Towards a metadata base of digital spatial data for Ireland

K. Rybaczuk
Trinity College Dublin

M. Ovington
Trinity College Dublin

W.P. Prendergast
Dublin Institute of Technology

Introduction

Until now, metadata for Irish digital spatial data has been unavailable. The Irish Geospatial Information Directory (Geo-ID) aims to promote co-ordination and awareness within the Irish GI marketplace through the development of a meta database of available geographical information resources. Specifically the Geo-ID project aimed to:

- Create a metadata template for Irish digital spatial data
- Collect metadata from Irish Digital data producers
- Assess methods of disseminating metadata
- Produce a pilot metadata system

The Geo-ID was created as a result of two projects investigating the creation of a national metadata base for digital spatial data in Ireland. The first, funded through Dublin Institute of Technology / TCD seed funding, focussed specifically on the methodological issues surrounding metadata base creation. The second, funded by the Irish Organisation for Geographic Information (IRLOGI), focussed on the practical issues of creating a pilot national metadata base for Ireland. This work was carried out as part of a tendered project proposal and therefore was structured around the specific requirements of IRLOGI.

What is metadata?

In the context of digital spatial data, metadata is the background information that describes the content, quality, condition, and other appropriate characteristics of the data. Paper maps contain metadata, primarily as part of the map legend. In this form, metadata is readily apparent and easily transferred between map producers and map users. When map data, or any form of geo-referenced data, are in a digital form, metadata is equally as important. For both digital and non-digital data, metadata is essentially “data about data” or “information about information”. This is not, as such, a definition. It is a statement of what metadata does. The US Federal Geographic Data Committee has adopted a standard for the contents of digital geo-spatial metadata that includes this definition:

“Metadata is data about the content, quality, condition, and other characteristics of data”.

FGDC

The creation of national digital spatial meta databases – the current perspective

Interoperability, i.e. getting GIS to work with each other and exchange data without unnecessary and expensive reformatting (Burrough & Longhorn, 1998), represents a huge barrier to the development of GIS both in Ireland and world-wide. Commercial GIS applications often have to work alongside proprietary database systems and read from the arbitrary data formats employed by various organisations. This is even before issues of data compatibility are addressed.

Metadata standards

This issue of interoperability has been addressed by the development of standards for metadata and data exchange. The development of divergent “standards” at a local, regional and global scale is inevitable. In the EU alone, the differences between the various national metadata services is immediately evident. The European Spatial Metadata Infrastructure is one attempt to unite these various metadata services under a common banner. Other developments, such as those of the Federal Geospatial Data Committee (FGDC) and the International Standards Organisation (ISO) appear confluent with European developments.

1 Definition available at http://www.fgdc.gov/metadata/
Towards a metadata base of digital spatial data for Ireland

K. Rybaczuk
M. Ovington
W.P. Prendergast

Survey Ireland - Winter 1999

Methodology

The first objective was to evaluate the metadata services and standards that already existed worldwide and to assess how they might be tailored towards a metadata service for the Irish GIS community. On the basis of this model a database would be developed from an existing standard or alternatively a composite of several metadata databases might be used.

This approach was pursued for two reasons. Firstly, the standards evaluated were rife with small semantic and terminological differences. While efforts are being made to unify the various standards and services, to add to this problem would have confused issues further. This leads into the second important objective - to ensure that when internationally accepted metadata standards (such as the draft ISO TC 211 15046-15) or services such as ESMI, (European Spatial Metadata Infrastructure) came into being, a degree of compatibility would exist with the Irish Geo-ID.

Evaluation of final and draft standards and guidelines

Evaluation began with a review of existing and draft metadata standards - primarily from Europe, the United States and Australia. Current European and International metadata services and standards were evaluated over a three-month period. Some of the guidelines evaluated are listed in Table 1. It should be pointed out that language was a significant barrier to access certain templates and standards, especially in Asia and some European countries. A comparative framework was developed to compare the various standards and services. A small sample from this ‘metadata mapping’ is given in Table 2.

In addition, the uptake and user acceptance of each standard was estimated. Metadata standards and guidelines that employed plain language and clear conceptual models for complex GIS issues were favoured. This factor alone is seen a major cause of the failure of the first version of the FGDC standard (Foresman et al. 1996).

Developing the Geo-spatial Information Directory

Those standards of most influence in the Geo-ID are highlighted in Table 1 and are briefly described below.

- **CEN TC 287 prENV 12657**
  The Comité Européen de Normalisation (CEN) technical committee standard (prENV 12657 Geographic information - Data description - Metadata) has been of considerable influence in the EU where many member states have provisionally adopted it as their de facto patriarchal standard for European metadata services. The standard is likely to merge with draft ISO TC 211 15046-15 metadata standard.

- **MEGRIN GDDD**
  MEGRIN represents 19 European National Mapping Agencies (NMAs) including the OSI. The aim of MEGRIN is to raise awareness of, and provide information about, geographic datasets and services available from NMAs and to facilitate the creation of a range of cost effective pan-European geographic datasets. The Geographic Data Description Directory (GDDD) was the first meta service to be derived from the initial draft CEN TC 287 to provide information about geographical information held by the various European NMAs.

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>NAME OF INFRASTRUCTURE</th>
<th>STANDARDS AND GUIDELINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Australian Spatial Data Infrastructure</td>
<td>Based on ANZLIC guidelines</td>
</tr>
<tr>
<td>Canada</td>
<td>Canadian Geospatial Data Infrastructure</td>
<td>Canadian Standard</td>
</tr>
<tr>
<td>Denmark</td>
<td>Indodatabase on Geodata</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>Finland</td>
<td>National Geographic Information Infrastructure</td>
<td>Policy document only</td>
</tr>
<tr>
<td>France</td>
<td>Policy Document Only by Afiego</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>Germany</td>
<td>Metainformation System</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>Holland</td>
<td>National Geographic Information Infrastructure</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>Hungary</td>
<td>National Spatial Data Strategy</td>
<td>Policy document only</td>
</tr>
<tr>
<td>Japan</td>
<td>National Spatial Data Infrastructure Promoting Association</td>
<td>Based on FGDC CSDGM</td>
</tr>
<tr>
<td>Portugal</td>
<td>National System for Geographic Information</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>South Africa</td>
<td>National Spatial Information Framework</td>
<td>Based on FGDC CSDGM</td>
</tr>
<tr>
<td>Sweden</td>
<td>Geodatabank</td>
<td>Based on CEN TC 287</td>
</tr>
<tr>
<td>UK</td>
<td>National Geospatial Data Framework</td>
<td>Based on ISO TC 211</td>
</tr>
<tr>
<td>USA</td>
<td>National Spatial Data Infrastructure</td>
<td>FGDC CSDGM</td>
</tr>
</tbody>
</table>

Table 1. National metadata services examined
Towards a metadata base of digital spatial data for Ireland

K. Rybaczuk
M. Ovington
W.P Prendergast

Survey Ireland - Winter 1999

• ANZLIC metadata guidelines
ANZLIC, the Australia- New Zealand Land Information Council was formed in 1986 as the result of an agreement between the Prime Minister, the respective heads of State Government and the government of New Zealand. ANZLIC, has since produced a number of reports on national strategies for the management of land and geographic information. The ANZLIC Metadata Guidelines have had a high degree of uptake in Australia and New Zealand and widespread acceptability as a result of the non-technical language and ease of use.

• FGDC content standard for digital geospatial metadata
The US FGDC (Federal Geographic Data Committee) was established in 1990 to co-ordinate the “development, use, sharing and dissemination of surveying, mapping and related spatial data” (Masser, 1997). Four years after the establishment of the FGDC President Clinton signed Executive Order 12906 entitled “Co-ordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure” in order to strengthen the policies adopted by the FGDC. The Federal Geographic Data Committee (FGDC) approved their content standard for digital geospatial metadata in 1994. This is a national spatial metadata standard developed to support the development of the National Spatial Data Infrastructure. The NSDI encompasses policies, standards, and procedures for organisations to co-operatively produce and share geographic data. The 16 federal agencies that make up the FGDC are developing the NSDI in co-operation with organisations from state and local governments, the academic community, and the private sector. The standard has also been implemented outside of the USA, for example for the South African spatial data discovery facility and has been a considerable influence on the Draft ISO TC 211 15046-15 metadata standard.

The structure of the Geo-spatial Information Directory

Once the fundamental components of the template were identified, they were organised into themes. For the most part, the thematic classification of the template has been structured on the CEN TC 287 standard. This proved to be the most prevalent standard in the above survey and has been adopted for use in metadata services by many member states of the EU². It is also likely to merge with the draft ISO TC 211 15046-15 metadata standard.

The themes finally selected for the Geo-ID were:
• Data set Identification – unique description of each individual data.
• Data set Extent – the spatial and temporal extent of the data.
• Data set Distribution – the medium, means and conditions of data distribution.
• Data set Quality – descriptive quality statements for the data.
• Data set Administration – organisational and administrative details for the data.
• Data set Contact – personal contact details for the data.
• Metadata Information – record of when the metadata was gathered.

The alignment of the Geo-ID with the CEN TC 287 metadata standard means it will be prepared to accommodate any future European spatial metadata development.

Table 2. Examples of semantic inconsistencies between various metadata standards and guidelines

<table>
<thead>
<tr>
<th>GEO-ID</th>
<th>CEN TC211</th>
<th>ANZLIC GUIDELINES</th>
<th>ISO 15046-15</th>
<th>FGDC CSDGM</th>
<th>NGDF GUIDELINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Dataset Identification</td>
<td>Title</td>
<td>Title</td>
<td>Title</td>
<td>Title</td>
</tr>
<tr>
<td>Abstract</td>
<td>Dataset Interview</td>
<td>Abstract</td>
<td>Abstract</td>
<td>Abstract/Description</td>
<td>Abstract</td>
</tr>
<tr>
<td>Theme</td>
<td>Data Definition</td>
<td>Search Word(s)</td>
<td>Keywords/Thematic Classification</td>
<td>Keywords</td>
<td>Keywords</td>
</tr>
<tr>
<td>Browse Graphic</td>
<td>No Direct Equivalent</td>
<td>No Direct Equivalent</td>
<td>Browse Graphic</td>
<td>Browse Graphic</td>
<td>Sample</td>
</tr>
</tbody>
</table>

² In addition, many European organisations (including CNIG and IGN, the national mapping agencies of Portugal and France respectively) are evaluating the CEN TC 287 to succeed their own ‘standards’ that were originally based on earlier drafts of the CEN standard.
Populating the Geo-spatial Information Directory

The Geo-ID is stored on a Microsoft Access 97 database. Reasons for this choice include the fact that it is one of the standard applications in the Microsoft Office 97 Professional Edition suite and therefore ubiquitous throughout the computer industry. In addition of the available database packages it was deemed one of the most ‘user friendly’ options.

Once the template was created it had to be populated. This would require information from all of Ireland’s digital spatial data producers. A list of potential data producers was compiled from several different sources. The final data producer list constituted individuals from approximately ninety organisations in Ireland, including government departments, semi-state bodies, local government, research institutes and the private sector.

The Geo-ID template was transposed into a questionnaire for data collection. This was piloted over a two-week period beginning on the 1st August 1998. Following the success of the pilot study, organisations that were identified as data producers were asked to contribute to the study. They were contacted over a 16-week period beginning on the 11th August 1998 and finishing on the 1st December 1998. Figure 1 describes the procedure of data collection.
Ninety-one organisations were each sent a letter describing the project and requesting an interview. Each letter was followed within a week by a phone call to ensure contact was made with the correct persons and describing the project in more detail. The initial phone call also ascertained whether the organisation was a data producer or not, and whether they could satisfy the four criteria for data entry into the Geo-ID. These were:

- The data must have some form of direct or indirect geo-reference.
- The data must be available in a digital format.
- The administrative organisation must have either copyright or distribution rights for a given dataset.
- The data must refer to some part of Ireland or her territorial waters.

Figure 1 shows that a little under half the organisations originally listed were found to be a) suitable and b) willing to comply. Non compliance from producers originally identified arose from a number of reasons.

- Organisation did not produce or use any form of digital geo-spatial data.
- Organisation produced data but it failed to meet one of three basic criteria (as described above).
- Organisation did not produce digital data or possess copyright of digital data and was exclusively a licensed third party user.
- Organisation produced data for a third party and did not retain copyright of the data it produced.
- Organisation data was strictly confidential and would never become available within the public domain.
- Organisation did produce digital geo-spatial data but did not want to contribute to the project.
- Organisation deferred participation as it was not at present a producer of digital geo-spatial data but would at some time in the future and wished to contribute at a later date.

Any further queries that interviewees may have had were answered and an interview date was provisionally set. Occasionally further difficulties were encountered before the interview. These problems reduced the anticipated 46 organisations to 31 contributing organisations (Figure 1). Goodwill is also a significant part of any data audit therefore there was no obligation to contribute to the Geo-ID.

Results – disseminating the Geo-ID

Once the data were gathered and the database populated, consideration needs to be given to the dissemination of the meta database. Dissemination methods were evaluated on two fronts. The primary means of evaluation involved a detailed review of current practice with regard to the ‘Information Clearinghouses’ of Europe, America and Australia. Secondly, a questionnaire to evaluate Irish user requirements was circulated at the IRLOGI GIS 1998 conference. The objective of this questionnaire was to determine the current use of metadata in the Geographical Information community in Ireland and to examine the awareness and perceived requirements of Irish users for a distributed meta database. A total of 44 questionnaires were received. Respondents ranged from local government, the state sector, and commercial and third level researchers. The results of the questionnaire were of particular interest as they showed the preferences of a representative sample of the Irish GI market. The findings indicated that in the past most Irish users collected information about potential data sources from personal enquiry and personal contacts. However, such methods of data collection have their drawbacks. For example, informal methods of data gathering do not offer comprehensive or consistent information about data. Therefore even though data may be located, further information may not be available. This became apparent when users were asked to comment on the data found through their current information search methods. Personal experience and enquiry with established data producers were the traditional means of finding data in the past. The remaining methods, library and database searches, conferences and the Internet remained low in usage among the GI community.

When users were asked what their preferred method of dissemination of Irish metadata would be, the majority of respondents said they would prefer to see the information distributed via the Internet (Figure 2).

Aside from user preferences, studies of other meta databases have revealed that factors of particular importance in the investigations of appropriate dissemination media include the complexity of the published data, the technological requirement and the means and logistics of dissemination. Taking these issues into consideration, it would appear that the Internet offered the most appropriate method for dissemination from the user perspective.

---

3 Under the terms of the 1937 Constitution of Ireland, which takes precedence over all other legislation, the name of the state is Ireland in the English language and Eire in the Irish language. The 1948 Republic of Ireland Act redefined the status of the state as a republic, but did not change the name or title of the state in any way from that set down in the constitution. Digital mapping activities within Northern Ireland were expressly excluded from this study.
Towards a metadata base of digital spatial data for Ireland

K. Rybaczuk
M. Ovington
W.P. Prendergast

Survey Ireland - Winter 1999

Other issues included security, cost, interoperability and update. The Internet offers the security of centralised database control, low cost for update and distribution and world-wide access. Furthermore, in order to come under the aegis of the proposed ESMI the database must be available on-line – thereby making the Internet the only real option for interoperable dissemination.

Other dissemination options included paper and CD ROM. Paper publications have a slow update cycle and tend to be comparatively expensive to produce and distribute, thereby making them unsuitable. CD-ROM offers significant storage capacity and easy access. However, the constant updates required with a valid meta database incur additional production and distribution costs. In both cases the proliferation of different versions of the Geo-ID database would lead to confusion.

A pilot of the Geo-ID database, the Geo-ID Pilot, has been developed for dissemination on the Internet and is available at http://www.tcd.ie/Geography/GIS/GeoID/.

Discussion and recommendations for Future Development

The pilot Geo-ID will need to be promoted, extended, maintained and developed. Currently, the onus for these tasks falls at the feet of IRLOGI. However, continued support of the initiative may be beyond the financial and political remit of the organisation. Whether through the auspices of IRLOGI or otherwise, awareness of the Geo-ID will need to be raised both within the GI community and to a more general audience.

Secondly, the data contained within the Geo-ID will need to be maintained. Original data collection occurred in the latter half of 1998, and is already nearly a year out of date. A programme of either proactive or voluntary maintenance must be encouraged to ensure the validity of the Geo-ID database.

Thirdly the Geo-ID needs to be further developed especially in the light of the increased influence of information and communications technologies (ICTs) on Irish society and commerce, especially e-commerce. It is common now for many countries including many European member states, America and Australia to have a national policy for spatial information, usually referred to as a National Spatial Data Infrastructure (NSDI). The Australia-New Zealand Land Information Council (ANZLIC) describes the major components of the NSDI as:

- Standards to facilitate data collection, documentation, access, and transfer.
- A basic framework of digital spatial data that meets the minimum needs of large numbers of data users over any given geographic area.
- Clearinghouses to serve, search, query, find, access, and use spatial data.
- Education and training in the collection, management, and use of spatial data.

The Geo-ID provides a nucleus around which an Irish NSDI could be developed. It is important that the momentum gained as a result of the development of the Geo-ID is not lost and it should be maintained, developed and published to be of benefit to the Irish GI community. It is anticipated that the initiative should eventually be lead by central government as part of a funded information infrastructure project consistent with government information policy.

Finally, the Geo-ID needs to be co-ordinated with European developments. The European Spatial Metadata Infrastructure (ESMI) aims to unite the NSDIs of EU member states. This will be achieved not through federal mandate, as has been the case in the United States. ESMI are in the process of developing a standard process of interrogation for the NSDI of member states, providing mechanisms for linking disparate metadata services via the Internet. By ensuring that metadata services are able to communicate with each other, a single access to the ESMI node will allow access to all connected metadata services. ESMI is seen as a major step toward realising a European Information Infrastructure and is part funded by DG XIII-E of the European Commission and is part of the INFO2000 programme. Therefore it is vital that Ireland maintains its links to European initiatives and also her commitment to interoperability. Devising a precise mechanism to realise these aims is yet to be defined.
Bibliography


