GI-META

Final Report

Prepared for

The European Commission

July 1997

Contents

Executive Summary

Background

This document is the final report to the GI-META project. The report brings together the results of interviews with key members of the Geographic Information (GI) community undertaken by GeoInformation International and MVA, a survey of metadata services undertaken by the Department of Town and Regional Planning, University of Sheffield, and the results of a questionnaire survey. It is also supported by the experience of MVA and GeoInformation International gained from the IMPACT2 project OMEGA, during which time research was undertaken, and a metadata service was configured.

This six-month study has been undertaken in order to better understand the requirements for metadata services for the European Geographic Information industry. Metadata is defined as information used to describe other information. Metadata services for geographic information have the potential to remove barriers of data awareness and accessibility and to encourage the wider and more effective use of GIS technologies.

The main findings of the study are summarised in the following sections:

1. Current Metadata Products

1.1 Most metadata is provided at a very simple level in the form of catalogues, indexes and paper directories.

1.2 There is an increasing number of digital metadata directories supplied on a CD-ROM or floppy disc and a growing range of on-line metadata services.

1.3 Existing metadata ranges from very simple data descriptions covering data type, geographic coverage and supplier, through to full technical descriptions including data specifications.

2. User Requirements

2.1 There is still a general lack of awareness of what is meant by the term metadata and a range of views of what constitutes a metadata service. Increased awareness of what metadata is and how it can be used is a necessary part of implementing any metadata service.

2.2 The broad user requirements for a metadata service are as folows:

- A metadata service must be comprehensive in its content, both in geographical coverage and type of datasets.

- The data descriptions must have a minimum level of detail which is available for all datasets, although this does not have to be very extensive.

- A metadata service must be constantly maintained and updated to provide users with the latest information about data availability.

- It is desirable to have a very detailed description of a dataset if the metadata service is to be used to select suitable datasets for particular projects or applications.

- A metadata service must include descriptions of data from governmental agencies, non-governmental organisations and commercial data providers.

- Multilingual interface to any metadata service is necessary if it is to be accessible to the broadest range of geo information users in Europe.

- The interface must be non-technical, simple to use and largely intuitive in style.

- The metadata interface must allow comparison of different datasets covering the same geographical area.

- The metadata interface must allow comparison of datasets with the same theme or similar content.

- A metadata service must contain either sample images for datasets or sample digital data that can be downloaded by the user.

- A comprehensive metadata service should include a catalogue of paper-based maps as many users require geographic data not currently in digital form.

- The metadata service should be available on-line but with both CD-ROM versions and paper version available for those with no on-line access.

- Users anticipate that any metadata service will not be comprehensive and thus should contain search capabilities to review other metadata listings available through the Internet.

- There should be the facility to order data on line, with the potential for on-line delivery of datasets in the future.

- Measures of data quality should be included in the data descriptions.

- Data descriptions should use non-technical language.

- Search mechanisms must accommodate the experienced geo information user and the novice entering the market place for the first time.

2.3 There are a number of existing on-line or CD-ROM based metadata services at present. There is however, a perceived need for a greater use on-line metadata services in the future.

2.4 There is a generally recognised requirement for metadata services but emphasis is on the need for awareness about national and regional data sets rather than pan-European data sets.

2.5 In the short term there is a continuing need for CD-ROM and paper-based metadata information.

3. Types of Metadata Services

3.1 The best delivery mechanism for future metadata services, especially at a pan-European level, is certain to be on-line via the World Wide Web (WWW).

3.2 Metadata services can be divided into two types:

"broad and shallow" databases, containing many entries with simple data descriptions.

"narrow and deep" databases, containing detailed data descriptions of fewer data sets.

3.3 It is possible to conceive of a 'meta-metadata service', consisting of linked databases through a common point of entry. Such a model could limit the amount of data capture required, but would rely on the acceptance and adherence to some form of metadata standard.

3.4 The effective use of metadata services is based on the search criteria provided. As a minimum these need to include data type based on a structured hierarchy of themes or subjects and geographical coverage. Further criteria such as supplier or publisher of the data set and product name could be usefully added. In existing services these criteria vary in both their format and comprehensiveness, with no use of standard nomenclature.

4. Cost/Benefit issues

4.1 Users are generally prepared to pay a limited amount to access a comprehensive metadata service.

4.2 Commercial data suppliers see a need to have a metadata service but can identify only limited benefits in paying to have their data sets included.

4.3 There is an opportunity to generate revenue from advertising on a metadata service.

4.4 At present the cost/benefit case for a fully commercial pan-European metadata service is not established. It is anticipated that some level of government funding will be required to set-up and run such a service in the medium term.

1. Introduction

1.1 Introduction

This document is the final report to the GI-META project. The report brings together the results of interviews with key members of the Geographic Information (GI) community undertaken by GeoInformation International and MVA, a survey of metadata services undertaken by the Department of Town and Regional Planning, University of Sheffield, and the preliminary results of a questionnaire survey. It is also supported by the experience of MVA and GeoInformation International gained from the IMPACT2 project OMEGA, during which time a metadata product was developed, and results from other metadata initiatives.

1.2 Objectives and Scope

The objectives of the GI-META project are to examine the feasibility of providing European geographic information metadata services by

- reviewing existing local, national and international metadata services
- examining the costs involved in creating and maintaining metadata services
- reviewing the interconnection options available via EU-wide information networks
- presenting a set of possible implementation scenarios to suit different levels of preparedness for both offering and using such services

This study was confined to metadata services about geo information or data with a spatial element, i.e. a location reference. The geographic scope of this proposal covers all current EU member States but it also takes account of developments outside of Europe.

1.3 Background to Study

Starting at the same time as this project were two others; GI-BASE, and GI-POLICY. The former is concerned with an assessment of the current market for geographic information in Europe, and has to survey existing data sources and services. It must also identify the core of geographic data themes that should be standardised across Europe, and identify the business case for doing so. GI-POLICY is concerned with the political instruments that will be required for the European geographic base. GI-META and GI-BASE are working closely to ensure consistency of approach in the issues common to both to metadata and data.

The ITT called for the GI-META study to focus on the following issues:

- What geographic information metadata sets already exist in Europe?
- What geographic information metadata sets are being planned in Europe?
- What major geographic information data sets are not currently covered by existing metadata services?
- What do potential users of geographic information want from a metadata service, i.e. what are the minimum and the ideal levels of information to be held about a data set?
- What information is being proposed to be held in metadata databases by initiatives such as GDDD (Geographic Data Description Directory), and how does this compare with perceived user needs?
- How can issues of quality of data be addressed, both practically in the short term and theoretically in the longer term?
- What mechanisms exist to facilitate and control asset trading, taking into consideration legal issues such as different copyright legislation across Europe, differences in national contract law, etc.?
- How can metadata best be delivered (technically) from different sources to different types of user, i.e. on-line (by what means), off-line (CD-ROM, diskette, paper)?
- Can a business case be established for creating new metadata services where these do not exist today, based on direct cost-benefit analysis?
- How should a potential user of geographic information located in any part of the EU be best enabled to access this information, both the metadata and the actual data?

- What are the barriers to making metadata available on a wide-scale basis - technical, legal, practical (i.e. linguistic, payment mechanisms, control of information following its release electronically, etc.)?

- How can these barriers be eliminated or their impact reduced?

Within the context of these issues the GI-META study undertook to determine what the user requirements for metadata are at a local, regional and European level. This list of user requirements is then used as the basis to discuss the issues raised in the ITT.

1.4 Definitions

For the purposes of this study the term metadata is taken to mean 'a standardised description of a data set'. A metadata service is used to capture, standardise, provide ongoing access to and disseminate metadata. A metadata service can include the actual database or catalogue or can provide access to a network of other metadata services or data catalogues. Geographic metadata therefore describes geographic data. The term 'geographic data' includes two broad categories of geographic data, 'geo-coded' data, and 'location referenced' data.

Some metadata services include a much broader range of information, not just the data descriptions themselves. Examples might include directories of organisations, references to publications and lists of meetings. These aspects are considered to be part of much fuller information service which is outside the scope of this study.

1.4.1 Geo Coded Data

A geo-code defines the position and shape of each object either against a reference grid, or on the Earth's surface. Metadata is needed to define the accuracy, resolution and history of this geodetic information. The geo-code by itself is of little value, as the nature of the referenced object and its characteristics must also be known. This is conveyed by the object classification (what type of object) and attributes or characteristics of that object.

A geo-coded data set may or may not possess sufficient attribute information to be of value in isolation. A good example of this is provided by digital administrative boundaries, which are often

supplied with only an area code, but without the Census information which gives them geographic and demographic meaning.

1.4.2 Locationally Referenced Data

The full value of geographic information systems (GIS) is realised when data carrying a location reference (for example, a place name, postal code, road number or Census code) is linked to geo-coded data. The great majority of digital information actually carries some form of location reference, which has the potential of linkage to geo-coded themes (for example, road networks, administrative boundaries).

It is the linkage between the geo-coded data and the locationally-referenced data that constitutes the enormous potential for geographic analysis, provided that the location references are standardised. For many datasets there is little standardisation in the way they are collected and referenced.

It is frequently the case that different organisations are involved in the supply of geo-coded and location referenced data sets. Metadata has a critical role to play in enabling a better association between these two types of data, their suppliers and potential users.

1.5 Study Methodology

The GI-META study undertook to evaluate the user requirements by a series of in-depth interviews with active users of geographic data to understand in detail how these individuals and their organisations currently use metadata and what further use they might have for a comprehensive metadata service.

The study carried out the interviews and then sought further advice on the use and type of metadata from leading players in the European GIS market. To this information was added the results of a questionnaire circulated to the readership of GIS Europe magazine, prepared in association with the GI-BASE study.

The study also undertook a review of some existing metadata services and data catalogues to determine how well they met the user requirements defined in the first part of the study. In particular the study looked at the CEO (Centre for Earth Observation) initiative and the results from this programme to see how their user requirements study varied from that drawn from this study.

Finally the study reviewed the ITT issues in the context of the information gathered from the above activities.

1.6 Structure of Report

The following sections of this report detail the results of the reviews and questionnaire undertaken to support the previous discussion.

Chapter 2 Interviews. This summarises the results form a number of interviews undertaken with users and data suppliers.

Chapter 3 Questionnaire Survey. This summarises the results from a questionnaire survey undertaken with the GI-BASE project.

Chapter 4 Other metadata initiatives. This briefly reviews some of the other metadata initiatives which were considered with respect to compiling a list of user requirements.

Chapter 5 User requirements. This is list of user requirements distilled from the surveys, interviews and reviews of existing metadata initiatives.

Chapter 6 Review of some existing metadata services. This comprises a discussion of the results from the review of a number of on-line services, some in detail and a few paper based data directory products, compared to the list of user requirements.

Chapter 7 Implementation models for a metadata service. This discusses some of the implementation models which might be used for a European metadata service.

Chapter 8 Business models for a metadata service. This section discusses some of the business issues which might influence the launch of a commercial metadata service.

Chapter 9 Cost Benefit Issues. This reviews some of the issues related to the type of metadata model that might be adopted for Europe and some of the funding options.

Chapter 10 Discussion of the ITT issues. This reviews the issues raised by the ITT in the context of the study findings.

Chapter 11. Recommendations to the European Commission This summaries the previous chapters by making recommendations on the possible approach for a European metadata service for geo-information.

Chapter 12. Bibliography

2 Interviews

2.1 Introduction

In order to gain a better understanding of the issues faced by a metadata service provider for Europe the study undertook to establish the basic user requirements for such a service. To determine the key requirements a series of interviews were undertaken with existing GIS users and potential users of metadata.

These interviews were conducted either face-to-face where possible or by telephone in some cases. The interviewees were selected because they were active geographic information users, geographic information publishers or GIS solution providers for whom data is an integral part of the solution supplied to the client.

The study team selected the interviewees based on their practical experience of using GIS in an organisation, often in circumstances where the GIS is only one of a number of active information systems. It was felt that these individuals better represented the typical potential user of a metadata service and as such they would be able to better represent the needs of the user community with regards to metadata.

The interviews were structured to gain insight into the possible ways in which metadata might be provided in the future. The most important aspect of the interviews was to understand what users viewed as being essential components to a metadata service from which a user requirements list could be created. The interviews also included discussion on whether or not enough revenue could be generated for a commercial geo-information metadata service at the European level.

To supplement the information gained from the interviews and to ensure the study results were balanced a number of individuals with experience and knowledge of the GIS industry were asked to provide their views through a more detailed written questionnaire.

2.2 Topics Discussed

The purpose of the interviews was to understand current usage of metadata, both internally within organisations and from existing products and services, and to assess the potential for future metadata services. An important element to the issue of future metadata services was the format of the data entry and willingness to pay. The interviews were therefore structured to answer the following questions:

- What is the understanding of the term metadata?
- Does the organisation create its own metadata, and if so in what form?
- As a data user, what metadata sources does the organisation use, what are the most useful and why?
- As a data supplier, what metadata are created?
- Does the organisation currently use, or will it in the future, an on-line metadata service?
- Would the organisation be prepared to pay for such a service and how much?
- Would the organisation advertise or be prepared to pay to have products included in such a service?

The interviews were based on a structured interview questionnaire, Appendix 1, but were generally free ranging and sort to extract general opinions and ideas. The individuals perception of what was metadata in its broadest sense was an important part of the discussion, as well as how metadata is currently used in their organisation and how a metadata service might be used in the future.

2.3 Organisations Interviewed

A total of 17 interviews were conducted. Each was a structured discussion of the issues, guided by the interviewer who used an interview question sheet as a guide (see Appendix 1).

The organisations in Table 2.1 were interviewed and those individuals in Table 2.2 were asked to provide their opinions in a detailed questionnaire similar to the interview questionnaire. In some cases the responses were by completed questionnaire in others by telephone

conversation. The results were used to give balance and a pan-European perspective to the detailed interviews.

1. English Nature	Jonathon Budd	UK	Data supplier and user
2. ETSU	Steve Dagnall	UK	Data user
3. Geolink Ltd	Micheal Brownen	UK	Data supplier and value-added reseller
4. Lovell Johns	Ben Hill	UK	Data supplier
5. Dataview Solutions	Geoff Kendall, Derek Prior & Ian Prentice	UK	Value-added reseller
6. Macauley Land Research Institute	Richard Aspinall	UK	Data user and supplier
7. TeleAtlas	Philiep de Sutter	B	Data supplier
8. Landmark Ltd	James Cadoux-Hudson	UK	Data user and value-added reseller
9. Geodan	Henk Scholten	NL	Data user
10. TED Atec	Phillip Milliet	F	Data user
11. SINES Ordnance Survey	Claire Hadley	UK	Metadata service provider
12. GISIG	Emmanuele Roccatagliata	I	Data user
13. IGGI	Alan Oliver	UK	Govt. co-ordinating committee
14. Local Government Management Group	Andrew Larner	UK	Government advisory
15. Centre for Earth Observation	Josef Aschbacher	Eur	Metadata service provider
16. Association for Geographic Information	David Green	UK	GIS Association
17. EUROGI	Mike Brand	Eur	GIS Association

Table 2.1. List of interviewees

2.4 Findings

It was clear from the results that there were a number of key issues which were consistently raised. The issues were common to users, data suppliers and GIS vendors and that these issues seem to be common across the GIS community.

These issues may be summarised as:

Autodesk Inc.	Gilles Albaredes	CH	Software vendor
New University of Lisbon	Dr Antonio Morais Arnaud	P	Academic
University of Utrecht	Prof. Peter Burrough	NL	Academic
Unisys Europe	Edwin Cobb	UK	Software vendor
Am/FM International	Hans Festen	Europe	Association
Delft University of Technology	Menno-Jan Kraak	NL	Academic
Scankort I/S	Vagn Laursen	D	Consultant
Intergraph	Tim Mahoney	NL	Software vendor
Free University of Amsterdam	Prof Peter Nijkamp	NL	Academic
Star Informatic	Manual Pallage	F	Software vendor
URSA-NET	Prof Nicos Polydorides	GR	Academic
PONAD BV	Jan Pongers	NL	Consultant
ESRI	Jorg Schaller	D	Software Vendor
GIS Consultancy Bureau	Marc Uffer	CH	Consultant
Bentley Systems	Jan Willem van Eck	NL	Software vendor

Table 2.2. List of GIS specialists asked to complete a questionnaire.

- The current understanding of the term 'metadata'
- Levels of use of metadata within organisations
- Current uses of metadata products and services
- The types of metadata created by data suppliers
- Use of on-line metadata services
- Willingness to pay for on-line metadata services
- Interest in advertising on a metadata service

The following is a summary of the key issues that were gained from the interviews. The order is not indicative of any perceived level of importance.

1. What was the interviewee's understanding of the term metadata?

It was clear from the range of individuals interviewed that the understanding of what is and what is not metadata varies considerably depending on the activities of the organisation and the experiences of the individual.

Organisations that were purely geographic information users regarded metadata as simple directory information about sources of such information. They typically included only a limited range of information such as the following:

- Type of data
- Owner
- Supplier
- Scale of data or source from which it was derived
- Currency
- Cost
- Licence arrangements
- Name and address of contact supplier

Such metadata sources would include catalogues and listings, like the OMEGA product created under the IMPACT programme or the UK's SINES service (see Sections 4 and 6 respectively). The two most important criteria about a metadata service for these users would be its comprehensiveness in terms of the range of data sets covered and the need to limit the amount of technical detail. Such a metadata source may be typified as being 'broad and shallow'.

Organisations that were involved in supplying data, either as a reseller in some form or as a primary data gatherer, had a very different view of what was metadata. They tended towards the view that metadata should include detailed information about a data set such as its structure, the formats in which it might be supplied and the technical specification for the product.

Such a definition of metadata is similar to the MEGRIN GDDD which provides a relatively high level of technical detail about digital data sets provided by National Mapping Agencies (see Section 6). Such metadata may be characterised as 'narrow and deep'.

2. Does the interviewee's organisation create its own metadata, and if so in what form?

Few of the organisations interviewed created their own metadata in any systematic way. This included those involved in creating data and supplying it to clients, although technical specifications for data sets were supplied and can be regarded as a form of metadata.

Few organisations who held data sets for internal use or for creating value-added products maintained any form of metadata or catalogue of what was held and who had responsibility and controls on the use of this data. There were one or two exceptions, but these held only the simplest of listings. No-one used any form of standard, or were generally aware of any that might be applied. MEGRIN and the Ordnance Survey of Great Britain were exceptions to this.

3. As a data user, what metadata sources does the organisation use, what were the most useful and why?

Organisations interviewed found that the lack of a comprehensive source of such data made it very difficult to find out what was available, and as consequence of this lack of a comprehensive source very few used metadata services in any systematic or regular way.

The main reason for the lack of use of metadata sources was that they were not comprehensive. For a metadata product or service to be of value it was felt that it must include, even in a very simple form, a very wide range of data products from all sources and suppliers. Such a metadata source must also be geographically broad, with a number of interviewees expressing the need for not just information at a European level but also at a global level.

4. As a data supplier, what metadata do you create?

Data suppliers viewed metadata primarily as the information which contained the technical specification of their products. This would be supplied with the product and might be made available to potential users who wished to review the product before deciding whether or not to use it.

Such metadata would be extremely complex and its format would be quite specific to the data set. Other than national mapping agencies, few of the suppliers were using or contemplating using any standard form of metadata, although a number expressed an interest in having some form of common standard.

5. Does your organisation currently use or would they be likely to use an on-line metadata service in the future?

Current usage of on-line sources for locating geographic information was very limited. However, in a few cases where both the hardware connection and the experience existed, the Internet was used extensively to search for data sets. This was particularly true for data sets outside the country in which the organisation was based, and in particular related to data from non-European sources.

6. Would you be prepared to pay for such a service and how much?

Organisations who were data users had two perspectives. The first was that if the source was comprehensive enough, that is it was a true 'one-stop shop', they would be prepared to pay for a service. The amount they would be prepared to pay depended on the amount of data they used, how obscure it was and the consequent length of time it might take to locate such a data set. Typically organisations might be prepared to pay between 200 ECU and 1,000 ECU for an annual licence fee.

The second group of individuals were those who felt that a single metadata source would never be comprehensive enough for their needs and that alternative ways of locating data sets, (especially non-European) existed. An example of this might be a MapInfo reseller, who could call upon resellers in other countries whose own national knowledge would probably be much better than could be achieved by a single metadata source.

7. Would you advertise or be prepared to pay to have your products included in such a service?

Data suppliers were very reluctant to pay to have their products listed on a metadata service, believing that there was little to be gained from doing so. This method of payment may also be self-defeating because if the service were not comprehensive then it would be difficult to get users to pay to use such a service.

There was some limited interest in advertising, although there would need to be a clear method of recording interest in a product and ensuring the advertiser received a tangible gain from being on such a service. There is growing evidence that these aspects can be achieved to an advertisers satisfaction and that it might be possible to part-fund a European metadata service through advertising.

2.5 Summary of interview results

As a result of the interviews and the comments from the GIS experts a number of key issued can be clearly flagged as being critical to any future metadata service. These are summarised as:

- Metadata is not a clearly understood term, with a wide variety of interpretations.

- Metadata sources/services divide into two distinct groups:
  - comprehensive broad-based information that contains little detail
  - narrowly focused, very detailed information.

- Use of metadata is at present limited.

- There is a demand for more comprehensive metadata services, with an interest in on-line delivery.

- Users might be prepared to pay a 'reasonable' licence fee to support such a service, although what is reasonable depends on the individual user. Data suppliers would not be willing to pay to have their data included, unless there was clear commercial advantage in doing so. Some revenue might be generated from advertising but on-line services are still a very immature form of advertising.

2.6 Further issues from follow-up questionnaires

A number of other issues were raised as potentially important within the context of a European metadata service by those GIS professionals with whom the interview results were discussed.

The issue of data quality and how this is collected and reported within a metadata system was not one raised by the interviewees and clearly does not feature highly as an issue in users minds at present. Many in the GIS community however, feel this will become a major issue and needs to be considered as part of the information collected about geo spatial datasets.

There are two aspects to the question of data quality within the context of the metadata system. The first is the accuracy, currency and quality of the metadata record itself and the second is the quality measures used for the individual datasets.

Clearly it is important from the users perspective to know that the data records retrieved from a metadata service are accurate and current. Information on quality of the metadata service itself should be included in order to establish user confidence.

Measures of quality, including horizontal and vertical accuracy, data capture accuracies, data conversion accuracies, data generalisation issues and data classification issues, are much more complex. The lack of any standard approach by those involved in data collection and data capture makes this a much more difficult aspect of the metadata record to standardise. There have been some national attempts to review the options for such standards but at present there are no initiatives by those working on pan European standards to set data quality and accuracy evaluation standards. While clearly a desirable aspect in any data description, it was felt that at present the lack of any widely used standards in this area mitigates against the inclusion of a very structured and prescriptive data description in a metadata database.

It was a generally held belief that metadata was relatively cost effective to collect at the time of creating a database, but much more costly to collect retrospectively. To this end at least one GIS software vendor, ESRI, has introduced the concept of automatic metadata collection as part of the creation of new datasets. The intention is to encourage metadata collection as a core part of any data collection or data capture exercise.

The issues of data quality are also of importance to the data suppliers as the collection and maintenance of this information can be costly. There is also a series of liability issues surrounding the declaration of quality, and in particular accuracy, which are of concern to data suppliers. The imposition of a standard approach within a metadata service might be counter productive with data suppliers declaring poorer than actual quality levels to insure against future litigation or might even lead data suppliers to withhold data sets from the metadata database.

3. Questionnaire Survey

3.1 Introduction

To compliment the interviews and follow-up questionnaires, the study was concerned to establish the current level of awareness and usage of metadata in the wider user community. To achieve this a number of questions were included in a general questionnaire distributed on behalf of the GI-BASE study run in parallel with the GI-META study.

3.2 Survey Rationale

While it was not a requirement of the GI-META study to undertake a widely based questionnaire survey of users, it was felt that more broadly based user information was needed to support the detailed interviews.

It was therefore decided to undertake a joint survey with the GI-BASE study, for the following reasons:

- It is essential to obtain an understanding of the use of geographic information, the business functions and organisation details of each respondent and it is clearly much more efficient to do this once

- A number of surveys have been undertaken of metadata and geographic information use already. It was believed that a better response would be obtained for the GI-BASE project if there was not a previous GI-META survey of the same organisations.

- The first phase of the survey could be disseminated through GIS Europe, as GeoInformation International is a partner of the GI-META project.

The GI-BASE survey was undertaken in two stages. Stage 1 was to intended to obtain a broad response to a limited set of questions, including general awareness of metadata. A second, in-

depth survey was undertaken within the GI-BASE project of respondents who indicated their willingness to participate.

3.3 Survey Design

The survey posed recipients with only 3 questions on the issue of metadata, see Appendix 2. These were:

- Are you aware of metadata services for digital geographical information and have they been used in your organisation?
- If metadata services have been used, please list upto three that you found helpful.
- How could geographical metadata best be made available to your organisation?

The questions were framed to firstly give an understanding of the current awareness of what metadata services exist and to quantify what types of metadata are currently used. Then secondly to identify what users perceptions were on future requirements for metadata and how might point to future needs for metadata services.

A key issue in obtaining a consistent response was the understanding of the term metadata by those completing the questionnaire. A short statement on what is meant by metadata was included in the questionnaire to try and ensure that a broad definition was used which would include paper based catalogues as well as on-line directories.

3.4 Survey Distribution

The stage 1 questionnaire was distributed with the June 1996 issue of GIS Europe, and was also made available as an interactive questionnaire at the Ordnance Survey of Great Britain WWW site.

Analysis of the distribution of GIS Europe gives an indication of the market penetration. GIS Europe is sent only to named individuals within the GIS audience across Europe. For 1996, 11,300 copies of the magazine were sent each month to the magazine's subscribers and carefully controlled circulation. Additional copies were sent to exhibitions and conferences throughout Europe. Market analysis indicates that the distribution by this means will result in the following market penetration by European country:

Benelux	19%	
UK		18%
Germany		13%
France		12%
Other W Europe	9%	
Scandinavia		9%
Italy		7%
Spain		6%
Eastern Europe		3%
Rest of the World		4%

Table 3.1 GIS Europe distribution by country

From surveys and reader registration forms covering the magazine's controlled circulation gathered over the last two years, GIS Europe's readership is shown to cover the complete spectrum of GIS usage, from traditional areas such as local government and the environment, to the rapidly emerging sectors of retail and business planning. The questionnaire therefore targets the following markets:

Industrial GIS user		16%
Educational establishment		16%
Business GIS user		15%
Consultancy		14%
Central government		11%
Local government		10%
Utility company (non-governmental)		8%
Vendor	4%	
Other		6%

Table 3.2 GIS Europe distribution by sector

Assuming a typical response of some 1-2%, this survey would be expected to return about 150-200 questionnaires. In the end a total of 223 responses were received.

3.5 Survey Responses

3.5.1. Sample

A total of 223 completed questionnaires were received, either in paper form (149) or through the Internet where the questionnaire was posted on a couple of web sites (74). Responses were

received from a wide range of countries, including a number from outside Europe. There was also a number that did not give details as to the source, although these were still considered to be valid. The largest number came from the UK followed by Benelux and Germany. Overall the response was very disappointing and heavily skewed towards the UK. This may in part be a reflection of the number of questionnaires which are routinely sent to GIS users through GIS Europe magazine by various organisations attempting to evaluate the size and potential of the industry. However there was also a very poor response to the questionnaire on the various web sites where despite a large number of people visiting the questionnaire few completed it. For the English version only 74 responses were made from 850 visits and the French version had 33 visits but without a single response. Based on this response it is not possible to draw any conclusions about regional variations across Europe.

BOX 1: SAMPLE

Total sample :		223
Country of Origin:		%
Benelux		10.1
France		3.9
Germany, Austria		7.2
Italy		1.0
Denmark, Finland, Sweden		5.8
UK and Ireland		49.8
Spain, Portugal, Greece	5.3	
Rest of Europe		7.7
USA		4.8
Other		4.4
TOTAL	100%	

Table 3.3 Sample size and geographical spread of responses

BOX 2: TYPE OF RESPONDENT

Central/local government	25.0%
Commercial	39.4%
Education/research	26.0%
Unknown	9.6%

Table 3.4 Survey responses by sector

3.5.2 Type of respondent

An important aspect of the responses is the distribution between government agencies, commercial organisations and education/research organisations. The distribution was relatively even, with the commercial sector supplying a higher than expected return. A significant number of respondents have not included their name and address.

3.5.3 Awareness and use of metadata services

The purpose of this question was to determine the current use of metadata within the geographic information user community and what form that usage took. The question made the distinction between on-line services (such as the UK SINES service), digital (i.e. disc or CD-ROM) products and printed products such as catalogues, and sought to identify which of these were used and how important they were to the user.

BOX 3: AWARENESS AND USE OF METADATA SERVICES

Question: Are you aware of metadata services for digital geographical information and have they been used in your organisation?

	Unaware	Aware but not seen	Seen but not used	Used for selecting
On-line	38.2%	25.1%	25.7%	11.0%
Digital media	32.7%	24.6%	23.1%	19.6%
Printed publication	29.7%	13.1%	19.8%	37.5%

Table 3.5 Awareness and use of metadata services

The main findings from the responses to this question are as follows:

1. A significant proportion of respondents claim not to be aware of metadata products. Of this group, 30% claim not even to be aware of printed products.

This result indicates two problems which have been borne out in the interviews.

The first is that metadata is poorly understood and users do not necessarily recognise certain products as being metadata sources, despite the description of what constitutes metadata in the introduction to the questionnaire. For example, some geographic information users were unaware of printed catalogues such as might be published by a national mapping agency. We may therefore interpret this answer as being an indication that such catalogues are not viewed by some GIS users as metadata sources.

The second issue is that a significant proportion of the respondents are either unaware of or have not seen any digital form of metadata, be that delivered in an off-line or on-line form. Despite the fact that a number of products do exist their distribution is clearly not wide enough and many potential users have not had an opportunity to access or use them.

2. The number of respondents who were aware of metadata products but either had not seen them or were not using them was significant for all types of metadata, but particularly so for digital and on-line options.

This suggests either that the demand for metadata services may not be as great as initially predicted or that the existing services do not meet current user needs. For a significant number of geo-information users the fact that such sources exist but are not used suggests that they are either not business-critical to the way their GIS operation currently works or they do not regularly change their data sets and therefore have no need to search for new geographic information products, either regularly or infrequently.

It is important to note that the on-line and digital products are less likely to be used than the paper products. This is probably because the paper products are the traditional form of catalogue and are currently more comprehensive and/or more easily available than the digital and on-line services, in addition to the lack of awareness discussed above.

3. The majority of respondents who used metadata products used printed products (37.5%) and were far less likely to use digital or on-line services.

Given that we have already suggested that the comprehensiveness and availability of current metadata products and services is limited this response is not too surprising. It does however indicate that the successful launch of any future metadata service will be heavily influenced by these factors, a point borne out by the interviews.

3.5.4 Future availability of metadata

In terms of the future aspirations for metadata, respondents indicated a wish to migrate from printed to digital forms of such information. Both digital media and on-line are seen as important means for the dissemination of such information in the future.

The on-line option is clearly regarded as the most likely long term delivery mechanism for a metadata service (Table 3.6). There is however still a demand for digital media in the form of CD-ROM and to a lesser extent a paper version of the database. In the immediate future it may be necessary to have complimentary media such as CD-ROM for any European wide metadata service.

BOX 4: AVAILABILITY OF METADATA

Question: How could geographical metadata best be made available for your organisation? (Mark one only)

On-line	51.0%	
Digital media		27.6%
Printed publication		10.2%
Don't know		7.7%
No foreseeable requirement for metadata		3.4%
TOTAL RESPONSE		100.0%

NOTE: Some respondents marked more than one category and these have been added to the sample.

Table 3.6 Supply of metadata

To some extent these results conflict with the detailed responses from the interviewees who while they envisaged the use of on-line services as the most important delivery mechanism, also had concerns about how these might be accessed within certain types of organisations. There is clearly a difference between individual aspirations and organisation acceptance of widely used on-line services.

In a few cases respondents selected more than one option for the future delivery of metadata. These were treated as legitimate requirements and were included in the sample. In each case on-line was one of the options selected but added to this were either digital media or printed publications. This is indicates a clear need to provide more than one form of metadata and a clear conflict within certain organisations as to how metadata is to be used in the future.

A small number of respondents either had no view on what the most important future source might be, or actually could not perceive of any future need for metadata. This supports the findings from the interviews where there was a clear demand for a metadata service of some sought from most users. The answers to this question do not however indicate any level of usage, and there may well be a big difference between the desire to have access to a metadata service and the amount of actual usage anyone individual might make of it.

4. Other metadata initiatives

4.1 Introduction

As part of the study a review of some other metadata initiatives was undertaken. This included the following:

- IMPACT 2 OMEGA project
- CEO metadata initiative
- US Federal Geographic Data Committee metadata initiative

These reviews were combined with the results of the interviews and questionnaire surveys to formulate a user requirements statement for a European metadata service.

4.2 IMPACT2 OMEGA project

4.2.1 Purpose of the Project

The OMEGA (Object Metadata for European Geographic Analysis) project was run under the auspices of the IMPACT2 Programme of the European Commission. The objective of the project was to develop an interactive database of metadata covering digital geographic data sets, events and conferences, GIS vendors and consultants, together with some interactive case studies. The OMEGA product was a CD-ROM version of the more traditional paper-based directories or sourcebooks available within the GIS industry (Reference 1).

The initial focus of the project was on the metadata aspects of geographic information and a pilot product was built around the ArcView2 GIS software with programming undertaken in Avenue, a fully object-oriented language. However this proved to be too complex as a front end for a broad based directory product and too expensive to license for a commercial version of the product.

The final project CD was re-engineered as a multimedia product using Asymetrix Multimedia Toolbook because it offered good database functionality via ODBC connectivity and powerful multimedia functions. The metadata component of the final CD was based on the GDDD (Geographical Digital Data Directory) created by MEGRIN, one of project partners (see Section 6.4).

The project created a directory product but in doing so found that GIS functionality was far too sophisticated for a mass market CD product. The reaction to the initial pilot product was good from the core GIS industry but was considered by a broader potential user group to be far too complex in terms of its functionality and interface.

The OMEGA project also found that it was very difficult to collect metadata for spatially referenced datasets in any standardised format across the broad spectrum of data owners and suppliers. As a result large amounts of effort were expended on collecting data and then standardising data sets to fit a data model created for the project. As a result it was determined that a different data model had to be adopted which had two levels of data entry. The first was a very simple data description to be used for all data sets and the second a more detailed description, more like that being used for the GDDD.

4.2.2 Follow-up market research

Once the OMEGA CD was completed a full market research study was undertaken to evaluate the potential market for such a product and how it might be run as a full commercial activity. The market research identified a number of key findings:

- Users of geographic data tend either know about the data they require or expect metadata to be freely available. Thus the market value of a CD-ROM for a metadata service was of the order of 25 ECU and not the price considered at the start of the project which was in excess of 100 ECU.

- It was also found that there was less demand than had been expected for directory style products at a European level with a number of the paper-based products which were first published in the early 1990's on an annual basis being discontinued.

- The other mechanisms for covering the cost of such a product would be through the use of advertising. While this seems initially attractive there is little experience in advertising through electronic media within the GIS industry and for many of the key software vendors this was not viewed as a very productive advertising channel. The potential to attract new

advertisers in the form of data suppliers who are traditionally not strong print advertisers was also considered to be limited as this group had little expectation that their marketing spend on advertising would increase significantly over the next 5 years.

- The rise in interest and usage of on-line services also undermined the potential market for a CD-ROM only based product. Experience in the US has shown how on-line services attract a much higher usage rate and as a consequence could be more suited to certain types of advertising. However experience to date is that specialised on-line services are not able to earn enough from advertising alone to cover their costs. The GIS World Inc. web site 'GeoPlace.com' is an example where advertising meets only part of the costs, although the site is set up to service other marketing needs.

The problem for the OMEGA project therefore became one of balancing the higher than expected costs of collecting the metadata and, more importantly, maintaining the database, against the lower than expected revenues from advertising and subscriptions. In the end it was deemed that at a European level this was not a commercially viable product.

However the OMEGA work was merged with that undertaken in GIS World Inc. in the United States to create the first global directory products. The first of these was published in 1996 with a data catalogue including some content for Europe. The data catalogue will move to an on-line environment in 1997 run through the GeoPlace.com web site.

4.2.3 Conclusion from the OMEGA Project

Overall the OMEGA project demonstrated that any metadata service will be faced by two conflicting issues. The commercial challenge is to balance relatively low revenue streams at least in the short term, where users expectations are that such data should be provided free or at a relatively low cost, with a high cost of collecting and maintaining the database at a level of comprehensiveness which meets user needs.

4.3 The CEO metadata initiative

The Centre for Earth Observation (CEO) is a major European Commission initiative aimed at providing support to individuals and organisations across Europe who use earth observation data and information. The CEO programme covers many aspects of earth observation information, including a metadata service.

The CEO has undertaken a detailed series of studies into all aspects earth observation data. These studies have included an assessment of user requirements within the broad context of CEO and with respect to a more defined metadata service (References 4,5). They have also included a survey of current practices and standards for metadata (References 6,7). Much of the survey work has been supported by an on-going series of workshops to evaluate user needs and the specifications for the CEO services, including the metadata database.

A number of the major findings from the CEO work with respect to metadata are borne out in the results from this study. The key findings are discussed in the following sections.

A particular finding was that "the GIS community are very advanced when it comes to metadata and data management. They are also very nationally oriented, meaning that many of them are reluctant to invest in transferring their information to formats that could be used on a European scale, which would erode their current market niche." (Reference 6, p8). The GI-META study suggests that usage levels of metadata are less than this statement implies and that the GIS user is less sophisticated in its awareness and adoption of metadata as a data management tool. The result is that GIS users often have difficulty in defining in detail what their requirements are with respect to metadata.

A key issue for CEO was the use of standards and, if this was desirable, which standard should be adopted. Different market sectors had different views on what, if any, standards should be adopted and how these should be implemented. Overall the conclusion was that CEO should promote existing European standard initiatives, CEN TC 287, and minimise its efforts to create new standards. The GI-META study results show a general lack of awareness about standards and therefore users are not in a position to make a judgement about what a standard should cover and if the CEN TC 287 is appropriate as a general standard for all the geo-Information community.

Overall there are a very wide range of standards being developed or that already exist for handling data and which contain a metadata component or which are specifically designed for metadata records. The importance of these various standards depends on the sector of the information community from which they have developed. Some of these standards are very tightly focused and apply to very specialist data types while others are intended to be more generic, such as GeoTIFF or the FGDC standard described in Section 4.4 below. The problems identified by the GI-META study and by the work of CEO show that it will be very difficult to establish a single generic standard that suits all the requirements of users and data providers alike.

A key issue for the development of a widely-based European metadata service is the issue of how much standards have to be put in place in order to generate the breadth needed for a metadata service for the whole geo-information community. CEO have adopted a metadata service model which provides an interface to a series of linked databases located within data organisations. They have created the Catalogue Interoperability Protocol (Reference 2), which is a protocol designed for managing, searching and ordering spatial data located at distributed catalogues. The whole metadata service model is based on Internet delivery mechanisms, and assumes users require on-line metadata services. The GI-META study revealed that this model becomes more difficult to implement as the range of data types to be included in a pan European metadata service increases because a generic standard becomes more difficult to establish.

A general finding of the CEO work was that commercial organisations are more advanced in the use of metadata management procedures internally and are more aware of the benefits. Generally they were happy with their own internal practices and had no particular requirement for CEO standards to allow data to be more widely accessible. In addition there was no standard approach to the quality control of metadata records in the organisations CEO surveyed. The GI-META study found that commercial organisations were less advanced in the use of metadata than government agencies or national mapping organisations, and where commercial organisations did create metadata it was to provide information about their own formats and specifications.

The CEO metadata service, the European Wide Service Exchange (EWSE) Software System, which uses the IMS Gateway, is described more fully in section 6.3.1.

There are a number of differences between the findings of the CEO and the results from this study. The authors believe that much of this is due the different way the earth observation sector and the wider GIS community manage their information. Certainly the earth observation community seems much more likely to have developed metadata procedures for handling their data than is common in the wider GIS community where volumes of data are less and update cycles much longer. The GIS user is more likely to integrate their data into other information systems, with no requirement for detailed metadata systems.

4.4 The US Federal Geographic Data Committee metadata initiative

In June 1994 the Federal Geographic Data Committee (FGDC) approved a metadata standard which included the description of the content, quality and condition of the data and other

characteristics. The purpose of the standard was to provide a common set of terminology and definitions for documentation related to metadata. This standard was driven by the April 1994 Executive Order 12906 "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure." This requires federal agencies to use the standard for all new geospatial data from the start of 1995 and to make the metadata available to the public through the National Geospatial Data Clearinghouse.

The characteristics of the standard are that it should address the following:

- Availability - data needed to determine the sets of data that exist for a geographic location
- Fitness for use - data needed to determine if a set of data meets a specified need
- Access - data needed to acquire an identified set of data
- Transfer - data needed to process and use a set of data

It is noted that "these characteristics form a continuum in which the user moves through a number of choices to determine what data are available, to evaluate the fitness of, to acquire, transfer and process the data. The order in which data elements are evaluated, and the relative importance of the data elements, will not be the same for all users or for all tasks that use metadata. In addition, users with different tasks or at different stages of evaluation may require (or prefer) that a set of information can be available at different levels of abstraction or in different forms." (Reference 8)

This definition of the FGDC metadata standard encompasses many of the service aspects identified in the user requirement study for CEO and integrated into the user requirements list in this report. The key is to achieve a level of standardisation in the metadata service which allows users to access and evaluate the data record in accordance with their application requirements.

4.5 Conclusions

Only the OMEGA project looked closely at the user requirements from a commercial perspective. This has been the only European project to date to evaluate the market demand and to develop revenue models on which to test the cost-benefit assumptions. The other, clearly user oriented services have been primarily focused on the technical infrastructure.

5. User Requirements

5.1 Introduction

The results obtained from the detailed interviews and supporting questionnaires were combined with a review of previous user requirement studies for metadata. The result is presented as a generic user requirements list which may be used as the basis for the specification of a metadata service.

This user requirements list should be regarded as a detailed list of what users of geographic information would like to obtain from a metadata service. It does not imply that these requirements are either technically or practically possible to achieve within a single metadata service or whether these requirements are financially justifiable from either a governmental or commercial stand point. These issues are discussed in the context of each requirement under Sections 7 to 9.

5.2 User Requirements

The following user requirements for a metadata service were identified from this study:

- A metadata service must be comprehensive in its content, both in geographical coverage and type of datasets.

- The data descriptions must have a minimum level of detail which is available for all datasets, although this does not have to be very extensive.

- A metadata service must be constantly maintained and updated to provide users with the latest information about data availability.

- It is desirable to have a very detailed description of a dataset if the metadata service is to be used to select suitable datasets for particular projects or applications.

- A metadata service must include descriptions of data from governmental agencies, non-governmental organisations and commercial data providers.

- Multilingual interface to any metadata service is necessary if it is to be accessible to the broadest range of geo information users in Europe.

- The interface must be non-technical, simple to use and largely intuitive in style.

- The metadata interface must allow comparison of different datasets covering the same geographical area.

- The metadata interface must allow comparison of datasets with the same theme or similar content.

- A metadata service must contain either sample images for datasets or sample digital data that can be downloaded by the user.

- A comprehensive metadata service should include a catalogue of paper-based maps as many users require geographic data not currently in digital form.

- The metadata service should be available on-line but with both CD-ROM versions and paper version available for those with no on-line access.

- Users anticipate that any metadata service will not be comprehensive and thus should contain search capabilities to review other metadata listings available through the Internet.

- There should be the facility to order data on line, with the potential for on-line delivery of datasets in the future.

- Measures of data quality should be included in the data descriptions.

- Data descriptions should use non-technical language.

- Search mechanisms must accommodate the experienced geo information user and the novice entering the market place for the first time.

5.3 Review of User Requirements

The following is a detailed discussion of the specific user requirements identified as part of the interviews and questionnaire process.

5.3.1 A metadata service must be comprehensive in its content, both in geographical coverage and type of datasets.

A core requirement for any metadata service identified by nearly all existing users of geographic information was the need for any metadata service to be comprehensive in its coverage if it is to be really useful. This extended to both geographical coverage and types of data.

Geographical coverage

User work at very different levels of geographical coverage. These ranged from local authorities who have a very well defined and limited geography with which they are concerned, to organisations such as Landmark which have a strategic objective to work at a national level, through to organisations such as ETSU, a division of the UK Atomic Energy Authority, who are working at local, national and pan-national levels, including global. The areas in which the organisations operate determine to some extent the scales which they tend to use and the importance to them of national, European or global metadata services.

A metadata service will need to be able to accommodate searches at all the different geographical levels from European to local. It will also need to be able to index the datasets based on commonly used classifications. Users identified the need for administrative boundaries, regional classifications but also local classifications which would accommodate area definitions which do not match the administrative classifications. Health authorities in the UK would be an example.

How the geographic coverage of a dataset is described will also determine how the search operations of a metadata service will operate. Many users expressed the wish to search geography by gazetteer entries.

Themes or type of data

Users had a very wide range of information needs. The types of data extended from topographic map data and large scale cadastral and map information produced to by national mapping

agencies through to a very wide range of data sets from government and non-government sources.

For many users the power of their GIS was in the integration of data from different sources for comparative studies or complex modelling. Often it was necessary to source data from various producers or data owners to meet the needs of their application. In this context it was important to have a metadata service that contains a very wide range of geo information including topography, all man-made features, environmental data, sub-surface and atmospheric information, administrative boundaries, postcodes and address data and census and other socio-economic datasets.

5.3.2 The data descriptions must have a minimum level of detail which is available for all datasets, although this does not have to be very extensive.

Those interviewed were aware that it would be very difficult to achieve a metadata service that had both breadth of content and detailed descriptions of every dataset. While a number expressed the desire to have both detailed data descriptions and the a comprehensive metadata listing, when asked which was the most important, it was clear that a detailed description of each dataset was less important than having a metadata listing with comprehensive coverage of all datasets available.

Users where then asked to determine what would constitute the minimum level of data description for a metadata service. The following criteria were identified as essential within any product description for a metadata service:

- Dataset name or brand name
- Type of dataset or theme of content
- Geographical coverage
- Nominal scale
- Currency and quality
- Cost
- Owner or supplier
- Data format

<u>Dataset name or brand name</u>

This was seen as becoming more important as more and more users were identifying datasets by the publisher or reseller and the brand name that was associated with that product. Product names were also viewed as potentially indicative of the content.

Type of dataset or theme of content

The metadata record should include not only the theme or data type as classified in the metadata service but a short general description of the product in simple, non-technical language.

Geographical coverage

The area covered by the dataset should be given, preferably in both a textual form and a graphical form. Geographic co-ordinates were seen as less important. Graphic indexes were thought to be very important in cases where datasets were supplied in blocks or subsets such as for imagery or series mapping.

Nominal scale

Users believed it was very important to have some understanding of the nominal scale of a dataset, even though digital data can be viewed at any scale. This was important so as to indicate the source from which the database was derived or the intended viewing scale of the dataset, thereby giving an indication of content, level of detail and generalisation.

A couple of those interviewed suggested that datasets should be marked with a 'warning' as to the publisher or owners maximum intended scale of use. This would prevent those with limited geo-information experience from acquiring two datasets at very different nominal scales and then trying to combine them without some awareness as to the problems this might cause.

Currency and quality

Users identified the need for currency information, at least at the level which would indicate if a dataset was being constantly updated or if it has a fixed publication date and when that would be. More detailed descriptions of the currency of the content and the sources from which it was derived was desirable but not seen as important in the context of the simple metadata description.

Some users, particularly those involved in combining datasets from many sources, wanted to see a level of description relating to quality. What this actually meant to the user was difficult to define, except that the key issue was that users often acquired data that was not of a resolution or consistency that met their expectations. One user suggested that suppliers might be asked to verify what Quality Assurance procedures had been adopted during the creation of the dataset.

Cost

Cost was an important issue for users. It was seen as a determining factor in many cases as to whether or not a dataset might be used for a particular application. Linked to this was the need to understand in very simple terms the licensing arrangements, for example is there an annual payment, or if use of the data was restricted in a particular way.

Owner or supplier

The final information required by the user in a simple metadata description is the contact details through which further information could be obtained. This should include telephone, fax and email, with a contact name where possible, and other services that data suppliers might offer such as web sites where data could be viewed and samples downloaded.

Data format

There was a wish to know about the format the data would be supplied in. While it was recognised that the GIS industry was moving to open standards for data transfer it was felt this would take some considerable time to achieve and for many users data transfer was still a considerable cost in any project. It was therefore important to know what data formats a dataset could be provided in and where there were several formats, if there were any differences in the data content.

5.3.3 A metadata service must be constantly maintained and updated to provide users with the latest information about data availability.

For most users one of the main concerns was that while they might not use a metadata service very regularly, when they did, there was the confidence that the database was as current as possible.

Typically users who expressed the need to use a metadata service thought they might access it as little as once every few months or when new projects were being planned. However confidence in the currency of the service was seen as very important.

To this end it was clear that users were interested in the metadata service stating information about the currency of the records it contains and when they were last updated. They were also interested in having functions that could indicate which dataset entries had changed or were new entries over a defined period of time. Some users expressed an interest in having a service that could identify new or changed dataset records since their last search.

Users were concerned that a large comprehensive metadata service could become very difficult to search, and as with some on-line library searches, lead to either very few 'hits' where the search was very tightly prescribed or many 'hits' where the search was very broad. A concern was that users would not wish to constantly re-enter lengthy searches to check on the currency of the information available, but would like to have options to store searches and possibly re-enter these searches automatically on a user-prescribed time cycle.

5.3.4 It is desirable to have a very detailed description of a dataset if the metadata service is to be used to select suitable datasets for particular projects or applications.

For certain applications users would like to have much more detailed data description, such as has been developed for the GDDD. These descriptions would allow users with a stronger GIS background or more technically-demanding requirements to make more valuable assessments as to the appropriateness of the dataset. Therefore several users thought that for certain types of data more detailed descriptions would be useful.

Unfortunately the types of data that were considered for a more detailed data entry varied depending on the users application. It was argued by some that such descriptions should be restricted to 'core' datasets such as those produced by the national mapping agencies to ensure the broad spectrum of users for these datasets are well provided for in terms of data descriptions. However several users wanted the detail to be for those datasets which were more limited in their likely user base or had a much more complex history.

As to what was considered appropriate at the more detailed level, few users had much experience with metadata standards and their views on what constituted a detailed data description depended on whether or not they were users or data suppliers and their level of GIS expertise. Some users thought the GDDD type of data description based on CEN TC 287 was

adequate while others wanted more technical specifications such as a data supplier might provide to a customer, including detailed transfer data formats.

Most users recognised the problems in collecting such large amounts of data for a wide comprehensive metadata service. When asked if they would prefer the detail or the breadth of information in a metadata service nearly all preferred to see breadth rather than depth.

5.3.5 A metadata service must include descriptions of data from governmental agencies, non-governmental organisations and commercial data providers.

In the context of creating a broad metadata service covering as many geographically related themes as possible, those interviewed expressed a need to have data from every source included. A concern was that a metadata service should firstly not focus on only one particular group of data owners but must be inclusive of all government departments and agencies, both national and at the European level, and secondly should include local government data and data from non-governmental organisations and commercial companies.

While different countries in Europe have different policies towards the dissemination of geographic data and the role of commercial companies varies from country to country it was felt that any metadata service should be structured to provide incentives to all these groups to participate in the service.

5.3.6 A multilingual interface to any metadata service is necessary if it is to be accessible to the broadest range of geo information users in Europe.

For those interviewed about a European wide metadata service language was an important issue. While many recognised that a certain level of GIS users were able to work within English, most users today and those who were likely to be geo-information users in the future, would require a metadata service in their own language.

While dataset records might be held in the language of the publishing or reselling agency, it was necessary to ensure the interface was in the users language. Most users thought that language would be a major barrier to its usage, to the point where the service would have little or no impact on the growth of GIS in those areas whose language was not included in the service. Some exceptions were noted where dual languages are more commonplace.

It was also thought to be important that a metadata service should be able to handle a database entry from more than one source in more than one language. This would accommodate those commercial organisations creating datasets across national boundaries and using national revellers as the sales channel in particular countries. This would of course considerably increase the cost of an such service.

5.3.7 The interface must be non-technical, simple to use and largely intuitive in style.

A key complaint by users of existing directory services, not necessarily geo information metadata services, was the complexity of the interface which users were presented with. It was felt that a metadata service which was to be widely used must have a simple and easy to understand interface.

The results from the OMEGA project showed that a GIS front end to a metadata listing was not a suitable interface for many users, hence a more multimedia environment was adopted. The key issues were the need for a non-technical interface that was simple and intuitive.

It is however difficult to match the requirement of a simple interface with the more complex geographic and theme search requirements expressed earlier. The interface of a metadata service was viewed by many as a key barrier to successful adoption by many users on a regular basis and requires extensive study to ensure it is properly configured for a European service.

5.3.8 The metadata interface must allow comparison of different datasets covering the same geographical area.

For some users a full metadata service was considered to include functions that would allow comparison of datasets for the same geographical area. These functions included indexing to show overlapping coverages and gaps within coverages. This was felt to be particularly important for large scale databases which were still not complete and for data such as aerial photographic surveys.

The simplest form of this comparison would be visual with the ability to view more than one graphics window showing an index map for a particular data set. This was seen as important for those interested in small areas and for datasets with irregular boundaries.

More complex forms of this comparison involved levels of GIS functionality upto and including

overlay of index maps in some form for comparison. Those interviewed agreed this was technically difficult to do and not practical in the context of a large metadata service for Europe.

5.3.9 The metadata interface must allow comparison of datasets with the same theme or similar content.

One of the comparisons thought to be most valuable in a metadata service was for users to be able to review datasets with similar content or the same theme classification for the same area. The purpose being to provide the user with the ability to make judgements through the metadata service as to the most appropriate dataset for a particular application.

With the increase in release of government data and the duplication of data collection across departments and agencies, and with the increase in commercial data collection it was felt that users would be faced with more choice in data over the coming years. A metadata service if it was to be useful should provide functionality that would allow the user to make comparisons. These were again considered to be both visual and through more complex GIS functionality.

An example of the importance of this requirement was raised by users who use image-based data, especially aerial photography, which is not currently covered by the CEO metadata service. Various coverages of different date, scale, image quality, cost and supplier are available and users would need to be able to make a choice on the most appropriate for their application. There are some limited metadata services for such imagery but with the growth of image-based GIS solutions, there is a perceived need to include such options in a metadata service.

5.3.10 A metadata service must contain either sample images for datasets or sample digital data that can be downloaded by the user.

Most users wanted to have access to data samples on a metadata service. These would either be in the form of simple graphics which could be viewed or possibly downloaded, but only as image files and not geo-referenced data sets, or would be fully geo-referenced sample datasets.

The ability to access data samples as well as viewing sample data sets was important to some users. It was also important to commercial data suppliers who saw the option of sample data being made available through a metadata service as being a useful component of their sales and marketing activity.

There are some examples of data suppliers making sample datasets available on-line and this is likely to increase as a sales and marketing channel. It was felt by some of those interviewed that to encourage active commercial involvement in a European metadata service it would be necessary to have this type of sample download functionality.

There was some discussion and concern about the technical issues surrounding the use of graphics and downloadable images or datasets. In particular files size, access times and image quality on screen were considered important. The key was that any metadata service could display datasets to show the level of detail and quality to the wide range of users who might access the system. Image data and files were of particular concern.

5.3.11 A comprehensive metadata service should include a catalogue of paper-based maps as many users require geographic data not currently in digital form.

While all those interviewed were existing GIS users many expressed concern that a metadata service should not be restricted to digital data only. It was felt that while the conversion of analogue data to digital data was progressing quickly across Europe there was still a considerable amount of data that had not been converted and much that would not be.

For many projects and applications users still need to undertake some level of data capture and for those applications which are undertaking some level of comparative studies with older data, much is still map based. It is therefore necessary to have access to some level of paper-based data at least for the foreseeable future. Several users wanted to see a catalogue of paper map products included in the metadata service.

This issue was of particular importance to those agencies looking beyond Europe, to either work at a global level or to focus on regions outside of Europe. There was in this context a keen interest to see a metadata service developed that was global in structure at least, if less so in content.

Results from the OMEGA project, which included a market research study into the inclusion of paper maps products into the catalogue concluded that the map publishing industry at present regarded the CD-ROM market as too small to warrant high levels of investment at present. As such it might be more difficult to get this sector to contribute to a metadata service, at least in the short term.

5.3.12 The metadata service should be available on-line but with both CD-ROM versions and paper version available for those with no on-line access.

While many of those interviewed were comfortable with using a purely on-line service, there were those who had no access to on-line services or would have to rely of access through a centralised library or support function. While this was viewed as likely to change as more systems became linked direct to the Internet, it was also thought that this might take many years, particularly in local government.

Realistically therefore it was felt that a metadata service that was to get to many GIS users and be widely adopted as a key resource needed to be supported by off-line versions of the database, published on CD-ROM, and possibly supported by paper versions of the database. The general consensus was that quarterly CD-ROM versions would be worth having and many users wanted an annual paper directory product.

This is not necessarily borne out by the US experience where CD-ROM based directory products in the GIS market have not been widely adopted. Paper-based products are still purchased but sales are in decline and on-line services are growing strongly.

5.3.13 Users anticipate that any metadata service will not be comprehensive and thus should contain search capabilities to review other metadata listings available through the Internet.

Some of those interviewed expressed the concern that any metadata service would only contain a proportion of the available datasets, possibly only a small proportion. It was also felt that even if there was a European initiative to create a centralised metadata service that there would still be organisations that ran their own metadata directories.

It was felt important that a metadata service should be able to link directly into other services or at lease provide a search function to access other services. The CEO model of directly linked services was considered a possible model in the form it operates for the Earth Observation community but for many geographic information providers the imposition of some level of standard metadata format was felt to be too restrictive, especially for those organisations who have already created a metadata listing.

An alternative suggestion was the inclusion of web browser functionality in the metadata service to allow the user to search across other web sites and to move onto those sites should other datasets be identified. The option of a metadata service which is primarily a catalogue of other

listings and services may be more attractive to commercial service providers, who might consider the data content collection process and subsequent maintenance process as being economically unviable.

Whether it would be possible to provide a comprehensive metadata service with such a structure is unlikely. However some of those interviewed suggested a hybrid of the two approaches might work best.

5.3.14 There should be the facility to order data on line, with the potential for on-line delivery of datasets in the future.

Linked to the ability to view data on-line and to download sample datasets, many users saw the long term potential of a European metadata service as providing an on-line ordering mechanism. This presents several problems to the service provider, not least of which are the technical issues of linking to a data owner and the licensing and copyright controls on being the data supplier.

However, users expressed an interest that any metadata project for European should at least consider these issues for the future enhancement of the service.

5.3.15 Measures of data quality should be included in the data descriptions.

Most users wanted to have measures of data quality included with the data description. The key concern was the ability to judge the appropriateness of the dataset for the application but it was not clear from most users exactly what data quality meant and how it would be shown.

For most users a qualitative statement was wanted giving a range of information, including:

- sources used
- control and geodetic information
- period over which the database was collected
- accuracy statements, both horizontal and vertical
- classification accuracy
- quality assurance procedures

Some users expressed an interest in more quantifiable quality statements. However most recognised that data suppliers had no standard approach to accuracy statements for their users and it would be difficult to include these in a metadata system, especially if these were intended for comparative purposes.

5.3.16 Data descriptions should use non-technical language.

A common issue for those who have GIS solutions implemented across the organisation is the use of technical language in the systems and the need to educate users to this. Many expressed the hope that a metadata service would use as little technical language as possible and would present information in a non-technical format wherever possible.

One suggestion was that a metadata service should include a dictionary or glossary of terms to help users with little or no expertise in the subject area understand the directory entries.

5.3.17 Search mechanisms must accommodate the experienced geo-information user and the novice entering the market place for the first time.

Any metadata service must include a range of search parameters that accommodate the wide range of user experience from the professional GIS user to the most casual and inexperienced geo-information user.

The key search criteria were identified as 'geography' and 'theme'. However users identified that in order to successfully use the search criteria the metadata service would need various search mechanisms.

The geographic searching options were identified as:

- Graphic search based on index map with point and click options, preferably showing existing coverage of datasets.
- Geographic co-ordinate search with the user inputting geographic co-ordinates for an area or point of interest.
- Gazetteer search with the user inputting geographic place or feature names or selecting from a listing. If a listing was used most users believed a hierarchical structure would be required based on some notion of the size of the area under consideration in order to prevent the list becoming too extensive.

A key potential weakness in the geographic search option was identified by users interested in small areas where national or regional datasets would be offered as covering their area of interest. Many of these datasets would be at a scale that was not suitable to the application. These users suggested that restricting the geographic search by some notion of scale may be necessary.

The ability to search by theme or content type were identified as key search parameters. However it was recognised that experienced users would wish to be more focused in their search criteria, defining more specific theme or topic subjects.

As an example of this the following was discussed with users and used to demonstrate the need to ensure the themes were both comprehensive and flexible, probably with a level of hierarchical structuring.

An organisation in the UK interested in road networks for Europe might use the metadata service to source road networks on a country by country basis. The search would depend on their existing knowledge and experience, so that for the UK the organisation might be aware that road network data is available through the Ordnance Survey, that the data bases are called OSCAR and that these come in three versions depending on the detail required. For France they might be aware that the Institut Geographique National is the national mapping agency but know nothing about any road network datasets produced by them and for Germany not even know who the mapping agencies are. A metadata service search must accommodate simply and effectively this type of user problem.

In services such as GDDD the themes are not structured hierarchically, leading to a subject listing which is long and difficult to search. For a more comprehensive metadata service it would be necessary to divide the theme list down into more manageable groupings.

A further aspect to the requirement for broadly specified search mechanisms is the idea that search times are important to the users perception of the value of the service. A number of those interviewed expressed an increasing frustration with the amount of time it takes to find and download data from the Internet. A key specification criteria should be the inclusion of target search and report times within a metadata service.

6. Review of some existing metadata services

6.1 Introduction

Following the definition of user requirements derived from interviews, questionnaires and reviews of previous studies, the study undertook to review a number of existing metadata services in the context of these user requirements. The objective was to assess how well or otherwise these services met the perceived needs of GIS users.

A number of sites were evaluated at a fairly simplistic level, followed by a more detailed review of three of the main metadata services currently operational in Europe. The lessons learnt from this review were used to re-enforce the user requirements list and to make recommendations with respect to the issues raised in the project objectives.

A third part of the review was an analysis of three commercially published data catalogues from the USA. These were selected because they contained extensive lists of data sets, two of which were published by software vendors in part to promote their software products. This was undertaken to show a different method of presenting and cataloguing spatial datasets, and to assess what value these might have over an on-line metadata service.

6.2 Review of on-line metadata services

The first part of the review was an analysis of the content and usability of a selected number of metadata services accessed through the World Wide Web (WWW).

The WWW sites selected differ in the type of service offered, the type of information available, and the type of service. Whilst the use of brochures for sales and marketing and technical metadata are well defined, the exact status and role of on-line metadata services is still uncertain. This chapter summarises the findings from the survey. The detail is provided in Appendix 3.

6.2.1 Methodology

The review considered in detail 18 WWW sites that offered metadata information on geographic data, for Europe, North America and Australia. These are given in Table 6.1.

WWW sites tend to have one characteristic in common; they are all distinctive. This makes their comparison quite difficult, unless based on reasonably generous criteria. Each of the sites was therefore evaluated using the criteria summarised in Table 6.2.

1.	MEGRIN (France)	http://www.ign.fr/megrin/gddd/gddd.html
2.	SNIG (Portugal)	http://snig.cnig.pt/
3.	MMH (Finland)	http://www.nls.fi/index_e.html
4.	NSDI (US)	http://www.fgdc.gov/
5.	Wisconsin LI System (US)	http://badger.state.wi.us/agencies/wlib/sco/
6.	AUSLIG (Australia)	http://www.auslig.gov.au/welcome.htm
7.	SINES (UK)	http://www.ordsvy.govt.uk/sines.html
8.	ESRC (UK)	http:/dawww.essex.ac.uk/
9.	NSD (Norway)	http://www.uib.no/nsd/
10.	GISDATA (UK)	http://www.shef.ac.uk/uni/academic/D-H/gis/gisdata.html
11.	AGI (UK)	http://www.geo.ed.ac.uk/agi/agi.html
12.	GISIG (Italy)	http://gisig.ima.ge.cnr.it/
13.	EUROGI (EU)	http://www.frw.ruu.nl/eurogi/
14.	Centre for Earth Observation (Italy)	http://ewse.ceo.org/
15.	CIESIN	http://www.ciesin.org/
16.	Ordnance survey	http://www.ordsvy.govt.uk
17.	Alexandria	http://alexandria.sdc.ucsb.edu
18.	CEOS	http://www.smithsys.co.uk

Table 6.1: World Wide Web sites reviewed as part of the GI-META Study

6.2.2 Discussion of On-Line Services

The primary focus of the review was to identify the level of detail contained in the databases and the ease of access to data sets of interest to the user. From the number of sites visited an overall view on their ease of use and accessibility to the data was determined.

1. **Name** of service.
2. **Organisation** providing the service.
3. **Internet Address**.
4. **Focus** - the main focus of the service in terms of the data being described.
5. **Product description**: A brief overview of the structure of the metadata service.
6. **Target group**: Whether the target group is the public sector, private sector or academia.
7. **Delivery**: The main features of the service were classified as follows:
   - Simple listing of products.
   - Yellow pages, i.e. structured list of organisations and products.
   - Hierarchical structure, richer in levels than the Yellow Pages.
   - Searchable databases which are richer still and allow more horizontal movement than hierarchical models.
   - Open shelf browsing, where the data itself is on display rather than a description of it.
8. **Searchable engines** which go beyond the 'Find' facility of the WWW interface.
9. **National links** indicating whether existing linkages to other national databases or metadata services are an important feature of this service.
10. **International links** indicating whether existing linkages to other international databases or metadata services are an important feature of this service.
11. **Degree of organisation** which indicates the degree of internal organisation of the service.
12. **Degree of Homogeneity** of the data.
13. **Currency** depending on the extent to which the metadata service is kept up to date.
14. **Language** being the main language of the service.

Table 6.2: Criteria used to evaluate WWW metadata sites

The overall impression was that metadata services vary enormously. There is little commonality in approach to either the type of information held on each data set or on the way the databases may be searched.

A number of key issues were identified from the review. These may be summarised as follows:

- Level of detail
- Search methods
- Language
- Flexibility

6.2.3 Level of detail

Overall the level of detail available for each data set varies quite considerably across the range of metadata services. However, a consensus seems to be developing to limit the amount of detail shown on such services to a basic description of the data. Should the user require further information about specific technical issues or products they are directed to the information owner.

As an example the SINES metadata service provides the user with a single screen of information on the data set. It is easy to read, provides enough background for the user to pursue if there is an interest in the data set but did not take long to upload. A screen of this data set is reproduced as Figure 6.1.

Not all the services provide such compact and easily digestible results from a search. The GDDD database is complex and provides a much greater range of information on each data set. There is an element here of assuming the user has some understanding of digital map data and can therefore interpret much of the information given.

There are without doubt serious limitations for on-line services at present. Transmission and display of images such as map samples or index maps is still relatively slow and liable to cause some level of user frustration. By comparison, browsing through a conventional paper catalogue or CD-ROM data source is quick and efficient. However, there are technical advances which will improve Internet access over the next few years including greater access to faster modems and cable networks. These will improve information retrieval times for sample data and make it a more viable alternative to paper-based catalogues. As an example GeoInformation International now provides sample images from its Cities Revealed product range through its web site.

It is clear that the size and complexity of the data set entry is key to both ease of access and interpretability of the record. This is particularly important in a large metadata service where there might be thousands of records and any search may yield a number of data sets that meet the search criteria.

Figure 6.1 Data set record from the SINES service.

6.2.4 Search Methods

It is clear from the metadata services used that ease of use would be a vital factor in the wide adoption of the service by users. In most cases ease of use related to the search method adopted.

It is apparent from the WWW sites that there is a need for two types of search, one geographical and the other based on theme or subject. A third and fourth may be added to these which are organisation and currency, although these were considered to be less critical. This finding is in-line with one of the main user requirements specified in section 5.2.

In large metadata services which potentially will hold many tens of thousands of data sets the ability to ensure that a search which may not be very tightly focused can identify all the appropriate records is critical to establishing user confidence in the system.

The first part of this search capability is to use geography as a means to prescribe an area of interest or a level or scale of interest. For example a user may only be interested in a country wide coverage, a regional coverage such as Benelux or a more local area. These may be defined by a name or a set of geographical co-ordinates either typed in or more likely defined using a box or area tool on a map.

The problem with only identifying an area is that all global coverages will contain detail for the area of interest even if that is very small. Geographical searches may therefore have to take into account a level of detail or scale which is appropriate to the search. As an example a search for road data sets of the UK should not necessarily identify a global road network data set such as the Digital Chart of the World.

In most of the examples viewed the geographical search capabilities were fairly limited and would therefore be likely to lead to problems in providing a suitable result from a search, particularly to a large database. The NSDI service was the best example of a geographical search.

The second area of search is likely to be by theme or subject. In the examples reviewed this type of search was fairly limited. The problem is that it is very difficult to create a set of themes that can accommodate a wide range of users needs. Inevitably the themes or subjects need to be structured in a form that allows for both unfocused searches and very specific searches.

An example of this problem might be at one level the user who knows he wishes to use a digital elevation model with a resolution of 50 metres and height accuracy of +/- 5 metres, and at the

opposite end of the scale a user who knows only that he wishes to utilise heighting information in his GIS.

The most successful search capabilities will be those that can integrate these two elements and allow the user to interrogate the database using organisations or product names. The SINES database, discussed more fully below, is an example of where these three search criteria are included.

6.2.5 Language

Some WWW sites allow multi-lingual services to be provided, but English was generally used as the de facto. Even some of the national sites had a large component of the database reproduced in English. The user requirements identified a need for services to be provided in a users first language as part of the process of stimulating usage. Clearly cost is an issue in establishing a multi-lingual service and for those services reviewed a multi-lingual format had not been a key criteria in the design.

6.2.6 Flexibility

The presentation of metadata information on WWW sites allowed for different national policies on such issues as copyright and charging to be presented without too much complication but their comparison required further study, due to lack of standardised terminology. It is noticeable however that most of the metadata presented is national and not pan-European.

WWW sites can offer considerable flexibility in searching, not only at specific sites as shown above, but for searching across different sites. However, only a few sites have the right balance between detail and content, and having browsed a well structured search engine as illustrated above, transfer to a different site is likely to present the user with a totally different environment.

6.2.7 Conclusions

Given the wide range of services considered, it is to be expected that the degree of user satisfaction also varies considerably depending on what one is looking for. If one is looking for tightly defined information such as "who is active in the field of GIS Diffusion" or "what maps are available from the national mapping agency in Greece" the services reviewed perform reasonably well as the information is stored in a mono-thematic structure and/or the whole database is small.

However, in most cases, users will simply not know what to look for in very precise terms and would be expected to use a metadata service to help identify what is available for a given area and/or topic and who has it. On this very simple criteria, the services reviewed generally fail the test.

The one service that comes closest to the user requirements is the NSDI. It limits searches first by geography and then by what is available. This is absolutely critical. As more and more databases become available, the ability to use geography as a searching mechanism must not be underestimated.

There are obvious reasons why the NSDI service score better than other services, such as having a common geographic framework for data across the US, but clearly there are also important implications from a European perspective as to the resources needed to geo-reference the metadata and enable the type of searching contained in this service.

In Europe, the nearest equivalent type of service is EWSE, the information exchange of the Earth Observation Community, based at the CEO, Ispra, which is reviewed in Section 6.3.1 below.

A key finding of this review is also the difficulty of creating meaningful search criteria. It is necessary to search by both geographical location and some theme or topic as a minimum. The two more detailed reviews are examples of how this has been attempted but not necessarily successfully. The importance of the search criteria is that they need to be structured in such a way as to allow the use easy access to the very specific data sets they are looking for even when the query is poorly defined. To achieve this it is necessary to have a tightly defined geographical search capability and a well structured set of themes or topics.

6.3 Detailed evaluation of selected metadata services

The second part of the review process was to look at three on-line sites in more detail and to consider some of the specific features that were highlighted in the user requirements.

The three sites selected were the EWSE service from the Centre for Earth Observation, the SINES service from the UK Ordnance Survey, run on behalf of UK Government, and the MEGRIN GDDD service.

The review focused on the type of data held in the service and the search mechanisms available for extracting information about appropriate data sets. All the systems were in English but were very different in nature. EWSE is a much more comprehensive database including reference to companies, products, academic and educational activities, services and new developments as well as a data directory. The SINES service is a simple on-line listing of UK Government geo-referenced data, while GDDD is a more detailed pan-European listing of datasets available through the national mapping agencies.

6.3.1 European Wide Service Exchange Software

European Wide Service Exchange (EWSE) Software System is published by the Centre for Earth Observation. The stated objective for the EWSE is as follows:

"The EWSE is an on-line information service for the Earth Observation community. Its goal is to help new and existing users find data and services and to allow service and data providers to find users. In this way the EWSE hopes to expand and promote the application of the wealth of data gathered by satellites. "

The EWSE site is much more than just a metadata directory in that it provides a whole range of services in addition to metadata, for the earth observation community based on an extensive database, including the following:

- A virtual Trade Show which organises services according to subject and presents them as a virtual shopping mall. Each floor of the Mall concerns a different subject. You may enter a shop, browse around, request data, run demonstrations, subscribe to update material and so on.

- Organisations may be found by Yellow Pages inquiries and users may be found by White Pages searches. In effect a range of powerful database searches are offered similar to a normal relational database.

- There is a calendar of EO events, listings of Educational Courses, Educational Events, and Educational Resources, in the form of educational support material. A document library provides a repository of Earth Observation material in machine readable form.

Use of the service is essentially free, although there is a registration system, which users may be require to complete for certain operations such as searches of on-line catalogues which may require later posting of results. Registered users are provided with a customisable personal page. Registered organisations are provided with free advertising and customisable organisation pages, or "virtual" shops.

The service was launched in September 1995 but has undergone continuous modifications and enhancements through 1996.

The metadata component of EWSE is found through the IMS catalogue which includes 17 product categories. For each of these data sets there is information such as archive, IMS data set ID, IMS parameter and the date the data is available between. The description also includes a short statement about the satellite system. The information obtained for a search can include details about processed and unprocessed or raw data and derived datasets. These categories are regarded as generic categories within the Earth Observation domain, but are very specific to this field of study.

Although EWSE is a single large database it also provides access to users own catalogues and data listings through the virtual trade shows. It is therefore difficult to determine the actual number of data sets which may be located through EWSE. At present there are six other databases linked to EWSE.

Currency and accuracy of information obtained through EWSE are the responsibility of the organisation who posted them. Information is updated continuously but the service provider does not require any quality statements for datasets.

There are no quality checks on the information available through EWSE other than a coarse check on formats. This is due to the difficulty of imposing standards on data suppliers, especially where EWSE is accessing other databases. There is also a policy issue related to the imposition of standards, in that it is difficult for CEO to impose quality standards on data suppliers.

The metadata is updated continuously and the whole EWSE service is under going a series of charges and enhancements through 1996 and into 1997, to improve the value of the service to users.

The metadata records depend on the format and metadata structure provided by the individual suppliers or organisations. The data entries do not conform to CEN TC 287, although CEO is currently reviewing metadata standards with other European organisations to select the most appropriate for the Earth Observation community. At present a number are under review such as the Catalog Interoperability Protocol (CIP) and the US Geoprofile standard. The CEN TC 287 standard is not one of those being considered at present, even though a recommendation of the standards report (Reference 6) was that CEO should adopt CEN TC 287 as a standard.

EWSE is an on-line dynamic database. There are several views of this database offered via the WWW interface. One of the main access mechanisms is a free text search facility that allows the user to query the database by typing in a query that may then, optionally, be expanded by the system to include (user selectable) synonyms and related words from an Earth Observation (EO) thesaurus.

EWSE allows users to access a number of IMS catalogues. The gateway can run queries in batch mode, to save searches, to repeat searches, and to schedule searches for execution at regular intervals.

One way of searching the data catalogue or any other part of the database is by geographical location. Users can select an area of the earth at different resolutions and list entries for that area. Alternatively you might select a country and look for products, services, or users in that country. The IMS catalogue can also be searched using the project themes or data types.

The system is simply designed and is easy to navigate through. Like many web sites it is very hierarchically structured and therefore there can be many pages which have to be viewed before finding the appropriate one. This can be very irksome if you have a slow modem or Internet traffic is high.

The types of international catalogues currently held is at present limited. However the links to other organisations web sites and therefore linked catalogues has a lot of potential for building an extensive metadata network without the need for cataloguing all the data in one organisation.

At present there are about 1,200 registered users of EWSE and a further 200 organisations or projects registered.

Considerable progress has been made over the last few months which has moved the service from being a useful collection of a very wide range of data sets, in which it was easy to get lost,

to a more structured approach. EWSE has started to provide a more structured interface for searching the data holdings.

EWSE is the most complex metadata service reviewed in that it is not a single database but includes links to other databases. This is why searches are essentially undertaken in batch mode. This method of searching may be regarded as slow and frustrating for some users but is necessary because of the technical problems in linking databases across Europe. The key technical problems that had to be solved were the search protocols and the definition of the attribute set.

In terms of the search parameters, there is still some way to go before comprehensive geographic searching is available. Although the service enables geographic searches, the mechanism for defining areas of interest is very coarse and in some cases there seems to be problems between the subsequent result of a query and the area selected.

Like all the metadata services reviewed this is a free service to everyone with Internet access and for those who register the full search capabilities are available. The opinion of those involved in the EWSE service is that there would be no demand, at least in the short term, for a service where the user pays.

Figure 6.2 Sample pages from the EWSE service.

6.3.2 Spatial Information Enquiry System

The Spatial Information Enquiry System (SINES) is run by the Ordnance Survey of Great Britain, on behalf of Central Government Departments. The SINES project is co-ordinated through the Inter-governmental Group on Geographic Information (IGGI) run by the Department of the Environment. It was launched in November 1995, although there was an earlier development project.

SINES is described as "a meta-database describing over 500 spatial data sets held in UK government departments and agencies". The purpose of SINES is to "provide summary information concerning the data sets that satisfy the criteria given including a map that indicates the cover. It will also give contact information and, where appropriate, the WWW URL of the organisation that owns the data".

SINES only includes data sets owned or managed by UK government departments which have a spatial component. Most data sets apply to a whole country - England, Northern Ireland, Scotland, Wales. A minority apply to specific English counties only. The database contains just over 500 records.

The purpose of SINES is to provide summary information concerning the data sets it contains and to include a map that indicates the cover. It also gives contact information and, where appropriate, the WWW URL of the organisation that owns the data.

The data record is displayed as a single scrollable window which contains the following records for each data set:

- Title
- Purpose
- Source
- Start Date
- End Date
- Data Update Frequency
- Explicit Spatial Reference
- Other Means of Referencing
- Smallest Spatial Unit
- Data Availability
- Organisation

- Department
- Name
- Title
- Address
- Telephone
- Date Summary Information was Entered
- Date Summary Information was Amended
- Contact URL
- Coverage Map

The database contains information relating to the last time the metadata record was updated but there is no information about the currency of the actual data. SINES entries do not use either a UK metadata standard or the CEN standards.

The user of SINES may search the database using three basic criteria: geographical area, theme or subject and government department or agency. This allows the search to be very specific if the user is looking for a particular data entry. However given that the database only contain some 500 entries, a specific search request is likely to yield few or no results. It is noted within the service that "For enquiries regarding English coverage, best results will be obtained by not restricting coverage at all".

SINES allows the user to search from a list of geographical areas. There are 17 options which include some counties but the list is not inclusive of all counties for the UK. The database can also be searched using a total of 846 themes or subjects. The range is extremely broad and was designed to cover all government activities.

While the theme list is very comprehensive, it is a simple listing without any hierarchical structure. This can make it difficult to use as some categories are very specific to geo-information, for example *contours*, while others are very generic, such as *abortion*. The list also includes general categories such as *animal* and many subsets of that category. The length of the list also makes it difficult to find appropriate themes to search by, especially in the way it is presented as a scrolling window which only displays a few themes at one time.

It is also possible to search by government department or agency within SINES. A total of 50 are included.

SINES is a very good example of a simple metadata listing, giving only a very high level data description and providing simple search routines. Despite the problems of extracting appropriate

data sets using the theme search option, it is easy to use and given the size of the database, provides a quick response to users enquiries over the web.

The structure of the presentation is simple and therefore easy to use and understand. It is easy to structure a search, even though the advice is not to limit the search options.

Since its launch in November 1995, the SINES web site has been visited by over 500 users.

Figure 6.3 Sample pages from the SINES service.

6.3.3 Geographic Data Description Directory

The Geographical Data Description Directory (GDDD) is described by MEGRIN, who run the service, as containing information on the availability of digital geographic information in Europe. It consists of contributions from more than 36 National Mapping Agencies (NMA), and supposedly contains all the data sets published in digital form by the National Mapping Agencies.

The GDDD exists as a relational database and an Internet service. The core database is designed to meet the needs of the NMAs, who use the GDDD for maintaining the geographic metadata and for analysing European data. The GDDD Internet service is meant to provide users of geographic information with summaries of the data sets available at the NMAs. It was launched in 1995.

It is maintained on a regular basis, with dataset entries updated as they are provided by the national mapping agencies.

The GDDD contains information on the following:

- Data quality
- Coverage (geographic extent)
- Distribution (commercial conditions)
- Keywords or descriptions

In the GDDD (hypertext version) the data is organised using three criteria structured into three levels:

- European country
- Supplier or National Mapping Agency
- Product category

The GDDD uses standard product category definitions, of which there are 13. The metadata record conforms to the CEN TC287 Draft European Metadata Standard.

The GDDD is organised into country units. Searches can be made by country for any particular product. Within a country the data are organised under the NMA or equivalent. Searches can

also be made by theme. Within this category the data sets are divided into scale groups and product types. The data sets are cross-referenced by product name.

For the specialist the GDDD (Hypertext version) is easy to use and informative. It does however use a product category classification that could be confusing for the non-specialist. It is structured with an assumption that the user knows the country and product group although it is well cross-referenced throughout.

The detailed product descriptions contain a reasonable amount of jargon which again might be difficult to interpret to the non-specialist. However the GDDD is essentially intended for use by the specialist and can be regarded as meeting their needs very well.

Figure 6.4 Sample pages of the GDDD service.

6.3.4 Results

While the three services selected were considered to be among the best reviewed from within Europe, there were problems associated with them. Overall the issues were as follows:

- Complexity of the search process
- Level of detail on each data set

6.3.5 Complexity of the search process

As mentioned above, the SINES service contains a simple search engine that does integrate geography, theme and organisation. It does not include product name. The search is simple to perform and relatively quick.

One weakness of the search process is the limited number of named geographical regions, although this is in part due to the fact that the data sets contained in SINES are intended to be UK coverage only, and the inability to search by a grid reference or other co-ordinate. This latter limitation is not a particular weakness for SINES given the intended coverage of the data sets but could be critical in a pan European database.

A second and more important weakness for SINES is the theme or subject listing. The listing is extremely long, over 800 subjects, and is not structured in any form. This makes it very difficult to identify particular subjects or sub-groups and to know how data sets may have been classified. Such a list may be used as a keyword search in a product description but this might take a long time to search in a large database. However it would probably be more appropriate if the listing could be hierarchically structured to allow users to identify their own area of interest and define the search as loosely or as tightly as they wanted.

The EWSE service has made considerable progress over the last few months in its search capabilities. The service has moved from being a useful collection of a very wide range of data sets in which it was easy to get lost to a more structured approach. EWSE has started to provide a more structured interface for searching the data holdings. The IMS catalogue which is the metadata component of EWES includes a geographic, keyword and catalogue search. It can be characterised as relatively narrow and deep search, which allows the user to be very selective in the type and number of records retrieved.

Having said that, there is still a long way to go before true geographic searching is available. Although the service enables geographic searches, the mechanism for defining areas of interest is very coarse and sometimes there seems to be little relationship between the subsequent result of a query and the area selected.

What this highlights is not only the need for good search capabilities but the difficulty in providing them. In particular there needs to be more effort put into defining search criteria and how to structure these within a broadly-based metadata service.

6.3.6 Level of detail

The differences between the three metadata services demonstrate the difficulties in meeting the user requirements for a comprehensive metadata service. The EWES service is a series of linked databases which provides a much larger number of potential data descriptions than can be accessed through the very narrowly focused GDDD and SINES databases.

The level of detail in each varies with the simplest records being in the SINES database while the GDDD conforms closely to the full CEN TC 287 metadata standard. The EWES service provides a structured database entry for data from all linked databases with the output including graphic indexes or maps to show coverage of satellite images which can be downloaded and saved. The entry also includes thumbnail images which can be downloaded and saved. The level of detail is however limited, as would be expected given the range of dataset entries that can be accessed.

The three systems demonstrate the conflicting problems of generating a comprehensive metadata service for Europe that contains enough detail for each entry and yet meets the key user requirement of comprehensiveness.

6.4 Data catalogues

The third part of the metadata sources review was to consider the content and format of a selected number of digital data catalogues in order to provide a comparative analysis with the on-line services.

There are a wide range of other types of metadata for the purposes of promoting awareness of various geographic data sets. Many of these are for internal use, while others are promotional material to support sales of data.

The most common forms of these metadata products are:

- **Printed Catalogues**

  Typical examples would be those created by many national mapping agencies which describe the range of maps, digital data and other products, how to purchase, and some indication of copyright and licensing restrictions.

- **Indexes**

  These may show the geographical coverage of the product, for example map indexes for series mapping or flight diagrams for aerial photography. Typically these are large format (A3 or larger) paper products.

- **Technical specifications**

  Covering the formal specification of the product, from map specifications to technical data format specifications for digital geographic data.

- **Data transfer specifications**

  These include specifications for the transfer of digital data between computer systems. A number of national Data Transfer Formats already exist and are being used within the EU, particularly by national mapping agencies. There are a number of initiatives to promote the use of a European Transfer Format and International Transfer Formats under the auspices of the International Standards Institute (ISO), as well as the software industry Open GIS initiative.

- **Directories**

  These are simple product listings, usually compiled by associations or other interested groups trying to provide broad-based information sources to their constituent members.

For the purposes of this simple review it was not considered appropriate to use product catalogues created for marketing purposes by a single data publisher or agency, such as a national mapping agency. Instead a review of more generic data catalogues was considered to be comparable with the on-line services.

Three such catalogues were considered as being representative of the types of product currently available. These were:

1. ArcData Catalog published by Environmental Systems Research Institute (ESRI),USA.
2. MapInfo Data Products Catalog published by MapInfo Corporation, USA.
3. BLR Data Catalog published by Business Location Research , USA.

These digital data catalogues can be divided into two distinct types. The first two are collections of data sets that are in a particular software vendors proprietary format and are sold to enhance the software sales. The last is a data catalogue from a company specialising in the publication of third party data for a particular user community, the Business Geographics community, in the US.

The three catalogues are evaluated on a number of criteria, including their size, content, structure and ease of use (see Appendix 3 for details on each catalogue).

A number of conclusions may be drawn from this simple review. Firstly there is a considerable amount of geographical data that is available in proprietary data formats. In many cases a single data set may be available in many formats, and awareness that these formats exist will be of interest to users, depending on the GIS software they are using and their ability to convert data from one format to another.

Secondly all the metadata data entries are relatively simple but a consistent amount of core data is included although this is often in a descriptive form. Thirdly, the inclusion of a graphic to show the content of the database is a useful addition to the product description.

Lastly, the way in which the data lists are sorted and cross-referenced is key to how easy the catalogues are to use. Both geography and theme are important in searching for appropriate data sets. There is also the need to cross-reference between these so that users may select on both criteria.

The software suppliers are providing their own on-line catalogues (either already in existence or planned in the near future). To provide a comprehensive metadata service it will be important to integrate such data directories.

7. Models for metadata services

7.1 Organisational types of metadata

Research undertaken during this project has emphasised that metadata can be used by organisations for both internal and external functions. Internal functions typically comprise data management functions. Many organisations involved in the management of geographic data will maintain some level of metadata internally about that information. Data management metadata is required in order to know what data sets are held, copyright restrictions, the individual responsible for the data set, update policy and security procedures. They may also include technical specifications for the transfer of the data within the organisation.

External metadata are the sources used by organisations to identify what data sets exist, who owns or publishes it, what is the geographical coverage etc. These types of metadata typically range from catalogues and indexes to structured directories distributed both through CD-ROM and on-line services.

7.2 External Metadata

Metadata created to aid organisations in locating suitable geo-information data sets
can be classified into a three-tier model, as shown in Figure 7.1. The three tiers represent three levels of complexity in the data description:

- Marketing - simple product outline
- Sales - includes details on ownership, content, licenses, costs
- Implementation - technical formats, data dictionaries etc.

These are not mutually exclusive categories but must be viewed as part of a continuum. Part of the GI-META study was to evaluate the complexity of data description required in a broadly based European metadata service.

In Figure 7.1, the arrows indicate the increase in complexity in metadata. The dotted line divides the pre-sales or marketing functions from post-sales or user support type of metadata. The

width of box denotes the quantity of metadata required, and is an indication of the investment required to create the metadata record. In this model the CEN TC 287 standard sits somewhere between the metadata for selling and the technical metadata in that it contains a component of both of these but is not comprehensive for either.

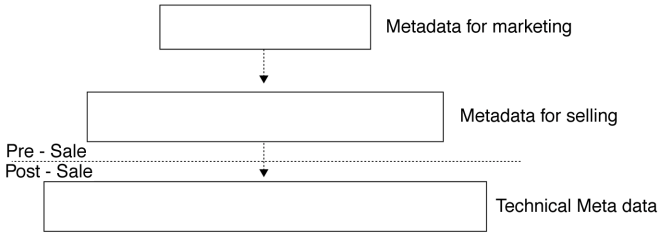

Figure 7.1 Metadata for External Activities

It is also possible to develop the CEO model and to envisage the development of a European wide metadata service which provides a single, easy to use interface and gateway to many individual metadata directories, shown in Figure 7.2. To make this a seamless service, standardised key words and classifications are needed. The dotted box represents metadata sites that are 'virtual' and actually provided by the meta-metadata services. Key elements in making such a service function efficiently are the metadata data formats and search parameters.

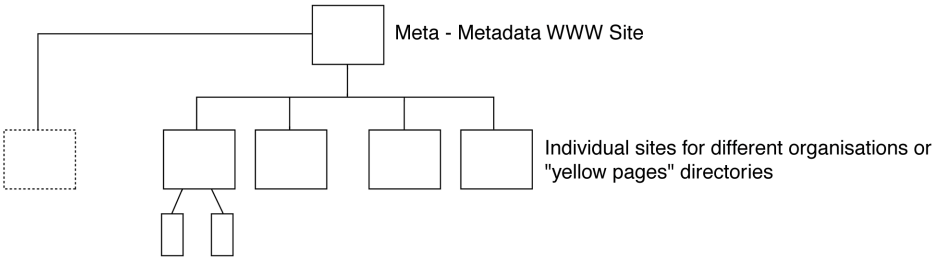

Figure 7.2 Model for a Meta-Metadata Service

A further model is the 'Yellow Pages' or classified type of service which promotes a wide range of products from many organisations as a central resource, see Figure 7.3. The service links to sales functions by fax or E-mail. Some services may also provide on-line supply of data. This may be regarded as more of a marketing and sales support type service rather a technical metadata service.

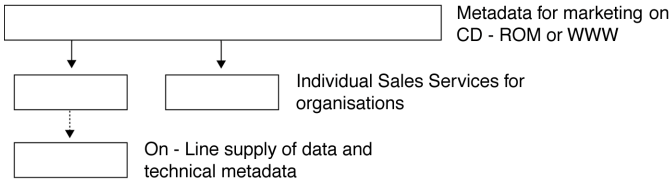

Figure 7.3 'Yellow Pages' type service

To-date, many available metadata services have tended to emphasise technical metadata when inspired by technical, rather than marketing and sales interests. However, marketing oriented services do exist in which the main objective is to identify data sources, and obtain contact details (a 'yellow Pages' approach), often in a paper format.

7.3 Delivery Mechanisms

Current metadata products are paper-based, on-line, or on a magnetic media such as CD-ROM or floppy disc. The last category is further divided into data only CDs, or interactive CDs which combine data and software. For example, most digital map data is supplied with a data dictionary, but it is not interactive, and is intended to be machine-readable for data transformation. In contrast, CD-ROM products such as OMEGA provide software to interrogate the metadata, as do on-line services.

The three categories of metadata identified in Figure 7.1 are shown in Table 7.1 with some indication of the type of delivery mechanism that is currently most commonly used. Potentially however, any delivery mechanism can be used for any type of service. The services that are available vary greatly in quality, and it is fair to state that many on-line services are launched on to the World Wide Web when still under development, with little evidence of quality assurance or standards. Interfaces for searching also vary greatly in quality.

Metadata Type	**Paper**	**Magnetic media**	**On-line**
Marketing	Corporate brochures	CD-ROM, floppy discs	Simple WWW pages

Selling	Product brochures	Data samples, dictionaries	WWW directories and product descriptions
Technical	Product specifications	Transfer formats, data dictionaries	Detailed product specifications

Table 7.1 Common delivery mechanisms used for metadata. The density of shading indicates frequency of use.

8. Business case for a metadata service

8.1 Introduction

This chapter examines the business case for and against pan-European metadata services and the possible form of a future metadata services. It draws upon the findings of the results from the various reviews and interviews.

8.2 Rational for a Metadata Service

There are several arguments that are put forward for establishing on-line metadata services. The underlying belief is that metadata is seen as an important reference for existing and future users of geo information and that with appropriate metadata there should be a growth in the use of geo information technologies.

There are several forms that metadata can be published in, but potentially the most beneficial is through the medium of the Internet using World Wide Web technologies. The benefits for creating an on-line metadata service based on the WWW include:

- Easy access to a broad metadata service
- Simple updating and maintenance mechanisms
- Links into future on-line selection, ordering and payment for data
- Standardised representation of data for comparison of different products
- Mechanisms to effect searches over many databases
- Enhanced public service by the increased availability of government data
- Better marketing and promotion for commercial data suppliers
- A mechanism to integrate information sources

There is no doubt a demand from existing GIS users to know more about what geo- information is available. There is also a need to provide basic information sources on geo-information for the new and emerging users of GIS such as the target analysis and market profiling communities.

Many users have access to, and make use of, existing published metadata sources in the form of catalogues, indexes and directories. However the rapid speed with which government data is being converted into digital form and made available to the public and the increasing range of commercial value-added data is making it more difficult for such metadata products to be comprehensive.

To-date, the great majority (if not all) on-line services have been established through public sector payment. The GDDD and EWSE are the only comprehensive pan-European metadata databases to have been assembled and the cost of initially setting up such a resource has been in excess of 0.5m Ecu, excluding the supplier costs of assembling metadata.

The fact that there are no private sector equivalents to the GDDD is a clear indication that the perceived business case for doing so is not strong. This contrasts with existing private metadata services for satellite-derived remotely sensed data, which are by their nature standardised internationally, frequently updated, with a very much simpler object hierarchy.

8.2.1 Data Content Issues

There is a clear need is to provide access to one or two comprehensive metadata services. The key to such services are their comprehensiveness in terms of the breadth of data sets that they contain. Of the utmost importance is the need for users to be able to obtain information outside of their own geographical region and their own subject area.

It is clear that most users require a metadata service that is broad and relatively shallow in terms of its data content. There is little demand in the first instance for metadata that contains large amounts of technical information or specification details. There may be a more limited requirement for such data once the user has determined that a data set is appropriate for their needs, however this might be effectively serviced direct from the data provider.

It is possible to define a metadata content model with two levels of data content. The first is very simple high level information about every conceivable data set, which can be used for marketing and simple searching to evaluate general availability. the second is very technical data sets to assess uses of a particular data set for a particular application or problem.

In Figure 8.1, the typical types of products are defined on the basis of level of detail and range of datasets included, and then divided into the three categories defined in Section 7. 'Marketing' metadata services are typified by the MapInfo Data Products Catalog, where emphasis is placed

on a wide range of information products with full contact details. The objective is to educate and inform the user, and provide maximum choice.

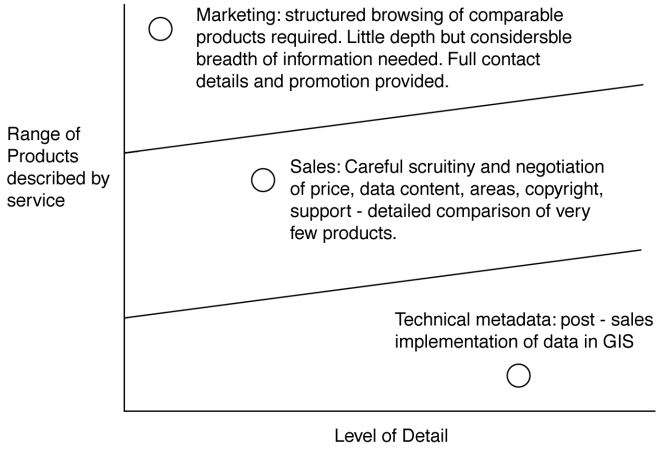

Figure 8.1 Metadata Content Model

At a slightly more detailed level information on content, geographical area, copyright, ownership, accuracy, currency and similar variables might be included to support accurate sourcing and sales processes. This might be viewed as the most practical level for a pan European or global metadata service.

At the most detailed level would be technical specifications and transfer formats such as those provided by national mapping agencies. This is much more to do with implementation procedures for a GIS, and might be supported by the data supplier and either the software vendor or consultant involved in the implementation process.

8.2.2 Data Quality Issues

In addition to the content issues any supplier of a metadata service has to be aware of the data quality issues. A key issue for users of such a service must be that the data sourced through the service is both appropriate and applicable. This is in part due to the searching mechanisms, but is also due to the need for accurate metadata records.

There is some need to educate the data supplier in the types of information needed and to impress on those creating the data records of the need for accuracy.

8.2.3 Future Options for a Metadata Service

It is likely that over the next five years there will be increased trade in per unit sales of geographic information, meaning that frequent, small volume sales and updates may contribute much more to the overall value of the market for geographic information than at present. If this is so then metadata for marketing and selling are likely to gradually converge into an integrated service through this period. Clearly, data capable of supporting transactions requires a consistent, assured level of quality, whereas metadata intended for browsing has a much lower quality requirement.

It is quite possible to conceive of a metadata service which is linked to real databases for the provision of sample data and also actual data sets, be they complete, or subsets defined by the user. This means the key component of the metadata service, that is the search functions and the data content, have a much increased significance since these will be vital to any enhancement to the service.

The key to future developments of this kind will be increased communications to allow rapid access to images, index maps, on-line geographic searching and on-line data delivery.

8.3 The Current Situation

The review of metadata services has made it clear that there are no standards in common practice for structuring and classifying geographic information, and so it is not easy to browse across sites and compare products from different suppliers. This makes it difficult to construct a business case to develop a comprehensive metadata service that does not incur large up front investment.

With separate classifications it is not possible to make comparisons between data at an equivalent level, or to search across the metadata for different products. If a common metadata standard is adopted at each site, or a service is provided to provide a translation between different classifications, then it becomes possible to search and compare using the same criteria. A meta-metadata service is a service that provides a common translation of other, varied metadata classifications, in order to provide a single view of the information to the user. In the figure, each of the three sites has unstandardised metadata, but the meta-metadata service

provides a uniform translation to the user. Were the sites themselves to standardise, many more metadata services could exist, with many metadata sites themselves linked to data descriptions of adjoining countries or regions. This would be a preferable situation.

When standards for location referencing are in place and observed, metadata searches will lead users to greater sources of location referenced geographic information. At present, the discovery of the existence of a data set is only the beginning of an often complex process of negotiation, and highly technical data integration issues all too frequently follow. Furthermore, It is also often the case that public sector bodies are not well prepared to supply data externally, or may do so in unsuitable formats.

Private sector suppliers must compete with the public sector, yet are not in the same position to invest in speculative or long term pay-offs, which is how metadata is often perceived. Whilst meta-metadata services and 'yellow pages' directories are an attractive concept, the funding models must be examined very closely.

8.4 Funding Models

There is doubt that the European market for geographic data will currently support a fully commercial pan-European metadata services. Public sector data providers are supporting the exercise at National level and, via MEGRIN, are moving towards a single European standard.

However, the greatest markets for metadata are national, as they are for spatial data itself. Trans-border markets are far smaller, and the least demand exists for pan-European data sets. There is however a growing demand for users to find out about national data sets beyond their own country, and often for a wide range of European countries. This offers the greatest potential to a pan-European metadata service, since users in one country are generally very unaware of geo-information activities in other countries.

A further factor which may influence the growth of European data sets, which are not contained by national boundaries, is the rapid growth in image-based data, especially that from the next generation of remote sensing satellites. As these sources of imagery grow there is likely to be less emphasis on the national borders, and more spatial data may be viewed as European. There are similar trends in other data sets, in particular that of road networks. The GI-BASE study has been established to review the feasibility of establishing a geographic base for Europe and to define its contents.

Current investment in metadata services is based largely on the premise that trans-border and pan-European data will grow in importance with both demand and awareness of data sources. Increased awareness of locationally referenced data sources may stimulate demand for this information.

There are four sources of revenue, at present, for funding metadata services:

1. User pays

The best example is provided by Minitel, where users pay per enquiry. Revenues from this model depend upon an extensive data content and a rapid response, to obtain the maximum 'hit rate'. Thus it is most likely to succeed if users can move seamlessly from one metadata directory to another. Current on-line metadata services require time-consuming browsing through different site architectures, and are unlikely to appeal as a chargeable service. As long as data suppliers provide free metadata, any commercial services must rely on the value that they can add. The added value needs to be: much faster access time, greater choice through meta-metadata, and the addition of private sector metadata. Purchase of low cost CD-ROMs may provide some revenue, but far higher data publishing costs are likely to be incurred.

The current perception is that users of geo information are prepared to pay a small amount for a license to access a comprehensive database. The numbers given ranged between 200 ECU and 1000 ECU for an annual licence, depending on the amount of searching an organisation undertakes.

Users were less interested in a pay-as-you-use options with either credit card or invoice billing. This was seen as too costly, even though no prices were mentioned and a potential service that could be cut when times were hard. To achieve a successful service of this sought it would be necessary to make sure users returned frequently to the service.

In many cases the pattern of usage of a metadata service was viewed as being intermittent with high levels of usage focused on the start of particular projects. This suggests an annual licensing model may be more appropriate.

2. Advertisement sponsors

The concept of electronic advertising is relatively new, and advertisers are far more likely to choose paper as the distribution medium at present if given a choice. Expenditure on

advertising is directly linked to the audience size, which is currently very small via electronic media (CD, on-line) when compared to paper. However, it is likely that on-line advertising will develop over the next five years. At present, advertisement sponsorship will not provide a sufficient business case for metadata services. Furthermore, there are very few advertisements for data or data providers, the majority being from software or hardware vendors.

It is possible to conceive of a 'Yellow Pages' form of advertising that could be linked to a metadata service. However this would require a considerable number of advertisements, possibly over 200, to make it viable and it is not clear that the GIS community can currently support such a large group of advertisers.

3. Supplier pays

The impetus for this to-date has generally been technical rather than led by marketing and public sector organisations have funded metadata services to promote information dissemination. It is debatable as to whether many of the services that have been developed are achieving their objective in attracting buyers, rather than other researchers and technicians. This is again supported by the lack of comparable activity from the private sector data suppliers.

For the users to pay or use a metadata service it must be comprehensive. To achieve this it is necessary to encourage all data suppliers to contribute to the service. This will not be achieved if data suppliers are asked to pay for their entries. At present this is not viewed as being a viable funding option for a metadata service.

4. Government pays

To-date, national-level metadata directories have been funded via supplier organisations, usually with participation from the Academic sector and there has been considerable interest and use of the services, exemplified by the UK SINES. In effect, the funding has to-date largely been directed toward more efficient dissemination of public sector data.

Given that it is expected that a commercial metadata service will find it difficult to create a comprehensive database and to fund the operation of the service, it is expected that for the foreseeable future some level of government funding will be required if there is to be a European metadata service. The EWSE service is a case in point.

8.5 Models for Future Metadata Services

The results of the user requirements analysis leads to the identification of two clear models for metadata which may be regarded as existing at either end of the continuum shown in Figure 7.1. These are:

- "broad and shallow" databases, containing many entries with simple data descriptions.
- "narrow and deep" databases, containing detailed data descriptions of fewer data sets.

In the first instance the user requires access to a very broad range of datasets, with only enough detail about each to make an informed decision about its general value. For many users this is the main purpose of a metadata service and is followed up by contacting the data provider directly. In the second instance the user is accessing the metadata service to obtain detailed information about a dataset that has been acquired or is being evaluated for purchase.

It is anticipated that the Metadata user migrates from using the 'broad and shallow' model to the 'narrow and deep' model as the data requirements are defined and the method of usage has to be established in detail.

The content of a metadata service can be very broad as the review of existing on-line sites has shown. This includes the range of products which can vary from traditional maps to digital geographic data sets, through to statistical databases and environmental data. It can also

include details about each product from simple descriptions to detail technical specifications, and information on currency, ownership, copyright, accuracy and extent.

Clearly the more detail that has to be collected about each data set the more costly it becomes to create the metadata. There is a point where the collection of data is no longer cost effective. This will depend on the usage of the data sets. For a data set used very widely it may be cost effective to collect lots of detailed information, while for the majority of data sets where usage is counted in tens only very simple data may be collected.

It is therefore possible to build a model for a metadata service that has in effect two levels of information detail. The first level serves the "broad and shallow" requirement and is a comprehensive database in terms of the number and types of datasets in contains references to. By definition this level will allow quick and simple search access to many datasets, with simple descriptions not dissimilar to the paper catalogues reviewed in Section 6.4. The second level meets the "narrow and deep" requirement, having information of far fewer datasets, quite likely those with the widest usage, such as the National Mapping Agency datasets currently in the MEGRIN metadata service.

The other problem is that to achieve a comprehensive metadata source, proportionally greater cost is incurred as the database becomes more comprehensive. That is to say, 10% of the total cost of collecting the database may yield 40% of the available data sets, while to collect the last 10% of the data sets may cost 40% of the total cost.

In this respect the content of the metadata service or product must be closely related to the funding model adopted and the process by which the metadata is disseminated.

8.6 Alternative Metadata Models

A number of organisations are also currently investigating the potential of CD-ROM catalogues, as a way of enhancing the information they can provide to existing customers. Ordnance Survey of Great Britain is converting its digital product catalogue into a CD-ROM version, using multimedia authoring tools. The advantages of this medium are that it is much more versatile in the ways in which products can be presented, and a greater level of interaction can be obtained with the customer.

It is pertinent to note that several organisations created a WWW site for promotion and awareness of their activities and products before they considered creating a CD-ROM

catalogue. However the greater access to CD-ROM players has encouraged such organisations to review their strategy, and the expectation is that more will publish catalogues on CD-ROM over the next 18 months.

The potential of CD-ROM based 'Yellow Pages' directories has been explored extensively in the OMEGA project. Unlike WWW sites, a CD-ROM such as OMEGA follows the 'Yellow Pages' business model in covering all aspects of the geographic information market, for all countries in a single data source. Other advantages from WWW sites are the ability to apply far more quality assurance to the information content, and to access information quickly through a single standardised interface.

It is attractive to think of 'meta-metadata' services providing an efficient gateway for moving between different metadata sites. However, the business case for establishing such services is difficult to justify, partly due to uncertain demand, uncertainty concerning willingness to pay, and the complexity of gaining acceptance of a suitable standard which could bring together such diverse sets of information. Such services could only occur if metadata sites were already moving towards standardisation, but such a move would be seen by some organisations to reduce their competitiveness.

In order to obtain as much benefit from metadata information as possible, organisations need to have a planned approach to creating and distributing metadata information. This information needs to support both marketing and sales activities and technical support activities.

8.7 Commercial issues

Once a metadata database has been created critical influences affecting its usage will be the following:

- Awareness that the database exists
- Ease of access at an appropriate price

Awareness of the database will have to created in a very broad and diffuse geographic information user community. This will require that the metadata database is supported by appropriate productisation or commercialisation that includes branding, an awareness campaign, marketing and sales support.

It is therefore likely that the desired commercial metadata model is an on-line metadata service from which more detailed CD-ROM versions of the service are produced at regular intervals,

such as every 6 months, for distribution to subscribers. It may also be necessary to support such a service with printed products (such as a newsletter) describing new products to be added to the service.

The commercial model that emerges is a three-layered set of metadata products. Traditional product catalogues will continue to be supplied but these will be complimented and enhanced by CD-ROM versions of the same product. At the same time there will be a growing use of WWW sites to provide timely information, news and product details, potentially with a centralised European meta-meta database.

9. Cost Benefit Issues

9.1 Commercial Viability

As has been identified in the discussion on the results from the OMEGA project, there are clearly problems of viability associated with a commercial European metadata service which is purely funded from user subscriptions and advertising. It is suggested that there will be a requirement to maintain a certain level of government funding if a metadata service is to achieve the objective of stimulating the European geo-information industry.

While it is not the purpose of this report to present a business case for the development of a metadata service, the authors believed it is worth reviewing some of the financial aspects which would affect the commercial viability of a European metadata service.

It is possible to build a business model for the cost of establishing and implementing an on-line metadata service. It is difficult to evaluate the likely levels of revenue from non-government sources without a clear understanding of the potential size of the geographic information market.

The GI-BASE project is attempting to evaluate the potential market for base mapping information. There is also a rapidly growing number of other data markets, many of which overlap with particular vertical sector user groups. These markets include those of business geographics, GPS data collection and image-based data.

The GIS market is also likely to become much less distinct over the next few years as system integration, open systems and client-server technologies take hold in the workplace. This will mean many more people will potentially be using geographic data.

9.2 Potential size of a metadata database

To establish a comprehensive cost/benefit case for the creation and maintenance of a metadata service it is necessary to understand the size of the geo-information industry in terms of the number of datasets and the rate of change that these datasets experience.

To date there has been very little research undertaken to establish the potential size of a European metadata database. Most of the existing databases are for limited numbers of datasets based on very well defined supplier groups, for example the UK SINES database is some 600 datasets and the MEGRIN GDDD is only 130 datasets. Similarly the published data catalogues are relatively small. The examples shown in Appendix 3 of this report contain only 250, 154 and 33 respectively.

This is however only a very small proportion of the total number of datasets that exist. If the total of UK government datasets published in SINES is matched by similar numbers of datasets for other European countries, and there is no reason to believe this is not the case, then we can extrapolate a minimum number of 10,000 central government datasets across Europe.

Commercial data suppliers have considerable database holdings. In the UK, The Data Consultancy has a catalogue of over 1,000 datasets and GeoInformation International publishes a total of 150 datasets. While the status of the commercial publishing activity varies across Europe it is likely that the total of commercially published datasets exceeds 100,000.

These numbers do not include local government and non-governmental data. It is possible to speculate that the total number of potential datasets in a comprehensive geo-information metadata service for Europe could exceed 500,000. Given the constraints of data quality, demand and cost, the actual number of datasets offered by a metadata service might be half this number.

Many of these datasets are maintained either on a continuous basis or on a regular update cycle. Therefore any metadata service must have in place a maintenance system for the collection of change information for these datasets.

9.3 Costs of creating and maintaining a metadata service

While it is difficult to be precise about the cost of establishing a metadata database there are certain fixed costs which will have to be incurred. These include hardware and infrastructure costs, systems development and data collection. The largest of these will be the data collection process given the number of potential datasets.

It is anticipated that any metadata project will take sometime to create the basic database. A minimum of two years is suggested but it is more likely to be about four years before the database can be regarded as truly comprehensive. A key investment decision will be to determine the size of the database necessary to attract and hold subscribers. It may be that a

database has to be created in sections, with an initial European level component and then region by region across Europe.

It is anticipated that any metadata service will incur set-up and development costs over a five year period. It will also incur maintenance costs for the database from an early time within the development cycle. The actual costs and the period of development depends on the precise specification of the database.

9.4 Subscription models for a metadata service

The most commonly anticipated revenue source for a metadata service is user subscription. This is despite CEO not believing that users of their metadata service will pay at present. Subscription models may be of two basic types:

- Annual or period based subscription with free usage
- Pay-as-you-use subscriptions, with a small initial fee and charges made for each database access or search.

The interviews indicated that if the user was required to pay a subscription fee, the first of these models was that preferred. It was anticipated that the normal subscription period would be annual, but that there was some demand for one-time access for companies involved in large projects for which they might wish to access a metadata service on a very occasional basis. Pay-as-you go options were felt to be difficult to manage in the context of a large organisations and often expensive to administer.

9.5 Potential subscription income

The interviewees were asked to quantify how much they would typically be prepared to pay for access to a metadata service. Most did not want a per access fee, as they felt would eventually be very expensive and in organisations with tight financial limits would be one of the services to be cut. They also felt that in governmental organisations in particular the organisation would prefer to pay an annual licence of some sort.

Typically the annual licence fee that users were prepared to pay was between 200 ECU and 1000 ECU. The number of subscribers is difficult to anticipate but from those interviewed less than 50% thought that they might subscribe and only three organisations declared that they

would definitely subscribe. This suggests that the potential subscriber base would be less than 5,000 across Europe, with an active subscriber base of no more than 1,250 organisations or 25% of the potential. This is not very different from the current level of registrations to the CEO service.

If the average subscriber is prepared to pay 200 ECU, the total subscription revenue is estimated at 250,000 ECU. At a subscription rate of 500 ECU the revenue would be 625,000 ECU.

9.6 Advertising models for a metadata service

A further source of revenue for a metadata service is potentially advertising. At present there is only a very limited experience of advertising in on-line services and many companies are not prepared to spend marketing budgets in this way. This may change as the user base increases and responses to advertising increases.

At present however a typical banner advertisement on the GIS World web page costs $3,000 for 200 click throughs. The total annual advertising revenue for the GIS World on-line web site was less than $50,000 for 1996 and is not anticipated to grow by more than 50% in 1997. While a metadata service might be argued to be a more targeted on-line service which would attract a greater level of advertising at more premium rates, there is little evidence that European advertisers will pursue this media for advertising any more strongly than in North America.

It is anticipated that the total revenue from advertising is likely to be less than 50,000 ECU in year One of an operational service and will not grow much beyond 250,000 ECU given the current size of the GIS industry. The one factor which affects these assumptions is the increased use of advertising by data suppliers, although there is little evidence of this in Europe at present.

9.7 Other Cost/Benefit Issues

The previous discussion assumes that a European metadata service will be run as a purely commercial undertaking. However there is a strong argument for financial support for a metadata service coming from either the European Commission or national governments.

These arguments surround the importance to the European geo information industry of initiatives to promote and stimulate the growth of the industry. For many users and data suppliers the availability of a metadata service is seen as important for the future growth of the industry.

A number of those users interviewed considered metadata was seen as an important part of stimulating the geo information industry. In particular some data suppliers saw it as an important part of promoting their products and services, possibly at little or no extra costs to themselves. Users on the other hand wanted to have access to a single comprehensive metadata service to make the data souring issue much simpler to solve. There were some however who did not believe that a metadata service would do much to stimulate the geo-information market because knowledge of specialist datasets would still reside with the specialists.

It is extremely difficult to estimate the benefits of a metadata service to the European geo information industry. There are no national or international services which have been active long enough to quantify the benefits that might have accrued. To a large extent the benefits of a metadata service are similar to that from developing a European standard geographic base.

The interim results from the GI-Base Study (Reference) and from the UK Standard Geographic Base Study (Reference) indicated that it is very difficult to produce a viable cost/benefit model for the adoption of a standard geographic base. Where this was attempted for the UK it was shown that it would take six years before the cumulative net present value became positive, i.e. the total benefits were greater than the total costs of implementing a standard geographic base.

Most GIS implementations are not justified on hard cost/benefit terms but are typically justified on 'soft' or difficult to measure benefits, such as the improved quality of information delivery or better decision making. It is almost certainly the case that the benefits that can be accrued for a European metadata service will be difficult to quantify and thus make it more difficult to cost/justify at a purely commercial level.

Therefore while there is user demand for a metadata service and data suppliers believe it would stimulate use of their products, the perception is that it should be provided at little or no cost. To compile a comprehensive metadata base will be a lengthy and costly exercise and the maintenance will be considerable. It is therefore very unlikely that it will be possible to establish a commercial metadata service without some level of government support.

10. Discussion of the ITT issues

10.1 Introduction

The original ITT for the GI-META project raised a specific set of issues (see section 1.3). Many of these have been addressed in the preceding discussion based on the information collected as part of this metadata study. However in order to make recommendations on the basis of these key issues the following is a consolidated summary drawn from the study as they apply to each issue in the ITT.

10.2 What European geographic information metadata sets already exist?

There are many metadata products in existence, mainly at a national level. Significant pan-European metadata sets in existence include:

- the GDDD
- EUROSTAT's GISCO database
- the OMEGA database, combining elements of the GDDD with private sector supplier metadata about data, software and services
- the European Territorial Database ETDB (an inventory of information held on analogue 1:25,000 maps from each European country, compiled by CERCO)
- CEO Satellite - derived data sources
- CORINE data, compiled by the European Environmental Agency
- GISIG databases (although these are not data directories)

The majority of databases are at a national level and are published by the public sector. There are also some CD-ROM based products and paper-based catalogues. Many of these exist as general marketing products to service particular supplier or data owner needs in the market place.

10.3 What European geographic information metadata sets are being planned?

There are proposals to enhance the GDDD database to cover more than just the datasets from the national mapping agencies. Likewise, the CEO as the largest European metadata initiative, has plans to further develop the existing metadata service, and there are plans to enhance existing Internet sites such as SINES. Many of the existing metadata directories have plans to enhance their offerings, either by extending the data content or by updating the data more frequently. The study revealed no concrete proposals to create a comprehensive European metadata service although organisations such as EUROGI are discussing the options.

10.4 What major geographic information data sets are not currently covered by existing metadata services?

For pan-European data, the most noticeable absence is metadata for public sector location-referenced data other than base mapping. This includes central government, local government and European Commission databases, with the exception of the GISCO administrative area codes. Whilst the GDDD covers public sector digital mapping, no comparable data exists for the private sector suppliers, who have been unwilling to make the same expenditure on meta data collection. The EWSE service provides a metadata service for the earth observation community but does not include many of the satellite sources or aerial photographic data, which is used widely in many parts of Europe. At a national level there are a number of examples of good metadata services such as SINES and SNIG but these are far from comprehensive as they tend to focus on government data.

10.5 What do potential users of geographic information want from a metadata service, i.e. what are the minimum and the ideal levels of information to be held about a data set?

Users identified a number of requirements for a metadata service, which encompassed the nature, range and detail of the content, how it should be delivered, what types of access should be available including types of search and how it might be financed.

The detailed review identified that an absolute minimum for a metadata entry would be the data set name, a brief description of content and area covered, and contact details.

The most practical amount of information required by a user is a simple classification of the main data objects (i.e. themes, such as properties, road centre lines, administrative boundaries), date of last update, completeness, coverage, resolution, GIS delivery formats and GIS compatibility, map projection, indicative costs and above all, good supplier contact details.

The ideal level of detail would be the above, with a clear hierarchic structure, enhanced where available with more detailed data descriptions such as those proposed in the Draft CEN TC287 standard and implemented in the GDDD.

10.6 What information is being proposed to be held in metadata databases by initiatives such as GDDD, and how does this compare with perceived user needs?

The information content in metadata services such as GDDD, SINES and CEO vary in their level of detail and types of data covered. In some cases they are more detailed than is necessary for a simple directory service while they are generally less detailed than technical metadata needed for implementation.

All of the existing metadata service or directories were created for a very specific user group or category of data. They largely fulfil their initial objectives well but do not meet all the user requirements that have been identified for a comprehensive European metadata service. Particular issues where they fail to meet user needs are the breadth and geographical range of the datasets included, the type of data description, the limited delivery models and the search mechanisms.

10.7 How can issues of quality of data be addressed, both practically in the short term and theoretically in the longer term?

Whilst standardised descriptions of data quality are of great potential value, experience shows that it is very difficult to reach agreement on interpretation of and adherence to the standards. Quality parameters need to be related to particular types of application, and in practice data samples must be assessed in order to determine the quality with this in mind. This is common practice with many other sales procedures, and is probably the most practicable solution. The CEN (Comitè Europèan de Normalisation) draft standard TC287 should address this issue.

It is noticeable that none of the metadata services reviewed attempted to include any form of comprehensive quality assessment. Given that such a requirement is not currently that important

to users, it is probably impractical to consider detailed quality assessments being included as a standard component of a metadata service.

10.8 What mechanisms exist to facilitate and control asset trading, taking into consideration legal issues such as different copyright legislation across Europe, differences in national contract law, etc.?

From this perspective, metadata is much less complex than the actual data. Most public sector metadata is already available in the public domain, and not subject to formal control. Metadata services would therefore trade in the value added by improved accessibility to free information. It is in the interests of data suppliers to provide this data at no cost.

It is clear that most of the issues raised regarding metadata are inextricably linked to the real data, and the nature of the market. For example, expensive data that is difficult to obtain will itself generate demand for metadata. This will be particularly true for data sources that lack adequate marketing channels, as is the case with much central European data. Under these circumstances, users can be expected to pay for up-to-date metadata. However, relatively inexpensive data already supported by suitable channels such as Value Added Re-sellers (VARs) is not likely to generate the same business case. This is because if the user has to pay, the metadata will add to the data cost. Thus metadata costs will be marginal, with most emphasis on metadata for marketing and paid for by the supplier.

In the context of this report it is very difficult to address the issues raised by the on-line supply of data through a metadata service. While this might be a long term objective and desirable from a users perspective, the issues related to licensing, reseller agreements, supplier liability and copyright controls are complex. It is not clear that data suppliers would wish to user a centralised supply channel as a core delivery mechanism, given the complexity of the sales and marketing porches for geographic data.

10.9 How can metadata best be delivered (technically) from different sources to different types of user, i.e. on-line (by what means), off-line (CD-ROM, diskette, paper)?

This depends on the type of metadata. Metadata for marketing should be delivered in the medium most used by the target market; for the technically advanced academic market, for example, on-line sources are suitable. For the health sector, this might not be so appropriate, and paper would probably be best. As the technical environment of the markets change, so should the delivery mechanism. For metadata such as the GDDD, either CD-ROM, floppy disk

or on-line would seem most appropriate. Hybrid CD-ROMs which give update access through the Internet are an alternative solution and have been used for information delivery outside of the geographic field. With quarterly updates, any delivery mechanism will do. For continuous updates, the delivery must be on-line, or down loaded on-line, to improve performance.

10.10 Can a business case be established for creating new metadata services where these do not exist today, based on direct cost-benefit analysis?

It would seem unlikely that a business case can be established for a commercial pan-European metadata service. This is because of the basis for payment for the services. Demand for trans-national or pan-European data is still weak compared to national markets. Services to-date have been funded publicly, and indicate an interest in disseminating public sector data rather than a market response to direct cost benefit forces. The benefits have to-date been essentially to provide better public service. Public sector organisations are more likely to make long term investments in the information infrastructure than the private sector, which requires a faster return. The lack of activity in the latter indicates the weak perceived cost-benefit ratio.

The greatest benefit to the GI community in terms of improved public service, and increased use of geographic information would be to encourage and support metadata services for local and central government location referenced data.

10.11 How should a potential user of geographic information located in any part of the EU be best enabled to access this information?

The process of accessing data is quite different from metadata. Potential users need, above all other information, contact details of who to negotiate with, and usable sample data. The best mechanism is probably the World Wide Web, via a 'meta-metadata' service using standardised search terms. However, in the short term any metadata service needs to be able to service those without Internet access.

The process of accessing actual geographic data will continue to be a cumbersome and expensive process until certain prerequisites are met. These are: standardised projections and an industry standard GIS format (not an exchange format, an import format). For location referenced data, adoption of suitable supply formats and standardised location references are needed. It is these issues that cause most problems with geographic data.

10.12 What are the barriers to making metadata available on a wide-scale basis - technical, legal, practical (i.e. linguistic, payment mechanisms, control of information following its release electronically, etc.)?

The main problem is the lack of any widely adopted standard data description at the simple product description level. TC287 standards need to provide standardised definitions of the data objects, and provide standardised keywords for searching. As long as metadata services are defined locally, it will remain very difficult to compare their resources in a seamless manner. The GDDD forms one core for a maintained pan-European meta-database, and it may be more practicable to develop this further as a centralised resource than to try to integrate many separate services.

There is a considerable barrier to overcome before private sector metadata is available to the same level of detail as public sector sources. It is likely that this will remain at the 'yellow pages' level of detail, directed at marketing and promotion, whilst public sector sources are classified and described in much more detail. Metadata will not be controlled when released electronically, any more than it is at present, when released on paper. It should devalue fairly rapidly as updates are released, and the minimal or non-existent market value of the metadata (as opposed to the data) means that control is probably not relevant, other than as time stamps.

A further dimension is provided by the type of data supply. Most geo-coded data is supplied in bulk, i.e. via magnetic media, and infrequently. At present, updates will be at intervals of several months or years, and will cover significant geographic areas although this may change with users having greater opportunity to purchase small sub-sets of data as and when required.

The metadata with which to de-cipher and interrogate the information is usually supplied with the data as a data dictionary. The scenario in which a user selects specific digital data (for example, property boundaries) for a specific area, and the data is downloaded on-line exists currently within organisations with distributed GIS on to which data are already loaded. It is not likely to be implemented between European data suppliers and users as long as digital map data is distributed as a GIS-independent transfer format, due to system and data incompatibilities.

11. Recommendations to the European Commission

11.1 Introduction

This chapter concludes the report by stating the key factors to be considered by the European Commission and makes recommendations for any future European metadata service.

11.2 Are Metadata Services Essential?

Marketing materials are essential, and technical metadata are essential in order to implement databases. Finding out about what is available is an essential part of using geo-information and as such there is a need for some level of metadata information.

Metadata becomes essential when it is one of the key drivers in the data acquisition or sales process. This is not currently the case, but may become more so as more data is supplied on-line. At present on-line sales are impeded by the often cumbersome procedures that are involved in actually procuring data. For example, transactions of digital mapping are usually based on batch data encoded using national transfer formats, a process that can take days to achieve. In many cases therefore, there is little added value in placing an order on-line, as opposed to a fax or letter.

The expenditure involved not only in buying digital mapping but in subsequently importing, reformatting and organising it means that orders are usually placed following a certain amount of human interaction. On-line ordering is therefore impeded by the absence of an agreed 'plug and play' exchange format. This largely explains why metadata is usually either for awareness and marketing, or post sales processing of digital mapping. There are signs that more and more data are being made available with viewing software for ease of use.

Digital mapping is vastly exceeded by the huge volume of location referenced governmental data such as census, health or road network related data. This includes geographic data generated by Local Government. Very often GIS costs are largely due to the need to find, integrate and manage this data. It is technically more plausible for this information to be traded on-line via metadata services for selling data, because of the generally simpler formats which

can be directly imported into spreadsheets or databases (although all too many exceptions exist). However, supplier organisations are far less commercially oriented that digital map suppliers.

In summary, the greatest justification exists currently for marketing-level metadata, with sufficient depth to facilitate comparison between products, to an agreed standardised classification. This form of metadata is essential but this does not imply that a European metadata service is essential. However if a centralised metadata service with more ordering and data supply capabilities were to be developed the existence of a metadata service might become a key component of the data supply process.

11.3 Standards - Appropriateness and Perceived Value

The main purposes of metadata are, at present, to provide marketing and implementation tools. The most effective means of reducing the costs of selecting and implementing geographic data is to introduce effective standards at both levels. The supply of standardised metadata for marketing provides a niche for data publishers, and makes it possible to search for products across directories and Web Sites. It also makes it possible to compare products. CEN TC287 has defined standards for geocoded data, but no standards are as yet defined for location references, with which to tie together the geocoded data and the geographic information. Where standards do exist, they are often not implemented.

This is illustrated by the inconsistent use of hierarchic area coding for administrative geography across Europe. It is therefore necessary to demonstrate the value of implementing and adhering to these standards, in order to gain the maximum benefits from metadata.

11.4 Summary of Conclusions

11.4.1 Current Metadata Products

1 Most metadata is provided at a very simple level in the form of catalogues, indexes and paper directories.

2 There is an increasing number of digital metadata directories supplied on a CD-ROM or floppy disc and a growing range of on-line metadata services.

3 Existing metadata ranges from very simple data descriptions covering data type, geographic coverage and supplier, through to full technical descriptions including data specifications.

11.4.2 Types of Metadata Services

1 The best delivery mechanism for future metadata services, especially at a pan-European level, is certain to be on-line via the World Wide Web (WWW).

2 The level of data description is critical to the way in which any service is used and to the cost of establishing the service.

3 Metadata services can be divided into two types:

"broad and shallow" databases, containing many entries with simple data descriptions.

"narrow and deep" databases, containing detailed data descriptions of fewer data sets.

4 It is possible to conceive of a 'meta-metadata service', consisting of linked databases through a common point of entry. Such a model could limit the amount of data capture required, but would rely on the acceptance and adherence to some form of metadata standard.

5 The effective use of metadata services is based on the search criteria provided. As a minimum these need to include data type based on a structured hierarchy of themes or subjects and geographical coverage. Further criteria such as supplier or publisher of the data set and product name could be usefully added. In existing services these criteria vary in both their format and comprehensiveness, with no use of standard nomenclature.

11.4.3 User Requirements

1 There is still a general lack of awareness of what is meant by the term metadata. Increased awareness of what metadata is and how it can be used is a necessary part of implementing any metadata service.

2 Use of on-line or CD-ROM based metadata services is limited at present. There is however, a perceived need for on-line metadata services in the future.

3 Users require 'comprehensive' metadata services covering a wide range of data sets for a wide geographical area. The required level of data description is generally low, with far less demand for detailed technical descriptions of data sets.

4 There is a generally recognised requirement for metadata services but emphasis is on the need for awareness about national and regional data sets rather than pan-European data sets.

5 In the short term there is a continuing need for CD-ROM and paper-based metadata information.

6 The broadly-based level of metadata service could be modelled on the UK SINES service as an example of good practice. Metadata services at the more detailed level could usefully standardise on MEGRIN's GDDD.

11.4.4 Cost/Benefit issues

1 Users are generally prepared to pay a limited amount to access a comprehensive metadata service.

2 Commercial data suppliers see a need to have a metadata service but can identify only limited benefits in paying to have their data sets included.

3 There is an opportunity to generate revenue from advertising on a metadata service.

4 At present the cost/benefit case for a fully commercial pan-European metadata service is not established. It is anticipated that some level of government funding will be required to set-up and run such a service in the medium term.

11.5 Recommendations to the European Commission

1. Use of suitable European metadata standards should be promoted as far as possible for both the simple, high level, data descriptions and the more complex technical descriptions. Awareness and use of the standards will be fundamental to data searching across different metadata sites.

2. Provision of metadata about government-held information that is location referenced should be given priority, in order to increase the availability of publicly held information. This includes data held by the EU, local, regional and national governments and their agencies. The UK SINES database provides a good example of this type of service.

3. Existing metadata services offered by National agencies could usefully standardise on the GDDD, in order to allow meta-metadata services and cross-border searching for public sector digital mapping. This would require a collaborative exercise involving MEGRIN and each of the main national providers. The value of the exercise would be increased by inclusion of private sector suppliers.

4. It is highly unlikely that existing demand can justify the development of a commercial pan-European or cross-border metadata services. The establishment of the infrastructure for such services will therefore require public sector investment, which could be supported in part by a low level of payment from users and some forms of advertising.

5. Quality standards or accreditation for the delivery of services needs consideration. Existing services vary greatly in quality of interface and functionality, and so does the quality of information. The lack of commercial drive and competition from many of the WWW metadata sites combined with a high degree of accessibility has resulted in uncontrolled quality.

6. The greatest demand for metadata is still at regional and national level, and the least demand is pan-European. However, this does not weaken the argument for standardised information, because the demand for regional or national-level data will increase from other regions and nations. Each national or regional information provider will be increasing their individual market, rather than meeting demand for their combined information.

12. Bibliography

1. OMEGA (Object Metadata for European Geographic Analysis) Final Report. April 1996.

2. M Krinitz and A. Kjeldsen 'A step towards Virtual GIS using the Catalogue Interoperability Protocol' in Proceedings of the 2cnd EC - GIS Workshop, Genova, Italy, 26-28 June 1996.

3. GI-BASE Interim Report for DGXIII of the European commission. August 1996.

4. Preliminary Assessments of User Requirements. CEO Programme. Document CEO/155/1995. July 1995

5. Action 2d: Assessment of User Requirements within the Services of the Commission. Centre for Earth Observation (CEO). Document CEO/156/1995 (Version 2.0) December 1995.

6. Survey of current practices and standards for metadata management; Volume A: Survey and analysis of practices in 20 EU Organisations. Document: CEO/US/1400/217 Version 1.1. November 1996.

7. Survey of current practices and standards fro metadata management; Volume B: reference document for major metadata standards & initiatives, major metadata attribute sets. Document CEO/US/1400/218. Version 1.1. November 1996.

8. Metadata Standards Development. report from the Mitre Corporation for the Federal Geographic Data Committee. Extracted from the FGDC World Wide Web Site http://www.fgdc.gv/metadata/metahome.html.

Appendix 1 - Interview Pro forma

GI-META PROJECT
European Commission Contract

METADATA SURVEY: INTERVIEWS RESULTS

Interviewer:		Date:
Type of interview:	Face-to-face: ___	Telephone: ___

Interviewee:

Position held:

Company/Organisation:

Address:	Tel:
	Fax:
	Email:

Company/Organisation's role:

Software Vendor ___ Data supplier ___ Data reseller ___

Commercial GIS User ___ Research Agency ___ Govt GIS User ___

QUESTIONS

Q1. What do you understand by the term Metadata in the context of Geographic Information?

Q2.1 Does your organisation create Metadata for geographic data sets it creates or derives, either for use internally or for your customers?

Q2.2 Do you use a nationally or internationally recognised standard and if so what?

Q2.3 What type of information does your organisation put into its Metadata database?

Q2.4 What future plans do you have for creating Metadata products for use internally or for your customers?

Q3.1 What Metadata sources does your organisation use in the course of its business? (Include on-line, digital, paper etc.)

Q3.2 Give an example of the most useful Metadata source used and the least useful, and why.

Q3.3 As a data user, what information would you like to see in a Metadata database? Given an idea of the detail you would require from simple indexes through to technical data formats.

Q3.4 Would you use an on-line metadata service? If so, what would would you like to see as the content and geographical extent of that service.

Q4.1 Would you pay for a metadata service? If so, how much?

Q4.2 How often would you use a metadata service?

Q4.3 Would your organisation advertise on a metadata service?

Q4.4 As a data supplier would you pay to have metadata for your products put onto an on-line service?

Q5.1 Do you have any comments or views on the role of Metadata for the Geographic Information community that might be pertinent to this review?

Q6.1 Are there any samples of metadata that your organisation/company has produced that you are able to supply as examples for this study?

List items:

Appendix 2 - User Questionnaire

Appendix 3 - Review of Metadata Products

Introduction

The main objective of GI-META was to examine the feasibility of providing geographic information metadata from a broad range of European sources by:

- reviewing existing local, national, and international metadata services;
- examining the costs involved in creating and maintaining metadata services;
- reviewing the interconnection options available via EU-wide information networks;
- presenting a set of possible implementation scenarios to suit different levels of preparedness for both offering and using such services.

For the purposes of this study metadata is defined as "structured data that provides a standardised description of another data set. A metadata service is used to capture, standardise, provide access to and disseminate metadata." (Project and Quality Plan, page 1).

The data collection component of the project envisaged the need to collect and review a range of existing European, and where relevant, non-European metadata services in order to assess their strengths and weaknesses. The purpose was to use existing practice to help identify the key criteria for any future European metadata service and highlight examples of good practice.

Appendix 3 fulfils this task by presenting in a comparable format a set of examples of best practice currently available to users of geographic information.

The review is divided into 2 parts. The first part is an analysis of 18 different World Wide Web sites selected for their declared objective of providing geographic metadata or directory services. A comparative evaluation of the 18 sites was made based on 14 different criteria. From this evaluation a series of key issues were identified.

Part 2 is a review of three commercial data catalogues published in support of either a particular GIS software product range or a particular user community. The three catalogues are the ARC Data Catalogue by ESRI which will be launched as an on-line service in October 1996, the MapInfo data catalogue and the Business Location Research (BLR) data catalogue, all published in the US but with content extending beyond the US.

PART 1: ANALYSIS OF METADATA SITES

The purpose of this review was to evaluate a range of existing on-line metadata services provided across Europe and from other areas around the world. A total of 18 sites were selected covering a range of organisations from central government to local or state government, pan-national organisations such as MEGRIN or national associations such as the Association for Geographic Information (AGI).

The full list of World Wide Web sites reviewed as part of the GI-META Study is given in Table A3.1. This table is also reproduced in the main text of the report.

The following sections outline the methodology adopted in selecting the sample sites and the variables used in the analysis.

Methodology and Definitions

The starting point for this study was provided by the provisional list of European metadata projects identified in the Project and Quality Plan (page 8). However, it was decided to focus specifically on those metadata services which are currently accessible via the World Wide Web as the Internet is likely to be one of the main vehicles for delivering any metadata service of geographic data to be set up at the European level.

The World Wide Web at this stage of technological development is considered to be the most appropriate interface for this type of data dissemination as well as the easiest to interrogate for a wide variety of users from different cultural backgrounds. Looking at different metadata services from a single user interface also adds an extra comparative dimension to the data set as well as raising technical issues for future European-wide services.

Whilst a broad definition of metadata and metadata service is provided in the Introduction above, it is clear that there is a wide range of possible interpretations of these definitions with services that range from the tightly focused, structured, and extremely detailed to those that consist of a broad collection of pointers to existing organisations and databases based on some subjective judgement of their relevance.

With this in mind, the examples included in this report fall into five broad categories have been chosen as representative of this broad universe of services:

- Services tightly focused on digital mapping data such as MEGRIN's Geographical Data Description Directory;

- Services forming the backbone of a (national) geographic data infrastructure (e.g. Portugal's SNIG, Finland's MMH, Great Britain's SINES, and examples from the US at federal (NSDI) and state level (Wisconsin), and from Australia (AUSLIG). Here the mapping dimension is still strong but a wider range of data sets are included and the aim is more strategic in character;

- Services which represent the well-established community of data archives (such as Norway's NSD and the ESRC data archive at Essex) where the main focus is on statistical data and the geographic dimension is less fully developed although often present;

- Services which aim at providing a resource base for a variety of users of geographic information, that is services which do not focus so much on geographic information itself but on research, projects, and initiatives related to geographic information such as those of the European Science Foundation (GISDATA), EUROGI, the AGI and GISIG;

- Services which have a strong environmental dimension and encompass a very wide range of data sets and topics, such as those provided by the Centre for Earth Observation and the Consortium for International Earth Science Information Network (CIESIN).

Analysis

For each of the services selected the following variables have been considered:

ID number (for internal purposes).
Name of the Service.
Name of the organisation providing it and country of affiliation.
Internet Address
Focus: main focus of the service in terms of data being described.
Product description: brief overview of the structure of the metadata service.
Target Group: whether the principal target group is the public sector, private sector or academia.
Delivery: Main feature of the service along a continuum which includes:

- Simple listing of products;
- Yellow Pages: i.e. structured list of organisations and products;
- Hierarchical structure, richer in levels than the Yellow pages approach, where the concept of hierarchy is dominant;
- Searchable database(s) which are richer still and allow more horizontal movement than hierarchical approach;
- Open shelf browsing, where the data itself is on display rather than a description of it.

The above classification is not meant in a strict database sense but more as a way of capturing the flavour of the service provided.

Searchable Engines: YES/NO field which indicates the presence in the service of an internal querying mechanism that goes beyond the Find facility of the WWW interface (i.e. ability to search a string of text in the database) or the WWW search engines such as Altavista, Infoseek, etc.

National Links: YES/NO field indicating whether existing linkages to other national databases or metadata services are an important feature of this service.

International Links: YES/NO field indicating whether existing linkages to other international databases or metadata services are an important feature of this service.

Degree of Organisation: HIGH, MEDIUM/HIGH, MEDIUM or LOW which indicates the degree of internal organisation of the service and is partly dependent on ownership of the data being described, resources available, and nature of the service.

Degree of Homogeneity of the Data: HIGH, MEDIUM/HIGH, MEDIUM or LOW where for example services focusing on mapping or statistical data will score HIGH and those encompassing a wide range of environmental data will score LOW.

Currency: HIGH, VARIABLE or LOW depending on the extent to which the metadata service is kept up to date.

Language: main language of the service as available on the WWW.

1.	MEGRIN (France)	http://www.ign.fr/megrin/gddd/gddd.html
2.	SNIG (Portugal)	http://snig.cnig.pt/
3.	MMH (Finland)	http://www.nls.fi/index_e.html
4.	NSDI (US)	http://fcdc.er.usgs.gov/
5.	Wisconsin LI System (US)	http://badger.state.wi.us/agencies/wlib/sco/
6.	AUSLIG (Australia)	http://www.auslig.gov.au/welcome.htm
7.	SINES (UK)	http://www.ordsvy.govt.uk/sines.html
8.	ESRC (UK)	http:/dawww.essex.ac.uk/
9.	NSD (Norway)	http://www.uib.no/nsd/
10.	GISDATA (UK)	http://www.shef.ac.uk/uni/academic/D-H/gis/gisdata.html
11.	AGI (UK)	http://www.geo.ed.ac.uk/agi/agi.html
12.	GISIG (Italy)	http://gisig.ima.ge.cnr.it/
13.	EUROGI (EU)	http://www.frw.ruu.nl/eurogi/
14.	Centre for Earth Observation (Italy)	http://ceo-www.jrc.it/
15.	CIESIN	http://www.ciesin.org/
16.	Ordnance survey	http://www.ordsvy.govt.uk
17.	Alexandria	http://alexandria.sdc.ucsb.edu
18.	CEOS	http://www.smithsys.co.uk

Table A3.1 List of on-line metadata services reviewed

Examples of On-line Metadata Services - Summary Analysis

ID:	1
Name:	Geographical Data Description Directory
Organisation:	MEGRIN
Internet Add:	http://www.ign.fr/megrin/gddd/gddd.html
Focus:	Mapping data
Product Description:	Hierarchical data base listing producers and products in 18 European countries following agreed metadata standards (TC287)
Target Group:	Public/Private/Academia
Delivery:	Yellow Pages. Data not available on-line. Searchable only by organisation and some standard products.
Searchable Engines:	No
National Links:	No
International Links:	Yes
Degree of Organisation:	High
Degree of Homogeneity of Data:	High
Currency:	Variable
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	2
Name:	CNIG
Organisation:	National Centre for Geographic Information CNIG (Portugal)
Internet Add:	http://snig.cnig.pt/
Focus:	Geographic and other data
Product Description:	Alpha numeric and geographic databases linked at national, regional, local level
Target Group:	Public/Private/Academia
Delivery:	Hierarchical set up of producers and products including non-geographical data such as organisation's structure
Searchable Engines:	No
National Links:	Yes
International Links:	No
Degree of Organisation:	High
Degree of Homogeneity of Data:	Low
Currency:	Unknown
Language:	Portuguese with English summaries

Examples of On-line Metadata Services - Summary Analysis

ID:	3
Name:	MMH
Organisation:	National Land Survey of Finland
Internet Add:	http://www.nls.fi/index_e.html
Focus:	Geographic Information
Product Description:	Wide range of data bases on property, buildings, infrastructure, maps, population, environment and natural resources
Target Group:	Public/Private/Academia
Delivery:	Description of data according to agreed national standards, owners and access rules
Searchable Engines:	No
National Links:	Yes
International Links:	No
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium
Currency:	High
Language:	Finnish and English

Examples of On-line Metadata Services - Summary Analysis

ID: 4

Name: NSDI

Organisation: Federal Geographic Data Committee (US)

Internet Add: http://fcdc.er.usgs.gov/

Focus: Geographic Data

Product Description: Wide range of databases linked following agreed national (NSDI) standards for metadata and transfer.

Target Group: Public/Private/Academia

Delivery: Searchable databases (also spatially) and lists. Preview of data possible. Data available on-line at federal level in many instances.

Searchable Engines: Yes

National Links: Yes

International Links: Yes

Degree of Organisation: High

Degree of Homogeneity of Data: Medium/High

Currency: High

Language: English

Examples of On-line Metadata Services - Summary Analysis

ID:	5
Name:	Wisconsin LI System
Organisation:	State of Wisconsin (US)
Internet Add:	http://badger.state.wi.us/agencies/wlib /sco/ pages/wis
Focus:	Geographic Data
Product Description:	Wide range of databases linked following agreed (NSDI) standards for metadata and transfer.
Target Group:	Public/Private/Academia
Delivery:	Searchable databases (also spatially) and lists. Preview of data possible. Data available in digital format but not on-line.
Searchable Engines:	Yes
National Links:	Yes
International Links:	No
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium/High
Currency:	High
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	6
Name:	AUSLIG
Organisation:	Australian Surveying and Land Information System
Internet Add:	http://www.auslig.gov.au/welcome.htm
Focus:	Geographic Information
Product Description:	Maps, GIS consultancy, aerial photos, geodetic services and remote sensing.
Target Group:	Public/Private/academic
Delivery:	Loose assemblage of advertising material.
Searchable Engines:	Yes
National Links:	Yes
International Links:	No
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium/High
Currency:	High
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	7
Name:	SINES
Organisation:	Ordnance Survey-Department of the Environment (UK)
Internet Add:	http://www.ordsvy.govt.uk/sines.html
Focus:	Geographic information
Product Description:	Metadata on some 500 data sets held by different government departments
Target Group:	Public/Private/Academia
Delivery:	Searchable database of data holdings by keyword and agency
Searchable Engines:	Yes
National Links:	Yes
International Links:	No
Degree of Organisation:	Medium
Degree of Homogeneity of Data:	Low
Currency:	Variable
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	8
Name:	ESRC Data Archive
Organisation:	ESRC (UK)
Internet Add:	http://dawww.essex.ac.uk/
Focus:	Statistical and other data
Product Description:	Some four and a half thousand data sets for the social sciences
Target Group:	Public/Academia
Delivery:	The data is searchable through an on-line catalogue (BIRON) and a WAIS system. 28 variables describe the data which can then be ordered from the archive
Searchable Engines:	Yes
National Links:	Yes
International Links:	Yes
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium/high
Currency:	High
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID: 9

Name: Norwegian Social Science Data Service

Organisation: NSD (Norway)

Internet Add: http://www.uib.no/nsd/

Focus: Statistical data

Product Description: Alpha numeric databases and statistical software for analysis

Target Group: Public/Academia

Delivery: Hierarchical database of data holdings and research projects. Only projects are searchable. Data not on line. Links to all major data archives across the world

Searchable Engines: No

National Links: Yes

International Links: Yes

Degree of Organisation: High

Degree of Homogeneity of Data: High

Currency: Variable with data

Language: Norwegian with English summaries

Examples of On-line Metadata Services - Summary Analysis

ID:	10
Name:	ESF
Organisation:	GISDATA
Internet Add:	http://www.shef.ac.uk/uni/academic/D-H/gis/gisdata.html
Focus:	Research resources
Product Description:	Reports of meetings, abstracts of papers, directory of addresses of researchers in GIS
Target Group:	Academia
Delivery:	Open Bookshelf Browsing.
Searchable Engines:	No
National Links:	Yes
International Links:	Yes
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium
Currency:	High
Language:	English
Metadata Standards:	No

Examples of On-line Metadata Services - Summary Analysis

ID:	11
Name:	AGI (UK)
Organisation:	Association for Geographic Information
Internet Add:	http://www.geo.ed.ac.uk/agi/agi.html
Focus:	Geographic Info Resources
Product Description:	Wide range of services including news of conferences, servers, and GI resources in general
Target Group:	Public/Private/Academia
Delivery:	Yellow pages. Some data available on line but depends on organisation being accessed
Searchable Engines:	No
National Links:	Yes
International Links:	Yes
Degree of Organisation:	Medium
Degree of Homogeneity of Data:	Medium
Currency:	Variable
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	12
Name:	GISIG (Italy)
Organisation:	GISIG
Internet Add:	http://gisig.ima.ge.cnr.it/
Focus:	Geographic information resources
Product Description:	Info on the organisation and 3 databases: research, data, GIS organisational profiles
Target Group:	Academic
Delivery:	Hierarchical database organised by country. Limited data. Open Bookshelf Browsing
Searchable Engines:	No
National Links:	No
International Links:	Yes
Degree of Organisation:	Medium
Degree of Homogeneity of Data:	Medium
Currency:	Variable
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	13
Name:	EUROGI (EU)
Organisation:	EUROGI
Internet Add:	http://www.frw.ruu.nl/eurogi/
Focus:	Geographic Information Resources
Product Description:	A series of directories being developed: education, research projects and data.
Target Group:	Public/Academia
Delivery:	Searchable databases, but not "open bookshelf" browsing
Searchable Engines:	Yes
National Links:	No
International Links:	Yes
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium
Currency:	Unknown
Language:	English

Examples of On-line Metadata Services - Summary Analysis

ID:	14
Name:	Centre for Earth Observation
Organisation:	EU-Joint Research Centre-Ispra
Internet Add:	http://ceo-www.jrc.it/
Focus:	(Broadly) Geographic Data
Product Description:	Very large collection of data sources relating to Atmospheric, Ocean, Land, and Interior-Crust Earth Data
Target Group:	Public/Academic
Delivery:	The databases are searchable by keyword, not spatially. Also list of data sets and themes
Searchable Engines:	Yes
National Links:	No
International Links:	Yes
Degree of Organisation:	Medium
Degree of Homogeneity of Data:	Low
Currency:	High
Language:	English mainly but some organisations have also multi lingual versions (e.g. World Meteorological Office)

Examples of On-line Metadata Services - Summary Analysis

ID:	15
Name:	CIESIN
Organisation:	Consortium for International Earth Science Information
Internet Add:	http://www.ciesin.org/
Focus:	Environmental and Socio-economic
Product Description:	Description and partial access to databases held by partners in this consortium focusing on the interdisciplinary study of global environmental change
Target Group:	Public/Academia
Delivery:	Hierarchy of databases held by partners described in standard format, access to public domain data sets, analytical and visualisation interfaces for users
Searchable Engines:	Yes
National Links:	Yes
International Links:	Yes
Degree of Organisation:	High
Degree of Homogeneity of Data:	Medium

Currency: Variable

Language: English

Examples of On-line Metadata Services - Summary Analysis

ID:	16
Name:	Spatial Information Enquiry Service
Organisation:	Ordnance Survey
Internet Add:	http://www.ordsvy.govt.uk (sines.bin/search.cgi)
Focus:	Wide and diverse range of geographically related data from central government organisations for the U.K.
Product Description:	Meta-database describing over 500 spatial data sets held in UK government departments and agencies
Target Group:	Public, private and academic
Delivery:	Three alphabetical scroll down lists including KEYWORDS, ORGANISATIONS and COVERAGE AREAS. The query lists all items with all of these characteristics which can each be viewed in turn
Searchable Engines:	Yes - Ability to query keywords, organisations and coverage areas. Usage of the GLIMPSE and ARCHITEXT search engines
National Links:	Yes - From county to country
International Links:	No

Degree of Organisation:	Low - The OS maintain the information created by all of the departments and agencies supplying details of their data sets
Degree of Homogeneity of Data:	Medium/High - Most of the data is mappable, however, not all of it is in a digital format
Currency:	Variable
Language:	English
On line Graphical Display:	Fair
Comments:	Why isnít the database relational?

Examples of On-line Metadata Services - Summary Analysis

ID:	17
Name:	Alexandra Digital Library
Organisation:	Alexandria
Internet Add:	http://alexandria.sdc.ucsb.edu/public-documents/metadata/
Focus:	Geographically referenced data
Product Description:	Descriptions of available data and papers that are hot linked with other similar categories
Target Group:	Public, private and academic
Delivery:	A distributed digital library for geographically referenced information and publications
Searchable Engines:	Yes - Able to search documents by title, abstract, author, client, organisation, bulletin, digital library ID, publication year
National Links:	No (US data)
International Links:	Yes
Degree of Organisation:	Low
Degree of Homogeneity of Data:	Low - Many references to documents
Currency:	Medium

Language: English

On line Graphical Display: Poor

Examples of On-line Metadata Services - Summary Analysis

ID: 18

Name: International Directory Network

Organisation: Committee on Earth Observation Satellites (CEOS)

Internet Add: http://www.smithsys. co.uk/dif

Focus: Geographical data - the CEO is on-line at: http://ceo-www.jrc.itl

Product Description:

Target Group: University departments, research agencies, government bodies, other world wide organisations

Delivery: 3000 references to data sets such as earth science (many in remote sensing), socio-economics, life sciences, space physics, planetary science, astronomy and solar physics

Searchable Engines:

National Links: Yes - National and regional data sets

International Links: Yes - Worldwide data sets

Degree of Organisation: Low - This is a directory of international data sets, with associated links

Degree of Homogeneity of Data:

Currency:

Language: English

Conclusions

The examples summarised in the previous listing indicate the range of metadata services and models already available on the Internet. These examples are considered to be indicative of the ways in which a metadata service may be provided using the Internet as a delivery mechanism.

A number of key issues were identified from the review. These were:

1. Language:

These metadata services do not appear to be overly constrained by language barriers as it is possible to have multi-lingual services using English as a de facto minimum common denominator. Some WWW sites allow multi-lingual services to be provided, but English was generally used as the de facto.

2. Flexibility:

The presentation of metadata information on WWW sites allowed for different national policies on such issues as copyright and charging to be presented without too much complication but there is a lack of standardised terminology which allows the user to compare sites with confidence. It is noticeable however that most of the metadata presented in the sites studied is national and not pan-European.

WWW sites can offer considerable flexibility in searching, not only at the specific sites as shown above, but for searching across different sites. However, only a few sites have the right balance between detail and content, and the necessary search engine to make browsing across sites easy.

3. Level of detail:

A consensus seems to be emerging that it is necessary to limit the amount of detail shown on such services to a basic description of the data. Should the user require further information about specific technical issues or products they are directed to the information owner.

There are without doubt serious technical limitations for on-line services at present. Transmission and display images such as map samples is extremely slow and liable to

cause considerable frustration. By comparison, browsing through a conventional paper catalogue or CD-ROM data source is quick and efficient. However, the steep demand curve for the Internet should result in improvement in service quality.

The one notable exception to the trend towards simple data descriptions is the Geographical Data Description Directory (GDDD) launched by Megrin as part of a project to amass European digital geographic data. The GDDD has been described by Peter Larson (GIS Europe, July 1996) and is in part reproduced in Part 4 of this report..

Given the wide range of services considered, it is to be expected that the degree of user satisfaction also varies considerably depending on what one is looking for. If one is looking for tightly defined information such as "who is active in the field of GIS Diffusion" or "what maps are available from the national mapping agency in Greece" the services reviewed perform reasonably well as the information is stored in mono-thematic and/or the whole database is small.

However, in most cases, users will simply not know what to look for in very precise terms and would be expected to use a metadata service to help identify what is available for a given area and/or topic and who has it. On this very simple criteria, the services reviewed generally fail the test.

The one service that comes closest to what a user might need is the NSDI. Its strength compared to many of the other services is that it is possible to limit searches first by geography, and then find out what is available. This is absolutely critical. As more and more databases become available, the ability to use geography as a searching mechanism must not be underestimated.

There are obvious reasons why the NSDI service score better than other services, such as having a common geographic framework for data across the US, but clearly there are also important implications from a European perspective as to the resources needed to geo-reference the metadata and enable the type of searching contained in this service.

In Europe, the nearest equivalent type of service is EWSE, the information exchange of the Earth Observation Community, based at the CEO, Ispra, which is reviewed in Part 2 of this appendix.

A key finding of this review is also the difficulty in creating meaningful search criteria. It is necessary to search by both geographical location and some theme or topic as a minimum. The

two more detailed reviews are examples of how this has been attempted but not necessarily successfully. The importance of the search criteria is that they need to be structured in such a way as to allow the use easy access to the very specific data sets they are looking for even when the query is poorly defined. To achieve this is necessary to have a tightly defined geographical search capability and a well structured set of themes or topics.

PART 2 - DATA CATALOGUES

To compliment the review of on-line services a comparative review was undertaken of three commercially produced data catalogues. These are as follows:

- ArcData Catalog published by Environmental Systems Research Institute (ESRI), USA.
- MapInfo Data Products Catalog published by MapInfo Corporation, USA.
- BLR Data Catalog published by Business Location Research , USA.

These three data catalogues come into two categories. The first two are collections of data sets that are in a particular software vendors proprietary format and are sold to enhance the software sales. The last is a data catalogue from a company specialising in the publication of data for a particular user community, the Business Geographics community, in the US.

The three catalogues are evaluated on the following criteria:

- Number of data sets included
- Geographical areas covered
- Scale range of data sets
- Currency of the data sets
- Type of metadata on each data set.
- Ease of use and clarity of the data set entry
- Indexing functions
- Currency and update cycles

In addition to these comments a sample is taken from each catalogue to show the format for the data entry.

ArcData Catalog

Publisher: Environmental Systems Research Institute (ESRI), USA.

Edition:	Third edition, 1994.
Number of data sets included:	250
Geographical areas covered:	Mainly US, but some global data sets
Scale range of data sets:	Global data sets to very detailed
Currency of the data sets:	Date of publication in some cases
Type of metadata on each data set:	General description, no defined format (see Figure A3.1)
Ease of use and clarity of the data set entry	Organised by theme, with cross reference to supplier
Indexing functions	Indexed by supplier, with summary tables by theme
Currency and update cycles	Published at least annually

Note: ESRI intend to put their data catalog on the ESRI web site from October 1996.

Figure A3.1 Sample extract from the ESRI ArcData Catalog.

MapInfo Data Products Catalog

Publisher MapInfo Corporation, USA.

Edition:	August 1995
Number of data sets included:	154
Geographical areas covered:	Global range
Scale range of data sets:	Small scale to very detailed
Currency of the data sets:	Publication date recorded
Type of metadata on each data set:	Structured data product description
Ease of use and clarity of the data set entry:	Ordered by country and then theme, consistent layout, easy to use. Metadata fields are: • Product name • General free text description • Product Number • Coverage • File size • Vintage • Scale • Source • Price
Indexing functions:	Index of all data sets by country and theme to cross reference with catalogue order
Currency and update cycles:	More than once a year

Figure A3.2 Sample extract from the MapInfo Data Products Catalog.

BLR Data Catalog

Publisher Business Location Research, USA.

Edition:	1996
Number of data sets included:	33
Geographical areas covered:	All USA
Scale range of data sets:	Continental to Site level
Currency of the data sets:	Not given
Type of metadata on each data set:	General description with some sample plots
Ease of use and clarity of the data set entry:	Ordered by subject
Indexing functions	List by theme and then each data set
Currency and update cycles	More than one a year

Figure A3.3 Sample extract from the BLR Data Catalog.

chapter one

chapter two

chapter three

chapter four

chapter five

chapter six

chapter seven

chapter eight

chapter nine

chapter ten

chapter eleven

chapter twelve

appendices

executive summary