Implementation of Recent Metadata Directives and Guidelines in Public Administration: the Experience of Sardinia Region

Luisa Manigas, Michele Beneventi, Luca Corvetto, Rita Vinelli and Marco Melis

Region of Sardinia (Italy), Local Authority for Municipalities, Finance and City Planning
Department for Cartography and Information Systems
{lmanigas, lcorvetto, rvinelli, melisma}@regione.sardegna.it
mbeneventi@sardegnait.it

Abstract
Since 2006, the Region of Sardinia (Italy) has been developing its Geographic Information System and the related Spatial Data Infrastructure, known as *Sistema Informativo Territoriale Regionale - Infrastruttura dei Dati Territorial* (SITR-IDT). The aim of SITR-IDT is creating and managing a spatial database and the technologic infrastructure, services and application needed for data access. This is in agreement with the principles of geographic data sharing and accessibility expressed by the INSPIRE Directive; for this purpose, metadata play a fundamental role. In SITR-IDT a high importance has been given to metadata definition, acquisition and management. Proper tools for metadata management, according to INSPIRE Implementing Rules (IR), are being developed. Both INSPIRE IR and the requirements coming from an appropriate daily spatial database administration have been considered in metadata organisation. In this article, the metadata organisation adopted in SITR-IDT for data of different typologies is illustrated. It presents the evolution of metadata organisation according to European Union Directives and technical guidelines. It also explains the switch from a quite rigid and constraining metadata schema to a more flexible and standards-compliant one. Different questions such as the metadata manager and the organisation of metadata for non-geographic data related to geographic data are discussed. The article presents a tool, the metadata manager, that should help to create, collect, and manage metadata at the appropriate levels of a Spatial Data Infrastructure.

Keywords: Geographic Information System (GIS), SITR-IDT, Sardinia, metadata, INSPIRE, ISO19115, metadata manager, non-geographic data.

1. INTRODUCTION

The most important target of SITR-IDT is to create a spatial database containing all geographic data produced by the Regional institutional agencies in Sardinia, Italy. The data quality should be granted by official validation and certification. Metadata play a fundamental role in arriving at this target.

Metadata are the data ‘passport’; they describe data features and therefore represent the data identity. When structured according to international standards, metadata allow data consultation, access and use by international stakeholders. In the definition of the metadata organisation initially chosen in SITR-IDT, the ‘IntesaGIS’ project played a fundamental role. This project, started in 1996 and nowadays promoted by the CNIPA (*Centro Nazionale per l’Informatica nella Pubblica Amministrazione*), aimed at defining and sharing operational guidelines for geographic data creation among the public administrations of Italian Regions. The main goal of this project was to make the implementation of spatial databases standardised at a national level (Intesa GIS, 2004a; Intesa GIS, 2004b).

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

Over the last few years, important in-depth analyses concerning geographic metadata were carried out, and several technical guidelines were delivered. Internationally, ISO19115 (ISO, 2003) and ISO19119 (ISO, 2005) standards, INSPIRE Implementing Rules (IR) (Drafting Team Metadata, 2007) and the European Commission Regulation 1205/2008 have been published. In Italy, operational guidelines have been delivered by the Italian national authority for geographic data (CNIPA, *Centro Nazionale per l'Informatica nella Pubblica Amministrazione*).

In 2006, CNIPA has delivered two drafts of technical guidelines (CNIPA, 2006a; CNIPA, 2006b) for creating and updating a National Register of Spatial Data (*Repertorio Nazionale dei Dati Territoriali*, RNDT). The RNDT should collect metadata of all geographic data produced by the Italian public administrations, in order to create a national catalogue. The drafts explained the metadata organisation that Public Administrations should adopt to feed the RNDT. They contained the mandatory plus some optional fields of the EN ISO19115:2005 core set of metadata. Therefore, in SITR-IDT all metadata were structured according to the ISO19115 standard.

In October 2007, the INSPIRE Draft IRs for Metadata, Version 3, were published. With their implementation in December 2008, CNIPA delivered the draft Decree of the President of the Council of Ministers 'Regulation governing the definition of the content of the National Register of Spatial Data, as well as the modalities of establishment and update of the latter'. This draft descended almost completely from the 2006 drafts. The only change was that this final version fully implemented the INSPIRE metadata IR, giving further restrictions concerning multitude and obligation and adding some further fields.

First, this article explains the geographic data organisation in IntesaGIS. Then the metadata organisation in SITR-IDT and the requirements of regional authorities for metadata are provided and a tool to manage metadata proposed. This metadata manager should help to create, collect, and manage metadata at the appropriate levels of a Spatial Data Infrastructure, including the regional level.

2. INTESAGIS GEOGRAPHIC DATA ORGANISATION

'IntesaGIS' laid down a hierarchical organisation of geographic data; according to this schema, data were classified in a pyramidal structure constituted by *Layers*, *Themes* and *Classes*. In particular, *Classes* coincided with the most atomic data level (usually corresponding to the shapefile), while *Themes* were aggregations of *Classes,* and *Layers* were aggregations of *Themes*. Initially, in SITR-IDT metadata organisation followed strictly the hierarchical data organisation of 'IntesaGIS'. The individuation of Series, Dataset and Tiles according to the schema reported in Figure 1 descended almost naturally.

However, it must be clear that the 'IntesaGIS' guidelines concerned geographic *data* structure, while no rules concerning *metadata* organisation were given.

3. CHANGES IN SITR-IDT METADATA ORGANISATION

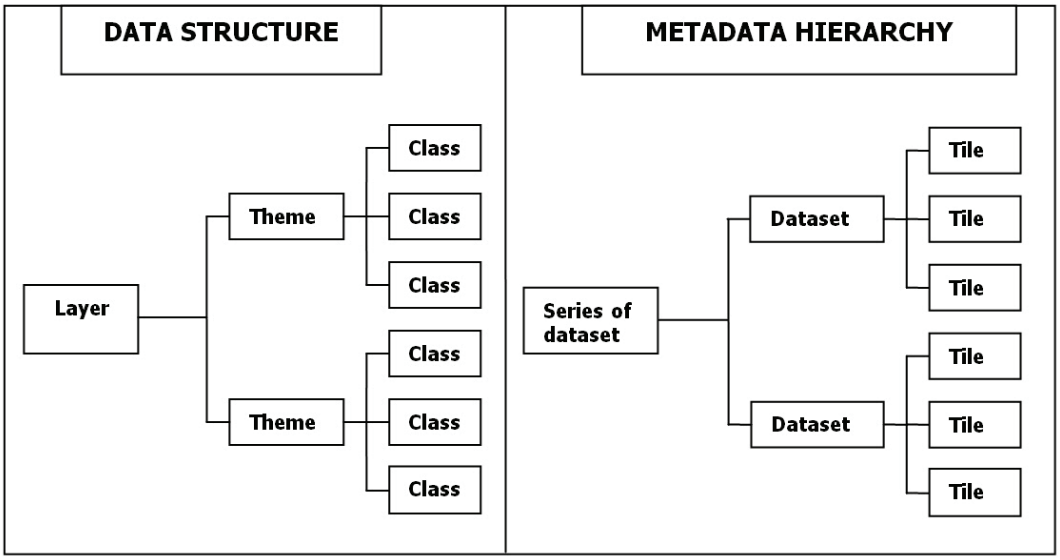
Some changes are going to be done in the near future in SITR-IDT metadata organisation, in order to fulfil the following requirements:

a) possibility of organising metadata out of the strictly hierarchical schema of 'IntesaGIS';

b) definition of a metadata organisation suitable with the requirements coming from the daily database management, and
c) compliance with CNIPA guidelines.

These questions will be better explained in subsection 3.2 (points a and b), and subsection 3.3 (point c).

Figure 1: Correspondence between data structure and metadata hierarchy in SITR-IDT, according to 'IntesaGIS' data organisation.



3.1 Current metadata organisation in SITR-IDT

In SITR-IDT, metadata of several cartographic products were successfully organised according to the hierarchical structure of 'IntesaGIS'. For instance, metadata of Sardinian Regional Technical Map were organised with the following criteria:

– the whole Regional Technical Map (as a cartographic product) constituted a *Series*;
– the single sections of the cartographic product, produced in different periods or covering different areas, coincided with *Datasets*, and
– further divisions of the sections into geographic areas where catalogued as *tiles.*

This metadata schema, following the concept of 'dataset and series' laid down by ISO19115, avoids the repetition of several metadata, which, due to the data typology, occurred several times in the whole cartographic product. For instance, different production sections of the Regional Technical Map (each of them constituting a Dataset) shared the same metadata concerning scope and use of the map, lineage, authors, points of contact, and similar information. Therefore, joining these datasets in a series avoided the repetition of those metadata shared by all of them.

Following the same structure, metadata of the Sardinian Regional Topographic Database were effectively organised. This spatial database, which contained the data of the Regional Technical Map organised in an RDBMS, was built according to the 'IntesaGIS' data specifications. For this product, the hierarchical metadata schema was easily identifiable and fitted very well with the data model, which was the following (see Figure 2):

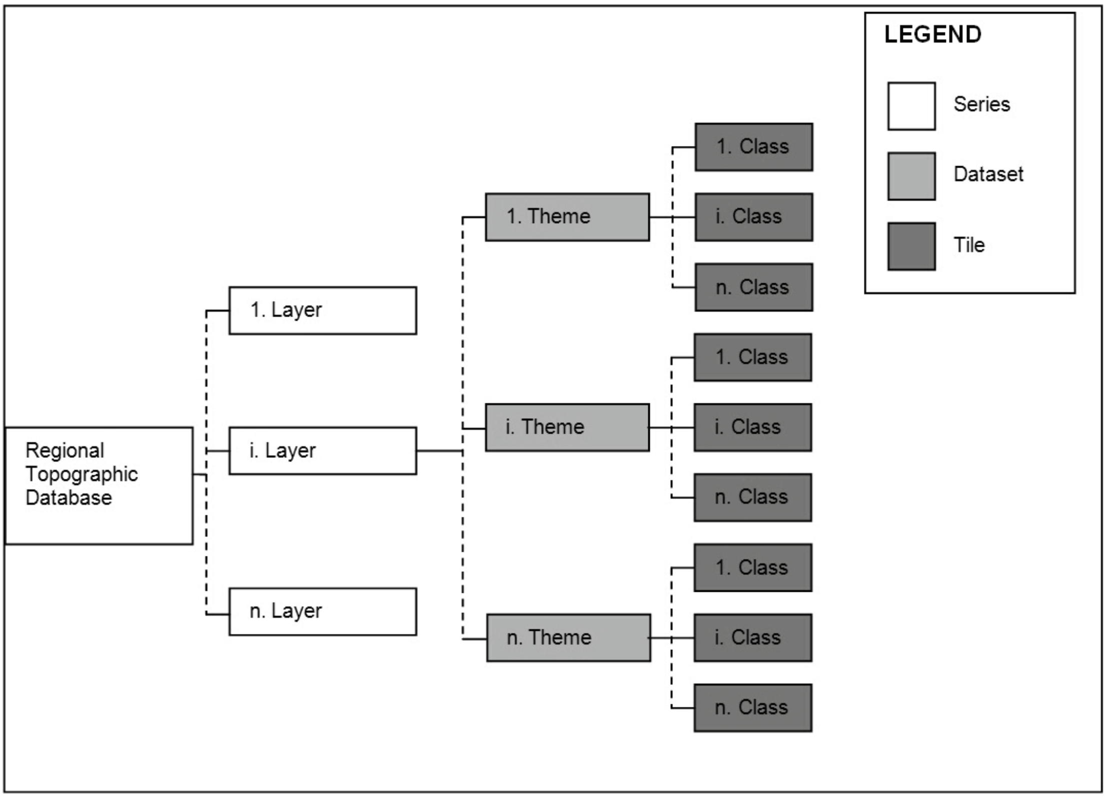
SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

– the Regional Topographic Database was considered as the whole cartographic product;
– the whole spatial database was divided into several *Layers* (such as Orography, Viability, and Administrative boundaries);
– for each *Layer*, several subsets, called *Themes*, were individuated; for instance, the *Layer* of Viability was divided into several *Themes*, each of them containing the data of streets, squares, bridges and similar objects;
– the smallest components of the cartographic product were the shapefiles which constituted each *Theme*; each shapefile was identified as a *Class*, i.e. the most atomic entity identifiable in the database.

According to this data structure, metadata were organised following this schema:

– each *Layer* constituted a Series;
– each *Theme* coincided with a Dataset;
– each *Class* was catalogued as a Tile;
– the whole cartographic product was considered as a series of series, constituted by all the *Layers*.

Figure 2: Data and metadata structures for the Regional Topographic Database.



The metadata organisation previously described showed to be effective for the categories of data quoted before, particularly for the Regional Technical Map and the Regional Topographic Database. However, different typologies of data, not organised in a hierarchical structure, can require a different metadata organisation, as explained in the next subsection.

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

3.2 Definition of a metadata schema suitable with the requirements of SITR-IDT management

If cataloguing only general metadata is the most proper choice for a register at national level, it is not suitable for cataloguing geographic data at regional level. In creating and managing a regional spatial database, data of very variegate typologies can occur, requiring specific metadata for their appropriate and complete description. Managing these metadata can be even more complex than what is required by standards.

In SITR-IDT, several non strictly geographic data are catalogued. For instance statistical, demographic, sanitary or environmental data, being geo-referenced, are part of the SITR-IDT spatial database. In comparison with strictly geographic data, different metadata are required to describe the alphanumeric information contained in these data. Also the detail level of information to be described can be different: for instance, specific metadata for non-geographic attributes, which have different lineages, are needed. In this case, a simple 'series-dataset-tile' schema is not sufficient; a higher detail is required.

For example, consider the metadata organisation adopted for cataloguing a set of statistical data concerning local job areas. The data on the local job areas consists of a shapefile, made both by the geographic data and by the attributes containing the alphanumerical information. The shapefile of all the job areas on the regional extent is catalogued as a dataset, while each single area is catalogued as a tile. Since each shapefile attribute is created by different authors, in different periods and with different lineages, each single attribute of the shapefile needs its own metadata. In this case, cataloguing metadata as series, dataset or tile is not sufficient. The value 'attribute', present in the 'MD_ScopeCode' list of the ISO19115 standard would be required to catalogue metadata at the most appropriate level.

If structuring metadata at a different hierarchical level than series, dataset or tile can be intuitive for non-geographic data, it can be necessary also when managing simple geographic data.

In SITR-IDT, this requirement occurred for instance when the Sardinia Regional Geological Map was catalogued. This geographic product was constituted by 40 distribution units, each of them covering a different area of the regional territory. Each distribution unit was constituted by four different shapefile and a *.mdb* file with the map legend. The four shapefile contained in each distribution unit were:

– shapefile of the geological surveys authors;
– shapefile of the geological surveys reviewers;
– shapefile of the polygonal geometries containing the geological information, and
– shapefile of the linear geometries containing the landslides information.

Each shapefile with the geological and landslides geometries was further divided into geographic sections, according to the local official cartography division. The shapefile of the authors and of the revisers contained the alphanumerical information in its attributes. Each attribute differed from the others for the creation period and for the lineage. With this foreword, metadata for the cartographic product of the Sardinia Regional Geological Map were organised in this way:

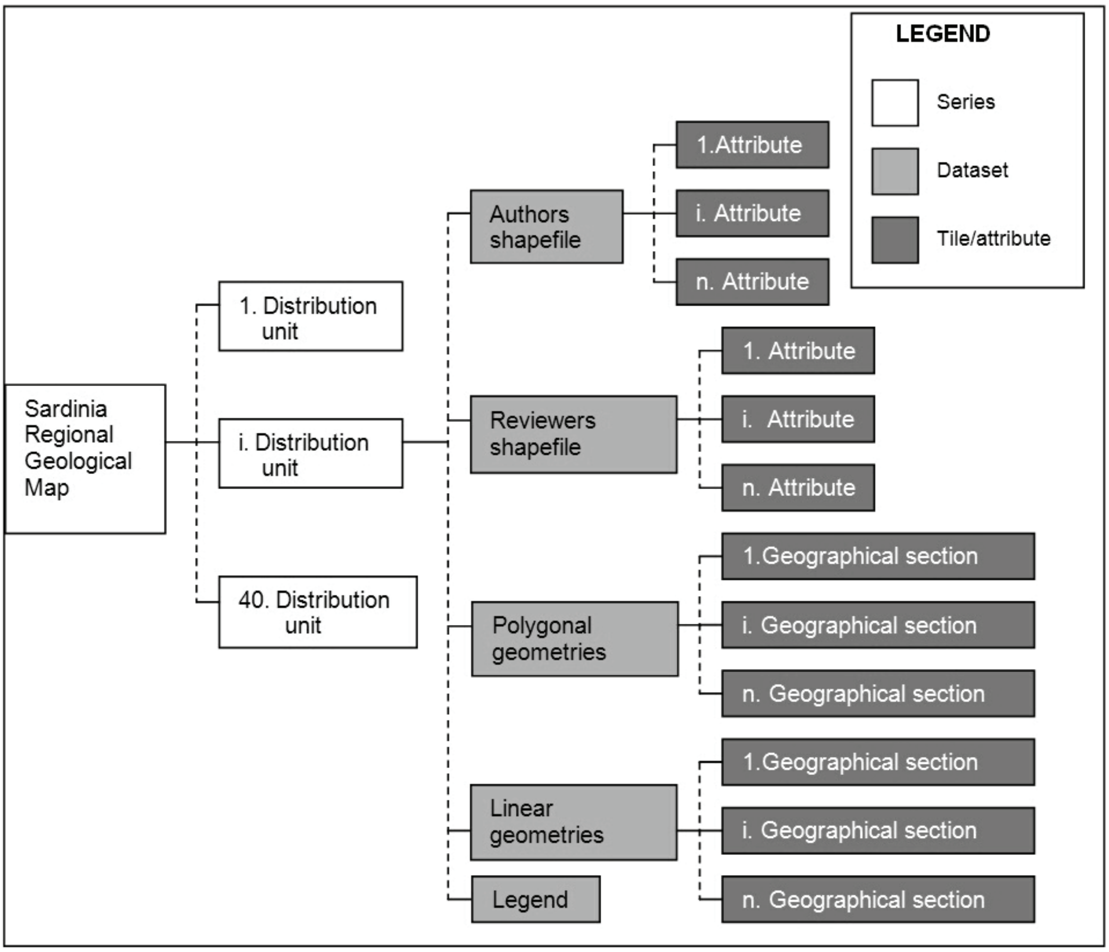
– the shapefile of authors and reviewers, for each single distribution unit, constituted two separate datasets;

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

– the attributes of each of these two shapefile should have been catalogued with the ‘attribute’ value of the ISO19115 ‘MD_ScopeCode’ list. Since, in SITR-IDT it is not possible to chose the value of ‘attribute’ as hierarchical level, the attributes were catalogued as ‘tiles’. Tiles were thus constituted by non-geometrical subsets of geometrical datasets, individuated on the basis of the differences of production periods and of lineage;
– each of the two shapefile containing the geology and landslides geometries was catalogued as a dataset;
– each of the geographic sections that constituted the shapefile was considered as a single tile;
– each distribution unit was considered as a series, and
– the whole Regional Geological Map was considered as a whole cartographic product, constituted by the mosaic of the forty distribution units (so covering the whole regional extent). It was considered as a series itself, made by the joint of all the forty series, as a sort of ‘series of series’.

In this way every components of the cartographic product were described by their own metadata. Sardinia Regional Geological Map data and metadata structures are schematised in Figure 3.

Figure 3: Data and metadata structures of the Sardinia Regional Geological Map.



SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

This is just an example of how complex a complete cataloguing of geographic data can be. Many other data present in SITR-IDT required as complex metadata structure. As above mentioned, some data require that metadata are classified into attributes, features or the other values of the ISO19115 ‘MD_ScopeCode’ list. At the moment, this is not possible in SITR-IDT, and therefore, in the next months, important changes in the metadata database conceptual model are going to be made.

3.3 Compliance with CNIPA (and INSPIRE) operational guidelines

The aim of the National Register of Spatial Data (RNDT) is to give a complete and general overview of all geographic data existing at a national level. In order to ensure its effective feeding, management and consultation, a geographic data register at a national level should contain a limited number of metadata for each single datum. A too detailed description of every data set would imply a great complexity in managing the whole national register. For this reason, the ISO19115 metadata set is not the most proper choice for the RNDT. Instead RNDT was structured according to the INSPIRE metadata requirements, plus some further fields.

Metadata in the RNDT should be organised in datasets and series of datasets, and, if needed, in subsets of datasets, i.e. in tiles. No further hierarchical levels should be individuated, because a higher detail level is out of the scope of a data catalogue at a National level.

Since CNIPA metadata structure is organised according to the INSPIRE IR, adherence to the CNIPA metadata schema ensures the compliance with INSPIRE. SITR-IDT metadata organisation does not need many important changes to reach this compliance. Since its beginning, for defining metadata fields SITR-IDT has followed CNIPA technical operational guidelines. Metadata elements are constituted by the ISO19115 core, plus some further fields of the same standard. In order to be fully compliant with the most recent version of CNIPA technical guidelines, these minor changes will be made in SITR-IDT:

– change for some existing fields that they do not have to be filled out (obligatory status), and
– introduction of some new fields, mandatory for the most recent CNIPA operational guidelines.

With these small modifications, SITR-IDT will be able to produce the *.xml* files necessary for feeding the RNDT, which will become mandatory in 2010.

4. METADATA OF NON-GEOGRAPHIC DATA

Textual documents, legends or other non-geographic data correlated with geographic data require appropriate metadata. INSPIRE IR and CNIPA operational guidelines do not give any precise indications about these data, which are, however, very significant and constitute a very important part of the geographic data.

However, ISO19115 standard recognises this kind of data, affirming that its criteria can be extended to many other typologies of geographic data such as maps, charts, textual documents and other non-geographic data (ISO19115, 2003). In SITR-IDT metadata of non-geographic data are described according to the Dublin Core set.

5. METADATA MANAGER

For metadata creation and management, a fundamental role is played by the tool adopted. In SITR-IDT this tool, which consists of dedicated software, is still being developed. It is a client desktop working on the local SITR-IDT intranet and connected via Open Database Connectivity (ODBC) with the Oracle metadata database.

Software was created at the beginning of the SITR-IDT project to access and manage the metadata database. Its conceptual model was structured according to the 'IntesaGIS' hierarchical data structure. SITR-IDT's metadata manager is tightly linked to this hierarchical metadata organisation, so that in creating metadata, its organisation in series, datasets and tiles is compulsory. Therefore, it is not possible to create a single dataset without the related series. This is clearly out of the ISO19115 conceptual model.

A new metadata manager is therefore under analysis. The new metadata manager should allow a flexible metadata organisation, and should work on a different metadata database conceptual model, according to INSPIRE and ISO19115. It should allow to catalogue metadata with all the values of the ISO19115 'MD_ScopeCode' list.

Currently, the SITR-IDT metadata manager can be accessed (both in consultation and in editing modality) only from the local SITR-IDT intranet. However, it might be helpful to extend this access in client-server mode to other Sardinian Public Administrations producing geographic data and provide these to the SITR-IDT database. It is nowadays well recognised that the most effective way to produce good quality metadata is to create it contemporarily to data production. This would be possible with a tool which allows any geographic data author, external to the SITR-IDT intranet, to access metadata manager and to create the metadata in parallel to data creation. After the data author has completed the metadata entry, the SITR-IDT metadata administrator should consult and validate the metadata. The implementation of this scenario implies to create different profiles for all the users of the metadata manager. The set up of different profiles for metadata consultation, editing, validation, publication, creation or elimination would then be necessary.

The possibility of creating metadata by several users implies that precise guidelines must be made available for all metadata creators. SITR-IDT will write down and publish among all the Regional institutional producers of geographic data, appropriate operational guidelines illustrating practical case-studies covering the largest range of data that are likely to be created at a regional level.

The new metadata manager should be able to export the *.xml* metadata files of different metadata sets, customisable depending on specific requirements (e.g. metadata set of the ISO19115 'Core', of the RNDT, or others).

The capability of connecting to different spatial databases (Oracle, PostgreSQL, HSQLDB) is being evaluated. Also compilation wizards could be implemented. This would allow the pre-compilation of some fields as the geographic bounding box for a shapefile, or as the distribution format for different typologies of file. A functionality for the versioning management of a metadata set should be also implemented. Finally the metadata manager should be so flexible to allow the implementation of all possible variations of national or international guidelines.

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

6. CONCLUSIONS

The Public Administration of Sardinia Region is making a great effort in terms of human and economic resources to create its Spatial Data Infrastructure (SITR-IDT). In the management of SITR-IDT, great importance is given to metadata. They are object of study concerning their creation, collection and management but also for more general questions such as the individuation of the most suitable database structure and of their appropriate organisation.

A good basis consisting of the individuation of the metadata structure compliant to ISO19115 standard and to INSPIRE IR has already been created in SITR-IDT. Feeding metadata to the National Register of Spatial Data will be possible without significant changes to the metadata structure adopted in SITR-IDT.

However, many further steps are still to be performed in SITR-IDT to develop the metadata database and to create the tools that allow for managing metadata of different typologies of geographic data in the most appropriate way, cataloguing the information at the most appropriate level of detail and allowing metadata management by different geographic data producers.

REFERENCES

CNIPA, Segreteria tecnica del Comitato tecnico nazionale per il coordinamento informatico dei dati territoriali (2006a). Repertorio Nazionale dei Dati Territoriali - Linee guida per l'applicazione dello Standard ISO 19115 Geographic Information - Metadata. Versione 0.2.

CNIPA, Segreteria tecnica del Comitato tecnico nazionale per il coordinamento informatico dei dati territoriali (2006b). Repertorio Nazionale dei Dati Territoriali - Linee guida per l'applicazione dello Standard ISO 19115 Geographic Information - Metadata. Versione 0.3.

Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:EN:NOT.

Draft Decree of the President of the Council of Ministers "Regulation governing the definition of the content of the National Register of Spatial Data, as well as the modalities of establishment and update of the latter", at: http://www.cnipa.gov.it/site/it-it/Attivit%C3%A0/Sistemi_Informativi_Territoriali/Specifiche_tecniche/Regolamento_Repertorio/.

Drafting Team Metadata (2007). Draft Implementing Rules for Metadata (Version 3), at: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/INSPIRE_Metadata_ImplementingRule_v3_20071026.pdf.

Intesa GIS (2004a). Specifiche per la realizzazione dei database topografici di interesse generale - Specifiche di contenuto: gli strati, i temi, le classi.

Intesa GIS (2004b). Specifiche per la realizzazione dei database topografici di interesse generale - Specifiche di contenuto: documento di riferimento.

ISO (2003). ISO 19115:2003(E); Geographic Information – Metadata.

ISO (2005). ISO 19119:2005(E); Geographic Information – Services.

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.

SDI Convergence. Research, Emerging Trends, and Critical Assessment. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors).
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009.