



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre



INSPIRE Metadata Survey Results

Joanna Nowak and Massimo Craglia

Institute for Environment and Sustainability

2006

EUR 22488 EN

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection the environment and sustainable development of the European and global environment.

European Commission
Directorate-General Joint Research Centre
Institute for Environment and Sustainability

Contact information: M. Craglia and J. Nowak
Address: Institute for Environment and Sustainability, Spatial Data Infrastructures Unit,
TP262, Via Fermi 1 , 20120 Ispra (Va) Italy.
E-mail: Massimo.craglia@jrc.it or Joanna.nowak@jrc.it

<http://ies.jrc.ec.europa.eu/>
<http://www.jrc.ec.europa.eu/>

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server
<http://europa.eu.int>

EUR 22488 EN

ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 2006

Reproduction is authorised provided the source is acknowledged

Table of Contents

Executive Summary	4
Objectives of this report.....	5
Glossary of acronyms and abbreviations used.....	5
Definitions and Methodology	6
Analysis of the Results	7
Development Phase.....	7
Metadata content standards followed	7
Standard for encoding MD	8
Organisation of the metadata holding	9
Discovery mechanism.....	9
Catalogue Service Standards.....	10
Protocols used for Search.....	11
MD Accessibility.....	11
Languages	11
Coverage.....	12
Number of datasets documented.....	12
Summary of Key Issues and Future Directions	13
Appendices.....	14
Appendix 1 - Guidelines on how to fill the template.....	14
Appendix 2 - Categorical Details.....	16

Executive Summary

This report summarizes the outcomes of a survey organized by the European Commission DG Joint Research Centre in Spring/Summer 2006 on the current availability and characteristics of online metadata for data sets. This was the first of a regular series of surveys to be undertaken in order to monitor change in practice and uptake of the INSPIRE implementing rules on metadata over time.

The metadata survey was administered to all the Legally Mandated Organizations (LMOs) and Spatial Data Interest Communities (SDICs) that have registered on the INSPIRE web site (<http://inspire.jrc.it>) to participate in the preparation of the INSPIRE implementing rules. In total, 68 SDICs and LMOs answered the survey providing information on 120 metadata holdings. The response rate for the survey was 26%.

The survey revealed that almost 80% of the metadata holdings are operational and 76% are accessible through the Internet. Some 60% of all metadata holdings allow users to both search and browse metadata.

Regarding standards followed, 65% of metadata holdings follow ISO standards (19115, 19119, 15836) to document their resources.

As far as the standards to encode metadata are concerned, the survey showed that more than half of the respondents currently follow some type of data standards, of which the majority are ISO standards (19139, 19115).

With respect to the catalogue service standards, the survey analysis shows that OGC standards are applied to 39% of all metadata holdings. Where catalogues are used, the SOAP and HTTP protocols are the most popular for metadata search.

The majority of metadata holdings are in one language, and only 11% are in more than one. The user interfaces can work in multiple languages in only 18% of cases, while the majority (66%) operates only in one language.

Objectives of this report

This report summarizes the outcomes of a survey organized by the European Commission DG Joint Research Centre. The objective of this initiative is to define a benchmark of the current availability and characteristics of online metadata for data sets held by the Spatial Data Interest Communities (SDICS) and Legally Mandated Organizations (LMOS) that have registered on the INSPIRE web site (<http://inspire.jrc.it>) to participate in the preparation of the INSPIRE implementing rules. These organisations are the forefront of the community of users, producers or transformers of spatial data and related services and therefore are expected to represent best practice in metadata creation and use. By repeating the survey on a regular basis it will be possible to monitor changes in practice and uptake of the INSPIRE implementing rules on metadata over time.

Glossary of acronyms and abbreviations used

LMO - Legally Mandated Organisations

LOI - Lack Of Information

MD - metadata

NSI - Not Sufficient Information

SDIC - Spatial Data Interest Communities

Definitions and Methodology

A Spatial Data Interest Community (SDIC) bundles expertise and resources of users, producers or transformers of spatial data and services. Therefore, SDICs typically cluster multiple organisations from both public and private sector and are organised by region, by societal sector and thematic issue. LMOs are the public authorities, institutions and bodies who have a legal mandate to set up and run one or some of the components of national and regional Spatial Data Infrastructures, and will therefore contribute to the implementation of INSPIRE in the Member States. Both SDICs and LMOs participate on a voluntary basis in the INSPIRE process, making available experts, projects and reference material. For more information, the reader is referred to the INSPIRE Work Programme Preparatory Phase 2005 – 2006, available from the INSPIRE website (<http://inspire.jrc.it>).

The metadata survey was submitted to all 266 registered LMOs and SDICs, during the period April-June 2006. A series of forms were designed as Excel sheets to collect the information from the individual member organisations of a SDIC/LMO, one for each metadata holding, together with guidelines on how to fill the forms (see Appendix 1). The questions to be asked by the survey were agreed with the INSPIRE Metadata Drafting Team to provide input to their activities, and included:

- Development Phase of the MD holding (planning, implementation, operational, or redesign)
- Standards followed to document the contents and to encode the metadata
- Organization of the MD holding (e.g. centralized database, or distributed catalogue)
- Extent of searching and browsing capabilities
- Mode of access (Internet, Intranet, off-line)
- Language(s) used to describe the metadata
- Languages of the interface
- Geographical coverage of the datasets documented in the holding
- Themes of the datasets documented (with reference to the INSPIRE Annexes)
- Temporal coverage
- Number of datasets documented in the holding

By metadata holding we mean a collection of metadata records that may refer to a dataset series or collection (for example a logical grouping of datasets sharing the same product specifications, like “Population Census 2001”), or to a system already documenting heterogeneous collections.

The nominated person for each SDIC and LMO was contacted by e-mail and asked to distribute the survey forms to each member organisation.

In total, 68 SDICs and LMOs responded to the survey providing information on 120 metadata holdings. The response rate for the survey was therefore 26%.

The completed questionnaires included free text answers which required some degree of harmonisation for the analysis. In the case of questions where the names of standards (for MD content, or MD encoding, or catalogue service) were required, the answers were harmonised according to the official names of the standards. The same applied to the question concerning the protocol used for catalogue search.

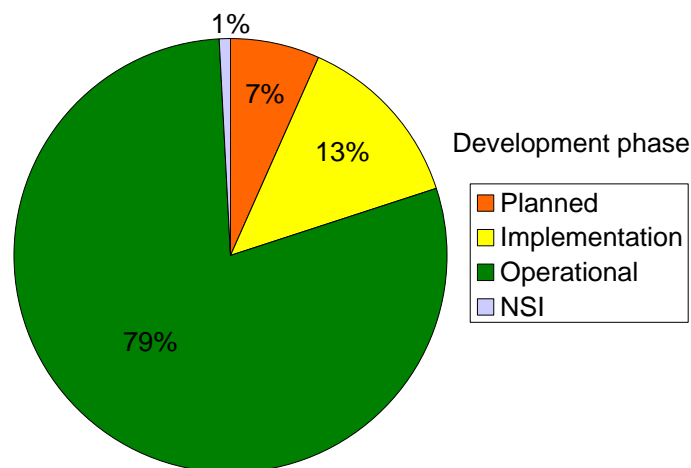
After harmonizing the data, to the extent possible, the answers were clustered and categorised and are presented in the next section through pie charts together with a legend. Some categorical details and the range of the different answers are presented in the Appendix 2.

In several instances, only a partial answer or no answer was provided to a question of the survey. For this reason, the analysis introduces two acronyms: namely: NSI = Not Sufficient Information (NSI) in case of partial answer, and LOI = Lack of Information. Void cells were considered as LOI.

Analysis of the Results

Development Phase

With regards to development phase, the survey revealed that almost 80% of the metadata holdings are in the operational stage. 13% of the metadata holdings are in implementation phase, while 7% are still planned. In 1% cases the information we received was not sufficient (NSI) to describe the metadata development phase.



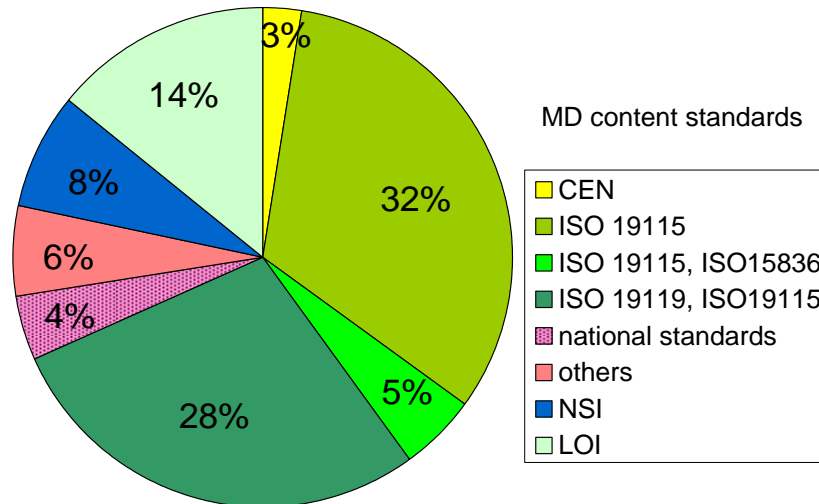
At the same time, 21% of metadata holdings were declared as being in the process of redesign, which would indicate the dynamic nature of this field, with changes in standards and technologies affecting the way in which metadata is created and made available.

Metadata content standards followed

The question regarding the standards followed to document the contents of the MD holding provided a range of different answers. The pie diagram illustrates that in total, 65% of metadata holding follow ISO standards (19115, 19119, 15836) with some internal differentiation:

- ISO 19115 (32%),
- both ISO 19115 and ISO15836-Dublin Core (5%),
- both ISO 19119 and ISO19115 (28%).

The survey also revealed that 3% of metadata holding follow the CEN¹ standards, 4% follow national standards, while 6% follow other standards (namely: AACR2, CEOS/CIP, XML/OIOXML, Content Standard for Digital Geospatial Metadata - FGDC, CSDGM, proprietary standards). In 8% of the cases the metadata holdings was indicated as using standards but the name was missing (NSI). In 14% of the cases the data was lacking (LOI).

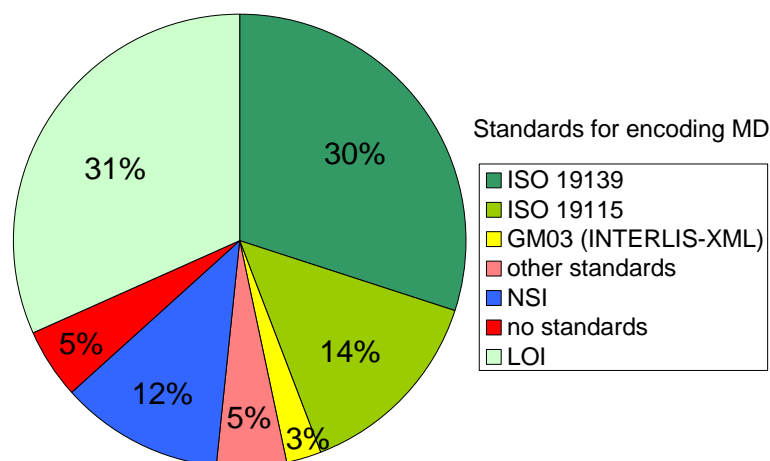


Standard for encoding MD

For the metadata encoding purposes the following standards are being applied: ISO 19139 (30%), ISO 19115² (14%), and GM03-INTERLIS (3%). The 'other standards' class (5%) encompasses the answers as follows: UKMARC, CEOS/CIP, ISO 9001, DIF, MS 1250 (see Appendix 2, Table 1).

5% of respondents use HTML which was considered to be a 'no standards' answer. In a great number of instances (31%) we received no information about standards which can be assumed as lack of standards but we classified such responses as a LOI.

To sum up, more than half of the respondents follow some standards to encode MD, of which the majority are ISO standards (19139, 19115).



¹ From the responses it is not clear whether this the old CEN 'pre-standard' or it is the CEN version of ISO standards.

² This answer may refer to the encoding used prior to ISO 19139.

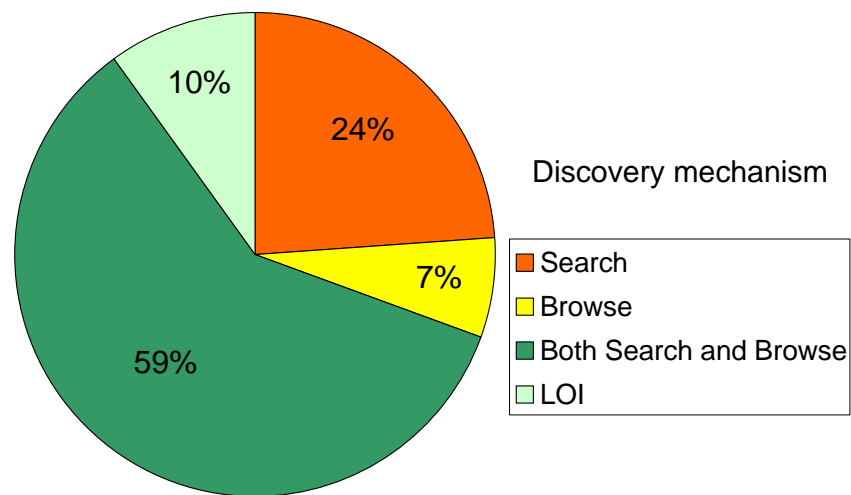
Organisation of the metadata holding

The most significant information is that metadata is searchable through catalogue almost in 60% of cases. Three-quarters of metadata holdings are in the form of internal databases.

As for other types of organisation we received 16 non blank answers, e.g. PDF, J2EE server, NetCDF, Excel file, html-files, asp, internet application, text files, word document, etc.

Discovery mechanism

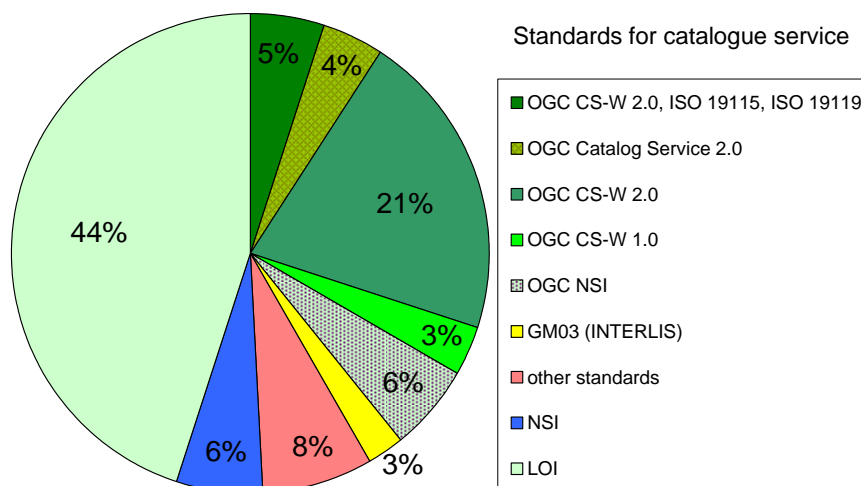
The most noteworthy finding is that 59% of all metadata holdings implement mechanisms for both searching and browsing metadata. In 24% of the instances the search mechanism exists, while in 7% - browse mechanism can be used. The other answers (voids) were classified as LOI (10%).



Catalogue Service Standards

The responses concerning catalogue service standards required harmonization due to the complex nomenclature of the OGC Catalogue Service. Thus, we created the classes as listed in the legend below (for details see Appendix 2). The analysis shows that OGC standards are applied to 39% of all metadata holdings. This number includes also 6% of the instances when metadata holdings did not sufficiently describe the standard's name (OGC NSI), and 3% of OGC CS-W 1.0. The GM03 (INTERLIS) is applied to 3% of metadata holdings. The category 'other standards' (8%) encompasses the standards not mentioned before, mainly national or proprietary ones (i.e. FGDC, OLIB, CEOS/CIP, ISO 19139-DE-Profile, NOKIS-Schema, CERA / html GUI, DIF, DC).

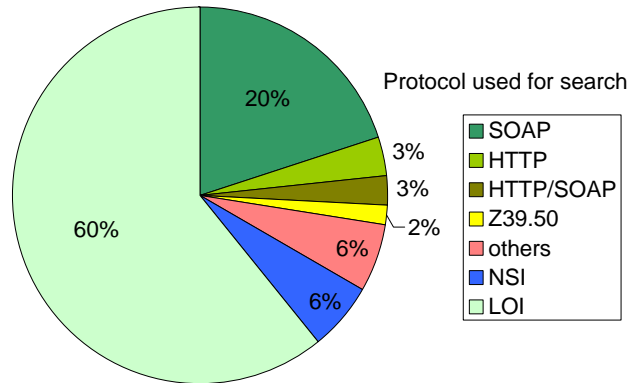
6% of the forms included information that was not sufficient (NSI), while voids appeared in 44% of instances.



Protocols used for Search

In case of catalogue existence the following protocols are used for search: SOAP (20%), HTTP (3%), HTTP/SOAP (3%), Z39.50 (2%), others protocols (6%).

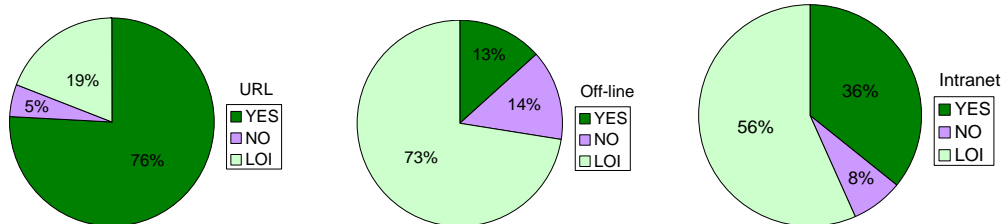
In 60% of all responses the protocol used for search was not specified (LOI)



MD Accessibility

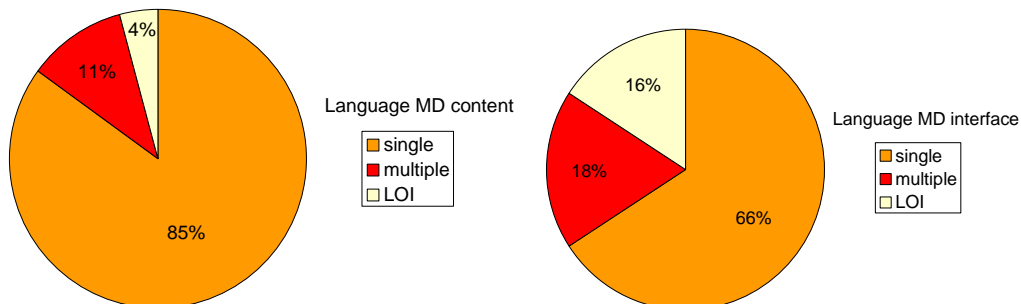
The survey revealed that three-quarters of the MD holding are accessible on-line. At the same time 36% are accessible via intranet, and 13% are accessible off-line (e.g. CD).

There is obvious overlap as regards to internet and intranet availability. That is because multiple answers were possible.



Languages

With regards to the languages of the metadata content, the majority of metadata holdings are in one language, and only 11% - in more than one. Moreover, the user interfaces operate in multiple languages in 18% of cases, while the majority (66%) operates only in one language.



Coverage

The geographical coverage of the datasets documented in the holdings is mostly regional (41%) or national (36%), while 16% have global coverage. Only 2% of the metadata holdings described has pan-European (or cross-border) coverage.

As for thematic coverage, the environmental type of data was listed the most often.

The question concerning the temporal coverage provided a range of answers that were difficult to categorise. The existing sets mostly hold data from the last 5 years but there are some metadata sets that cover historical data as well.

Number of datasets documented

The issue concerning the number of datasets documented in a holding appeared to be confusing to the respondents, who in several instances must have understood “number of datasets” as number of records in the dataset. Thus the responses ranged between 1 and 400,000. Future surveys will have to articulate this more clearly.

Summary of Key Issues and Future Directions

The survey aimed at defining a benchmark of the current availability and characteristics of online metadata for data sets held by the INSPIRE Spatial Data Interest Communities (SDICS) and Legally Mandated Organizations (LMOS). These organisations are the forefront of the community of users, producers or transformers of spatial data and related services and therefore are expected to represent best practice in metadata creation and use.

The survey revealed that:

- the vast majority (80%) of metadata holdings are operational and mostly (76%) available through the Internet while some 20% of metadata holdings are currently being redesigned indicating the state of transition at the current time.
- 65% of metadata holdings follow ISO standards and more than half of the respondents follow some standards to encode MD, of which the majority are again ISO standards.
- OGC Catalogue Service standards are applied to 39% of all metadata holdings and in the in case of catalogue existence the SOAP and/or HTTP protocols are used for search the most often (26%).
- 11% of metadata holders are already created in more than one language, and moreover, the user interfaces operate in multiple languages in 18% of all cases.

These findings indicate that organizations understand the necessity of providing and organizing metadata in a way that is compliant with international standards, accessible through the Internet and increasingly made available in more than one language.

The survey results will provide an input to INSPIRE Metadata and Network Services Drafting Teams.

Further surveys are planned after the approval of the INSPIRE Directive to monitor the evolution of metadata availability over time. A separate survey for services will be undertaken at a later stage if needed.

Appendices

Appendix 1 - Guidelines on how to fill the template

A series of forms have been designed as Excel sheets to collect the information. The forms should be completed by the individual member organisations of a SDIC/LMO

The first form includes General Information about the organization filling in the information, such as:

- Name of the SDIC or LMO to which the organization belongs
- Name of the Organisation
- Do you have metadata holding(s) describing your datasets?
- Willing to collaborate in with the JRC in testing the searching and discovery facilities of your metadata (which in the forms is sometimes abbreviated to MD)?

Then we have a series of forms (Excel sheets) which should be completed one for each metadata holding. In the template provided there are 2 identical sheets for separate holdings. In case of more holdings, please create new sheets by cutting and pasting the BEFORE you start filling them.

What do we mean by Metadata (MD) holding? We mean a collection of metadata records that may refer to a dataset series or collection (for example a logical grouping of datasets sharing the same product specifications, like “Population Census 2001”), or to a system already documenting heterogeneous collections. For example, EUROMAPFINDER is an initiative by Eurogeographics to provide documentation and discovery facilities to the datasets of its member National Mapping Agencies, while GEIXS is a similar initiative by the geological surveys members of EuroGeoSurveys. We would consider both GEIXS and EUROMAPFINDER a metadata holding for the purpose of this survey.

So for each metadata holding, please fill in an Excel sheet (navigate using the tabs on the bottom left of the workbook) with the following information:

- Development Phase (whether in planning, implementation, operations, or redesign)
- Standards followed to document the contents of the MD holding (e.g. ISO 19115) and to encode the metadata (for example ISO 19119)
- Organization of the MD holding (is it a centralized database, is it a distributed catalogue able to search different databases containing the metadata, is it a digital document organized as text or other means, and so on)
- How do users of your MD holding interact with it: are they presented with a searching interface only, or can they browse through the metadata (or both)?
- Is the MD holding accessible on-line via Internet? Is it only internal to the organization or via (secure) intranet (not open to the public), or is it off-line (e.g. in CD)?
- In what language(s) is the metadata? (e.g. English, French, German, etc.)
- Does the interface to the user have multiple languages?
- What is the geographical coverage of the datasets documented in the holding (e.g. Catalunya, Italy, EU25, EU25 + Turkey, Bulgaria, Romania, etc.)

- What themes do the datasets documented in the holding cover (for guidelines, please refer to the INSPIRE themes as in Annex 1-2-3)
- What time interval do the datasets cover (for example 1991, 2001, or from 1950 onwards, etc.)
- How many datasets are documented in the holding?

Thank you for your help.

Appendix 2 - Categorical Details

The input data were subjects of clustering and categorising.

In some cases too many categories were created whereas some of them could be aggregated. However, the intention of the authors was to present as much unchanged answers as it was possible. Some more complex categorical details are described below.

The table 1 presents categorical details for standards for MD content and for MD encoding. Our intention was to show how we categorised 'others' or 'national standards' based on original responses in the context of MD standards followed (column 1,2), and 'others' in the context of the standards for encoding MD (column 3,4).

TABLE 1. Categorical details for MD content and MD encoding.

MD content standard(s) followed		Standard for encoding MD	
ORIGINAL ANSWER	OUR CLASS	ORIGINAL ANSWER	OUR CLASS
proprietary standard	others		LOI
National Geographic Data Framework (NGDF), now moving to UK GEMINI which is the UK profile of ISO 19115	national standards	XML	XML
AACR2	others	UKMARC	others
CEOS/CIP	others	CEOS/CIP	others
XML / OIOXML	others		
Content Standard for Digital Geospatial Metadata (FGDC)	others	ISO 19115 (GI - Metadata)	ISO 19115
ISO 9001	NSI	ISO 9001	others
ISO 9001	NSI	ISO 9001	others
CSDGM, ISO	others	ISO, DIF	others
The old national recommendation on GI metadata	national standards	Encoded as html-txt	no standards
Standaard Boor Beschrijving NL	national standards		LOI
Internal format	national standards		LOI
conversion to ISO planned	national standards		
CEN TC 287	CEN	MS 1250	others
ARINC, AIXM,	others		LOI

The table below (Tab.2) presents categorical details for standards for catalogue service and for the protocol used for search. In particular we focused on the 'other standards' category based on original responses in the context of catalogue service standards (column 1,2), and 'others' and 'Z39.50' in the context of the protocol (column 3,4).

TABLE 2. Categorical details for catalogue service standards and search protocol.

if catalogue service, standards used		If catalogue, protocol used for Search	
ORIGINAL ANSWER	OUR CLASS	ORIGINAL ANSWER	OUR CLASS
proprietary standard	other standards	SOAP	
FGDC	other standards	Z39.50	Z39.50
OLIB	other standards	OLIB	others
CEOS/CIP	other standards	CEOS/CIP	others
ISO 19139, NOKIS-Schema (xsd)	other standards	OGC CS-W 2.0	
		jsp	others
GM03 (INTERLIS-XML)	GM3 (INTERLIS)	jsp	others
OGC CS-W 2.0	OGC CS-W 2.0	Z39.50 – Next Generation Search/Retrieve – Web Service (SRW)	Z39.50
CERA / html GUI	other standards	partly OAI	others
Catalog Gateway Protocol of COGIS, the GIS Coordination Group of Swiss federal administration, http://www.cosig.ch/	GM3 (INTERLIS)	Catalog Gateway Protocol of COGIS, the GIS Coordination Group of Swiss federal administration, http://www.cosig.ch/	others
(ISO19115, DIF, DC)	other standards	OAI-PMH	others

European Commission

EUR 22488 EN – DG Joint Research Centre, Institute for the Environment and Sustainability

Title: Inspire Metadata Survey Results

Authors: Joanna Nowak and Massimo Craglia

Luxembourg: Office for Official Publications of the European Communities

2006 – 17 pp. – 21 x 29.7 cm

EUR - Scientific and Technical Research series; ISSN 1018-5593

Abstract

This report summarizes the outcomes of a survey organized by the European Commission DG Joint Research Centre in Spring/Summer 2006 on the current availability and characteristics of online metadata for data sets. This was the first of a regular series of surveys to be undertaken in order to monitor change in practice and uptake of the INSPIRE implementing rules on metadata over time. The metadata survey was administered to all the Legally Mandated Organizations (LMOs) and Spatial Data Interest Communities (SDICs) that have registered on the INSPIRE web site (<http://inspire.jrc.it>) to participate in the preparation of the INSPIRE implementing rules. In total, 68 SDICs and LMOs answered the survey providing information on 120 metadata holdings. The response rate for the survey was 26%. The survey revealed that almost 80% of the metadata holdings are operational and 76% are accessible through the Internet. Some 60% of all metadata holdings allow users to both search and browse metadata. Regarding standards followed, 65% of metadata holdings follow ISO standards (19115, 19119, 15836) to document their resources. As far as the standards to encode metadata are concerned, the survey showed that more than half of the respondents currently follow some type of data standards, of which the majority are ISO standards (19139, 19115). With respect to the catalogue service standards, the survey analysis shows that OGC standards are applied to 39% of all metadata holdings. In case of catalogue existence, the SOAP and HTTP protocols are most popular for metadata search. The majority of metadata holdings are in one language, and only 11% are in more than one. The user interfaces can work in multiple languages in only 18% of cases, while the majority (66%) operates only in one language.



EUROPEAN COMMISSION
DIRECTION GENERALE
Joint Research Centre

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.



Publications Office
Publications.eu.int