Metadata: The First Step to Democratization Access to Public Sector Geographical Information

Math. Angela IONITA Ph.D.
National Institute for Research and Development in Informatics - ICI Bucharest

Abstract: *There is a consensus within the Geographical Information (GI) Community in the metadata and data capture and management process and of technological development - in particular those concerning intelligent tools, intelligent search engines - that metadata content and services are really necessary to increase the visibility of the existing content and to facilitate access by users to geographical information.*
In our paper dedicated to "MEDIAGIS Project" presented in 7th EC GI&GIS Workshop, we discussed about On Line Service for Geographical Information focused on a general architecture. In this context we mentioned the four levels of complexity of this service: metainformation, transactional, integration, new opportunities. The development of Metadata is very important for all levels presented here.
This paper is not an original paper in the real sense of word but try to present a synthetically point of view regarding the metadata and their development in the framework of On Line Service in the context of Virtual Communities specific for Information Society. The fourth sections try to answer what is metadata, why metadata, how to create metadata and what kind of information resources it will be necessary and/or recommendable to use in the framework of metadata services according to users requirements and why use standards and make a short overview of geospatial standards and initiatives. and what tools are available to do this. The last section is dedicated to the conclusions and to the situation at European level and try to answer.
Keywords: *data access, data dissemination, metadata culture, metainformation, Geographical Information (GI), Geographical Information System (GIS), clearinghouse, education, social networks, level of users, coverage of data and quality of metadata, updating, searching and navigation facilities, best practices, metadata service providers, geospatial standards.*

1 Introduction

In our paper dedicated to "MEDIAGIS Project" (Ionita, A., Pribeanu, C. (2001)), presented in 7th EC GI&GIS Workshop, we discussed about On Line Service for Geographical Information focused on a general architecture (figure 1, 2). In this context we mentioned the four levels of complexity of this service (figure 3): *metainformation, transactional, integration, new opportunities.*

Metainformation: an On Line Service for GI which does no more than provide information across the Internet. This can be done via web pages or simply through attaching files to be read. The key thing is that it is basically web presence and publishing, and no more. We see many such sites in government - from advice on completing tax returns, to lists of government contacts, and so on.

Transactional. This is characterized by systems which provide a degree of transaction or application capability. The transaction involves the exchange of value - the value might be financial, such as paying for a license renewal, but it might be data .

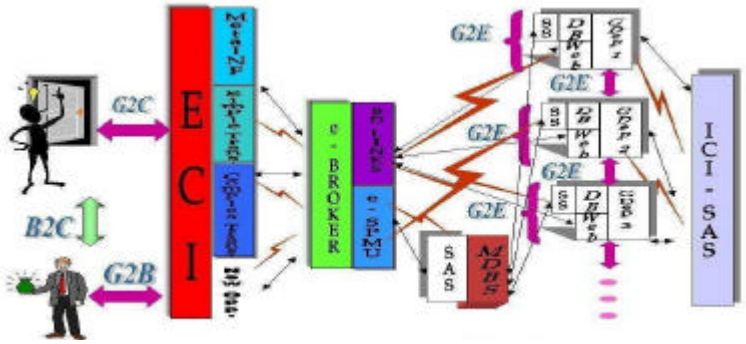


Fig.1. The On Line Service for GI architecture (adapted from R. Ilie, A. Ionita, *"E-Government Toolkit for Developing Nations"*, WB proposal, May, 2001)

where:

ECI	"Every Citizen" Interface
MetaINF	Simple information content request (access to .html like pages)
SIMPLE TRANZ.	tranzactions with a single governmental agency/department
COMPLEX TRANZ	tranzactions with multiple governmental agencies
BD LINKS	DataBase with "Links" to the governmental web sites
e-SPMU	electronic service "provider" management unit
SAS	Secure and Authentication System
MDBS	Meta Data Bases Server
SS	Security System
DB	Data Base
Web	governmental institution web site
ICI	Information Core Interface
e-BROKER	electronic "balance" between citizen's demand and governmental offer

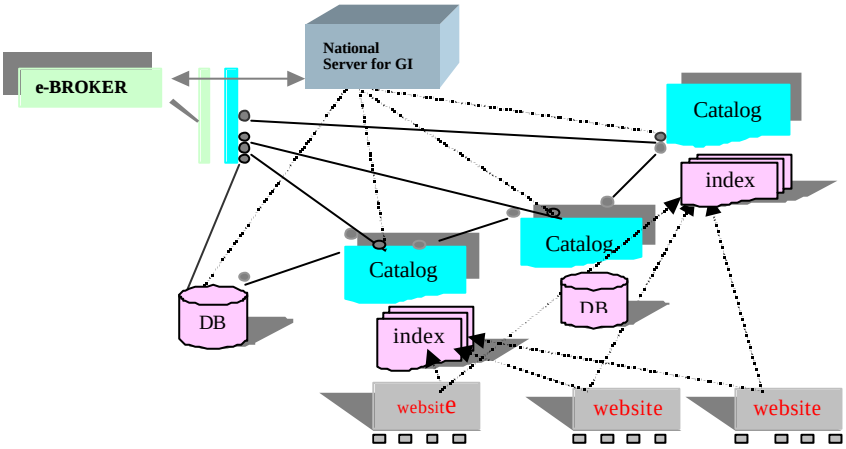
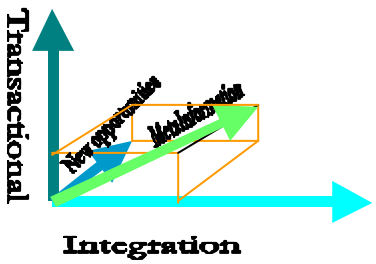


Fig.2. The Clearinghouse in the framework of MEDIAGIS Project (adapted from Raper, Jonathan: *Information and Interoperability*, Workshop on Virtual Interoperability Laboratory, Ispra, Italy; February 1999, **URL:** http://ams.emap.sai.jrc.it/dg3gis/vil/**)**



Integration: where the On Line Service for GI is used to deliver information and services across multiple local and central government departments. This is the most complex approach, but paradoxically it is the one about which most visionary leaders speak. This is what people really mean when they talk about the true "one stop shop". It is where e-business adds the most value for government.

It is the easiest to understand for the users, and the most difficult to deliver. Integrating user access to the multiple functions of

government, from the user's perspective, requires significant commitment and reengineering of process and data across multiple public organizations. It is no longer a question of just putting together a good-looking website.

New opportunities. The generation of data from new sources is an on going development.

Application developers have attempted to research and implement new data source into their work. Most of these new data sources are based strictly on scientific and technological developments. Finally, progress could be measured through the new opportunities that it creates. This is more consistent with contemporary notions of performance measurement

through *outcomes*. The most compelling case for true progress would be made, for example, by being able to point to new scientific insights that resulted from interoperability, or new collaborative policies and plans, or new organizational forms.

The development of Metadata is very important for all levels presented here. This paper is not an original paper but try to present a synthetically point of view regarding the metadata and their development in the framework of On Line Service in the context of Virtual Communities specific for Information Society. The next sections try to answer what is metadata, why metadata, how to create metadata and what kind of information resources it will be necessary and/or recommendable to use in the framework of metadata services according to users requirements and what tools are available to do this. The last section is dedicated to the European context and try to answer why use standards and make a short overview of geospatial standards and initiatives.

1.1. What is metadata?

Metadata is the information and documentation, which makes data understandable and shareable for users over time (ISO 11179 Annex B). We can distinguish different levels of Metadata of increasing detail: *Metadata for Inventory* (i.e. internal to an organisation), *Metadata for Discovery* (i.e. that necessary for external users to know who has what data, where to find it, and how to access it), and *Metadata for Use* (i.e. a fuller description of an information resource that enables users to make a judgment about the relevance and fitness-for-purpose of the resource before access it).

In this paper the term refers to geospatial metadata, meaning information about geospatial data such as images of the earth, maps and the geographical features that maps represent (http://geonsdi.er.usgs.gov) Geospatial metadata are structured in a consistent way so that once you get used to the format, like cards in a library catalog, it becomes easy to find what you need. People who produce geospatial data have to be concerned about how to document the data resulting from their studies. But people who manage a collection of metadata have a number of different concerns that often are not discussed in metadata-training materials, and extend beyond the relatively simple matter of making a collection accessible through the Clearinghouse. These resources are intended to assist those who see the larger issues surrounding geospatial metadata collections.

1.2. What is Clearinghouse?

Figure 2 present a *clearinghouse* in the framework of On Line Service. In this context the Clearinghouse is a distributed catalog of metadata where *distributed* means the information is kept on many different machines instead of gathered into one big database and *catalog* means it can search it for interested in things, and when it find something that looks interesting it can read more about it, and will learn what it's about, where it came from, who made it (and why), and where to get it (according to Peter Scweitzer, http://geology.usgs.gov/tools/metadta/tools/doc/opinion/clearinghouse.html, Feb. 2001). *Metadata* are documentation of geospatial data written in a consistent way.

1.3. Why metadata?

Because many different people can use geographic information in a variety of ways, it's important to know more details of the data's origin, history, and characteristics than you would need for a library book. So metadata tend to be long descriptions of the data. But it's hard to read long descriptions, especially if the descriptions made by one organization look radically different from those produced by another. Most geospatial data have a lot in common; using standard structures and formats for documenting them makes it easier for people to find the characteristics of the data that will help them understand it quickly. The same structures and formats help us write software to search for key characteristics of the data and to present the documentation in a consistent manner. The Clearinghouse is a way for abilitated organizations to make it easier for people to find, get, and use their information effectively.

The biggest change the scientific community has undergone in the last twenty years is the improvement in the public's ability to use technology to see what we're doing and why. But that technological capability does not bring understanding by magic. People still need to figure out what they have, what they need, and how any data that are offered to them might help meet those needs. The most dramatic reversal of recent time is that in order to work well, regular people (including regular scientists) have to become data managers. They need to use some systems for keeping track of the data they have and the data they need, in the context of the problems they want to solve. We cannot expect people to use our information just because it's available, or because it's free, or because we work for the government.

Regarding the significance and benefits of metadata the following table shows that there are might be least two perspectives on the usage of metadata, that of the user and that of the producer.

User Perspective	Data Producer Perspective
Discovery if data exists	Document what you have with minimum effort
Identify source	Test the value of the data to others
Make a judgement about data usability BEFORE getting it	Bridge the gap between data owners and users
Minimise costs (time, money, staff, resources, hassle) in search, retrieve, the data integrate, and use the data	Educating users about the characteristics of the data

Table 1: Two perspectives of Metadata (adopted from MADAME Report 04 /PUB 1108 MADAME 2485910, Dec. 2000)

1.4. How metadata is created

In the Report of the European Project called MADAME (www.info2000madame.org) the generation and maintenance of accurate metadata is the cornerstone of information discovery. Users cannot find resources if metadata is not present. Generating and maintaining high quality, accurate and consistent metadata is an expensive task requiring a mix of human expertise and special purpose tools. The main issue is the trade-off between the quality of the metadata and its costs. The criteria for creating high quality metadata are (figure 4): Simplicity, Cost, Accuracy, Consistency, Feasibility.

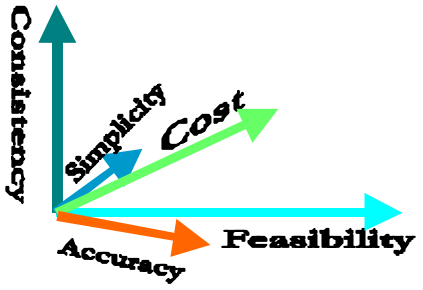


Fig.4. The criteria for creating metadata

An example of good practice is the INFO2000 CLEAR Project (http://carto.ped.org) which aims to put online a cata-

logue of metadata concerning geographical information from the Saar-Lor-Lux region. Compliant with Dublin Core and IS19115 standards, this online clearinghouse allows users on a multilingual way (French-German): to search GI datasets using keywords with their common web browser; to get information about the identified data (extend, producer, etc.); authorizes the data producers to remotely generate and to update their metadata. Software that can help creating metadata exists, but it is not necessarily easy to use or conforming to a given standard (see for example http://geology.usgs.gov/tools/metadata/tools/doc/faq.html#2.5). However, the Open GIS Consortium is working very closely with the ISO/TC211 to develop the software tools to help. Check also the web site of the FGDC (http://www.fgdc.gov) as there is already considerable operational experience that is of relevance also for European users. This will increase as the FGDC endorses the ISO 19115 standard.

2. Information resources

The definition of what is essential and what is not varies from country to country at the present time although there are indications that these differences may in future level out as more experience is shared and a degree of uniformity is deemed necessary to level the playing field across Europe. Individual public sector organisations, which are the focus of these guidelines, may therefore find themselves in significantly different contexts, which in turn affect their flexibility in what they may do.

Some countries like Portugal define three types of Public Sector Information (PSI): "*information for the citizen*" that should be free of charge and universally available, "*information for development*" normally of interest to economic or social actors or institutions, which can be charged at a symbolic rat or be free of charge, and "*value-added information*" to be charged at market prices.

Similarly in France, the Mandelkern Report (available at http://www.ec-gis.org), identifies "essential" data as those needed for the fulfillment of the democratic rights of French citizens or residents. These data should be universally available and free of charge. Other type of data, including "raw" and Value-added data can be charged for with different modalities. However, the report argues that the overall objective must to maximise the use of data, and that therefore charges should not deter usage.

In Finland, The Pricing Basis Act defines which products and services provided by the government agencies are subject to charge and which should be free of charge. The latter is the exception to the general rule and is usually implemented when individuals are accessing some of their personal information held in public records. Furthermore, the Pricing Basis Act defines which products and services are governed by public law and which are subject to market conditions. For a product or a service to be governed by public law it must meet three conditions (figure 5):

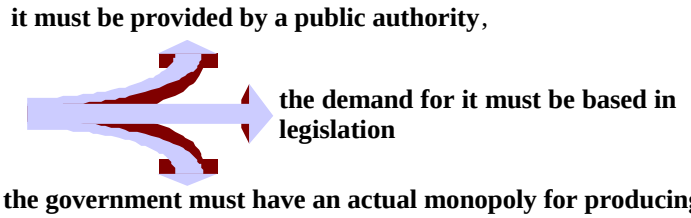


Fig.5. The three conditions for a product or service to be governed by public law

If all these three conditions are met, the price should be based in cost recovery.

A public sector organisation considering increasing access to its information resources may therefore find itself in three different positions: it operates in a context where clearly defined policies and guidelines are in place; a policy exists but the lack of guidelines leaves much room for interpretation; no policy exists.

Each case leaves more and more responsibility on the individual organisation to define its own policy.

3. Metadata Service and User Requirements

A metadata service is therefore essentially one that enables data producers to document and market their data sets, and data users to identify data sets that are relevant to their needs, and assess their fitness for purpose. This is the major, and sometimes only, function that users require from a metadata service. However, what also emerges forcefully is that this is a narrow perspective of user needs as much more is required from a service of this kind.

According to MADAME Report 4 (www.info2000-madame.org) these needs include: *education, social networks, level of users, coverage of data and quality of metadata, updating, searching and navigation facilities, best practices.*

Education is a key function required by users. Metadata is an educational tool and there is a strong demand for a metadata service to recognise and satisfy this need. Of course the educational demands vary with the characteristics of the users. Some are more introductory to novice users, others are more focused on experienced or commercial users. It may be useful nevertheless, to recall some of these demands to appreciate their range. Users of the services analysed expressed demands in the following areas:

- to know more about Geographical Information System (GIS) and remote sensing: this may require links to educational resources, both on-line and off-line, and/or the development of educational modules from introductory to advanced. There is a wealth of resources already available to support this need, including GIS dictionaries such as that provided by the AGI in the UK (http://www.agi.org.uk/pag-es/diction/dict-agi.htm), and distance-learning courses in GIS provided by vendors and academic institutions.
- to know the basics about particular data types, such as orthophotos: what are they? how can they be used? what are the limitations? Or processes: how are maps made?
- to know more about data integration: how can alphanumeric and geographical data be integrated? What are the limitations? to have examples of applications for which particular data sets were used. This would clearly involve feedback from users to document the extent to which certain data sets have been fit for the purpose or the development of chat lines to bring users together.
- to be guided through complex data environments such as the European statistical system. How does it work? How is data harmonised? How is it made available? What data is missing and why? What data exists but is not published? Where can alternative sources be found? Is it possible to find the raw data even if methodologically impure?
- to use the metadata as a research tool to analyse different pricing practices and data policy approaches taken by the data providers,
- to use the data descriptions and the practices of alternative suppliers as tools for negotiation with data providers

From the perspective of *the data providers*, the educational dimension of metadata services is very important:

- to manage user expectations
- to develop niche services, such as providing examples of applications of their data
- to develop and maintain a relationship with the users, which is very important in an increasingly competitive market.

From the perspective of the *metadata service providers*, these same needs also apply with the additional elements of signposting, to both users and providers, where they are differing and adding value to the services developed by data owners. This is crucial if they have to justify their charging. In this respect, it is important to note the extent to which users may have

the erroneous impression that data sets are described in greater or more consistent detail by the data providers than by a third-party metadata service.

Another important aspect is to help develop a *"metadata culture"* among data providers, that is help them see metadata as an important element of maintaining, developing, exploiting, and generally managing their information assets. It is striking the extent to which this metadata culture is still low even in environments with a long tradition of data management and information services.

Social Networks: this is another key area of need that emerges strongly from the analysis. Users do not just want information about data but want contacts with individuals. Again the range of needs varies with the level of experience in the field. Novice users want to become part of a social network that gives them support, and maybe status. They want to be able to talk to other people who are facing their same problems, or recognised experts in the field, or individuals and companies that can provide a service or a solution to their need. More experienced users also want direct contact with the data providers to conduct negotiations on conditions of access and use of the data, including pricing, and contact with those technically responsible for a dataset to discuss in more detail the characteristics of the data, or unpublished "raw" data that may be relevant to their needs.

Level of Use: The success and importance of a metadata service are not necessarily reflected by the frequency of use. In some instances, the services are used heavily, and benefit users in a tangible way with "real savings in working hours ". However, as shown, users who work mainly with data internal to their organisations do not use the metadata services much, while those who work all the time with the same external data sets tend to go directly to the data providers. Moreover, whilst most users use the metadata service to get an overview of what is available, they will often then deal with the data providers either because they want to negotiate directly, or because they have special conditions of access, as in the case of academia and government organisations, or because they are required by administrative procedures to deal with the data owners directly.

Coverage of Data and Quality of Metadata: Almost all the users of the more established metadata services are satisfied with the coverage of the most important data sets and with the documentation provided. In many cases, users were very impressed with what was available, in others they were even surprised, as they had not by themselves discovered all that these services could offer. As highlighted earlier many users went directly to the data providers after a brief visit to the services being analysed.

The views of the users are very much influenced by their level of experience of using metadata services, and the overall maturity of the data environment in which they operate. In other words, in those environments in which the availability of digital data is still a major issue, then users have low expectations, and are happy with any data and metadata they find. As users gain in experience, they also become more demanding. Ideally, users would like one single service answering all their needs. This is clearly very difficult given the extent of variations of these needs. There is also a conflict between the users expectations for detailed and timely metadata, and the data providers' relatively low metadata culture, and willingness to provide this service. Hence the need to manage user expectations, obtain regular feedback, and develop constructive relationships.

Updating: this is a crucial issue emerging in all the services analysed. A centralised metadata service that collects information from the data producers, develops it into a consistent format, and makes it available for dissemination is the obvious starting point for the "first generation" of services of this kind in which raising awareness among producers and users, and building

operational experience are key concerns. Maintaining the system thus developed up to date is however a major undertaking, particularly if data sets or contact details of individuals change frequently.

The process of maintaining metadata involves therefore two different tasks: the first involves updating the metadata during the data production cycle. The second involves making visible the updated metadata to potential users either directly, or through an intermediary such as a national node.

Searching and Navigation Facilities: Whilst the panels of users were generally satisfied with the searching and navigating facilities provided, they also offered detailed comments which refer specifically to each service being analysed. There are however, also some findings of a more generic nature worth highlighting:

There is a balance to be struck between using simple structures like ASCII files that are efficient and fast but may not cater for more complex or dynamic searches, with structure that cater for more sophisticated searching mechanisms, for example by geographical area, but may be slower and require more work to set up and maintain. Obvious as it may seem, the general rule of starting simple and develop as you go along seem to be still valid. By and large, users would prefer to spend more time searching but find good quality, up-to-date information, rather than arrive very quickly and user-friendly to poor quality or out-of-date information. Hence, if a choice has to be made, better to invest in quality than fancy searching tools and applets. Users would also not be averse to seeing advertising on the site if this helped generate revenue to improve the quality of metadata further. Searching by place name through a gazetteer is a facility finding a high level of favor with users. Less enthusiasm is expressed for mechanisms that require plug-ins, or for nice visualisations where users can get easily lost. The use of icons to represent landmarks or places that are likely to be known by users to help them navigate through the data might be an idea worth following up.

There may be some useful lessons to be learned from multimedia educational products for children to facilitate navigation in complex data environments.

Best practices: The experiences of the services analysed include:

- Segmenting the service for different users.
- Providing samples of the data to assess fitness for purpose, preview facilities of the data selected, and of the charges that are going to be incurred before downloading.
- Well developed audit trails of who accesses the service for what purpose. This is necessary for reporting back to the data providers, but is also helpful to develop a knowledge base of users and customers.
- Web mapping facilities that allow selection of different layers of information without the need for specialised knowledge or software. Simple interfaces to modules to increase the range of applications relevant to users, and increase the client base. This maybe even more relevant as office software suites increasingly include mapping facilities (e.g. Office 2000).
- Provide facilities, and incentives, to develop a constructive dialogue with users and data providers. If it is true that the quality of the data, and the metadata, increases with shared use, there is clearly a need to develop mechanisms to report back about inconsistencies, or mistakes. Equally, developing a range of examples for which a data set has been used, has been one of the strong requests from users. How can this be done without a feedback from the users themselves? There have to be incentives though to make this happen so that users feel a sense of ownership of the service and are rewarded for their contribution.
- Mapping the metadata elements adopted by on-line services to the element of the most common GI metadata standards, or emerging international standards

and publishing this so that is clearly visible by users and other providers.

3.1. Standardization aspects

To be truly useful, the metadata must be clearly comparable with other metadata, not only in a visual sense, but also to software that indexes, searches, and retrieves the documents over the Internet. For real value, metadata must be both parseable, meaning machine-readable, and interoperable, meaning they work with software used in the Clearinghouse (figure 6).

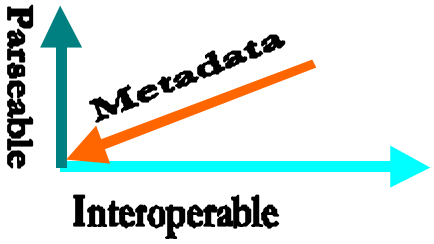


Fig.6: Real value of Metadata: parseable and interoperable

Parseable

To parse information is to analyze it by disassembling it and recognizing its components. Metadata that are parseable clearly separate the information associated with each element from that of other elements. Moreover, the element values are not only separated from one another but are clearly related to the corresponding element names, and the element names are clearly related to each other as they are in the standard.

In practice this means that metadata must be arranged in a hierarchy, just as the elements are in the standard, and they must use standard names for the elements as a way to identify the information contained in the element values.

Interoperable

To operate with software in the Clearinghouse, the metadata must be readable by that software. Generally this means that they must be parseable and must identify the elements in the manner expected by the software.

The FGDC Clearinghouse Working Group for example, has decided that metadata should be exchanged in Standard Generalized Markup Language (SGML) conforming to a Document Type Declaration (DTD) developed by USGS in concert with FGDC.

Geospatial Standards and Initiatives

In the US the main initiative is:

- **FGDC-CSDGM**: Content Standard for Digital Geospatial Metadata. Developed in 1994 by the U.S. Federal Geographic Data Committee (FGDC) as part of the National Spatial Data Infrastructure. It is a full metadata standard with 219 fields to describe a resource. It is mandatory for federal agencies (http://www.fgdc.gov/ metadata/contstan.html).

In Europe, an initiative on this front was led by CEN, the European Committee on Standardisation

- **CEN/TC 287 Env 12657**: it is the European equivalent to CSDGM. It is a voluntary pre-standard (i.e. not enforceable) developed in 1997-98 for the European Committee for Standardisation (CEN). It is again a full metadata standard, which has provided the basis for many European initiatives. A fuller range of initiatives is discussed in (Craglia and Evmerfopoulou, 2001). The point though is that in the absence of an agreed international standard, many different approaches have been taken thus far. This is not necessarily all bad news. The most important aspect is that people and organisations have started to document their information resources. That they have used different ways of doing it has drawbacks but less than no documentation at all. More recently we have seen the emergence of two important developments. The first, is the convergence of expert opinion towards an agreed international standard:

- **ISO/TC 211**: the International Standardisation Organisation (ISO) is developing in its Technical Committee 211 a family of standards related to geo-spatial information (http://www.statkart.no/isotc211/pow.htm), including one for metadata, IS 19115. Work on this standard has been going on for five years and is in its final stages. Version 3 of the Draft standard was

published in June 2000, and comments are expected by the 1st September. A final draft will be prepared in November and the International Standard is expected to be issued in July 2001. Work on this standard has brought together the experiences of the FGDC, the CEN Technical Committee 287, which had developed a pre-standard on GI metadata in 1997, and similar activities that have taken place in Australia-New Zealand and Canada. All together 33 countries and 12 observer organisations are participating in the development of IS 19115. It is expected that those organisations that had already developed their standard on metadata will make the modifications necessary to become a recognized profile of the ISO standard. So for example Version 3 of the FGDC standard will become a profile of IS 19115. Similarly, it is expected that CEN will endorse this standard, thus making it a point of reference for all European organisations. The second important development is the recognition that different organisations have different resources, and level of experience in documenting their resources. Therefore, using very lengthy and complex standards may result in incomplete or outdated information very quickly, ultimately defeating the whole purpose of documentation. Hence, the more recent focus in the international community on simple “discovery” metadata, the purpose of which is to enable potential users to find who has what information while at the same time enabling producers and intermediaries to sustain the effort. An important initiative in this respect is:

- **Dublin Core (DC)**: this is not a standard but an international initiative coming from the library and publishing communities (i.e. not explicitly geographic). It is based on consensus and is entirely open. The Dublin Core is only for discovery purposes, but is increasingly being looked at by a wide range of industries. (http://purl.org/DC/). The Dublin Core has also been recently endorsed at the European level by CEN through a Workshop Agreement, i.e. a voluntary mechanism involving a wide range of partners from different sectors. (http://www.cenorm.be/news/press_notices/metadata.htm). The DC does not replace sector-specific standards such as that of ISO 19115. The relationship is therefore one of complementarity between sector-specific standards and DC which helps discover information resources across disciplinary or sectoral domains

3.2. What tools are available to check the structure of metadata?

An 'Intelligent' metadata tool is one that automatically extracts some metadata from data files that it supports. Under Metadata Storage Structure, 'Discrete' means that each document exists as a standalone (probably ASCII) document. 'Database' means at least some of the metadata is stored in a database so sections which are common to several documents are not redundantly stored. Any metadata which accompanies these tools as sample output should not be taken as 'example' metadata, i.e. they are samples of tool output. On Internet we discovery two dedicated pages containing: "Historical Metadata Tools'" (http://www.state.wi.us/agencies/sco.metatool/mtoolhis.htm) and "Metadata Tools for Geospatial Data" (http://www.state.wi.us/agencies/sco.metatool/mtools.htm) both maintained by Hugh Phillips, Madison, WI)

Regarding the tools who check the accuracy of metadata: no tool available. Moreover, no tool can determine whether the metadata properly include elements designated by the Standard to be mandatory if applicable. Consequently, human review is required. But human review should be simpler in those cases where the metadata are known to have the correct syntactical structure.

Another problem: tools cannot be said to conform to the Standard. Only metadata records can be said to conform or not. A tool that claimed to conform to the Standard would have to be incapable of pro-

ducing output that did not conform. Such a tool would have to anticipate all possible data sets.

4. Conclusions

Any public sector organization that planning to use Internet to increase access to their information resources must start by creating their own metadata services. It is useful to consider the nature of these services from a user as well as a producer perspective from the outset.

Central and local authorities and any other public sector organizations intending to make their information resources available via Internet must recognize that there is no an unique receipt of providing access to their information resources on the Internet. In fact their Internet strategy will reflect the specific organizational requirements, the needs for potential users, the resourcing and pricing strategy that they decide to follow and how to avoid the data development of data products and data services that bring them into direct competition with private sector companies.

As Blakemore and McKeever state, Central and local authorities will take care about the advantages in charging extend beyond mere economic return, which is a point made by the current Director General of the British Library, "*Western experience suggests that often information is more highly valued and appreciated if it is paid for; the willingness to pay is dependent on a culture which understands the value of information resources, and this understanding takes a considerable time to develop." Indeed, free data can arouse scepticism, as one academic librarian has highlighted, "services may be offered because they are available, not because they are needed*". On the other hand, as recognized by Mandelkern Report, the predominant cultural paradigm of the Internet is one of "free" access to information leading to natural tensions between different cultures and traditions.

The issue of charging for data products and services should therefore not be based on dogma, a sort of war of religion between believers in “free” data and non-believers, but a pragmatic and reasoned argument based on the objectives to be reached and the means to achieve them

Besides the institutional and organizational issues providing on line access to Public Sector Geographical Information requires a technical infrastructure including:

- Internet
- structured databases
- data catalogue services.

There are several styles of spatial metadata collection and management:

- Personal collection (on PC) forwarded to server
- Metadata centrally collected and managed
- Metadata as integrated with data within GIS (database)

The collection of metadata is still primarily seen as a task external to the development of the basic data or the regular activities of a GIS. Consequently, metadata is usually acquired from the GIS software system and user information or is performed as a task completely outside the GIS. Although this may parallel the role that a "*catalog*" or abstracting session plays in the library world, without rigorous rules of classification and focus of approach, the metadata are collected with different levels of skill, detail, and intent. This is perhaps unavoidable and some form of collaboration between librarian and the GIS professional is suggested to help get the most representative information possible.

With the help of GIS vendors, metadata can someday be managed as part of the spatial data base, alongside the spatial features and their attributes. Until it is fully integrated with other aspects of GIS, the development of consistent, basic metadata will never be widely practiced.

The tools available for the collection and management of metadata fall into all three categories, with most tools being developed for standalone PCs to permit the author to input and process metadata entries. Entries from PC metadata editors are usu-

ally sent to central servers for search or straight Web service. Entries developed within a GIS are usually used to compile metadata to accompany a data transfer to populate catalogs. Web-based centralized metadata services are typically designed to collect minimal metadata for general purpose use but cannot support the documentation of "live" data sets. Such centralized systems are particularly well suited to describe legacy data sets with little known documentation. There are exceptions to all these general statements, and one must evaluate the requirements and the benefits of each before selecting a system.

Metadata can be managed for data sets in a number of ways depending on whether

- the metadata is dynamic or relatively static
- there is a large amount of re-usable metadata among potential entries
- there is an existing system that collects or manages at least some of the metadata

Store the metadata in a database if the data are dynamic (prone to frequent changes), there is a large potential for re-using metadata elements, or some of the metadata are already managed in a system. Databases can be quite complex to establish, update, and administer and the large number of tables that may be required can result in slow query or presentation times. Basic metadata may already be managed within systems such as GIS. The choice is yours whether to augment the existing metadata system, use supplemental tools, or offload all the metadata into a true database. The principal advantage in using a database is that skills and software are likely to exist in the same office. The principal disadvantages are that complex multi-table joins are very costly in terms of search times, and the ability to perform a full-text search across many items (as is the trend in Internet searches today) is not supported in RDBMS.

Create and manage the data in text files if the metadata are relatively static, their content doesn't change frequently, and the content is reasonably unique. The advantages to a text-based metadata system are that word processors or text editors can be used at a minimum, metadata output does not require report- writing software, and full-text search is well supported. Fielded search will only work in those cases where the metadata are rigorously structured. As such, the use of SGML is an ASCII-based alternative to pure "text" metadata.

A transition from a file-based to database system is possible through use of SGML-tagged or other highly structured metadata. Conversion programs can be written to convert the SGML contents into the appropriate instructions to load them to a database. More easily, for example the FGDC-compliant metadata can be written from database implementations that can be indexed by text-based systems. Either "system" is an appropriate metadata management system if written properly and will support metadata service via Z39.50 and the Web.

Although access to data and dissemination of data are often used interchangeably, there is some difference between the two:

- *Data Access* can be viewed as a more reactive policy: if a potential user makes a request to access a dataset, this is considered, and access may be granted subject to a number of conditions and possible charges. Although there may be a moral onus on the data owners to document their data resources and make visible to potential users what resources they hold, there is often no obligation on the data owners to do so as exemplified by Freedom of Information legislation in some European countries like Ireland and the UK.
- *Data Dissemination* can be viewed as a more proactive policy the objective of which is to encourage the use of data. It therefore includes Data Access but extends it with a more positive effort, for example to document the data, provide examples of applications, develop user support services, and seek to extend the user base.

Both Access and Dissemination require resources, if nothing else to document the data or administer requests of access. Dis-

semination requires additional resources for example to develop user support services. Internal re-organisation of the responsible agency may also be required together with extensive training and efforts to change some aspects of the organisational culture The latter may include a shift throughout the organisation in valuing information as an important asset of which the agency is not the "owner" (this is my data) but the custodian on behalf of the community as a whole. Identifying the real costs involved in the implementation of a policy, whether limited to access or full dissemination, is absolutely crucial and should come well before formulating possible strategies for funding such activity.

The Situation at European Level

According to MADAME Report 4, the Internet continues to grow at a phenomenal pace. Internet users have soared from 171 million in 1999 to 377 million in September 2000, of whom 160 million in North America, and 105 million in Europe (Nua Surveys 2000). Current predictions indicate that Internet users could exceed the 1 billion mark by 2005, with 700 million located outside North America (CITU, 2000). Whilst North America remains in the lead, the variations within Europe are very significant as shown in Table 2-1 with Internet penetration ranging between 49% in Norway to 6% in Portugal (Nua Surveys 2000). These differences reflect different stages of economic developments, and historical and cultural traditions. However, governments also play an absolutely crucial role not only in the provision and regulation of the underlying infrastructure, including direct public investment and education, but also in creating the policy framework within which increased use of information and communication technologies takes place.

Norway	49%	Slovenia	23%
Sweden	44%	Germany	19%
Finland	38%	Italy	16%
Denmark	35%	France	15%
UK	26%	Greece	12%
Benelux	24%	Spain	9%
Switzerland	23%	Portugal	6%

Table 2: National Internet Penetration Rates (Source: Nua Survey 2000)

Moreover, public sector organisations are also the major single holders of information, and hence of the essential resource upon which information-based industries and services can develop. A recent study for the European Commission (Pira et al. 2000) estimates the total value of PSI in Europe at £ 68 billion annually, a substantial part of the total economic activity within the European economy. The report distinguishes between investment value, i.e. what government invests in the acquisition of PSI, and economic value, i.e. the part of national income attributable to industries and activities on the exploitation of PSI. European governments invest annually some £9.5 billion in PSI, the largest sector of which is represented by geographic information (mapping, land and property, meteorological services, environmental data), followed by cultural and company information. The economic value of such investment is estimated to be in the range £ 28-130 billion, with sectoral breakdown shown in Figure 7:

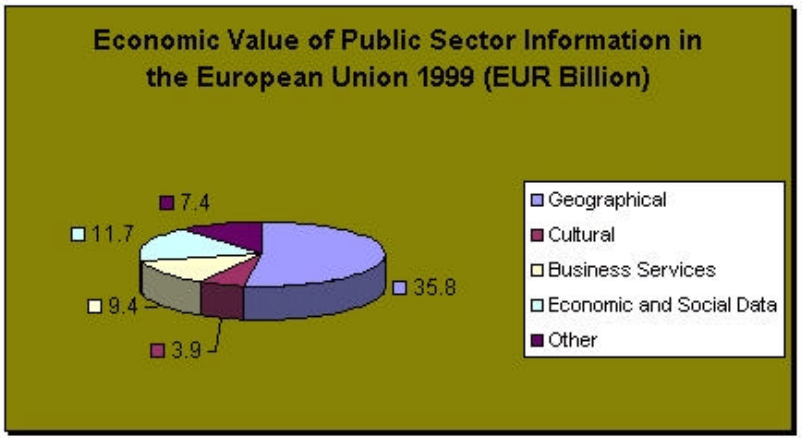


Fig.7. Economic Value of Public Sector Information in the EU (adopted from MADAME Report 4)

Given the importance of PSI to develop the Information Society, a number of initiatives at national and European level have started to take place to increase access to this vital resource to both citizens and business. At the European level, the most recent initiative launched by the European Commission and agreed at the highest political level at the Lisbon Summit in 2000 is *"e-Europe: An Information Society for All"* setting out the European agenda for the further stimulation and growth of the Information Society. It identifies ten priority areas to bring Internet access to the reach of all, and develop key applications in the fields of education, health, transport, and access to government information. At the Feira Summit in June 2000 the e-Europe Action Plan was agreed. Within the action plan *"Government on-line: electronic access to public services*" focuses on the extent to which digital information can transform old public sector organisation and provide faster, more responsive services. It can increase efficiency, cut costs and speed up standard administrative processes for citizens and business.

National initiatives are also moving along similar lines. As a matter of example, the UK government has set a target of all interactions between government departments and citizens to be on-line by 2005 (CITU, 2000). Similar initiatives are also taking place in other countries. What we are witnessing therefore is an increasing pressure on public sector organisations, both at central and local level to use Information Technologies to become more effective and efficient, and more responsive to the needs of citizens. At the same time they are also often asked to become more open in respect to the information they have, share it with others, and in many instances also exploit this information to recover some of their costs.

References

- Blakemore, M, McKeever, L (2000), *"r-code user needs analysis"*, in: "Comparative evaluation of online metadata services and user feed back", MADAME Report 02, www.info2000-madame.org;
- CITU (2000), *"e-Government. A strategic Framework for Public Services in the Information Age*", London, Cabmet Office;
- Chenez, Chr., Kermarerec, G., (1999), "On Going Metadata Initiatives in Europe", in 5th EC-GIS Workshop, Stresa, Italy, June, 1999;
- Civi, E (2000), "Kno*wledge management as a competitive asset a review"*, in Marketing Intelligence & Planning 18(4);
- Craglia, M, Annoni, A, Masser, I (2000), *"Geographic Information Policies in Europe: National and Regional Perspectives"*, (www.ec-gis.org);
- Harreld, H (2000), *"Governments are offering their services online, but will they truly streamline processes or will people end up waiting in line on-line?"* CIO

Magazine Sept 15, http://www.cio.com/ archive/091500-redtape.htm);

- Ilie, R., Ionita, A., (2001), "*e-Government Toolkit for Developing Nations*", WB proposal, May, 2001;
- Ionita, A., Moise, E., (2001), "*The Specifications for the Components of County On Line*", in "Definition and the development of County On Line Concept", ICI, 2001
- Ionita, A., Pribeanu, C., (2001), "*MEDIAGIS Project - An approach for On Line Service for Geographical Information*", in 7th EC GI&GIS Workshop, Potsdam, Germany, June 2001;
- Nedovic, Z, (2000), *"US Experiences in Staring Geographic Information"*, Practitioners' Forum on Institutional and Organizational Access to Public Sector Information, MADAME Report 4,
- Nua Surveys, http://www.nua.ie/surveys;
- Pira International Ltd University of East Anglia and Knowledge Ltd (2000), *"Commercial Exploitation of Europe's Public Sector Information"*, Luxembourg, EC-DG INFSO;
- Raper, J. (1999): *Information and Interoperability*, Workshop on Virtual Interoperability Laboratory, Ispra, Italy; February 1999, **URL:** http://ams.emap.sai.jrc.it/dg3gis/vil/
- Stiglitz, J.E., Orszog, P.R. et al (2000), *"The Role of Government in a Digital Age"*, Washington DC, Computer and Communication Industry Association;
- Van Boxtel R.H.M.L. (2000), *"Towards the Optimum Availability of Public Sector Information"*, The Hague: Ministry for Urban Policy and Integrator of Ethnic Minorities.
- Mandelkern Report: http://www.ec-gis.org
- http://www.w3.org
- http://www.xml.org
- http://cigref.fr
- http://xmltechno.com
- http://www.mutu-xml.org/i
- http://www.mu.niedersachsen.de/cds (Eurpean Topic Centre on Catalogue of Data Sources, Online Catalogue of Environmental Data Sources (European Environmental Agency).
- http://geo-nsdi.er.usgs.gov
- http://geology.usgs.gov/tools/metadta/tools/doc/opinion/clearinghouse.html
- http://carto.ped.org
- http://geology.usgs.gov/tools/metadata/tools/doc/faq.html#2.5
- http://www.fgdc.gov
- www.info2000-madame.org
- http://www.agi.org.uk/pag-es/dict-ion/dict-agi.htm
- http://www.state.wi.us/agencies/sco.metatool/mtoolhis.htm
- http://www.state.wi.us/agencies/sco.metatool/mtools.htm
- http://www.fgdc.gov/metadata/contstan.html
- http://www.statkart.no/isotc211/pow.htm
- http://purl.org/DC/
- http://www.cenorm.be/news/press_notices/metadata.htm