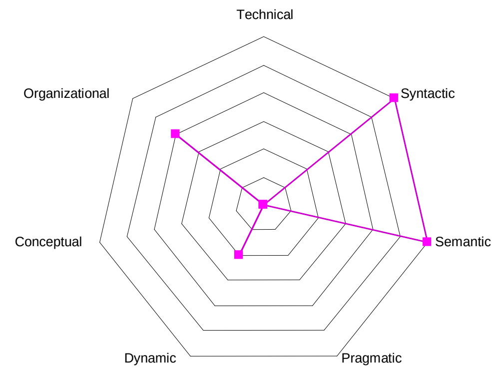
Table 4.5. Interoperability levels provided simultaneously

The main differences between Tables 4.3, 4.4 and 4.5 in regard to the number of items providing certain levels of the model have their origin in the form of counting. While Table 4.4 shows the number of items providing only one number of levels, the other two show all the possible combinations of levels. As an example, if a certain item provides three levels in the model (1, 2 and 3), it will be computed only as "item providing three levels" in Table 4.4, while in Table 4.3 the combinations of these three elements taken in pairs will be counted (1,2 + 1,3 + 2,3).

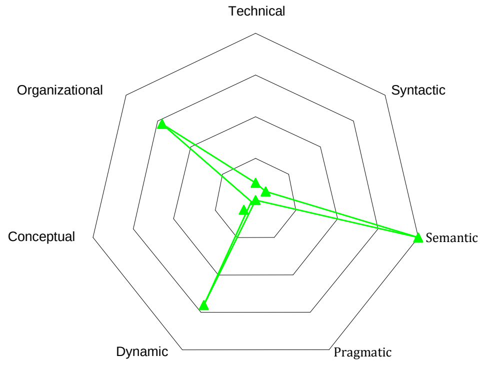
The disparity of the magnitudes shown in Table 4.3 makes the interpretation and description of the interrelations particularly difficult. This fact has motivated the search for a visualization technique for representation of relations graphically, easy to read and to interpret. A number of tests have been tried before the use of radial diagrams was adopted as the representation technique. First all the values of the table were represented in one single diagram but the result did not show more information than the table. Finally it was decided to show one radial diagram for each level and the relations of this level with the others. The results are displayed in Table 4.6.



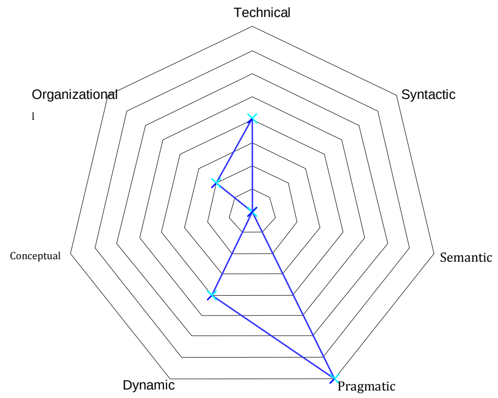
Technical



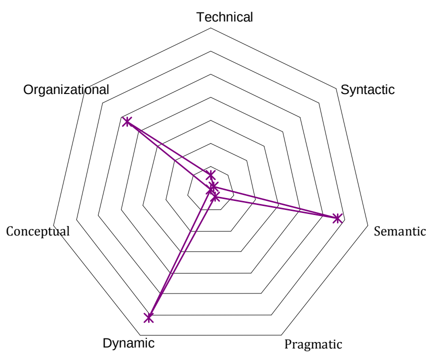
Syntactic



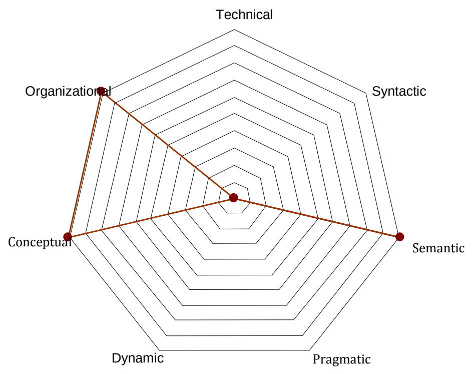
Semantic



Pragmatic



Dynamic



Conceptual

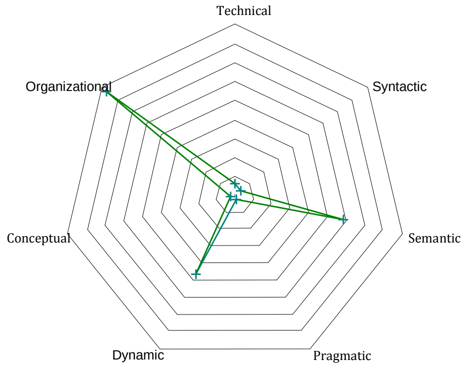
 Organizational	

Table 4.6. Graphic representation of the number of items providing two levels of the model

Every radial diagram represents in percentages the number of items which, in addition to the interoperability level appearing on the bottom of every figure, provides the remainder of levels. The similarity of the diagrams corresponding to the semantic, dynamic and organizational interoperabilities is observed, meaning that the same items provide the three levels simultaneously. In the remainder of the diagrams the relations established between pairs of interoperability levels without any reciprocity is also shown.

The following conclusions can be drawn from the diagrams of Table 4.6:

- The three diagrams associated to the semantic, dynamic and organizational interoperabilities are similar because the number of items providing them is very high compared to the remainder of the levels.
- Seven pairs of interoperability levels are not provided by the core items of the standard. In some cases this can be explained arguing the difficulty to identify a metadata item that might facilitate interoperability levels as different as technical-conceptual or syntactic-conceptual.
- A very low number of core items facilitating the technical, syntactic and pragmatic interoperabilities have been identified. It is difficult to draw conclusions from this fact because the core of the standard has been defined according to the roles played by data, not by interoperability criteria.

The same methodology has been used to analyze all the items of the metadata standard; the results are shown in Table 4.7. As can be seen, in the elements of the main diagonal of this table, the number of items analyzed to classify the relations between interoperability levels differs from the ones shown in Table 4.2. In a previous case the 400 possible items had been taken into account (ANZLIC, 2005); in this study the items belonging to the CI_Citation and CI_ResponsibleParty packages/classes have been computed only once.

	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Technical	**37**	3	12	8	11	0	26
	100%	50%	9,3%	53,3%	11,1%	0%	12%
Syntactic	3	**6**	3	3	4	0	3
	8,1%	100%	2,3%	20%	4%	0%	1,3%
Semantic	12	3	**128**	4	83	6	118
	32,4%	50%	100%	26,6%	83,8%	85,7%	54,3%
Pragmatic	8	3	4	**15**	12	0	10
	21,6%	50%	3,1%	100%	12,12%	0%	4,6%
Dynamic	11	4	83	12	**99**	0	90
	29,7%	67%	64,8%	80%	100%	0%	41,4%
Conceptual	0	0	6	0	0	**7**	7
	0%	0%	4,6%	0%	0%	100%	3,2%
Organizational	26	3	118	10	90	7	**217**
	70,27%	50%	92,2%	66,6%	90,9%	100%	100%

Table 4.7: Relations between the levels of the model according to the items of the standard

It may be noted that in Table 4.7:

- There are four pairs of interoperability levels (conceptual-technical, conceptual-syntactic, conceptual-pragmatic and conceptual-dynamic) in the model that have not been provided by the items of the metadata standard.
- The items favoring the conceptual level also provide the organizational level (7) and the semantic level (6), therefore it may be concluded that there is a close relation between the conceptual level and the organizational and semantic levels. This relation is not symmetrical since only 7 out of the 217 items providing organizational interoperability favor conceptual interoperability.
- The low number of items providing the conceptual interoperability (7) persists, as occurred when analyzing the core items.
- A representative number of items providing both pragmatic interoperability (15) and dynamic interoperability (12) are detected; hence this may be

interpreted as a close relation between both levels of the model, although this relation is also asymmetrical since only 12% of the items providing dynamic interoperability favor pragmatic interoperability.

- As can be seen in the last row of Table 4.7, the percentages of items providing at the same time organization interoperability for every level of the model is high: 50%, 66%, 70%, 90% and 100%. A possible interpretation is that ISO 19115 has been mainly defined to meet the needs of interoperability at the organizational level.
- The computed mathematical correlation between the number of metadata items providing each level in the scopes of the study (Table 4.3 and 4.7) results in a value of 0.9436, indicating there is a high degree of association between the outcome of both analyses.

The number of items only providing simultaneously one, two, three… levels of the model has also been analyzed; the results are shown in Table 4.8 and Figure 4.7 respectively.

# simultaneous levels	1	2	3	4	5	6	7
# items favoring them	80	46	101	6	2	0	0

Table 4.8. Number of items of the metadata standard providing only one, two, three… levels of the model

Another fact adding weight to the interpretation of the organizational level as the center of gravity in the definition of the metadata standard is that 74% of the 80 items (92%) providing one single interoperability level favor that level, as shown in Table 4.8,

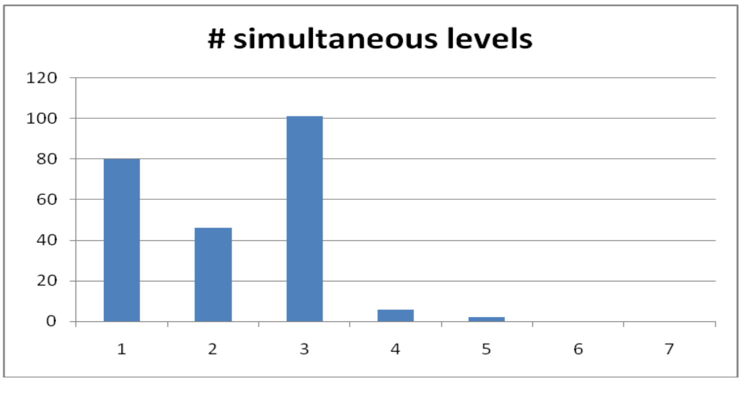


Figure 4.6: Distribution of the number of items of the metadata standard providing only a number of levels of the model

Next we will interpret the relations between the levels of the model based on the metadata items favoring them, with the purpose of describing the benefits they provide to the other level

Technical and pragmatic interoperability. The items data access point, service, format, file version, character set and communication protocol to be used enable technical interoperability and facilitate the use of data/services in a practical manner.

Technical and dynamic interoperability. Certain items such as data compression technique, file size, service type or version, in addition to enabling technical interoperability, allow evaluating whether the data or the service may be used in a dynamic context by other systems.

Syntactic and pragmatic interoperability. The pragmatic interoperability takes advantage of the syntactic aspects that allow description of attributes as important as the names of the parameters of a service, the typology of data, mandatory nature and cardinality.

Semantic and technical interoperability. The technical interoperability takes advantage of the metadata items contributing semantics, e.g. controlled term lists identifying file format (MD_MediumFormatCode), storage support (MD_MediumNameCode) or character set used in the data (MD_CharacterSetCode).

Semantic and pragmatic interoperability. The elements providing semantic interoperability also provide the pragmatic level, e.g. the metadata items using controlled lists to define GeoService categories, as is the case of the INSPIRE (2008) metadata implementation rules, the type of spatial representation (MD_SpatialRepresentationTypeCode), the status of the data (MD_ProgressCode) or security constraints (MD_SecurityConstraints) are all items that favor the pragmatic interoperability in addition to strengthening semantic interoperability.

Semantic and dynamic interoperability. The dynamic interoperability is favored by the metadata items that provide semantic interoperability. These items are, among others, the term lists allowing classification of the resources in topics or keywords belonging to domain multilingual dictionaries (topic, place, and layer). These items provide the semantic level and improve the quality of the results of data or service searches that may be used in a dynamic environment by the systems.

Semantic and conceptual interoperability. The most representative item favoring both levels is the item defining the scope of metadata (MD_ScopeCode).

Semantic vs organizational interoperability. The organizational interoperability takes advantage of some aspects of the semantic interoperability, e.g. definition of the role played by party in charge of data, types of legal constraints for access and use of data or services and security constraints (CI_RoleCode, MD_RestrictionCode, MD_ClassificationCode). As may be seen, all the mentioned examples are also controlled term lists.

Organizational and dynamic interoperability. The dynamic interoperability relies on aspects related to organization, such as sources of data, constraints and/or limitations of access or use of data. The metadata items that describe those organizational aspects facilitate reuse of data and service exploitation.

Pragmatic and dynamic interoperability. The dynamic interoperability is helped by pragmatic aspects to exploit data or services after these have been identified and their suitability for the pursued objectives has been assessed. It may be stated that this relation between the pragmatic and dynamic levels is not symmetrical.

4.5 Conclusions

- A new methodology for automatic metadata creation without human intervention has been proposed. It may also be used as a first step of a later refinement development on the part of users expert in the knowledge domain or information cataloging and finally, it may be used to update metadata as the data change.
- The new methodology:
    - automates the extraction of both implicit and explicit information from the data stores;
    - performs reasoning and translations to identify coordinate reference systems;
    - carries out coordinate conversion/transformation calculations;
    - locates geographic identifiers relevant for the geographic boundingBox?
    - represents data graphically;
    - infers data content based on stored statistical data and rules that allow data cataloging;
    - infers the data model and represents it in formats used in engineering, to finally
    - pack all the information obtained by various methods in XML format metadata, gathering together the remainder of the information in a metadata exchange format (MEF).
- As far as interoperability is concerned, an integrated model for SDIs has been defined based on seven levels: technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational. The choice of these levels has been made taking into account that SDIs are a particular case of systems of systems and there are interoperability models in this context. Since the aspects related to organization are important for SDIs, in this model the organizational level has been included in addition to the appropriate levels for SoS.

- The influence of the items belonging to the international metadata standard on the levels of the model has been analyzed. The scope of the study has been double: core items and total items of the standard. The results regarding interoperability in both cases are similar and indicate that the most favored levels of the model are the organizational, dynamic and semantic levels.
- When analyzing how the metadata items of the standard provide interoperability simultaneously to several levels, it was observed that an important percentage of items providing a certain level also provided the organizational level. 92% of the items only providing one level of the model do it at the organizational level. Our interpretation is that the ISO 19115 metadata standard to a large extent meets the interoperability needs at the level of organization and has not been designed to provide other interoperability levels appropriate for SDIs.
- The graphs representing the number of items simultaneously providing one, two, three… interoperability levels indicate that the items provide few interoperability levels simultaneously. Considering all items of the standard, there are 80 of them (37%) providing only one level and 101 items (46%) providing three levels simultaneously. 80% of the latter correspond to the semantic, dynamic and organizational levels, prevalent in this study.
- In our view syntactic interoperability is guaranteed by the standard when it defines the encoding rules of the items that make it up.
- The technical and pragmatic interoperability must be guaranteed by other standards of the same family, defining the communication protocols and data transfers or defining the service interfaces enabling access, treatment, conversion or visualization of data.
- An important lack of items providing conceptual interoperability in the standard is detected; it has only seven items that could describe the information data model and their definition is vague.
- The proposed interoperability model based on seven levels may be used to define metadata profiles providing for the different interoperability levels of the model in a balanced manner and maximizing the interoperability levels/ metadata item ratio. This is possible if a new metadata standard was designed meeting the interoperability needs and requirements.

5 IMPLEMENTATION AND RESULTS

5.1 Introduction

The objective of this chapter is to present the implicit metadata contained in the GI as sources of items for automatic metadata creation and the heterogeneity of the formats of representation of the Spatial Reference Systems (SRS). First we study the typology and number of metadata items that may be obtained from the data by means of utilities and libraries extracting the information stored in formats or through tables identifying the encodings used, with the purpose of expressing the SRS for the coordinates.

In the second place the proposed methodology to create metadata automatically is analyzed, taking into account the results of previous studies and identifying the remainder of items that may be created through calculation and inference or those that may be deduced from the context. Here the items that may be obtained with the methodology are described individually, identifying the typology of the data to which those items apply.

In the third place we should answer the fifth research question posed in the PhD thesis: What are the strengths and weaknesses of manually and automatically generated metadata in terms of interoperability of the systems that will exploit them (SDI)? With this purpose the items that may be created by the methodology of automatic creation of metadata are analyzed from two points of view: the interoperability they facilitate and the functions they have. In order to do that, data have been classified according to their nature: raster, DEM and vector. The results achieved with the methodology are compared with the results obtained by the analysis of the items' interoperability levels (both core items and total items of the standard). Finally, some reflections are presented.

In the fourth place we should answer the fourth research question posed in the PhD thesis: What proposal is most appropriate to validate a system interoperability model within the SDI context? For the validation of the interoperability model, an inquiry has been designed headed by a group of metadata experts to whom the needed information has been provided to identify the interoperability levels provided by the core items of the metadata standard. The analysis of both individual and aggregate results is used to detect disparities with the proposed model and finally to accomplish the validation.

Finally conclusions are presented.

5.2 Storage formats used in SDIs

GI, as a function of its nature, may be stored in different ways: rasters (images), vectors, DEMs or databases.

In every country, either because of cultural factors or available technology, a certain number of formats for GI exchange has been adopted as de facto standards. In many cases these formats are owned by companies developing technology in this sector and there is no public information describing the structure of the format; in other cases formats are the outcome of efforts at building consensus and standardization of organizations collaborating worldwide.

The last objective of this chapter is to identify the metadata items that may be extracted from the diverse storage formats. Formats have been classified according to the nature of the data they store. Finally, the metadata that can be recovered from the formats are described.

- General purpose raster (matrix) formats. Graphic formats were developed a long time ago within the context of computerization to store aerial images. Their main advantage was the availability of computer tools and libraries that allow graphic manipulation. There were also many drawbacks: number of bands, limited radiometric resolution or number of bits per pixel and inability to store metadata. The emergence of new standardized private products has caused those formats to be increasingly in disuse. They are general purpose raster formats: BMP, PNG, RAS, TIFF, JPEG, GIF, IFF and PCX.

- Raster formats used to store digital aerial photographs, orthophotographs and rasterized cartography; these formats allow storing a large number of data, they support a larger number of bands, they have more data types for the pixels and they provide a larger compression capability and multiresolution. In certain cases they also enable storing metadata about spatial reference systems (SRS). Some of the raster formats with these characteristics are: GeoTIFF, MrSID, ECW, JPEG2000, GeoJP2, INGR and NITF.
- Raster formats used in remote sensing: In order to store satellite images, it is necessary to use formats enabling storage of hyperspectral images, with radiometries stored as real numbers and storage of a large number of additional metadata. Some of the raster formats used in remote sensing are: IMG (or HFA), PIX (or PCIDISK), ERS, IMG (Idrisi), NOAAL1B and EOSAT.
- Computer-Aided Design vector formats (CAD). CAD is a computer tool used to edit digital cartography, whose purpose is printing as a map or plan. It is used in the domain of engineering and architecture and, although it has tools to store GI, it is more commonly used to store geometries and visualization styles. Same as for general purpose raster formats, the CAD formats do not usually store metadata, i.e. feature types, SRS, etc. Some of the CAD formats are: DGN (ISFF), DWG and DXF, FHX (Macromedia), AI (Adobe) and BIN (DIGI21).
- Vector formats used by GIS. The GIS usually set apart geometry, alphanumerical information and feature visualization. Some formats used by GIS are: ADF and E00 (ArcInfo), SHP, MIF and DAT (MapInfo), VEC (Idrisi), to which other GIS raster formats should be added.
- Digital terrain models (DTM) stored as vector format. They are formats storing TINs vectorially. Some of these formats are: TIN (Intergraph), MDT (MDTop).
- DTMs stored as matrices. They are formats supporting radiometric data and many other real type cases to store terrain elevations by means of points. The chief deficiency of these formats is that they do not usually store metadata; hence they do not store SRS or other important information. Some of the matrix formats for DTM are: ADF, GRD (ESRI), GRD (GoldenSoft), DEM

(USGS), DEM (MicroDEM), DTE (Socet Set), DT0 (DoD), HGT (Shuttle Radar Topography Mission), BIL, BIP and BSQ (MapInfo).

- GI stored in spatial databases. The DBMS have evolved and in addition to store alphanumerical objects, they are capable of storing geometric objects, carrying out topological operations with them and spatially indexing contents. In other cases some GIS implement middleware between relational DBMS and GIS to supplement the databases with those capabilities. In both cases DBMS store metadata in the shape of ancillary tables. Some examples of DBMS with native spatial extensions are: Oracle Spatial, PostGreSQL+PostGIS, MySQL, DB2 (IBM) and SQL2008 (Microsoft). Other applications capable of storing GI in relational DBMS are: MGE, Geomedia, ArcSDE, and MapInfo.

5.2.1 General purpose matrix formats

After reviewing the technical documentation of the raster formats BMP, PNG, RAS, TIFF, JPEG, GIF, IFF, PCX and PSD, it has been verified that the following set of characteristics may be read in the files:

- Dimensions of the image (width and height in pixels);
- Number of bands or components;
- Number of bits used to represent each pixel;
- Type of compression used.

The georeferencing information is managed by means of an ancillary text file containing the parameters needed to carry out an affine transformation (rotation, translation and scaling) "*World file*" (World file, 2007).

The inconvenience of these formats is the unawareness of the SRS to which the coordinates contained in the georeferencing file relate. Some formats, e.g. PNG and TIFF may contain additional metadata such as author, content description, date of creation and information sources. Table 5.1 shows the metadata items that may be extracted from this type of files.

Format	Width/Height	# bits/pixel	# bands	Compression type	Geographic extent	Pixel resolution	Other
BMP	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	**Printing resolution**
GIFF	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	
IFF	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	
JPEG	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	
PCX	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	
PNG	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	**Author, description, date of creation, sources, processing steps and legal constraints**
PSD	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	
TIFF	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	**Author, description, date of creation, sources**
RAS	**X**	**X**	**X**	**X**	**X** (1)	**X** (1)	

Table 5.1. Generic raster formats

5.2.2 High compression raster formats

The main difficulty to study these formats is the lack of public technical documents describing them, especially with the private MrSID and ECW. These difficulties have been mitigated by analyzing the information showing tools developed by the same commercial firms and identifying the ones having metadata items.

The information that may be obtained for the analyzed formats of this type (GTIFF, MrSID, ECW, JPEG2000, GeoJP2, INGR and NITF), in addition to the common information obtained for the general purpose raster formats, is as follows:

- Pixel resolution in each axis;
- Measure units;
- SRS.

Some formats, e.g. GeoTIFF encode the SRS through EPSG-defined numerical identifiers. Others identify reference systems through text, mnemonic or individual codes. Besides, some formats contain other metadata such as date of creation, compression quality, data source, constrains of use and/or access to information and more detailed parameters of cartographic projections. The capability of the JPEG2000 format of metadata storage stands out by including an information block at the header for this purpose. Table 5.2 shows the metadata items used to store this type of information.

Format	Width/Height	# bits/pixel	# bands	Compression type	Geographic extent	Pixel resolution	Spatial Reference System	Measure units	Other
ECW	X	X	X	X	X	X	X	X	
GeoJP2	X	X	X	X	X	X	X	X	**Author, description, date of creation, sources and other projection parameters**
GTIFF	X	X	X	X	X	X	X	X	**Author, description, date of creation, sources and other projection parameters**
INGR	X	X	X	X	X	X	X	X	
JPEG2000	X	X	X	X	X	X	X	X	**Open. GML_JP2K Spec.**
MrSID	X	X	X	X	X	X	X	X	**Compression quality, date of creation**
NTIF	X	X	X	X	X	X	X	X	**Date of creation, title, author, source, legal constraints**

Table 5.2. Raster format for large volumes of data

5.2.3 Raster formats used in remote sensing

After review of the formats we have sufficient information to identify the metadata items stored in the storage formats. This analysis has been made for LAN & IMG (ERDAS), PCIDISK, ERS, IDRISI, NOAAL1B and F-EOSAT formats. All of them have in common that they store a large number of metadata, many of which do not fit into the ISO 19115 standard, although they could be partially included in ISO 19115-2. This standard is supposed to store the metadata needed to describe both regular information grids and satellite imagery.

In addition to the common metadata of the previous formats, these types of files store:

- Type of data used to store pixels;
- Date of acquisition;
- Satellite platform;

- Reception station applying preprocessing;
- Preprocessing level;
- Statistic parameters of the pixel digital values;
- Parameters of the acquisition system (position and angles)

and others that may be hard to locate in other metadata items. Table 5.3 shows the metadata items of the formats used to store this type of information.

Format	Width/Height	Data type	# bands	Geographic extent	Pixel resolution	Statistics: max, min max, min	Statistics: other Other	Spatial Reference System	Other
LAN (Erdas)	X	X	X	X	X			X	
IMG (Erdas)	X	X	X	X	X	X	X	X	**Content type**
PIX	X	X	X	X	X			X	**Date of creation, processing date, process**
ERS	X	X	X	X	X	X		X	**Sensor, band spectral resolution**
DOC (Idrisi)	X	X	X	X	X	X		X	**Title, information source, categories**
NOAAL1B	X	X	X	X	X			X	**Platform, control points, receptor and format**
F-EOSAT	X	X	X	X	X			X	**Acquisition date, sensor, satellite, reception station**

Table 5.3. Raster formats used in remote sensing

5.2.4 Digital Terrain Models stored as matrix

The identified formats for storage of matrix DTM are: BIL, BIP & BSQ (MapInfo), Gtopo30, Export Raster (Erdas), HGT (SRTM), ADF & GRD (ESRI), Grid (Surfer), DEM (USGS), DEM (MicroDEM), DTED (DoD), DOQ2 and DTE (Socet Set). It has been verified that the obtainable metadata items are similar for all of them. The main difference concerns SRS storage (datum, ellipsoid, projection, zone, parameters), which is implicit for GTOPO30 and HGT format while the DTE (Socet Set) and Grid (Surfer) do not store the spatial extension. The information that can be extracted from the formats is basically the same as in the case of the previous ones, including max and min height. Table 5.4 shows the metadata items used to store this type of information.

Format	Width/Height	Geographic extent	Pixel resolution	# bits/pixel	# bands	Statistics:max, min max, min	Statistics: other Other	Horizontal units	Spatial Reference System	Other
DOQ2 (USGS)	X	X	X	X	X			X	X	
DTED	X	X	X	X	X			X	X	
Gtopo30	X	X	X	X	X	X[2]		X	X	
HGT (SRTM)	X	X	X	X	X	X[2]		X	X	
BIL, BIP, BSQ (Mapinfo)	X	X	X	X	X	X[2]		X	X	
DTE (Socet set)	X	X	X	X	X			X		
GRD (Esri)	X	X	X			X[2]		X	X	
GRD (Surfer)	X	X	X			X				
DEM (USGS)	X	X	X	X		X		X	X	
DEM (MicroDEM)	X	X	X	X	X	X		X	X	
E00 grd	X	X	X[3]		X	X[3]	X[3]	X[3]	X[3]	
ADF grd	X	X	X[3]		X	X[3]				

Table 5.4. Raster formats used in DTM

[2] The whole file may be read and the values calculated

[3] If all the file sections are present

5.2.5 CAD type vector formats

The following CAD formats have been identified: DGN (ISFF), DWG & DXF, BNA (Atlas BNA) and BIN (DIGI21). The main metadata items that may be read or calculated in this type of formats are:

- Max and min coordinates of the geographic bounding box.
- Altimetry (max, min).
- Number and type of existing geometries.

In addition, some formats cluster the information by layers; in this case the name of the layers may be useful to identify the feature types in the catalog. Table 5.5 shows the metadata items of the formats used to store this type of information.

Format	Geographic Extent	# layers	Name of layers	# feature types	Feature name & number	Other
DGN	**X**	**X**		**X**	**X**	May be accompanied by the CSF file containing information of spatial reference system
DXF	**X**	**X**	**X**	**X**	**X**	
DWG	**X**	**X**	**X**	**X**	**X**	There may be information regarding spatial reference system if dealing with Autocad Map
BIN	**X**	**X**(4)	**X**(4)	X	X	
BNA	**X**	**X**	**X**	X	X	

Table 5.5. CAD type vector formats

(4) Semantics equivalent to name of layer

5.2.6 GIS type vector formats

In contrast with the CAD formats GIS formats can link table information on the geometries and keep visualization apart. A first approach to this solution consisted of associating the geometries with a table row. This is the case of the MGE projects (Intergraph) or the shapefile. In the former the table is stored in a database manager and in the latter it is a DBF file.

Other formats store the geometries and the associated attributes jointly. The spatial databases do so as well. The file formats analyzed have been: E00, SHP & ADF (*Esri*), MIF & TAB (*MapInfo*), GML & KML (*OpenGIS*), GMT, GRASS, SDTS, UK-NTF, Tiger-Line, Interlist, GeoConcept, GeoJSON and SDF.

The main items that may be obtained from this type of formats are:

- Max and min coordinates of the geographic bounding box
- Number of features for every type of geometry (points, arcs, polilines, polygons, texts, etc.);
- Name of layers in which the features are organized (feature names);
- SRS

Some formats also contain information concerning the processes applied, dates and sources of information used. Table 5.6 shows the metadata items used to store this type of information.

Format	Geographic extent	# of layers	Name of layers	# feature types	Feature name & number	Spatial Reference	Name & type of attributes	Other
E00 arc	X					X	X	
ADF arc	X			X	X	X	X	
SHP	X	X	X	X	X	X	X	
MIF	X			X	X	X	X	
TAB	X			X	X	X	X	
VEC (idrisi)	X			X	X	X		
GML	X	X	X	x	X	X	X	
KML	X	X	X	X	X	X fixed	X	
GMT	X	X	X	X	X	X		
GRASS	X	X	X	X	X	X		
SDTS	X	X	X	X	X	X		
UK-NTF	X	X	X	X	X	X		
TIGER	X	X	X	X	X	X		
GeoConcept	X	X	X	X	X	X		
GeoJSON	X	X	X	X	X	X		
SDF (MapGuide)	X	X	X	X	X	X	X	

Table 5.6. GIS type vector formats

5.2.7 Databases with spatial extensions

As mentioned above, the present trend is for the most commonly used DBMS to have data types capable of storing geometries (points, lines, polygons) and processing them through spatial operators. These operators are considerably more complex than the alphanumerical ones (Yeung and Brent, 2007). The spatial objects that are commonly known as “geometries” are the mechanism allowing representation of spatial data. From a mathematical viewpoint, the geometry concept is related to the properties and relations between points, lines, angles, surfaces and solids in one or two-dimensional spaces. From the standardization viewpoint, the OGC Simple Feature Specification, together with the OpenGIS Simple Feature Specification for SQL, identifies the text and binary formats for the representation of objects and they define the table structure that should give support to the minimum metadata. Common reference is made to the GeoDatabase model that takes advantage of the database capabilities to handle indexes, define constraints and keep the integrity of spatial data, at the same time providing with mechanisms to manage transactions. Table 5.7 shows the metadata items of the formats used to store this type of information.

DBMS	Geographic extent	# layers	Name of layers	# feature types	Name & number of features	Spatial Reference System	Name & type of attributes	Other
PostGis	X	X	X	X	X	X	X	
MySQL	X	X	X	X	X		X	
Oracle	X	X	X	X	X	X	X	
DB2	X	X	X	X	X	X	X	
Informix DataBlade	X	X	X	X	X	X		
Microsoft SQL 2008	X	X	X	X	X	X	X	
SQLite (spatialLite)	X	X	X	X	X	X	X	
INGRES	X	X	X	x	X	X		
ArcSDE (MsAccess, Oracle, SQL Server, DB2 e Informix)	X	X	X	X	X	X	X	Feature catalog
MapInfo (spatialware) SQL Server	X	X	X	X	X	X	X	
SQLite + SpatiaLite	X	X	X	X	X	X	X	
SQLite + OGR-FDO	X	X	X	X	X	X	X	
H2 + Spatial DB in a Box	X	X	X	X	X	X	X	
HSQLDB + Spatial DB	X	X	X	X	X	X	X	
Derbi + Spatial DB	X	X	X	X	X	X	X	

Table 5.7. Databases with spatial extensions

5.2.8 Heterogeneity in the storage of Spatial Reference Systems

As mentioned in previous chapters, there are different manners of representing and identifying the SRS the coordinates are referred to. This diversity is the consequence of the large amount of available GI storage formats; it is a problem of syntactic and semantic heterogeneity related to the parameters univocally identifying an SRS. These are (a) the reference geodetic datum, defined by the ellipsoid and the coordinate origin and (b) the cartographic projection, together with the parameters characterizing it. In order to illustrate this heterogeneity, the different forms used by the different formats to express SRS will be enumerated and described next. Tables 5.8 to 5.13 show some specific cases of SRS associated to those forms, and then the storage format of the files is described in basic outline:

- The GeoTIFF, GeoJP2 and MrSID formats store standard numerical values as defined in a public dictionary (e.g. EPSG) as described in Table 5.8.
- Other formats use numerical values defined in private dictionaries: this is the case of MIF, DAT, DEM, CSF, DGN, IMG, RAS and LAN (Table 5.9).
- A third type of formats use mnemonics defined in private dictionaries: ECW, ERS, F-EOSAT, NITF and PIX. (Table 5.10).
- A fourth groups of formats, among them ADF, E00, GRD and Proj4, use text format in a quasi-structured representation (Table 5.11).
- A fifth group of formats, among them SHP and GRD ASCII use the structured syntax WKT defined by OpenGIS. (Table 5.12).
- And finally there is a sixth group in which we find the databases with spatial extensions: PostGIS, Oracle, DB2, SQL Server 2008, MapInfo SpatialWare (Table 5.13) that use standard numerical values and numerical or private text encoding, trying to mitigate semantic heterogeneity and make its interpretation easier.
-

Coordinate Reference System	EPSG	TIFF tag
European Datum 1950, UTM Proj. Northern Hemisphere Time Zone 30	23030	3072
Geographic Coordinates 2D, Datum ETRS89	4258	3072
Datum ETRS89, UTM Projection Northern Hemisphere Time Zone 30	25830	3072

Table 5.8. Examples of standard numerical description: GeoTIFF, MrSID

Coordinate system	**Format**	**Datum**	**Ellipsoid**	**Projection**
European Datum 1950 UTM	Intergraph	4	5	7
European Datum 1950 UTM	Erdas	Text	5	1
European Datum 1950 UTM	Mentor	Text	Text	46
European Datum 1950 UTM	Mapinfo	28	Text	8

Table 5.9. Examples of non-standard numerical description

Projection	**Proj4**	**PCI**	**FME**	**Ermapper**
OBLIQUE MERCATOR	OMER	OM	HOM10V	obmerc_b
LAMBERT AZ EQUAL AREA	LAEA	LAEA	AZMEA	lambazea
EQUIDISTANT CYLINDRICAL	EQC	ER	EDCYL	-

Table 5.10. Example of mnemonic description

DB2 & Esri	ECW	MapInfo & Oracle
GCS_European_1950	European_datum_1950	Longitude /latitude (ED50)
GCS_North_American_1927	North_American_1927	Longitude /latitude (NAD27)
GCS_WGS_1984	WGS_1984	Longitude /latitude (WGS 84)

Table 5.11. Examples of description by means of quasi-structured text

WKT
PROJCS["ED50 / UTM zone 30N", GEOGCS["ED50", DATUM["European_Datum_1950", SPHEROID["International 1924",6378388,297]], PRIMEM["Greenwich" ,0], UNIT["degree",0.01745329251994328]], PROJECTION["Transverse_Mercator"], PARAMETER["latitude_of_origin",0], PARAMETER["central_meridian",-3], PARAMETER["scale_factor",0.9996], PARAMETER["false_easting",500000], PARAMETER["false_northing",0], UNIT["metre",1]]

Table 5.12. Examples of description by means of structured text (WKT)

<table>
<tr><th>Database</th><th>Table diagram and value examples</th></tr>
<tr><td>POSTGIS</td><td><table>
<tr><th>srid</th><th>auth_name</th><th>auth_srid</th><th>srtext</th><th>proj4text</th></tr>
<tr><td>32636</td><td>EPSG</td><td>32636</td><td>PROJCS["WGS 84 / UTM zone 36N", …</td><td>+proj=utm +zone=36 +ellps=WGS84 +datum=WGS84 +units=m +no_defs</td></tr>
</table></td></tr>
<tr><td>Oracle</td><td><table>
<tr><th>Cs_name</th><th>Srid</th><th>Auth_srid</th><th>Auth_name</th><th>Wktext</th></tr>
<tr><td>WGS 84 / UTM zone 36N</td><td>32636</td><td>32636</td><td></td><td>PROJCS["WGS 84 / UTM zone 36N", ...</td></tr>
</table></td></tr>
<tr><td>DB2</td><td><table>
<tr><th>Coordsys_name</th><th>Organization_coordsys_id</th><th>Organization</th><th>Definition</th></tr>
<tr><td>WGS_1984_UTM_ ZONE_36N</td><td>32636</td><td>EPSG</td><td>PROJCS[WGS_ 1984_UTM_ Zone_36N",</td></tr>
</table></td></tr>
<tr><td>SQL Server 2008</td><td><table>
<tr><th>spatial_ reference_id</th><th>authority_ name</th><th>authorized_ spatial_ reference_id</th><th>well_known_text</th><th>unit_of_ measure</th><th>unit_ conversion_ factor</th></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td><table><tr><th></th><th></th><th></th><th></th><th></th><th></th></tr><tr><td>4326</td><td>EPSG</td><td>4326</td><td>GEOGCS["WGS 84", ..</td><td>metre</td><td>1</td></tr></table></td></tr>
<tr><td>MapInfo SpatialWare</td><td><table><tr><th>cs_name</th><th>srid</th><th>Auth_srid</th><th>Auth_name</th><th>srtext</th></tr><tr><td>UTM Zone 36, Northern Hemisphere (WGS 84)</td><td>82356</td><td>82356</td><td>MapInfo</td><td>PROJCS[UTM Zone 36, Northern Hemisphere (WGS 84)", GEOGCS..</td></tr></table></td></tr>
</table>

Table 5.13. Examples of numerical descriptions in spatial databases

After having described the issue of heterogeneity in SRS representation in GI storage formats, some quantitative data are now presented illustrating the magnitude of the question. Table 5.14 outlines the information of formats, encoding type used in the SRS and number of elements (different datums, ellipsoids, projections, parameters, etc,) or full definitions of available SRS in each case.

Means of SRS identification	**Source**	**Number**
Numerical	Database EPSG (v6.18.2)	~ 4.362
WKT + numerical + proj4	PostGIS (v1.3.5)	~ 3.162
Numerical + WKT	Microsoft (SQL Server 2008)	~390
Numerical + WKT	Oracle (v10g)	4.384
Numerical + WKT	IBM-DB2	2.360
Numerical + WKT	MapInfo SpatialWare (for SQL Server)	950
Mnemonic + WKT	Esri	2.400
WKT	Ermapper	875
Mnemonic	Ermapper	165
Numerical	Intergraph	190
Numerical	MapInfo	270
Mnemonic	PCI Geomatics	290
Mnemonic	FME	338
Mnemonic	Proj4	193
Mnemonic	Erdas Imagine	254
Numerical	Idrisi	430
Structured text	Mentor	~1890
Numerical + Mnemonic for projections	GCTP (General Cartographic Transformation Package)	54
Text Datum + projections	TouratechQV	~280
Text Datum + projections	OziExplorer	~150
Text Datum + projections	CompeGPS	~150

Table 5.14. Types and numbers of different identifiers in the formats

In order to put together the issue of SRS, the standard organizations in this field have been looked into, being the ISO 19111: 2007 Spatial Referencing by Coordinates and OGC Spatial Referencing by Coordinates Abstract Specification the documents addressing this topic. Both organizations also deal with this issue in other standards among which the data schema MD_Reference System defined in ISO 19115: 2003 Geographic Information – Metadata, and the data schema coordinateReferenceSystems, defined in Geographic Mark-up Language (GML) of OGC (v 3.1.1) stand out.

Public dictionaries describing SRS with accuracy have also been looked into; we should mention the pioneer work undertaken by the European Petroleum Surveyor Group (EPSG) with the database of geodetic parameters. EPSG is presently integrated in the International Association of Oil & Gas Producers (OGP) that maintains and provides a free database containing this dictionary, at the same time facilitating access to SRS definitions through a catalog accessible in the Internet as a Web service.

From a practical point of view, OGC and ISO, in the regulations concerning the GeoServices for which it is necessary to concisely define the SRS, propose the use of the numerical values of the dictionary or codeSpace EPSG (e.g. EPSG:23030).

5.3 Automatic metadata creation

Now, the metadata items that can be automatically created are identified on the basis of implicit contents in the GI, the calculated items, the items that can be inferred and the ones that may be obtained from the context.

MD_Metadata

---*fileIdentifier*. The Universal Unique Identifier (UUID) is usually built following the encoding rules that enable identification of the organization, and within the organization, the product and the element. For this reason it may be automatically generated on the basis of the context information identifying the organization (configurable parameter of the method, inverse query to the domain names service, etc.)

---*language.* Language in which metadata is expressed. As in the case of the previous item, it may be obtained from a configuration parameter of the metadata generation system or a specific language may be assumed (spa, eng). If the generation system is provided with language translation capability, metadata may be created in several languages, thus tearing down another additional barrier.

---*characterSet.* The type of character encoding to be used may be set up as a configuration parameter of the metadata generation system to be realized in a file. The use of *utf8* as default value is proposed since it is the most extended and used encoding.

---*hierarchyLevel.* The value of this item depends on the content of the URI provided as input of the method. If it makes reference to a file, its value will be a dataSet. If it provides the data of access to a database or makes reference to a directory of vector data, the metadata generator will deal with each table or file of the directory as a featureType, and in addition to generating metadata for each table or file, another one may be generated describing the feature set and its relations, thus realizing the achieved , deduced or inferred data model for the URI.

---*contact (CI_ResponsibleParty).* On rare occasions the contact information concerning metadata is implicit or may be calculated, although it is also true that it is usually common to the entire GI of an organization, hence it may be obtained as a configurable parameter of the metadata creator. Besides, these values will be used if new metadata are created and no template is provided.

------*individualName.* idem.

------*role (CI_RoleCode).* The use of the value "*author*" is proposed.

---*dateStamp.* The value of the item will be set up or updated with the date and time of the system in which the generator is being used.

---*metadataStandardName.* Standard name: ISO19115/19139

---*metadataStandardVersion.*Standard version: ISO-19139:2007

---*datasetURI.* The value of the URI of access to data may be incorporated since it is the mechanism the creator uses to access them. Depending on security levels and data access/use policies, certain URI information (e.g. user and password) may have to be omitted.

---*distributionInfo (MD_Distribution)*

------*transferOptions*

---------*unitsOfDistribution.* It is not an implicit metadata item and it is difficult to calculate or infer on occasions. This is so because certain formats are used to store datasets that may represent a full layer, a tile or just a geographic area. As a first approach, utilization of the usual values is proposed. As an example, if there are several files with images of regular sizes, we would be dealing with tiles; the same reasoning is valid in the case of CAD type files, e.g. dgn, also with regular sizes. The shapefiles generally contain full thematic coverages, hence we would be dealing with layers; databases do not use tiles and the distribution units are geographic areas or layers.

---------*transferSize.* When dealing with files, the metadata creator may access the file system to request their size and transform the units into Mbytes. In the case of databases, queries may be made depending on the DBMS with more or less complexity that allow finding out the hard drive physical size occupied by the table, although it is not possible to know the file size for its transfer.

------*distributionFormat*

---------*name.* For files it is possible to indicate the name of the format; for databases the name of the distribution format is unknown

---------*version.* Some file formats include information items allowing identification of the format version used, hence the distribution format.

---------*fileDecompresionTechnique.* Some formats have information items identifying the compression/decompression technique used. This is the case of the GTIFF, MrSID, JP2 or ECW formats in the image context.

---*contentInfo (MD_ContentInformation)*

------*MD_CoverageDescription.* This descriptive information package is applied to file formats storing images or raster databases.

---------*contentType.* Some formats, such as Erdas and Idrisi, allow distinguishing content type by naming the stored thematic information categories, hence they allow identification of their content as *thematicClassification*. Most formats used for images only support the *image* type. The file formats containing multispectral information, commonly used in remote sensing, may obtain data such as reflectance or transmittance which may be considered *physicalMeasurement.*

---------*dimension (MD_RangeDimension.* The coverage is made up of one or more bands. For each band the following descriptive information is incorporated:

------------*sequenceIdentifier*. Sequence number of the band in the coverage.

---------------*descriptor.* Name or number assigned to the band/layer.

---------------*MD_Band*

------------------*maxValue.* Max value of the pixels. If the format does not store it, the calculation may be made.

------------------*minValue.* Min value of the pixels. If the format does not store it, the calculation may be made.

------------------*units.* Measure units for max and min values. It is a value difficult to obtain.

------------------*bitsPerValue.* Number of bits used to encode the values of each pixel.

---*MD_FeatureCatalogueDescription.* Object catalog applied to vector data, databases or thematic images.

------*complianceCode.* At the present time formats do not include information indicating if the stored data correspond to specifications or standards. The implementation rules of the INSPIRE Directive indicate that this item is mandatory. In the case of a directory or a database, once the data model has been inferred, a comparison may be made with the descriptions of the objects in a catalog of feature types.

------*Language.* The language of the data may be inferred from their analysis using dictionaries.

------*includedWithDataSet.* It is uncommon for the object catalog to be stored together with the data, taking on the *false* value. Some image formats like Erdas or Idrisi or the geodatabases can store it. Once this has been verified, the value *true* is assigned to the item. The formats used in CAD (dgn, dwg, dxf, and MIF) usually name the layers of the files with the name or mnemonics associated to the type of stored feature type. In these cases the value *true* may also be assigned to this item.

------*featureTypes.* If the names of the feature types contained in the dataset can be determined, they will be included as its feature types.

---*spatialRepresentationInfo (MD_SpatialRepresentation)*

------*MD_GridSpatialRepresentation.* Applied to images and matrix data.

---------*numberOfDimensions.* Number of dimensions in space and time. In images and bands, it usually coincides with the number of bands.

---------*cellGeometry.* Geometry type: *area* or *point*.

---------*transformationParametersAvaibility.* Some formats, in addition to storing the rectified image, also store the information that allows carrying out that transformation. This is the case of Erdas and others such as LIB; in these cases if the transformation parameters are available, *true* will be indicated; otherwise *false* will be indicated.

---------*axisDimensionProperties (MD_Dimension)*

------------*dimensionName.* Name of the dimension: *row, column, vertical* or *time*. Most common are the first two; information will be exceptionally available to complete the others.

------------*dimensionSize.* The count of image rows and columns should be indicated as well as the number of additional dimensions: vertical or temporal, as may be the case.

------------*resolution*

---------------*resolution.* Pixel size according to the dimension: row, column, etc.

---------------*unitsofResolution.* Measure units of resolution.

------*MD_VectorSpatialRepresentation.* It is used in the case of vector data.

---------*topologicLevel.* Topologic level of data. In the vector formats used in CAD topology is not usually treated, in whose case it would be indicated that *geometryOnly* is considered. The topologic level may take on other values such as *topologyID, planarGraph, fullPlanarGraph, surfaceGraph, fullSurfaceGraph, topology3D* or *abstract*, depending on the content type. This type of information is not described in the implicit metadata

---------*geometricObjects*

------------*geometricObjectType.* The geometry types contained in the data are identified through the codes *complex, composite, curve, surface, point* and *solid*.

------------*geometricObjectCount.* The number of objects of each geometry type is indicated.

---*identificationInfo (Md_Identification)*

------*citation (CI_Citation).* Citation that allows indicating the resource title.

---------*title.* Proposed title for the resource. The generator of metadata should propose a title for the resource based on the obtained implicit information, the calculated data and the inferred data.

---------*date* Date of title creation.

---------*dateType (CI_DateTypeCode).* Code identifying the date type. If the metadata did not exist, the code will be *creation*, and if it already existed, it will be *revision*.

---------*otherCitationDetails.* Although this item is not mandatory, it may be used to indicate how the metadata title has been generated, i.e. to describe the elements that have been considered to generate the title. Reference can also be made to the metadata creator.

------*pointOfContact (CI_ResponsibleParty)*

---------*individualName.* The identification of the responsible party or author of the data will be included if known –or information may be obtained. Some formats contain that information and in other cases it could be completed with default parametrized information as adopted by the metadata creator.

---------*role (CI_RoleCode).* Information will be generally available about the author of the data and the code to be used will be *author*.

------*graphicOverview (MD_BrowseGraphic)*

---------*fileName.* The generator of metadata will calculate a data preview that will be stored in a file. It is the name of the file generated.

--------- *fileDescription.* In the same manner that other details of the citation, associated to the title of the resource it may be described how the title of the metadata has been generated, in this case the graphic and the manner of its generation may be described as well.

---------*fileType.* The file type will be the one the metadata creator uses to store the preview. Web-supported, widespread formats will be used, e.g. PNG or JPEG.

------*resourceFormat (MD_Format)*

---------*name.* The name of the format may be indicated in the case of a file and *unknown* in the case of databases.

---------*version.* Some file formats include information items allowing identification of the used version, hence of the distribution format.

---------*fileDecompresionTechnique.* Some formats hold information items identifying the compression/decompression technique used. It is the case of the GTIFF, MsSID, JP2 or ECW formats in the imagery context.

------*resourceSpecificUsage (MD_Usage)*

---------*userDeterminedLimitations.* Some formats store information advising on use limitations.

------*resourceConstraints (MD_Constraints)*

---------*useLimitation.* This is an item that may parametrize in the metadata creator setup to respond to the data use policy of the organization. Thus a predetermined value would be established for the item.

---------*accessConstraints (MD_RestrictionCode).* This is about knowing the code identifying data access constraints. Some formats store information defining access policy. The possible values of this item are *copyright, patent, patentPending, trademark, license, intellectualPropertyRights, restricted* and *otherRestrictions.*

---------*useConstraints (MD_RestrictionCode).* Some formats such as PNG enable storing use constraints.

---------*otherConstraints.* As with the item *useLimitation*, its value may be obtained by parametrization of the metadata creator, in agreement with the organization's data policy. In this case it may be a message of the *disclaimer* type on the inappropriate use of the data.

---------*classification (MD_ClassificationCode).* Classification code of data confidentiality. In some cases formats contain this type of stored information. The following values are used: *unclassified, restricted, confidential, secret,* and *topSecret.*

------*descriptiveKeywords (MD_Keywords)*

---------*keyword.* After the metadata creator module has identified the content type of the data, the cataloging module may select a set of *keywords* that will identify the content more easily.

---------*type.* The type of keyword proposed should be indicated; it may belong to one of these classes: *discipline, place, stratum, temporal,* and *theme.* The cataloging module selects the keywords and also identifies the type and the thesaurus they belong to.

---------*thesaurusName (CI_Citation)*

------------*title.* Title of the thesaurus containing the keyword used to catalog the resource.

------------*date.* Thesaurus date, creation, revision, etc. as the case may be.

------------*dateType (CI_DateTypeCode).* Code identifying the type of date: *creation, publication* and *revision.*

------*spatialRepresentationType.* Depending on the format and content type, the type of spatial representation of the data may be classified into *vector, grid, textTable, TIN, stereoModel* and *video.*

---------*spatialResolution (MD_Resolution). Data scale.*

------------*equivalentScale (MD_RepresentativeFraction)*

---------------*denominator*. The denominator of the equivalent scale may be calculated based on the pixel size for images and matrix data

---------------*distance (Distance).* Pixel size in images and matrix data.

------*language.* Language in which data are expressed may be inferred by using dictionaries whereby values of the data attributes may be compared.

------*characterSet.* Information may be obtained to identify the character set used in databases and some file formats.

------*topicCategory.* As the metadata creator determines the type of data content, the cataloging module may select one or several topics of the list of categories used by the metadata standard to classify data.

------*environmentDescription.* The metadata creator may access the information provided by the operative system in which it is being run in order to describe the processing environment of the data, the type of operative system, file system, database, etc. It is context information.

------*supplementalInformation.* Some file formats used in remote sensing contain additional information concerning the platform and the sensors that have captured the data. Part of this information may be stored in the new metadata standard (ISO19115-2) that will be hopefully approved in the near future. Other information may also be included in this item.

------*extent (Ex_Extent)*

---------*description.* Descriptive text explaining how the stored values for the *geographic extent have been obtained and/or calculated.*

---------*geographicElement (EX_GeographicExtent)*

------------*EX_GeographicBoundingBox*

---------------*extentTypeCode*. The *true* value is assigned to indicate that the geographic extent corresponds to the interior of the rectangle described next.

---------------*westBoundLongitude*. The values of the North, South, East and West coordinates are obtained from the conversion/transformation module once the SRS has been identified.

---------------*eastBoundLongitude* idem

---------------*northBoundLatitude* idem

---------------*southBoundLatitude* idem

------------*EX_GeographicDescription*

---------------*extentTypeCode*. The *true* value is assigned to indicate that the geographic extent corresponds to the place name as shown below.

---------------*geographicIdentifier (RS_Identifier)*

------------------*authority (CI_Citation)*

---------------------*title*. Name or title of the thesaurus of places used.

---------------------*date*. Date of the thesaurus.

---------------------*dateType(CI_DateTypeCode)*. Code identifying the date type of the thesaurus. The possible values are *creation*, *publication* and *revision*.

---------------------*code*. Value of the place name. It will be obtained by using an inverse geocoder, so that given the BBOX coordinates, the most relevant toponym of the place is requested. Names of cities or small towns will be searched for.

---------*verticalElement (EX_VerticalExtent)*. Usually for 3D datasets and for grids of regular data containing DTMs or DEMs.

------------*minimumValue*. Min height contained or calculated in the format.

------------*maximumValue*. Max height contained or calculated in the format.

------------*unitOfMeasure*. Units of measures. Usually the meter, but it may depend on the format.

---*dataQualityInfo (DQ_DataQuality)*

------*scope (DQ_Scope)*

---------*level (MD_ScopeCode)*. Generally the formats that include this type of information describe the steps of the process, indicating the affected datasets. For this reason the level to which they are applied is *dataSet*, although the following levels of detail may also be applied: *attribute, series*, *features*.

------*linage (LI_Lineage)*

---------*source (LI_Source)*

------------*description.* Some formats include a description of the data sources used, e.g. IDRISI.

---------*processStep (LI_ProcessStep)*

------------*description* Some formats store the steps taken on the data.

------------*dateTime (DateTime).* Some formats also include date and time.

---*applicationSchemaInfo (MD_ApplicationSchemaInformation)*

------*name (CI_Citation)*

---------*title.* Description of the method used to create or infer the application schema.

---------*date.* Creation date of the application schema.

---------*dateType(CI_DateTypeCode).* Date type of the application schema. If there was no metadata and the application schema has been generated, the code will be *creation*, otherwise it will be *revision*.

------*schemaLanguage.* Languages used by the generator of metadata to define the schemas: *UML* and *GML-Schema*).

------*constraintLanguage.* The generator of metadata uses *UML* and *GML* as constraint languages.

------*schemaAscii.* The generator of metadata, with the assistance of the module determining the data model, will store the data application schema in *ASCII* format.

------*graphicsFile.* The generator of metadata will create a preview of the application schema and will store it in a graphic format.

------*softwareDevelopmentFile.* XML file containing the application schema, either UML model, in whose case it is stored in XMI format or GML application schema, in whose case it is stored in XSD format.

------*softwareDevelopmentFileFormat.* In one case it will be XMI (UML) and in another case it will be XSD (GML Schema).

---*referenceSystemInfo (MD_ReferenceSystem)*

------*referenceSystemIdentifier (RS_Identifier)*

---------*codeSpace.* *EPSG* is set as dictionary of code identifying the SRS

---------*version.* Version of the EPSG database used.

---------*authority (CI_Citation)*

------------*title.* Refers to the International Association of Oil and Gas Producers (OGP) and more especially to the European Petroleum Surveyor Group (EPSG) which has defined a database of geodetic parameters.

------------*date (CI_Date)*

---------------*date.* Date of the EPSG database used.

---------------*dateType (CI_DateTypeCode) revision*

------------*code.* The module identifying the SRS should get the EPSG code which identifies the coordinates.

5.4 Performed functions and interoperability levels favored by items of automatically created metadata

In the previous section a detailed description has been made of the possible metadata items that may be extracted, calculated or inferred for a particular dataset in an SDI. It is difficult to determine the exact number of items that may be automatically generated for a GI repository since every type (aerial images, multispectral images, DTM, drawing files and vector layers) has a different set of metadata items associated. In addition the information that can be extracted from the files and data stores closely depends on the storage format used.

Table 5.15 identifies (optimistically) the metadata items that could be automatically created with the proposed methodology. The first column shows the class to which the item belongs and the second column identifies the item. The third column classifies items as “C”, created (extracted, calculated or inferred), “N”, with cardinality, depending on the dataset and “F”, setting up a fixed value by agreement or depending on the context. The fourth column identifies those items that are only applied to a certain type of GI (“R” raster data, “D” digital terrain models and “V” vector data). The fifth column identifies the function performed by the metadata item (“L” location, “E” evaluation , “A” access and “U” use). Columns 6 to 12 describe the seven interoperability levels: *technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational,* defined in the integrated interoperability model for SDIs.

MD_Metadata: *Metadata Package*	**Metadata Item**	**(C) created. (N) depends on data**	**(R: Raster, V:Vector, D: DTM)**	**Function: (L: location, E: evaluation, A: access, U: use)**	**Technical**	**Syntactic**	**Semantic**	**Pragmatic**	**Dynamic**	**Conceptual**	**Organizational**
	language	C					X				
	characterSet	F			X		X				
	hierarchyLevel	C					X			X	X
	metadataStandardName	F				X	X				X
	metadataStandardVersion	F				X	X				X
	dataSetURI	C			X		X	X			
	dateStamp	C					X		X		X
distributionInfo:transferOptions	unitsOfDistribution	C		A			X	X			X
distributionInfo:transferOptions	transferSize	C		A	X		X				X
distributionInfo:distributionFormat	name	C		A	X		X		X		X
distributionInfo:distributionFormat	version	C		A	X				X		X
distributionInfo:distributionFormat	fileDecompressionTechnique	C		A			X		X		X
contentInfo:MD_CoverageDescription	contentType	C	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	sequenceIdentifier	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension:MD_RangeDimension	descriptor	N	R	U							X
contentInfo:MD_CoverageDescription: dimension: MD_Band	maxValue	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension: MD_Band	minValue	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension: MD_Band	units	N	R	U			X		X		X
contentInfo:MD_CoverageDescription: dimension: MD_Band	bitsPerValue	N	R	U			X		X		X
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	complianceCode	C		U			X				X
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	language	C		U							X
contenInfo:MD_ContentInformation: MD_FeatureCatalogueDescription	includedWithDataset	C		U			X		X		X
contenInfo:MD_ContentInformation	featuresTypes	C		U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	numberOfDimensions	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	cellGeometry	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation	transformationParameterAvailability	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension	dimensioName	N	R	U			X		X		X
axisDimensionProperties:MD_Dimension	dimesionSize	N	R	U			X		X		X
spatialRepresentationInfo: MD_GridSpatialRepresentation: axisDimensionProperties:MD_Dimension: resolution	value	N	R	U		X	X		X		X
axisDimensionProperties:MD_Dimension: resolution	units	N	R	U		X	X		X		X
spatialRepresentationInfo: MD_VectorSpatialRepresentation	topologyLevel	N	V	U			X		X		X
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects: MD_GeometricObjects	geometricObjectType	N	V	U			X		X		X
spatialRepresentationInfo: MD_VectorSpatialRepresentation: geometricObjects:	geometricObjectCount	N	V	U			X		X		X

MD_GeometricObjects											
identification:MD_Identificacion: MD_DataIdentification: citation: CI_Citation	title	C		L				X	X		X
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	date	C		L			X		X		X
identification:MD_Identificacion: MD_DataIdentification:citation: CI_Citation: date: CI_Date	dateType	C		L			X		X		X
identification:MD_Identificacion: MD_DataIdentification	credit	C		L							X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	filename	C		L	X						X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileDescription	C		L							X
identification:MD_Identificacion: MD_DataIdentification: graphicOverview: MD_BrowseGraphic	fileType	C		L			X				X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	name	C		L	X		X		X		X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	version	C		L	X				X		X
identification:MD_Identificacion: MD_DataIdentification: resourceFormat: MD_Format	fileDecompressionTechnique	C		L			X		X		X
identification:MD_Identificacion: MD_DataIdentification: resourceSpecificUsage: MD_Usage	userDeterminedLimitations	C		E			X		X		X
identification:MD_Identificacion: MD_DataIdentification: resourceConstraints: MD_LegalConstraints	useConstraints	C		E			X		X		X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	keyword	C		L			X		X		X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords	type	C		L			X		X		X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	title	C		L				X	X		X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	date	C		L			X		X		X
identification:MD_Identificacion: descriptiveKeywords: MD_Keywords: thesaurusName:CI_Citation	dateType	C		L			X		X		X
identification: MD_Identificacion	spatialRepresentationType	C		L			X		X		X
identification: MD_Identificacion: spatialResolution: MD_Resolution	distance	N	R	L			X		X		X
identification: MD_Identificacion: spatialResolution: MD_Resolution: equivalentScale: MD_RepresentativeFraction	denominator	N	R	L			X		X		X
identification: MD_Identification	languaje	C		L			X				X
identification: MD_Identification	characterSet	C		L	X		X				
identification: MD_Identification	topicCategory	C		L			X				X
identification: MD_Identification	environmentDescription	C		L	X						X
identification: MD_Identification	supplementalInformation	N	R	L							X
identification: MD_Identification: extent: EX_Extent	description	C		L							X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	extentTypeCode	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	westBoundLongitude	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	eastBoundLongitude	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	southBoundLatitude	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicBoundingBox	northBoundLatitude	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation	title	C		L				X	X		X

identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation : date : CI_Date	date	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: geographicElement: EX_GeographicExtent: EX_GeographicDescription: geographicIdentifier: RS_Identifier: authority: CI_Citation : date : CI_Date	dateType	C		L			X		X		X
identification: MD_Identification: extent: EX_Extent: verticalElement:Ex_VerticalExtent	minimumValue	N	D	L			X		X		
identification: MD_Identification: extent: EX_Extent: verticalElement:Ex_VerticalExtent	maximumValue	N	D	L			X		X		
dataQualityInfo: DQ_DataQuality: scope: DQ_Scope	level	C		E			X				X
dataQualityInfo: DQ_DataQuality: lineage: LI_Lineage: source : LI_Source	description	C		E							X
dataQualityInfo: DQ_DataQuality: lineage: LI_Lineage: processStep: LI_ProcessStep	description	C		E							X
dataQualityInfo: DQ_DataQuality: lineage: LI_Lineage: processStep: LI_ProcessStep	dateTime	C		E			X				
applicationSchemaInfo: MD_ApplicationSchemaInformation: name: CI_Citation	title	C		U				X	X		X
applicationSchemaInfo: MD_ApplicationSchemaInformation: name: CI_Citation: date: CI_Date	date	C		U			X		X		X
applicationSchemaInfo: MD_ApplicationSchemaInformation: name: CI_Citation: date: CI_Date	dateType	C		U			X		X		X
applicationSchemaInfo: MD_ApplicationSchemaInformation	shemaLanguaje	C		U			X			X	X
applicationSchemaInfo: MD_ApplicationSchemaInformation	constraintLanguaje	C		U			X			X	X
applicationSchemaInfo: MD_ApplicationSchemaInformation	squemaAscii	C		U						X	X
applicationSchemaInfo: MD_ApplicationSchemaInformation	graphicsFile	C		U							X
applicationSchemaInfo: MD_ApplicationSchemaInformation	softwareDevelopmentFile	C		U	X			X			X
applicationSchemaInfo: MD_ApplicationSchemaInformation	softwareDevelopmentFileFormat	C		U	X		X			X	X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier	codeSpace	C		U			X		X		X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier	version	C		U			X		X		X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier	code	C		U			X		X		X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier: authority: CI_Citation	title	C		U				X	X		X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier: authority: CI_Citation: date: CI_Date	date	C		U			X		X		X
referenceSystemInfo: MD_ReferenceSystem: referenceSystemIdentifier: identifier: RS_Identifier: authority: CI_Citation: date: CI_Date	dateType	C		U			X		X		X
		C:64 F:3 N:21		L:34 E:6 A:5 U:36	12	4	68	8	57	5	82

Table 5.15. Classifications of automatically created metadata items

From an optimistic viewpoint the method of automatic generation of metadata may compile 83 items for images, 69 for vector data and 68 for DTM. The actual number of items will depend on the format used and on the information contained. This number of items may increase if the GI store contains more than one band or layer, with 21 more items identified for the three classifications (R, V and D) with cardinality ≥ 1. In addition, there is another set of items that will be produced by the cataloger ((*keyword, theme, thesaurus name title, date* and *topicCategory*) whose cardinality *n* may increase the mentioned values remarkably.

Although there are few metadata items specific for DTM (2) and for vector data (3) and many more for raster data (17), this should not be considered conclusive since for vector data the UML diagram of classes is obtained which contains the definition of features and their relations, contributing more elaborate, useful information that will balance the results.

Regarding the function performed by metadata items, 34 perform the *location* function (42%), 6 the *evaluation* function (8%), 5 the *access* function (6%) and 36 the *use* function (44%); it may be concluded that the functions most favored by automation are location and use.

By analyzing the interoperability provided by the metadata items created with the methodology, in addition to classifying them by the interoperability levels they provide, the obtained values have been analyzed and compared with the ISO 19115 Standard in all its extent and with the items belonging to the core of the standard. Table 5.16 shows the number of standard items, the number of core items and the items that may be obtained for each GI type, classified by interoperability levels.

	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
ISO-19115	37	6	128	15	99	7	217
Core ISO-19115	8	3	40	3	31	1	43
Automatic Raster	11	3	56	6	46	5	67
Automatic DTM	11	3	59	6	49	5	67
Automatic Vector	11	2	45	6	35	5	54
Automatic ISO Core	6	2	23-25	2	20-22	0	26

Table 5.16. Number of items providing interoperability

As shown in Table 5.16, the number of items providing organizational interoperability is very high as is also the case for the number of items of the metadata standard defined to describe data from the organizational and life cycle viewpoint. It should be pointed out that the number of items providing the different interoperability levels is similar for both the raster data and the DTM data categories, and they differ for the vector data category in that there are fewer items providing semantic, dynamic and organizational interoperabilities. It is confirmed that the number of automatically created metadata items surpasses the thresholds defined for the core items, with the exception of syntactic interoperability. Their relevance is scarce for the GI stores since 50% of the items providing this type of interoperability are linked to services.

With the purpose of analyzing the efficiency of the automatic metadata generation method and facilitate its interpretation, the numbers of created items have been normalized with respect to the values of the ISO 19115 Standard as a whole. Thus the comparison of the interoperability frequencies provided by the created metadata is made easier, as shown in Figure 5.1.

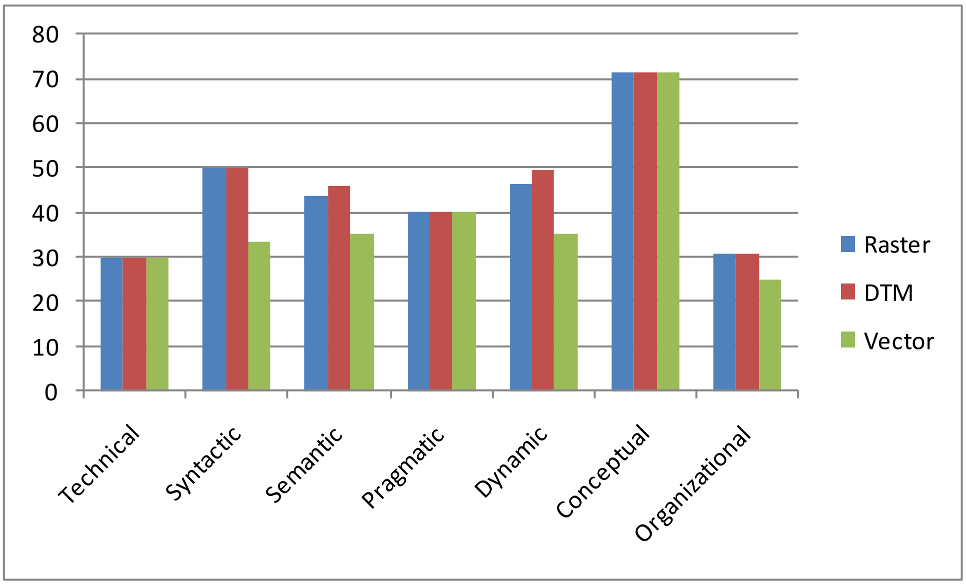


Figure 5.1. Percentages of automatically created items by interoperability level.

Figure 5.1 shows that:

- the percentages of automatically items created by the methodology are sufficiently homogeneous for the different interoperability levels;
- the mean values of those percentages are comprised between 30% and 40% with the exception of the organizational interoperability for vector data which is somewhat lower and the conceptual interoperability which is noticeably higher;
- it may be stated that the automatic generation method pays more attention to the interoperability levels less favored by the ISO 19115 metadata standard (syntactic and conceptual), although the number of items is scarce, 3 and 5 items respectively.

5.5 Validation of the interoperability model

After having defined the interoperability model applied to SDI which is made up of seven levels (technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational), and after having analyzed the interoperability the items of the ISO 19115 metadata standard can provide (see Chapter 4, Section 4), in this chapter the actions taken are shown to validate the results of the analysis through counsel and deliberation with an expert group coming from different knowledge domains, skilled in the area of metadata standards and directly involved in SDI. The purpose of the validation work is to quantify the subjectivity of the classification of the interoperability levels provided by the metadata items imputable to the author of the model.

To carry out this validation we have relied on five experts involved in SDI and skilled in the area of ISO 19115 metadata standard, ascribed to the Computer Science, Geodesy, Cartography and Surveying domains, native of Argentina, Colombia and Spain. They have been asked to identify and assign the interoperability levels provided by the core items, by means of a spreadsheet, including or excluding levels of the model. In addition they have been asked to expose descriptively the reasons for their decisions.

The decision of using only the core items to make the validation is justified by the fourth conclusion of Chapter 4 in which it is pointed out that the relations between interoperability levels are similar when analyzing both the core items and all the items of ISO 19115. Thus the effort of the collaborators is made simpler.

After having defined the extent, scope and purpose of the validation, the methodology used is described: in the first place the design of the validation process, the selection of the collaborators and the manner of collecting their interpretations are described; in the second place the processing of the surveys is explained, and in the third place the validity of the relations between the levels of the model and the classification of the levels attributed to the different items is explained. The results and conclusions will appear in the last section of this chapter.

5.5.1 Definition of the surveys

The survey used to carry out validation of the model has been designed in a spreadsheet, so that the first column contains the core metadata items of ISO19115/19139. The next seven columns identify the interoperability levels selected in the definition of the model. An additional column shows the criterion used by the author to select the levels as the model was designed. Finally two columns have been included: the first column in which the collaborator shows his (her) agreement or disagreement with the levels of the model, and the last column in which the collaborator has to contribute his (her) comments and justifications. To illustrate graphically the table, Figure 5.2 is presented containing a screen capture of the spreadsheet used.

Metadata Element	Mandatory/conditional	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational	Results and Comments	Do you agree?	Would you remove any type of interoperability? Would you add any other type?
fileIdentifier		x			x				**Technical:** identifier associated to the file and metadata file system. Intended to avoid maintenance and management problems in		
									Pragmatic: allows dynamic handling of metadata by identifiers: collection, maintenance.		
Language				x				x	**Semantic:** term of a list identifying language of metadata.		
									Organisational: enables language negotiation in business models.		
characterSet		x		x					**Technical:** allows correct handling of bytes from the metadata		
									Semantic: identifies univocally encoding type on the base of a list of		
parentIdentifier		x			x				**Technical:** identifies univocally the parent node in the hierarchical relation and makes access to metadata possible.		
									Pragmatic: allows automatically browsing through metadata		

Figure 5.2: Appearance of the survey spreadsheet

A document with definitions of the interoperability levels of the model has been distributed along with the spreadsheet.

The next step has been to select the collaborators who have participated in the survey according to these requirements: 1) to know the ISO 19115 metadata standard, 2) to be working in institutions managing GI, and 3) to belong to a diverse geographic environment and to come from different knowledge domains.

The involved institutions are:

- National Geographic Institute of Spain (IGN-E),
- Geographic Institute Agustín Codazzi of Colombia (IGAC),
- Institute of Regional Development of the University of Castilla la Mancha (IDR-CM),
- Military Geographic Institute of Argentina (IGM),
- Polytechnic University of Madrid (UPM).
- The knowledge domains involved are:
- Geodetic and Cartographic Engineering,
- Geographic Engineering,
- Surveying Engineering,
- Computer Engineering

5.5.2 Processing of surveys

After having received the documents collecting the answers of the collaborators, the next task has been their processing. Since one of the collaborators has preferred to answer in one text document, the first treatment of this survey has consisted of transferring his answers to the spreadsheet. The second stage of the common treatment of the data has been gathering together all the answers in one single document, on which to carry out the remainder of the treatments.

Once the data gathered together, the treatment given during the definition of the interoperability model, as applied to metadata, has been reproduced at the individual level. The purpose of this individual treatment of the results is to analyze and interpret the counts of items providing the levels of the model obtained by each collaborator. Thus the values that are presented numerically and graphically in Table 5.17 and Figure 5.3 respectively have been obtained. To keep the anonymity of the respondents their contributions have been named "respondent n".

	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Validate Model	7	4	39	2	31	1	42
Respondent 1	12	16	39	9	36	1	44
Respondent 2	2	7	33	1	12	3	24
Respondent 3	9	5	40	16	31	3	43
Respondent 4	8	5	52	3	31	1	47
Respondent 5	1	26	39	7	12	1	31

Table 5.17. Frequencies of interoperability levels per respondent when analyzing the metadata items of the ISO 19115 standard

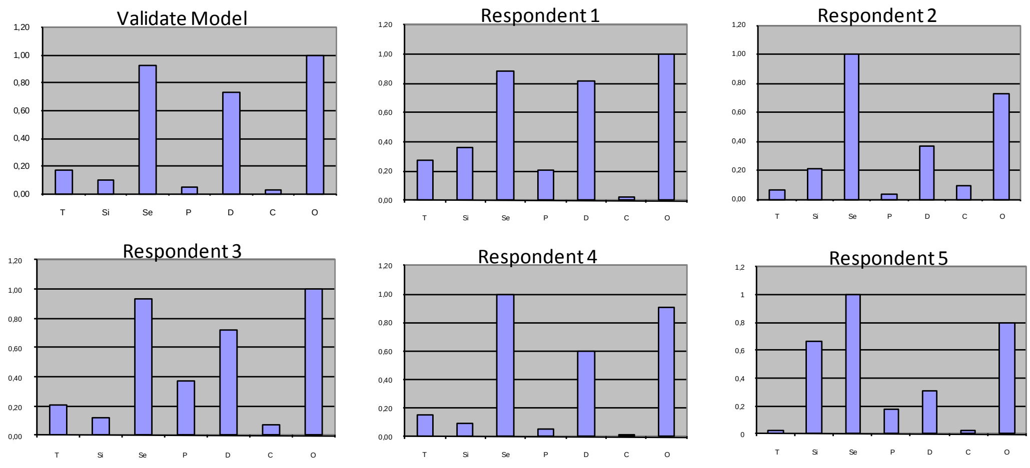


Figure 5.3. Histogram of the frequencies of interoperability per respondent

The histograms of Figure 5.3 allow the following interpretations:

- For respondents 1, 2, 3 and 4 the frequencies of the semantic, dynamic and organizational levels are high, with values well differentiated from the remainder of the levels.
- Respondent 5 considers as syntactic aspects of interoperability many aspects that the other respondents consider dynamic.
- Respondents 2 and 5 identify more items promoting semantic interoperability than the items promoting organizational interoperability, in disagreement with the others.
- Every respondent coincides with the proposed model in the low number of items providing conceptual interoperability.

Continuing with the method of analysis carried out for the model, the frequency with which the metadata items simultaneously promote several interoperability levels has been analyzed. The tables of Figure 5.4 show the frequencies both for the model to be validated and for every collaborator.

Validate Model	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	7	0	4	2	3	0	3
Syntactic	0	4	4	0	2	0	2
Semantic	4	4	39	0	28	0	30
Pragmatic	2	0	0	2	1	0	1
Dynamic	3	2	28	1	31	0	23
Conceptual	0	0	0	0	0	1	1
Organisational	3	2	30	1	23	1	42

Respondent 1	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	12	5	8	4	6	0	6
Syntactic	5	16	14	3	11	0	11
Semantic	8	14	39	6	31	1	29
Pragmatic	4	3	6	9	5	0	8
Dynamic	6	11	31	5	36	1	27
Conceptual	0	0	1	0	1	1	1
Organisational	6	11	29	8	27	1	44

Respondent 2	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	2	0	0	0	1	0	0
Syntactic	0	7	1	0	0	1	1
Semantic	0	1	33	0	10	1	14
Pragmatic	0	0	0	1	0	0	0
Dynamic	1	0	10	0	12	0	3
Conceptual	0	1	1	0	0	3	2
Organisational	0	1	14	0	3	2	24

Respondent 3	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	9	0	5	3	3	0	4
Syntactic	0	5	5	2	2	0	3
Semantic	5	5	40	12	28	1	28
Pragmatic	3	2	12	16	14	0	12
Dynamic	3	2	28	14	31	0	23
Conceptual	0	0	1	0	0	3	3
Organisational	4	3	28	12	23	3	43

Respondent 4	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	8	1	7	2	3	0	7
Syntactic	1	5	5	0	3	0	3
Semantic	7	5	52	2	31	1	22
Pragmatic	2	0	2	3	1	0	3
Dynamic	3	3	31	1	31	0	23
Conceptual	0	0	1	0	0	1	1
Organisational	7	3	22	3	23	1	47

Respondent 5	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organisational
Technical	1	1	1	0	0	0	0
Syntactic	1	26	26	0	4	0	9
Semantic	1	26	39	4	9	1	18
Pragmatic	0	0	4	7	1	1	6
Dynamic	0	4	9	1	12	1	9
Conceptual	0	0	1	1	1	1	1
Organisational	0	9	18	6	9	1	31

Figure 5.4. Tables showing the relations between the levels of the interoperability model according to the core items for every respondent

The tables of Figure 5.4 contain the count of items providing the interoperability levels indicated by the column, concurrently providing the levels identified in the rows, for example for respondent 5, of the 26 items providing syntactic interoperability, 4 provide dynamic, 1 technical and all of them semantic interoperability.

Figure 5.5 shows graphically the frequency of the relations between the levels of the model for every collaborator, resulting from interpreting the interoperability of the core items of the metadata standard. Due to the great differences in the number of items in the relations of the levels of the model, the criterion of using the maximum values of each combination has been adopted to define the isolines; therefore the figures are not comparable with one another.

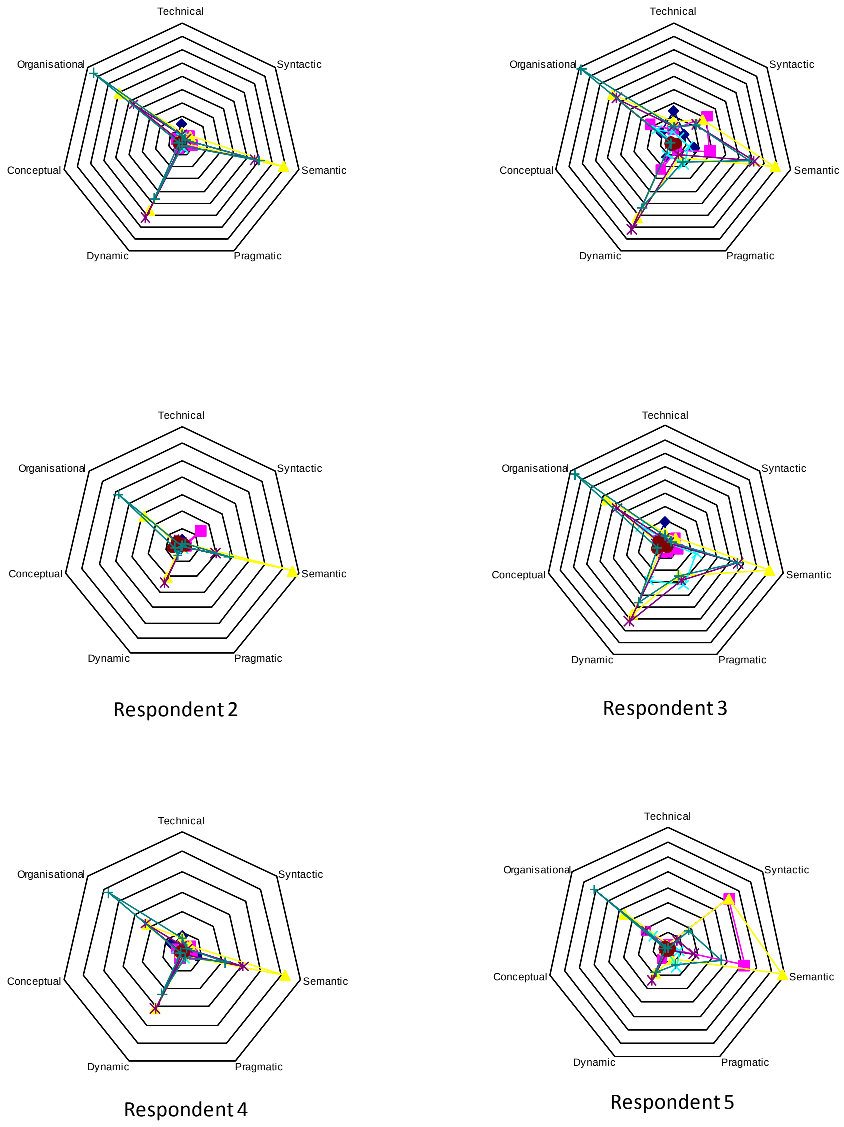


Figure 5.5. Radial diagrams with the relations between the levels of the interoperability model

The diagrams of Figure 5.5 allow the following interpretations:

- The distribution of the frequencies in the relations of respondents 1, 2 and 4 is similar to the relations observed in the model to be validated.
- A larger scatter is observed in the relations of respondent 1. This is due to the fact that this collaborator has considered that the metadata items contribute simultaneously several interoperability levels.
- Less relations and with lower frequency are observed in the diagram of respondent 2. Contrary to respondent 1, respondent 2 has considered that the items do not provide interoperability in certain levels of the model.
- For respondent 3 more relations are observed between the pragmatic, dynamic and semantic levels. This is due to the fact that she holds the view that a significant amount of items contribute simultaneously pragmatic, dynamic and semantic interoperabilities.
- Higher frequencies are observed between the semantic and syntactic levels for respondent 5. It is due to the fact that she has interpreted that the items do not provide interoperability at the technical or dynamic levels; yet she suggests that many items do provide interoperability at the pragmatic and syntactic levels.

After analysis of the model has been remade with the collaborators' interpretations, we proceed to identify the discrepancies with the proposed model. We intend to identify:

- The permutations of the interoperability levels proposed by the collaborators for one item;
- The discrepancies (too much or too little) for each item, relating it to the model to be validated.

Figure 5.6 show graphically the work carried out with the surveys. In the first two columns the names of the packages and analyzed items are shown. The following five blocks show the discrepancies found with the proposed model so that the items providing one more interoperability level than the model are highlighted in green and the items not providing any interoperability level in the model are highlighted in orange.

Figure 5.6: Graph of conformity and discrepancies between respondents and model to be validated

The graph of Figure 5.6 may be interpreted as follows:

- Respondents 2 and 5 interpret that the interoperability provided by the metadata items responds in a different way than it was interpreted when defining the model.
- Respondents 1, 3 and 4 indicate that the metadata items provide some interoperability level in addition to the one proposed en the model.
- For respondent 2 the metadata do not provide interoperability at the dynamic or organizational levels.
- Respondent 5 disagrees with the other collaborators and with the model to be validated when considering that a set of items provide syntactic interoperability while for the remainder they provide the dynamic level.
- Respondent 5 thinks there are not so many items providing the organizational interoperability level.
- For respondent 2 it is not easy to identify changes from one level to another in an item. In some cases she proposes not to consider more than one level in

each item; in others she proposes to add one level and only in four of them she feels there are some aspects of the organizational level in the model that are considered dynamic aspects (language and topicCategory).

The next step has been to join the proposals of the collaborators to subsequently filter the results so as to get rid of the random discrepancies of one single respondent. The purpose of this analysis is to obtain the aggregated model and compare it with the proposed model.

Figure 5.7 shows the results of this analysis. It shares with Figure 5.6 the two first columns. The third column shows the result of aggregating the five surveys being carried out and the model being validated. By means of a color legend the agreement between the five collaborators is indicated, and by means of text marks (x) the levels of the model that are being validated are shown. The fourth column repeats the reasoning aggregating the three most discordant or critical surveys corresponding to respondents 1, 2 and 5. The last column shows the aggregation of the five collaborators.

The color legend used to present the results is as follows:

- Dark green cell background: every respondent expresses the same opinion.
- Middle green cell background: only one collaborator disagrees.
- Light green cell background: two collaborators disagree.
- Red cell background: three respondents disagree with the proposed model.
- The text marks (x) of the third column identify the interoperability levels defined in the model to be validated.
- The text marks (A) of the fourth column identify the items and interoperability levels the most critical collaborators agree with, though disagreeing with the model.
- The red box marked with “!” is the metadata item in which two of the three most critical respondents disagree with the model. This will be analyzed in further detail later.

	Metadata Attributes	Aggregated results	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational	Respondents 1,2, 5	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational	All respondents	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Md_Metadata	fileIdentifier		X			X																			
	Language				X				X																
	characterSet		X		X							A													
	parentIdentifier		X			X			!								A								
	hierarchyLevel				X			X	X						A										
	hierarchyLevelName								X																
	dateStamp				X		X		X																
	metadataStandardName			X	X				X																
	metadataStandardVersion			X	X				X																
PT_Locale	language				X				X																
	characterEncoding		X		X							A													
CI_ResponsibleParty	organizationName								X																
	positionName								X																
	role				X				X																
RS_Identifier	code				X		X		X																
MD_Identification	abstract								X																
MD_DataIdentification	spatialRepresentationType				X		X		X																
	language				X				X																
	characterSet		X		X							A													
	topicCategory				X				X						A										
	distance				X		X		X																
MD_RepresentativeFraction	denominator				X		X		X																
CI_ResponsibleParty	organizationName								X																
	positionName								X																
	role				X				X																
CI_Citation	title					X	X		X																
CI_Date	date				X		X		X																
	dateType				X		X		X						?										
Ex_Extent	description								X																
Ex_BoundingPolygon	polygon			X	X		X								?										
Ex_GeographicBoundingBox	westBoundLongitude				X		X		X			A			?										
	eastBoundLongitude				X		X		X			A			?										
	southBoundLatitude				X		X		X			A			?										
	northBoundLatitude				X		X		X			A			?										
MD_Identifier	code				X		X		X																
Ex_TemporalExtent	extent				X		X		X																
Ex_VerticalExtent	minimumValue				X		X								?										
	maximumValue				X		X								?										
	unitOfMeasure				X		X								?										
MD_Format	name		X		X		X		X																
	version		X				X		X																
CI_OnlineResource	linkage		X				X		X																
DQ_DataQuality.LI_Lineage	statement								X																
LI_ProcessStep	description								X																
LI_Source	description								X																
Ex_Extent	description								X																
Ex_ GeographicExtent	polygon			X	X		X								?										
	westBoundLongitude				X		X		X			A			?										
	eastBoundLongitude				X		X		X			A			?										
	southBoundLatitude				X		X		X			A			?										
	northBoundLatitude				X		X		X			A			?										
Ex_TemporalExtent	extent				X		X		X						?										
Ex_VerticalExtent	minimumValue				X		X								?										
	maximumValue				X		X								?										
	unitOfMeasure				X		X								?										
MD_Identifier	code				X		X		X																

Figure 5.7. Degree of agreement with the interoperability levels by items and respondent groups

The graphs of Figure 5.7 allow the following interpretations:

- A close graphic correlation is observed between the results of the three aggregations. This means that the three groupings present similar results.
- Regarding the central column that shows the interpretations of the three most critical respondents, it would appear that some metadata items provide some more interoperability levels than proposed in the model. For example, the maximum and minimum latitudes and longitudes or the character sets provide interoperability at the syntactic level, the items topicCategory and

hierarchicalLevel provide dynamic interoperability and the item *parentIdentifier* would provide the organizational level. To facilitate identification of these items, they have been labeled with the text “A”.

- Items have also been identified which, for the most critical respondents, do not contribute any dynamic interoperability. These are the items of the package *EX_Extent* (maximum and minimum longitudes and latitudes, temporal and vertical extensions) and they have been identified with the label “?”.
- Once discrepancies identified, we have proceeded to analyze them. As a result decisions have been taken leading to either consider or reject them. Next the decisions and their justification are shown.
- Respondent 5 considers that the items containing the maximum and minimum longitudes and latitudes provide syntactic interoperability. In the proposed model the syntactic interoperability defines how data are encoded to facilitate their use, while in these items numerical values defining the geographic context are stored. For this reason this particular opinion is not considered valid.
- The character sets used to create metadata or the character set used in the data favor syntactic interoperability since they define the syntax of the data and enable interpretation of the information.
- We consider that the items containing horizontal, vertical and temporal geographic extent favor the dynamic level since knowledge of these items enable locating and selecting a dataset or service meeting certain needs. Therefore they allow replacing the data source if needed by changes in service availability. We agree with the most critical respondents in that items *topicCategory* and *herachicalLevel*, in addition to the proposed levels in the model, provide dynamic interoperability because their knowledge enables selection of other dataset or service meeting the need derived from eventual availability changes.
- We also agree with the most critical respondents regarding the item *parentIdentifier,* since we think it may provide organizational interoperability. This is due to the fact that the metadata of the elements of a series, e.g. a cartographic series, may point to the parent metadata describing it, hence enabling access to another type of information describing that series.

Finally, after interpreting and discussing the discrepancies between the proposed model and the collaborators' contributions, arguing for inclusion or exclusion of interoperability levels assigned to items, the final model including the changes is shown. Figure 5.8 shows both the model to be validated and the final model; the interoperability levels contributed by the items are labeled with "X". On the last column on the right containing the final model, the levels incorporating some metadata items in the validation process have been highlighted by filling the cells with a cream color.

	Metadata Attributes	Validate Modeldel	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational	Final Model	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Md_Metadata	fileIdentifier		X			X					X			X			
	Language				X				X				X				X
	characterSet		X		X						X	X	X				
	parentIdentifier		X			X								X			X
	hierarchyLevel				X			X	X				X		X	X	X
	hierarchyLevelName								X								X
	dateStamp				X		X		X				X		X		X
	metadataStandardName			X	X				X			X	X				X
	metadataStandardVersion			X	X				X			X	X				X
PT_Locale	language				X				X				X				X
	characterEncoding		X		X						X	X	X				
CI_ResponsibleParty	organizationName								X								X
	positionName								X								X
	role				X				X				X				X
RS_Identifier	code				X		X		X				X		X		X
MD_Identification	abstract								X								X
MD_DataIdentification	spatialRepresentationType				X		X		X				X		X		X
	language				X				X				X				X
	characterSet		X		X						X	X	X				
	topicCategory				X				X				X		X		X
	distance				X		X		X				X		X		X
MD_RepresentativeFraction	denominator				X		X		X				X		X		X
CI_ResponsibleParty	organizationName								X								X
	positionName								X								X
	role				X				X				X				X
CI_Citation	title					X	X		X					X	X		X
CI_Date	date				X		X		X				X		X		X
	dateType				X		X		X				X		X		X
Ex_Extent	description								X								X
Ex_BoundingPolygon	polygon			X	X		X					X	X		X		
Ex_GeographicBoundingBox	westBoundLongitude				X		X		X			X	X		X		X
	eastBoundLongitude				X		X		X			X	X		X		X
	southBoundLatitude				X		X		X			X	X		X		X
	northBoundLatitude				X		X		X			X	X		X		X
MD_Identifier	code				X		X		X				X		X		X
Ex_TemporalExtent	extent				X		X		X				X		X		X
Ex_VerticalExtent	minimumValue				X		X						X		X		
	maximumValue				X		X						X		X		
	unitOfMeasure				X		X						X		X		
MD_Format	name		X		X		X		X		X		X		X		X
	version		X				X		X		X				X		X
CI_OnlineResource	linkage		X				X		X		X				X		X
DQ_DataQuality.LI_Lineage	statement								X								X
LI_ProcessStep	description								X								X
LI_Source	description								X								X
Ex_Extent	description								X								X
Ex_ GeographicExtent	polygon			X	X		X					X	X		X		
	westBoundLongitude				X		X		X			X	X		X		X
	eastBoundLongitude				X		X		X			X	X		X		X
	southBoundLatitude				X		X		X			X	X		X		X
	northBoundLatitude				X		X		X			X	X		X		X
Ex_TemporalExtent	extent				X		X		X				X		X		X
Ex_VerticalExtent	minimumValue				X		X						X		X		
	maximumValue				X		X						X		X		
	unitOfMeasure				X		X						X		X		
MD_Identifier	code				X		X		X				X		X		

Figure 5.8. Comparative table with the interoperability levels of the core items of the metadata standard before and after validation

After having included the modifications on the model, we proceed with the analysis of how the changes affect the frequencies of the relations between levels.

To do this the number of items contributing other interoperability levels has been calculated for every level of the model. Table 5.18 shows the number of items providing other interoperability at every level and the number simultaneously provided by a pair of them.

Final Model	Technical	Syntactic	Semantic	Pragmatic	Dynamic	Conceptual	Organizational
Technical	7	3	4	2	3	0	3
Syntactic	3	7	7	0	2	0	2
Semantic	4	7	40	0	30	1	28
Pragmatic	1	0	0	3	1	0	2
Dynamic	3	2	30	1	33	1	24
Conceptual	0	0	1	0	0	1	1
Organizational	3	2	28	2	24	1	43

Table 5.18: Relations between the levels of the interoperability model after validation

Figure 5.9 shows the frequencies of every level of the model reflecting the values as shown in Table 5.18.

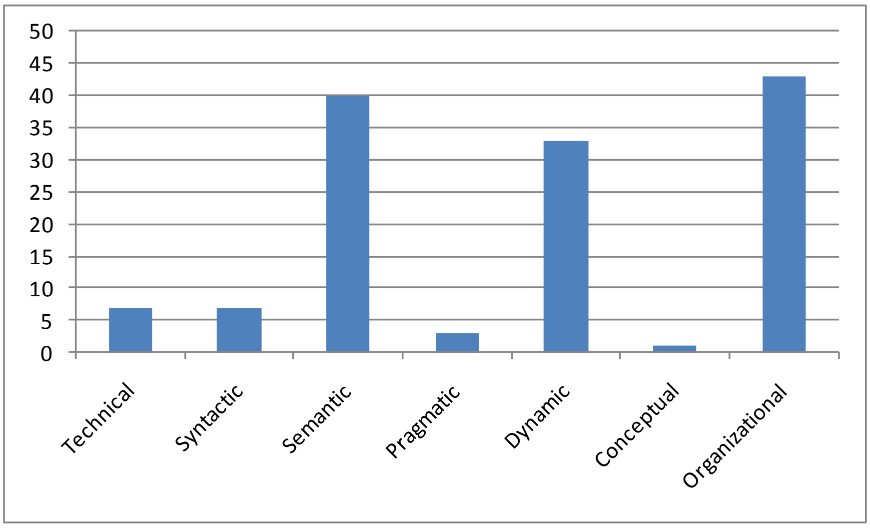


Figure 5.9. Histogram of the interoperability levels in the core items of metadata standard after validation

Figure 5.10 shows graphically the frequencies of the relations between the levels of the model after validation.

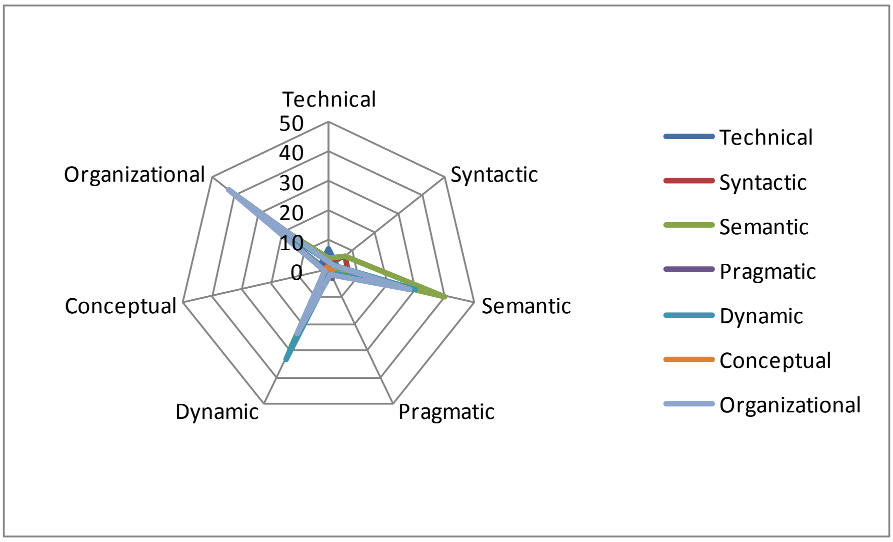


Figure 5.10: Frequency of the relations between the levels of the model after validation.

The mathematical correlation has been calculated between the number of metadata items providing interoperability at the different levels of the model, comparing the initial and final values of the study, after having discussed and readjusted the model during validation. To carry out this calculation, the values of Table 5.17 and Table 5.18 have been used. The result is 0.9940; this indicates the high degree of agreement between both tables representing the model to be validated and the already validated model. The correlation has also been calculated between the count of items of the full metadata standard (Chapter 4, Table 4.7) and the validated model (Table 5.18), the result being 0.9235; this value indicates a high degree of agreement between both studies.

Taking into account the interpretations, statements and considerations being made along the validation process, the results of this process may be summed up as follows:

- Two of the five respondents coincide in identifying the interoperability levels provided by metadata; the other three collaborators differ and enrich the model with their criteria.
- The aggregation of the results of the surveys has been carried out following different criteria: (1) merging the results of the five respondents and the model; (2) considering the results of the five respondents or (3) considering only the results of the most critical respondents.

- It has been verified that the results are similar and more relevance has been given to the sum of the results of three most critical collaborators.
- The discrepancies have been analyzed when all or two of them coincided, consequently some changes in the interoperability levels provided by some core items have been taken into account and in other cases they have been dismissed justifiably.
- The model as a whole and the frequencies of the relations between interoperability levels promoted by the core items have been analyzed and it has been confirmed that they are basically the same as the model being validated.
- It may be just mentioned –rather than highlighted– that by including the items providing more interoperability levels, the slightly favored one has been the syntactic level.
- Above all it should be emphasized that metadata items may be used to support interoperability levels.

5.6 Conclusions

Previous studies carried out about metadata –implicit or explicit– stored in the information and the forms of identifying the spatial reference system used by the coordinates allow drawing the following conclusions:

- In the first place it should be pointed out that there is a great variety of formats used by the industry and the scientific community to store geographic or space-related datasets. Seventy-one different formats grouped in 7 categories by context of use and types of data have been identified, analyzed and described.
- For every one of these formats, 7-8 metadata items can be obtained, many of them simple, some of them complex like the maximum and minimum coordinates, the band statistics, the count of rows and columns or the count of geometries of each type. Some types of formats deserve attention due to their capacity, unknown in many contexts, and their availability to store metadata, such as the jpeg and geotiff formats. In other cases the format allows inserting in the file header an information block which could be full metadata, such as

is the case with JPEG2000. Databases with spatial extensions also surprise in some cases by their capacity to store object catalogs or to include metadata in the geographic databases, such as is the case of ESRI.

- The second conclusion is related to the detailed study of the interpretation difficulty of the SRS used by the coordinates in the different data store formats. Instances of heterogeneity of the representations used have been identified and presented: non-structured text, semi-structured text, structured text, shared or agreed-on numerical and mnemonic representations. Another distinguishing aspect detected between heterogeneity and the analyzed formats is that in some cases this information is stored as a single attribute or metadata and in others a set of items is stored individually identifying the ellipsoid and the origin of the datum, the projection and the set of parameters defining it. These facts prove that the type of users to whom this knowledge is destined is the experts in Geodesy and Cartography. Fortunately standardization and normalization in this field (e.g. EPSG and WKT encoding) are enabling to identify the SRS univocally and as automatically as possible.
- The automatic interpretation of the different types of encoding used by the storage formats is a large task difficult to achieve; a possible solution would be to identify for every type of encoding its equivalent in a standardized encoding like the one proposed by the EPSG.
- Regarding the metadata items that may be automatically created, we should point out that the number depends on the type of geographic information analyzed and on the storage format used, as shown in the section on previous studies. Some indicative values are presented:
- Metadata associated to raster data or images made up of 83 items may be automatically created. This number of items could be increased if the number of bands of the format is higher than that of an image belonging to the visible spectrum (RGB) with 3 bands.
- Metadata associated to vector data made up of 69 items may be automatically created. This number of items could be increased if the format is capable of storing more than one type of geometry.
- Metadata associated to DEM made up of 68 items may be automatically created.

- The study of the functions and levels of the interoperability model applied to SDI, favored by the items of automatically created metadata has led to the following conclusions:
- The metadata items that may be created by extraction, calculation or inference (C) have been identified, also those having a cardinality >1 (N) or those that may be established by parameterization of the method or obtained from the context (F) for the different categories of geographic data; rasters, vectors and DTM. The functions performed by every item and the interoperability levels they provide have also been identified.
- The count of items providing the different levels of the interoperability model for every type of data and for the core items of the standard has been carried out (results in Table 5.16). The percentages of those values against the maximum numbers that may be obtained taking into account all the items of the standard have been represented in Figure 5.1 and it may be concluded that they favor all the levels of the model in a homogeneous manner. It may also be observed that the number of the core items that may be automatically created is high, about 60%. The values that stand out related to the conceptual and syntactic levels are caused by the scarce number of items in the standard providing them.
- The analysis of the functions performed by the metadata items that may be automatically created indicates that the most favored functions are the location and the use, with 34% and 36% of the items respectively, followed by evaluation and access, with 8% and 6% respectively. These values indicate that the automatic creation of metadata favors their two extreme, basic functions, first to locate, finally to use, while the intermediate functions of evaluating and accessing data are not favored to the same extent.
- Finally in this chapter the validation work is presented. The purpose has been to carry out this task about the identification of the different interoperability levels described in Chapter 4, provided by the core items. In addition to describing the methodology followed to carry out the validation, the results obtained in every analysis have been interpreted. Next we expose the results and conclusions drawn from the validation:
- Four out of five collaborators in the validation have made a constructive criticism, identifying or changing the interoperability levels provided by the

metadata items. Two of them have been more critical and for this reason a discriminating analysis has been made with their responses.

- It has been verified that the interpretation of the interoperability that may be contributed by metadata items is not subjective, as might have been expected. The number of items provided by the different levels of the model is similar to one another and to the model object of validation. It has been found that the aggregated analysis of the responses from several respondents regarding coincidences and discrepancies in interpretation of the interoperability gives off values similar to the individual ones, another reason backing the validity of the work presented in Chapter 4.
- As a consequence of this validation work, some changes in interpretation of the interoperability have been proposed (Figure 5.8). The result of analyzing the mathematical correlation on the count of the metadata items belonging to the core of ISO 19115 standard providing the interoperability levels [level-to-level or level pairs (Table 5.18)] with the values prior to validation, indicates that there is a degree of association of 0.9940; this fact validates the previous model.
- The correlation between the results of the analysis of the interoperabilities favored by all the items of the standard against the core now validated has been calculated and a degree of association of 0.9235 has been obtained; this validates our work.
- As a consequence of the validation of our analysis of the interoperability provided by the metadata items, it appears evident that their use as a mechanism to enable SDI interoperability has its strengths and weaknesses. The strengths are on the side of the organizational, semantic and dynamic interoperability levels while the weaknesses are on the side of the syntactic, pragmatic and conceptual interoperability levels. This conclusion is aligned to the other two described in Chapter 4 in which it is indicated that the syntactic interoperability is ensured by the standard itself and that the technical and pragmatic interoperabilities have to be guaranteed by means of other standards for technologies and service interfaces.

6 CONCLUSIONS

In the last chapter of this PhD thesis the scientific contributions provided by this work are exposed; the main research objectives and the issues posed in the first chapter are reviewed and it is shown how they have been answered. Finally new unsettled issues are put forward as future research work.

In the first place the scientific contributions are highlighted:

1. **Design of a new interoperability model between systems applicable to SDI:** A new interoperability model has been designed based on the research of the available models in the systems of systems context, into which their own organization aspects have been incorporated; the semantics of the levels has been described.

2. **Design of a new methodology for automatic GI metadata creation:** A new method has been devised to automate GI metadata creation based on the research of the existing methodologies in addition to a new one arising from the analysis of the implicit and explicit metadata contained in GI storage formats. This new methodology extracts, calculates and infers metadata in order to structure and store it in such a way as to support its exchange, exploitation and interoperability.

3. **Study of GI metadata from the interoperability perspective:** The items defined by the ISO 19115 international metadata standard have been analyzed from a new viewpoint: the interoperability levels provided by metadata. This analysis has been applied at two levels of detail: the items of the metadata core as defined by the standard and the entirety of items of the standard; similar results were obtained in both cases. The outcome is a new method to analyze the interoperability propitiated by metadata which is applicable to new profiles or standards.

4. **Application and validation of the analyses of the interoperability contributed by the items of the metadata standard:** A survey has been carried out under the author's supervision, geared to validate the

interpretation of the interoperability levels contributed by the metadata core items of the standard. The results of the survey validate the interoperability typification facilitated by the items of the standard.

5. **Application of the methodology of automatic metadata creation for different GI types and storage formats:** The theoretical results of applying the methodology of automatic metadata creation have been studied; the items that would be generated have been identified. The diversity and heterogeneity of formats and GI typologies precludes offering close results on the items that could be created; in the absence of an accurate outcome, aggregated level results by GI typology have been provided, identifying the items with cardinality > 1.

6. **Evaluation of the methodology of automatic metadata creation from the interoperability perspective:** The combined study of the metadata items that may be automatically created with the methodology and with the analysis of the interoperability provided by those items enables evaluation, from this perspective, of the degree of interoperability that will be provided by the methodology of automatic metadata creation.

6.1 Review of the research questions

The questions posed in Chapter 1 are now formulated anew; they will be answered later:

1. Is it possible to formalize an interoperability model of systems for SDI?
2. What is the contribution, in terms of interoperability, of the information contained in metadata?
3. Is it possible to create useful GI metadata automatically and efficiently?
4. What proposal is most appropriate to validate a system interoperability model within the SDI context?
5. What are the strengths and weaknesses of manually and automatically generated metadata from the point of view of the system interoperability that will exploit it (SDI)?

6.1.1 Interoperability model for SDI

First research question: *Is it possible to formalize an interoperability model of systems for SDI?*

Regarding this question, I propose the extension of the LCIM defined for SoS with an additional level to support the legal and organizational aspects of SDI. The bases for the formalization of the model has been: (a) the hypothesis that an SDI, analyzed as an information system, is an individual case of SoS; (b) the review and analysis of the available models within both the SoS and GIS contexts; (c) the analysis of the objectives proposed in the reviewed literature for every interoperability level.

Nine models (LISI, EIMM, OIMM, OIAM, LCIM, Goodchild et al., 1997, Bish, Intermodel5 and InterOP) and the classifications of interoperability proposed by a total of 27 authors (sources) have been reviewed and 15 different interoperability levels have been identified. Over 100 definitions or interoperability objectives have been reviewed and their analysis has allowed me to reject some and reclassify other related levels to finally propose a model applicable to SDI. In addition to analyzing interoperability within the SDI context, the literature concerning its measure and verification has been reviewed to conclude that the main objectives are detecting barriers and ensuring a certain level. In order to achieve these objectives, it is necessary to define the indicators on which to carry out the interoperability measures and a standard or measure scale for comparison of results. This has been achieved by analyzing the interoperability provided by the total number of items of the metadata standard, and it may be used as a reference to measure the interoperability provided by metadata.

The first research question is answered affirmatively and an interoperability model for SDI is proposed made up of seven levels coming from LCIM and from the above-mentioned detailed review and analysis. The proposed levels are: technical, syntactic, semantic, pragmatic, dynamic, conceptual and organizational; the next step has been to analyze the role played by metadata in the model.

6.1.2 Analysis of the interoperability provided by metadata

The second research question is: *What is the contribution, in terms of interoperability, of the information contained in metadata?*

The initial hypothesis was that metadata favor interoperability in one or several levels simultaneously. This premise has encouraged me to analyze the interoperability provided by metadata items at different granularity levels. In order to answer this second question, the metadata items belonging to the metadata core and the entirety of the items making up the standard have been analyzed and the interoperability levels provided by each have been identified.

The count of items favoring each level or groups of levels (pairs, trios, etc.) has been performed through a survey carried out by five different experts. The comparison of results achieved for the core items and for the entire standard are similar (correlation 0.943), the most favored levels being the organizational, semantic and dynamic levels; 46% of the items favor the three of them simultaneously. It should also be noted that the standard emphasizes the aspects of organizational interoperability since 92% of the items favoring only one level do so for the organizational level. The low impact of metadata on the syntactic level is equally remarkable, our opinion being that the actual metadata standard provides syntactic interoperability to the content defining the encoding rules. Regarding the technical and pragmatic levels, we think that these aspects of interoperability should be ensured by way of norms and standards defined by protocols, interfaces, etc. Finally, regarding conceptual interoperability, the important lack of items providing this level should also be mentioned.

We should point out that the results of the analysis of the interoperability provided by the items of the metadata standard as a whole may be used as a pattern or reference scale to carry out measures of the interoperability provided by metadata.

6.1.3 Methodology for automatic metadata creation

The third research question is: *Is it possible to create useful GI metadata automatically and efficiently?*

A new methodology suitable for automating the creation of metadata is proposed after the implementing system has been built. Our initial hypothesis is that the creation of metadata may become automatic, thus avoiding routine and wearisome tasks predisposing organizations and operators in charge against them. This hypothesis has prompted me to review the existing methodologies, to identify the information contained in the GI and the one implicitly related to storage format.

The proposed methodology consists of a number of stages for extraction of the information explicitly stored in the GI: interpretation or identification of data (spatial reference systems, formats), calculation on data (coordinate conversion/transformation, getting toponyms from a geographic nomenclator), identification of the stored content type (through statistical values in the case of images, with the attribute names and the feature types in vector data), derivation of categories from the topics and keywords for content cataloging, suggesting title for metadata, identification of the information structure in the formats to argue for a data model and finally, packaging the information in new metadata or on a supplied template.

The proposed methodology may be incorporated into the different flows of creation and updating of metadata described in the literature: (a) purely automatic creation; (b) automatic creation supplemented by GI expert; (c) automatic creation supplemented by the cataloging expert or (d) automatic creation supplemented by both experts.

The effectiveness of the method may be analyzed from different viewpoints. We have considered two aspects: the functions performed by automatically created metadata items and the interoperability levels provided by them. Regarding the functions they enable, the items created with this methodology favor the functions of location and use of data distinctly and the evaluation and access functions to a lesser extent. The second aspect (the interoperability levels provided by items) is dealt with later, with the fifth research question.

6.1.4 Validation of the interoperability model

The interoperability model for SDI, based on metadata and formalized as a consequence of the answers to the first two research questions, could be challenged by adducing subjectivity in the interpretation of the interoperability provided by the items of the metadata standard. In order to give an answer to the fourth research question, namely *"What proposal is most appropriate to validate a system interoperability model within the SDI context?"*, a survey has been designed with the purpose of validating the interpretations, hence the proposed model.

A controlled and representative sample of collaborators has been selected and they have been requested to identify the interoperability levels provided by the core items; they have been supplied with the descriptions of the interoperability levels of the model. The core has been utilized because it had been shown that there was a high correlation (0.943) between the results achieved when analyzing the core and the entirety of items making up the metadata standard. The analysis of the survey shows: (a) four of the five respondents have contributed comments and changes (adding or withdrawing interoperability levels provided by the items); (b) the results have been analyzed after having carried out several aggregations in the surveys (the three most critical and the five respondents) and the analysis of the interoperability provided by the core metadata items has been reproduced after justification and incorporation of changes; (c) A very high degree of correlation (0.994) with the model to be validated (core) has been found as well as a high degree of correlation (0.9235) with the complete model (all the items of the standard).

These results endorse the conclusions drawn from the answer to the second research question and they validate the model in view of the lack of important discrepancies in the interpretation of the interoperability.

6.1.5 Interoperability provided by automatically created metadata

In order to analyze the strengths and weaknesses of manually or automatically generated metadata, thus to answer the fifth research question, namely *"What are the strengths and weaknesses of manually and automatically generated metadata from the point of view of the system interoperability that will exploit it (SDI*)*?*, the metadata items stored with GI have been reviewed and identified; the representation forms of SRS used by the industry and academic environments have also been reviewed.

Based on the proposed methodology, the metadata items that may be automatically created have been identified for each GI typology. The results of this analysis indicate that an average of 83 items for raster data, 69 for vector data and 68 for DEMs may be created; these values may increase if raster data contain more than 3 bands or if vector data store more than one type of geometry.

When analyzing the interoperability provided by automatically created metadata, it has been verified that approximately 60% of the items belonging to the core of ISO 19115 metadata standard may be automatically created. It has also been confirmed that the percentage of automatically created items with respect to the entirety of items of the metadata standard is uniform for every level of the model.

These results show the strengths and weaknesses of metadata automatically created with the proposed methodology. Relative to manually created metadata this thesis provides a framework to analyze strengths and weaknesses. This framework is made up of the interoperability model for SDI, the methodology enabling analysis of the levels favored by metadata items and the patterns of theoretical interoperability provided by the items (of the core and the entire standard) that make their measure possible.

6.2 Future lines of research

The results of the research about interoperability models and the automatic metadata creation are novel and useful for the systems of an SDI to interoperate. However, along the development course of the thesis, other research issues have come up leading to future lines of work.

Some of the questions that have cropped up are:
How could the categories of GI stored as vector data be inferred or determined? To determine or infer those categories, research should be undertaken about the reasoning techniques, based on stored rules and on data mining algorithms to determine their applicability within this context. These techniques would be fed on the object catalogs stored together with data, the layer names and the data attributes. Their applicability to select descriptive keywords belonging to multilingual thesauri helping to catalog the resource should also be analyzed.

Which should be the metadata core describing datasets that would help SDI interoperation with respect to the proposed model? The ISO 19115 metadata core has been defined to support the functions of location and use of data. It has been shown that the interoperability provided by those items is not homogeneous and is centered on the organizational, semantic and dynamic levels. Therefore definition of a metadata core maximizing interoperability at their different levels is a research topic for the future.

Would it be possible to define a core of metadata useful for SDI services uniformly maximizing the interoperability provided by the items? The investigation carried out within the framework of this thesis has been centered on the metadata that describe datasets. The interoperability of SDI, in addition to be promoted by those metadata, should also be promoted by the services that offer or exploit them.

What and how should be the metadata items providing conceptual interoperability when describing services? Within the context of data their conceptual models described with engineering languages seem to be valid. Is this reasoning extensible to service models?

7 REFERENCES

ALCTS (2004). ALCTS/CCS/Committee on Cataloguing: Description and Access Task Force on Metadata, Summary Report. Web document. Visited 10-1-2008.
Website: http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html

Antonovic, V. and Novak, I. (2006). GISOSS - One-Stop-Shop GIS. Shaping the Change, *XXIII FIG Congress*, *Munich*, *Germany*, October 8-13, 2006. Web document. Visited 10-1-2008.
Website: http://www.fig.net/pub/fig2006/papers/ts37/ts37_02_antonovic_novak_0569.pdf

ANZLIC (1996). ANZLIC Guidelines: Core Metadata Elements Version 1 Report: ANZLIC Working Group on Metadata, July 1996,

ANZLIC (1997). Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand. The Australian New Zealand Land Information Council. Web document, http://www.auslig.gov.au/pipc/anzlic/metaelem.html.

ANZLIC (2005). Draft Anzlic ISO metadata profile. Web document.
Website: http://www.gsdi.org/SDI-AP/docs2005/drftmtdt.pdf

Arms,W., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J. Y Terrizzi, C. (2002). A Spectrum of Interoperability: The Site for Science Prototype for the NSDL. D-Lib Magazine, 8(1), January 2002. Web document. Visited 10-1-2008.
Website: http://www.dlib.org/dlib/january02/arms/01arms.html

Assche, F. (2006). An Interoperability Framework. (Learning Interoperability Framework for Europe). Web document. Visited 10-1-2008.
Website: http://www.intermedia.uio.no/confluence/display/life/An+Interoperability+Framework

Athena Integrated Project (507849) (2005). Framework for Establishment and Management Methodology, Deliverable DA1.4 Web document. Visited 10-1-2008.
Website: http://modelbased.net/aif/methodology/eimm.html

Bailer, W. and Schallauer, P. (1998). Metadata in the Audiovisual Media Production Process. Studies in Computational Intelligence (SCI) 101, 65–84 (2008).

Baird, K. (2006). Automated Metadata. March 2006, *Jorum Team*.

Balfanz, D. (2002). Automated Geodata Analysis and Metadata Generation. Society of Photo-Optical Instrumentation Engineers -SPIE-, Bellingham/Wash.: *Visualization and Data Analysis*

2002. Proceedings : 21 - 22 January 2002, San Jose, USA Bellingham/Wash.: SPIE, 2002 (SPIE Proceedings Series 4665) ISBN: 0-8194-4405-7

Batcheller, J. (2008). Automating geospatial metadata generation—An integrated data management and documentation approach. *Computers & Geosciences 34* (2008) 387–398

Beard, K. (1996). A Structure for Organizing Metadata Collection. *Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, January 21-25, 1996

Béjar, R., Nogueras-Iso, J., Muro-Medrano, P. y Zarazaga-Soria, F. (2008). Systems of Systems as a Conceptual Framework for Spatial Data Infrastructures. Article Under Review for the *International Journal of Spatial Data Infrastructure Research*.

Bermudez, L. (2004). PhD Thesis: ONTOMED: Ontology Metadata Framework, Drexel University, December 2004. Web Document, http://dspace.library.drexel.edu/bitstream/1860/376/8/Bermudez_Luis.pdf

Bernard, L. et al. (2005). Towards an SDI Research Agenda. Proceedings of 11th EC-GIS, Alghero, Sardinia, Italy

Bernard, L., Kanellopoulos, I., Annoni, A. and Smits, P. (2005). The European Geoportal - One step towards the Establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, 29: 15-31.

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *Int. Journal for Geographical Information Science*, 12(4):299–314, 1998

C2SS WG (1996). C2SS Working Group, 1996 C2 Support Study Phase 1 Report.

Campbell, T. (2008). Fostering a Culture of Metadata Production. GSDI10: *Tenth International Conference for Spatial Data Infrastructure*, St. Augustine, Trinidad February 25-29, 2008. Website: http://www.gsdi.org/gsdi10/papers/TS8.2paper.pdf

CAPLAN, P. (1995). You call it corn, we call it syntax-independent metadata for document-like objects. The Public Access Computer Systems Review, v. 4, n. 6, 1995.

Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago: American Library Association.

Carnegie-Mellon University Software Engineering Institute (CMU-SEI). Online Interoperability Guide. Web document. Visited 10-1-2009. Website: http://www.sei.cmu.edu/isis/guide/engineering/procurement.htm.

Carney, D., Smith, J. and Place, P. (2005). Topics in Interoperability: Infrastructure Replacement in a System of Systems (CMU/SEI-2005-TN-031). Pittsburgh, Pa: Software Engineering Institute, Carnegie Mellon University, November 2005. Web document. Visited 10-1-2008. Website: http://www.sei.cmu.edu/pub/documents/05.reports/pdf/05tn031.pdf

CGIAR-CSI (2004). Metadata Tips. Web document. Visited 10-1-2008. Website:http://www.csi.cgiar.org/metadata/Metadata_tips.asp

Chen, D., Daclin, N. (2007). Barriers Driven Methodology For Enterprise Interoperability. Virtual Enterprises and Collaborative Networks, 2007, pp:453-460

Chen D, Dassisti M, Elvesaeter B, (2007). Interoperability Knowledge Corpus, intermediate report IST-508 011. Web document.

Clark, T. and Jones, R. (1999). Organisational Interoperability Maturity Model for C2. Web document. Visited 10-1-2008. Website: http://www.dodccrp.org/events/1999_CCRTS/pdf_files/track_5/049clark.pdf

Codd, E. F. (1990). "The Relational Model for Database Management", Addison-Wesley, 538, 2 2.

Colleman, A. (2002). Metadata: The Theory behind the Practice. *4th State GILS Conference*, April 2002, Scottsdale, AZ.

Crompvoets, J. and Bregt A. (2003). World Status of National Spatial Data Clearinghouses. URISA Journal. Urban and Regional Information Systems Association, Vol. 15, No. 2.

Currier, S., Barton, J. et al. (2004). Quality assurance for digital learning objects repositories: issues for the metadata creation process. *ALT-F*, *Research in Learning Technology*, Vol.12, No. 1, March 2004

Daclin, N., Chen, D. and Vallespir, B. (2006). Enterprise interoperability measurement - Basic concepts. Proceedings of the *Open Interop Workshop on Enterprise Modelling and Ontologies for Interoperability*. ISSN 1613-0073.

Daclin, N., Chen, D. and Vallespir, P. (2008). Methodology for Enterprise Interoperability. Proceedings of the 17th World Congress: The International Federation of Automatic Control Seoul, Korea, July 6-11, 2008. Web document. Visited 10-1-2008.

Website:http://www.nt.ntnu.no/users/skoge/prost/proceedings/ifac2008/data/papers/2896.pdf

Dangermond, J. (1991). The Commercial Setting of GIS in Maguire, D.J., Goodchild, M. F. & Rhind, D.W. Geographical Information Systems: Principles and Applications. Longman.

Danko, D. (2002). ISO/TC 211 Geographic information/Geomatics The Standards in Action Workshop in Gyeongju, Korea: Implementation. Web document. Visited 10-1-2008.
Website: http://www.isotc211.org/WorkshopGyeongju/Presentations/Metadata.ppt

Danko, D. (2005). Interoperability & Metadata. At the *19th ISO/TC 211 Plenary meeting* in Palanza, Italy, the TC decided to install the Focus Group on Data Producers (FGDP). Website: http://www.isotc211fgdp.info/docs/figws/session2/Dave_Danko_%20Metadata.ppt

Dekkers, M. (2007). Metadata and modelling for Interoperability. Web document. Visited 10-1-2008. Website: http://library2.nalis.gov.tt/Portals/0/cdl_Makx_Dekkers_20070712.pdf

Díaz, L. Granell, C. Beltrán, A. Llaves, A. and Gould, M. (2008a). Extracción Semiautomática de Metadatos: Hacia los metadatos implícitos. *II Jornada de SIG Libre*. Universidad de Girona.

Díaz, L., Gould, M., Beltrán, A., Llaves, A. y Granell, C. (2008). Multipurpose Metadata Management in gvSIG. Proceedings of the academic track of the *2008 Free and Open Source Software for Geospatial (FOSS4G) Conference*, 29 September – 3 October 2008, Cape Town, South Africa. ISBN: 978-0-620-42117-1, pp 90-99

Ding, H. (2005). Challenges in Building Semantic Interoperable Digital Library System. Web document. Visited 10-1-2008. Website:http://www.idi.ntnu.no/grupper/su/courses/dif8901

Duval, E., Hodgins, W., Sutton, S., y Weibel, S. (2002). Metadata Principles and Practicalities. D-Lib Magazine. Web document. Visited 10-1-2008.
Website: http://www.dlib.org/dlib/april02/weibel/04weibel.html.

ECNBII, (2003). FGDC Biological Data Profile As it maps Dublin Core. GeoConnection. Web document. Visited 10-1-2008. Website:
http://www.geoconnections.org/developersCorner/devCorner_devNetwork/meetings/2003.06.10/Present/ECNBII.ppt

Ercegovac, Z. (1999). Introduction. *Journal of the American Society for Information Science*, v. 50, n. 13, p. 1165-1168, 1999

Flater, D. (2002). Impact of Model-Driven Standards, *Proceedings of the 35th Hawaii International Conference on System Sciences – 2002*. Web document. Visited 10-1-2008.

Website: http://csdl2.computer.org/comp/proceedings/hicss/2002/1435/09/14350285.pdf

FGDC (2000). Content Standard for Digital Geospatial Metadata Workbook, version 2.0. Federal Geographic Data Committee (USA), 2000.

Gayatri and Ramachandran, S. (2007). Understanding Metadata. The Icfai Journal of Information Technology, March 2007.

Georgiadou, Y., Puri S. and Sahay S. (2005). Towards a potential research agenda to guide the implementation of Spatial Data Infrastructures - A case study from India. *International Journal of Geographical Information Science* 19(10): 1113—1130

Gilliland-Swetland, A. (2000). Setting the Stage, in Introduction to Metadata: Pathways to Digital Information. Los Angeles: Getty Research Institute. Web document. Visited 10-1-2008. Website: http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/index.html

Goh, C. (1997). Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Sources, Ph.D. Thesis, MIT Sloan School of Management, 1997

Goodchild, M., Egenhofer, M., Fegeas, R. (eds.) (1997). Interoperating GISs: Report of a specialist meeting held under the auspices of the VARENIUS project. *Panel on computational implementation of Geographic concepts*. Santa Barbara (California), 5-6 December 1997. Web document. Visited 10-1-2008. Website: http://www.ncgia.ucsb.edu/Publications/Varenius_Reports/Interop.pdf

Goodchild, M. (2007). Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0, International Journal of Spatial Data Infrastructures Research, 2007, Vol. 2, 24-32.

Gordon, D. (2003). What is Interoperability and why does it matter? *MapInfo Magazine Volume 7, nº4. 2003*. Web document. Visited 10-1-2008. Website: http://resource.mapinfo.com/static/files/document/1074288321775/mapinfomag_summer2003.pdf

Greenberg, J. (2004). Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. *Journal of Internet Cataloging* 6 (4), 59–82.

Greenberg, J., Spurgin, K. and Crystal, A. (2006). Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. *Int. J. Metadata, Semantics and Ontologies, Vol. 1, No. 1, 2006.*

Groot, R. and J. McLaughin, Eds. (2000). Geospatial data infrastructure - Concepts, cases, and good practice. Oxford, Oxford University Press.

Guptill, S.G. (1999). Metadata and data catalogues. In: Longley, P., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), *Geographical Information Systems*. Wiley, Chichester, pp. 677–692

Guy, M., Powell, A. and Day, M. (2004). *Improving the Quality of Metadata in Eprint Archives*. Web document. Visited 10-1-2008. Website:http://www.ariadne.ac.uk/issue38/guy/

Hamilton, J., Rosen, J. and Summers, P. (2004). Developing Interoperability Metrics, in joint command and control interoperability: cutting the gordian knot, Chapter 6 (2004).

Harvey, F., Kuhn, W., Pumdt, H. and Bishr, Y. (1999). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science*, Vol. 33, No. 2. (10 May 1999), pp. 213-232

Hedorfer, M. and Bianchin, A. (1999). The Venice Lagoon Experimental GIS at the IUAV. *Interop99: The 2nd International Conference on Interoperating Geographic Information Systems*. Web document. Visited 10-1-2008. Website:http://www.hedorfer.it/docs/rsalv/rsalv1io-ENG.pdf

Heery, R. Biblink: LB4034 D1.1 metadata formats. [S. l.] : Biblink, 1996.

IDABC (2006). Interoperability for eGovernment Dec 2006. European Interoperability Framework for pan-European eGovernment services. Office for Official Publications of the European Communities, Luxembourg. Web document. Visited 10-1-2008.
Website: http://ec.europa.eu/idabc/en/document/3473/5585.

INSPIRE Metadata IR (2008). Draft Guidelines – INSPIRE metadata implementing roles based on ISO 19115 and ISO 19119. Web document. Visited 10-1-2008. Website: http://www.ec-gis.org/inspire/reports/ImplementingRules/metadata/Draft_Guidelines%20_INSPIRE_metadata_implementing_rules.pdf

IEEE (1990). Standard Computer Dictionary—A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.

ISO-14258 (1999). Industrial automation systems - Concepts and rules for enterprise models.

ISO 19101 (2002). Geographic Information – Reference Model. International Standard Organization

ISO 19115 (2003). Geographic Information – Metadata

ISO 19119 (2005). Geographic Information – Services. International Standard Organization

ISO 19128 (2005). Geographic Information – Web Map Service Interface

ISO 19142 (2008). Geographic Information – Web Feature Service

ISO 23950 (1998). Information and documentation -- Information retrieval (Z39.50) -- Application service definition and protocol specification. Web document. Visited 10-1-2008. Website:http://www.iso.org/iso/catalogue_detail?csnumber=27446

Janowicz, K., Raudal, M., Schwering, A. and Kuhn, W. (2008). Semantic Similarity Measurement and Geospatial Applications. *Transaction in GIS* December 2008, Vol. 12, Issue 6, pp 651-659.

Johnson, F. (1995). Automatic abstracting research, *Library Review*, Vol. 44, No. 8, pp.28–36.

JORUM (2004). The JISC Online Repository for [learning and teaching] Materials: *JORUM Scoping and Technical Appraisal Study*, Volume V: Metadata

Johnston, P. (2005). Good Practice Guide for Developers of Cultural Heritage Web Services. Research Officer, UKOLN. Enlace: http://www.ukoln.ac.uk/interop-focus/gpg/Metadata/

Jones, M. and Taylor, G. (2003). Metadata: Spatial Data Handling and Integration Issues. School of Computing Technical Report. Issued: Febrary 2003.

Kalantari, Mohsen, Rajabifard, Wallace and Williamson (2006). An interoperability toolkit for e-Land administration, Williamson, Enemark and Wallace (eds), Sustainability and Land Administration Systems, Department of Geomatics, Melbourne , 213-222

Kasunic, M. and Anderson, W. (2004). Measuring systems interoperability: challenges and opportunities, Software engineering measurement and analysis initiative, Technical note CMU/SEI – 2004 – TN – 003, 2004.

Kildow, M. (1996). The value of Metadata (An NSDI report). US Fisheries and Wildlife Services. Web document. Visited 10-1-2008. Website: http://www.r1.fws.gov/metadata/meta.html.

Kingston, G., Fewell, S. and Richer, W. (2005). An Organisational Interoperability Agility Model. Web document. Visited 10-1-2008. Website:http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA463924&Location=U2&doc=GetTRDoc.pdf

Korth, H. F. and A. Silberschatz (1991). "Database Systems Concepts". Singapore, McGraw - Hill, 694, 2 2.

Kuhn, W. (2003). Semantic Reference Systems. International Journal of Geographical Information Science 17(5): 405-409

Kuhn,W. and Raubal, M. (2003). Implementing semantic reference systems. The *6th AGILE Conference in GIScience,* Lyon France, April 24-26 2003.

Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics III* (2005), pp. 1-24

Lemmens, R. (2006). Semantic interoperability of distributed geo-services. Publication on Geodesy 63. NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission. Delft September 2006.

Lewis, G. and Wrage, L. (2005). Approaches to Constructive Interoperability (CMU/SEI-2004-TR-020 ESC-TR-2004-020). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2005. Web document. Visited 10-1-2008. Website:http://www.sei.cmu.edu/publications/documents/04.reports/04tr020.html

LISI (1997). Levels of Information Systems Interoperability. Software Engineering Institute. Web document. Visited 10-1-2008. Website:http://www.sei.cmu.edu/isis/guide/introduction/lisi.htm

Longley, P., Goodchild, M., Maguire, D. and Rhind, D. (2000). Geographic Information Systems and Science, 2nd Edition, Wiley Europe.

Longhorn, R. (2008). Burning Issue. Web document. Visited 10-1-2008. Website:http://www.geoconnexion.com/burning_issue_int.php

Losee, R. (2003). Adaptive organization of tabular data for display, *Journal of Digital Information,* Vol. 4, No. 1, Web document. Visited 10-1-2008. Website:http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Losee/.

Maier, M. (1996). Architecting Principles for Systems-of-Systems. In *6th Annual International Symposium of INCOSE,* Boston, MA, USA, p. 567-574.

Manso-Callejo, M., Nogueras, J., Zarazaga, J. and Bernabé-Poveda, M. (2004). Automatic Metadata Extraction from Geographic Information. *AGILE 2004. 7th Conference on Geographic Information Science, conference* Proceedings, 379-385.

Manso-Callejo, M. and Bernabé-Poveda, M. (2005). CSCAT: Translation Web Service & Coordinate Systems Catalogue. *GISPlanet 2005 Conference*, 30 May - 4 June 2005, Estoril, Portugal

Miller, P. (2000). Interoperability. What is it and Why should I want it? Ariadne Issue 24. 21-Jun-2000. Web document. Visited 10-1-2008.
Website:http://www.ariadne.ac.uk/issue24/interoperability/intro.html

Milstead, J. and Feldman, S. (1999). Metadata: Cataloging by any other name. Online 25-31. Web document. Visited 10-1-2008.
Website: http://www.onlineinc.com/onlinemag/OL1999/milstead1.html.

Moellering, H. and Brodeur, J. (2006). Towards a North American Profile of the ISO 19115 World Spatial Metadata Standard. *GSDI-9 Conference Proceedings*, 6-10 November 2006, Santiago, Chile. Web document. Visited 10-1-2008.
Website: http://gsdidocs.org/gsdiconf/GSDI-9/papers/TS12.4paper.pdf

Mohammadi, M., Binns, A., Rajabifard, A. and Williamson, I. (2006). Spatial Data Integration. *17th UNRCC-AP Conference and 12th Meeting of the PCGIAP*, Bangkok, 18-22 September 2006. Web document. Visited 10-1-2008. Website:
http://www.geom.unimelb.edu.au/research/SDI_research/publications/files/Spatial%20Data%20Integration.doc

Morris, S., Nagy, Z. and Tuttle, J. (2007). North Carolina Geospatial Data Archiving Project. *NCSU Libraries and North Carolina Center for Geographic Information & Analysis*. Web document. Visited 10-1-2008. Website:
http://www.digitalpreservation.gov/partners/ncgdap/high/NCGDAP_InterimReport_June2008_final.pdf

Mustacoglu, A. (2007). Event-Based Model for Reconciling Digital Entries. Ph.D. Proposal Report. Department of Computer Science, Indiana University, Bloomington. May 3 2007.

Nadkarni, P., Chen, R. and Brandt, C. (2001). UMLS concept indexing for production databases: a feasibility study, *Journal of the American Medical Information Association,*Vol. 8, N. 1, pp.80–91

Najar, C. (2006). A model-driven approach to management of integrated metadata – spatial data in the context of spatial data infrastructure. Doctoral Thesis. Web document. Visited 10-1-2008.
Website: http://e-collection.ethbib.ethz.ch/view/eth:28733

Najar, C. and Giger, C. (2006). Spatial Data and Metadata Integration for SDI interoperability. Under review for the *international Journal of Spatial Data Infrastructures Research*, submitted 2006-09-25.

NATO (2004). NATO C3 Board: NATO C3 System Architecture Framework (NAF) (Document AC/322-D (2004)0002 (INV)) NATO, 2004

Nebert, D. (2004). Developing Spatial Data Infrastructures: The SDI Cookbook. Web document. Visited 10-1-2008. Website: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Nedovic-Budic, Z. and Pinto, J. (2001). Organizational (soft) GIS interoperability: lesson from the U.S. International Journal of Applied Earth Observation and Geoinformation, Volume 3, Number 3, 2001, pp. 290-298(9). Elseivier.

Nell, J. D. (1996) TC184 SC5 WG1 convener Report.

Nogueras-Iso, J., Zarazaga-Soria, F. and Muro-Medrano P. (2005). Geographic Information Metadata for Spatial Data Infrastructures. Resources, Interoperability and Information Retrieval. Springer editions. ISBN:3-540-24464-6.

Nowak, J. and Nogueras, J. (2005). Issues of multilingualism in creating a European SDI – The perspective for Spatial Data Interoperability. 29 June- 1 July 2005 Sardinia. 11th EC-GI & GIS Workshop. Web document. Visited 10-1-2008. Website: http://www.ec-gis.org/Workshops/11ec-gis/

NISO (2004). National Information Standards Organization. (2004). Understanding Metadata. Web document. Visited 10-1-2008. Website: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

OpenGIS Guide (1998). Open GIS Consortium Technical Committee, The OpenGIS Guide: Introduction to Interoperable Geoprocessing and the OpenGIS Specification, Third Edition, Draft, June 3, 1998.

Oosterom, P. (2004). Geo-information Standards in Action. ISO TC211/Metadata (Danko, D.). Web document. Visited 10-1-2008. Website: http://www.ncg.knaw.nl/Publicaties/Groen/pdf/42Standards.pdf

Ostensen, O. and Danko, D. (2005). Global Spatial Metadata Activities in the ISO/TC211 Geographic Information Domain. World Spatial Metadata Standards: Scientific and Technical Descriptions, and Full Descriptions with Crosstable. H. Moellering, H.J.G.L. Aalders & A. Crane (Editors). Elsevier Ltd.

Ouksel, M. (1999). A Framework for a Scalable Agent Architecture of Cooperating Heterogeneous Knowledge Sources in Intelligent Information Agents: Agent-Based Information Discovery and Management of the Internet, Matthias Klusch Ed., Springer Verlag, 1999.

Ouksel, A. and Sheth, A. (1999). Semantic Interoperability in Global Information Systems: A Brief Introduction to the Research Area and the Special Section. SIGMOD Record 28(1): 5-12(1999)

Pasquinelli, A. (2003). Information technology directions in libraries: a sun microsystems white paper. Web document. Visited 10-1-2008. Website:http://wwwsun.com/products-n-solutions/edu/libraries/libtechdirection.html.

Pierkot, C. (2006). Using Metadata to Help the Integration of Several Multi-source Sets of Updates. *GSDI-9 Conference Proceedings*, 6-10 November 2006, Santiago, Chile. Web document. Visited 10-1-2008. Website: http://www.gsdi9.cl/english/papers/TS12.2paper.pdf

Phillips, A. Willianson, I and Ezigbalike, C. (1998). The importance of Metadata Engines in Spatial Data Infrastructures. The 26th Annual Conference of AURISA Perth, Western Australia, 23-27 November 1998. Web document. Visited 10-1-2008. Website:http://www.sli.unimelb.edu.au/research/publications/IPW/phillips_NZ_981.htm

Pokraev, S., Reichert, M., Steen, M. and Wieringa, R. (2005). Semantic and Pragmatic Interoperability: A Model for Understanding. *Proceedings of the CAiSE*. Web document. Visited 10-1-2008. Website: http://ceur-ws.org/Vol-160/paper21.pdf

Pridmore, J. and Rumens, D. (1989). Interoperability-how do we know when we have achieved it? Command, Control, Communications and Management Information Systems, May 1989, 192-205, ISBN:-85296-380-7

Probst, F. (2006). Ontological Analysis of Observations and Measurements. *International Conference on Geographic Information Science* No4, Münster, Germany (2006), vol. 4197, pp. 304-320. ISBN 3-540-44526-9; 978-3-540-44526-5

Rajabifard, A. and Binns, A. (2004). Facilitating Virtual Australia Utilizing an SDI Enabled Platform. December 2004. Web document. Visited 10-1-2008. Website: http://datasmart.oesr.qld.gov.au/Events/datasmart.nsf/0/F0E6094FCB6E7F124A256FA2007C2274/$FILE/QLD%20Seminar-CRC-Project%203.1%20-%20Dec%202004.pdf

Rawat, S. (2003). Interoperable Geo-Spatial data model in the Context of the Indian NSDI. Web document. Visited 10-1-2008. Website: http://www.itc.nl/library/Papers_2003/msc/gfm/sujata.pdf

SAGA (2006). Standards and Architectures for eGovernment Applications SAGA. October 2006. Web document. Visited 10-1-2008. Website: http://www.epractice.eu/document/3210

Schekkerman, J. (2004). President of the Institute for Enterprise Architecture Development (IFEAD). Roadmap to implementation. Web document. Visited 10-1-2008.
Website: http://web-services.gov/region4soa10104.ppt

Senso, J. and Rosa Piñero, A. (2003). El concepto de metadato. Algo más que descripción de recursos electrónicos. Ci. Inf., Brasília, v. 32, n. 2, p. 95-106, maio/ago. 2003.

Shanzhen, Y., Qi, L. and Jicheng, C. (1999). An interoperability GIS model based on the spatial information infrastructure. Geoinformatics and Socioinformatics. The Proceedings of Geoinformatics'99 Conference. Ann Arbor, 12-21 June, pp, 1-5.

Shekhar, S. (2004). Spatial Data Mining and Geo-spatial Interoperability. Report of the NCGIA Specialist Meeting on Spatial Webs, Santa Barbara, December 2-4 2004, National Center for Geographic Information and Analysis, University of California

Sheldon, T. (2001). Linktionary. Enter «Metadata». Visited May 29, 2006

Steinacker, A., Ghavam, A., and Steinmetz, R. (2001). Metadata Standards for Web-Based Resources. IEEE MultiMedia, January-March 2001.

Swick, R. (2002). Metadata Activity Statement. Visited May 29, 2006

Stroetmann, K. (2005). Towards an Interoperability framework for European e-Health Research area. Brussels, Feb. 14-15, 2005. Web document. Visited 19-1-2008, Website: http://www.i2-health.org/i2-h-presentations/WHO-WS_Semantic-IOP_KAS_2005-02-14.pdf

Struetmann, K. (2005). Towards an Interoperability framework for European e-Health Research area. Locating the Semantic Interoperability Domain. WHO/EC Workshop on semantic interoperability, Brussels, Feb. 14-15, 2005. Web document. Visited 19-1-2008, Website: https://www.who.int/classifications/terminology/stroetmann.pdf

SUMM (1991). Semantic Unification Meta Model, ISO/IEC JTC1 SC2 WG3, N1360, 1991-Oct

Taylor (2004). The Organization of Information. 2nd ed. Westport, CN: Libraries Unlimited

Tolk, A. (2003). Beyond Technical Interoperability—Introducing a Reference Model for Measures of Merit for Coalition Interoperability, Proceedings of the 8th ICCRTS, Washington, D.C., June 17-19, 2003

Tolk, A. and Muguira, K. (2003). The Levels of Conceptual Interoperability Model (LCIM). Proceedings of the 2003 Fall Simulation Interoperability Workshop, Orlando, FL, September 2003

Turnitsa, C. and Tolk, A. (2006). Battle Management Language: A Triangle with Five Sides Proceedings of the Simulation Interoperability Standards Organization (SISO) Spring Simulation Interoperability Workshop (SIW), Huntsville, AL, April 2-7, 2006

Tolk, A., Diallo, S. and Turnitsa, C. (2007). Applying the levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering. International Journal Systemics, Cybernetics and Informatics. Volume 5 – Number 5

Vckovski, A. (1998). Interoperable and Distributed Processing in GIS (Research Monographs in GIS). Taylor and Francis.

West, J. and Hess, T. (2002). Metadata as a knowledge management tool: supporting intelligent agent and end user access to spatial data. Decision Support Systems 32, 247–264

Wieringa, R. (2003). Design Methods for Reactive Systems: Yourdon, Statemate, and the UML. Morgan Kaufmann, 2003. Web document. Visited 19-1-2008, Website: http://www.mkp.com/dmrs

William A et al. (2002). A Spectrum of Interoperability: The Site for Science Prototype for the NSDL. D-Lib Magazine, 8(1), January 2002. Web document. Visited 19-1-2008, Website: http://www.dlib.org/dlib/january02/arms/01arms.html

Williamson, I., Rajabifard, A. and Feeney, M. (2003). Developing Spatial Data Infrastructures: From Concept to Reality, ISBN 0-415-30265-X, Taylor & Francis, U.K.

Williamson, I. (2004). Building SDIs—the challenges ahead. In Proceedings of the *7th International Conference: Global Spatial Data Infrastructure*, 2–6 February, Bangalore, India.

Wilson, E. (1998). Manuals Go Click. The Age. Melbourne, Australia

Whitman, L., Santanu, D. and Panetto, H. (2006). An enterprise model of interoperability. Information Control Problems in Manufacturing, Volume 12, Part 1. Web document. Visited 19-1-2008, Website: http://hal.archives-ouvertes.fr/docs/00/08/20/37/PDF/366a.pdf

Woodley, M. S., Clement, G. and Winn, P. (2003). DCMI Glossary. Web document. Visited 19-1-2008, Website: ttp://dublincore.org/documents/2003/08/26/usageguide/glossary.shtml

Woodley M, Clement, G and Winn, P. (2005). Dublin Core Metadata Initiative Web document. Visited 19-1-2008, Website: http://dublincore.org/documents/usageguide/glossary.shtml

Woodley, M. (2004). Dublin Core Metadata Initiative. Web document. Visited 19-1-2008, Website: http://dublincore.org/documents/usageguide/glossary.shtml

Wright-Patterson, T.P. (1936). Factors Affecting the Cost of Airplanes, *Journal of Aeronautical Sciences*, 3(4) (1936), p. 122-128.

Wyoming University. Metadata Resources at the University of Wyoming. http://www.sdvc.uwyo.edu/metadata/why.html

Yilmaz, L. and Tolk, A. (2006). Engineering Ab Initio Dynamic Interoperability and Composability Via Agent-Mediated Introspective Simulation. In *Proceedings of the 2006 Winter Simulation Conference*, p. 1075-1082. Monterey, CA, USA Dec 2006. ISBN: 1-4244-0500-9.

Zeigler, B. and Hammonds, P. (2007). Modelling and simulation-based data engineering. Introducing Pragmatics into Ontologies for Net-Centric Information Exchange. Elsevier Academic Press. ISBN:9780123725158