Go-Geo! resources: supporting UK academic spatial data management and sharing

Welcome to the third issue of Go-Geo! Metadata News, the biannual newsletter for UK academics, researchers and students interested in issues and news surrounding spatial data management and sharing. The newsletter also covers geospatial standards and metadata, and announces service updates for the Go-Geo! portal, Metadata Editor tool and resources for supporting metadata creation.

Go-Geo! is an online service provided by EDINA, a National Data Centre at the University of Edinburgh. It hosts a suite of JISC-funded online resources supporting geospatial metadata creation and spatial data discovery for the UK academic community. The Go-Geo! Metadata Editor tool, Academic Geospatial Metadata Application Profile (AGMAP), geospatial metadata guidelines and online teaching materials aid in the creation of metadata records and the Go-Geo! portal makes publishing records easy. Cross-searching other data catalogues is possible with Go-Geo!, which also provides access to a full range of online GI resources including current news, conferences, other events, links to free GIS software and online services and more.

Go-Geo! Metadata News circulation is intended to reach the wider geographic information community in UK academia.

Readers are invited to submit articles or comments for publication in future newsletters. If you have anything to share, please contact the EDINA helpdesk (edina@ed.ac.uk) or call the helpdesk at 0131 650 3302. The submission deadline for the next issue is 16 November 2007.

Name the Go-Geo! Metadata Editor tool and win a £25 voucher

Readers of the Go-Geo! Metadata News newsletter are invited to submit a new name for the Go-Geo! Metadata Editor tool. Suggestions already received include ‘GeoMetaMaker’, ‘DataDoc’, ‘GeoDoc’ and ‘MetaMagic’ tool. If you can come up with something better, let us know your ideas.

Please email your suggestion to Tony Mathys (tony.mathys@ed.ac.uk).

All names submitted will be judged 24 September 2007 with the winner receiving a £25 voucher.
Metadata teaching and learning modules are one step closer to completion

Work is nearing completion on the first of two teaching and learning modules on metadata. The first module introduces the concept of metadata and provides examples to highlight how metadata is relevant to everyday life ranging from book searches at a library to information about products.

The second module looks at metadata in a geospatial context. Users of this module will be introduced to geospatial metadata and standards as well as an overview of the benefits of creating metadata.

Both modules are Flash-animated and include soundtracks to make the content interesting for the user. Versions of the modules will be provided to support users with disabilities.

Metadata in the news

The EU INSPIRE directive was officially published 25 April and is now in force.

INSPIRE aims ‘to make interoperable spatial information readily available in support of both national and Community policy and to enable the public to access this information.’ Initially the Directive will focus on information needed in order to monitor and improve the state of the environment and the underpinning ‘multi-purpose’ spatial data. However, it is expected to widen to include other types of geospatial data.

Key components of the directive include the following:

“Member States shall ensure that metadata are created for spatial data sets and services, and that those metadata are kept up to date.”

“Member States shall establish and operate a network of the following services for the spatial data sets and services for which metadata have been created in accordance with this Directive: (a) discovery services making it possible to search for spatial data sets and spatial data services on the basis of the content of the corresponding metadata and to display the content of the metadata; (b) view services making it possible, as a minimum, to display, navigate, zoom in/out, pan, or overlay spatial data sets and to display legend information and any relevant content of metadata; (c) download services, enabling copies of complete spatial data sets, or of parts of such data-sets, to be downloaded; (d) transformation services, enabling spatial data sets to be transformed; (e) “invoke spatial data services” services, enabling data services to be invoked.

Those services shall be easy to use and accessible via the Internet or any other appropriate means of telecommunication available to the public.”

Go-Geo! portal is getting a new look

Visit Go-Geo! in several months time and you will discover that the portal has a new look. Go-Geo! is now being redesigned to enhance and improve existing features. Colour schemes will be applied to Go-Geo! to make navigating across the site easier for users. GI-related resource channels are being reorganised to give more prominence to books and other resources which users frequently access. New theme categories are being introduced for software, books and online resources to organise the 100s of entries. Icons have also been created to serve as visual aids for users to quickly assess cost, accessibility, service-type and other criteria for GI-related resources such as online services and data providers.

The Go-Geo! portal will also offer users a new mapping interface for the UK, a dedicated page with metadata resources including the metadata teaching and learning modules, information about metadata workshops, links to standards and support documentation for UK AGMAP, the metadata application profile created for UK academia to support the documentation of spatial data.

The Go-Geo! portal will continue to be the main source for users to search and discover spatial data in the academic community and across the UK. The portal will continue to serve as the access point to the metadata editor tool, which has been created to support spatial data documentation. Records created using the tool can be published on Go-Geo!.

Following the release of the new look Go-Geo!, the team will be working to extend the portal to hold international data and GI-related resources in support of academics and students conducting research in other countries. This is an important step towards the realisation of Go-Geo! becoming a major component of the Spatial Data Infrastructure (SDI) for UK academia.
GRADE: Scoping a Geospatial Repository for Academic Deposit and Extraction - Findings from the GRADE Project

The GRADE Project (December 2005 – April 2007) investigated and reported on the technical and cultural issues around the reuse of geospatial data within repository infrastructures. In this article we report the main findings from the exciting and innovative project.

The GRADE geospatial repository demonstrator (orange button link on the left menu, via the GRADE project web site home page: http://edina.ed.ac.uk/projects/grade) was set up to elicit feedback from the geospatial community as to their requirements for a repository capable of managing geospatial data. However, the GRADE repository hosts 162 GIS-research datasets to date, covering a huge variety of topics from: epidemiology studies of disease/health issues, archaeological sites, political and historical boundary datasets from the GB Historical GIS Project & other contributors, ecology species and biodiversity mapping, land-use and land-cover, Digital Elevation Models, glacial features mapping data, satellite imagery… to many individual social-based datasets such as ‘Locations of graffiti in Hackney’. There are many subject areas of GIS data that would be useful to other researchers in the UK. To register go to: http://gradedemo.edina.ac.uk/dspace/register. If you think also that you have GIS-research datasets that you would like to publish to the GRADE geospatial repository that would be great! Do get in touch and email grade-support@ed.ac.uk or Rebecca Seymour at bex.seymour@ed.ac.uk.

The project also undertook various surveys and consultations with geospatial data users. One of these looked at the key functions a repository should offer for those sharing and seeking geospatial data, a repository capable of absorbing complex GIS data (storage and retrieval and ingest). In terms of searching, users wanted to be able search via a map or location and via research subject, with support for both semantic (ontological) and standard geographic keywords. Being able to search from one’s academic Institution or within established GIS/geographical peer groups was also important. Users also wanted the deposited data to be grouped into collections by continent and country (the most important to the UK researcher being the UK and EU). For data discovery purposes, the basic Dublin Core metadata standard was sufficient but needed to be enriched with ‘geographical’ elements (such as, coordinate and projection datum, viewable extents, licensing/IPR rights, and data sources and lineage information).

GRADE geospatial repository demonstrator surpassed expectations by generating significant user activity. There are now 149 registered active users from a total of 40 HE-FE institutions in the UK. In addition to this there have also been an equally high number of users who were non-Digimap registered users who could not be permitted to use the demonstrator repository. Registrations are now on average one per day.

Downloading data straight to a desktop GIS was the favoured method for accessing spatial data in a repository, with the data extracted being...
accompanied by detailed metadata, ideally a full ISO 19115 metadata file.

When depositing data in the repository, there was an expressed need for the format of the GIS files to be verified at the upload/data-submission stage (GIS file types have complex arrays of component files and wide number of GIS file types and sizes, there needs to be automatic control of this). There also needs to be the ability to deal with upload of ‘packaged’ datasets, to retain the internal integrity and links of multi-layered and inter-connecting data within a project – while allowing data to be searched as single entities. A geospatial repository would also have to be able to deal with GI-data-versioning and time-series.

Further elements and issues that would be important for any repository were also flagged. For example, how to cite a GIS dataset stored in a geospatial repository in a resultant publication that a reader of the publication could link to the data. There are also issues relating to the following:

- sustainability such as data archiving and curation, and of trust between the depositor and the repository owner – that data will be looked-after, curated and maintained;
- metadata being harvested from the repository;
- machine to machine interoperability between repositories;
- IPR and licensing and data-sharing rights need to be formalised and qualified. (A report detailing the full findings of this work is available on the GRADE project web site).

Another deliverable produced for the GRADE project was the legal work on copyright with derived geospatial data completed by Dr Charlotte Waelde at the AHRC Research Centre for Studies in Intellectual Property and Technology Law, University of Edinburgh. To access the complete report visit: http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf.

The report was also picked up by the Free Our Data Campaign in the Guardian on April 05 2007: http://technology.guardian.co.uk/weekly/story/0,,2049772,00.html.

The main thrust of this investigative work has been the proposition that the EU Database Directive is the applied law which should govern digital GIS data; not Copyright Law in the UK. This is a controversial concept - whether one can have copyright within a digital database, defined on the condition that the database owner has 'contributed his own artistic effort' in creating it. However, the counter-argument is that geographical data, such as mountain height as an example, is a phenomenon that exists to all, hence absent of any creativity in surveying and recording this. For a useful synopsis, you can read more on Mike Smith (Kingston Centre for GIS) blog: http://www.journalofmaps.com/cgi-bin/blosxom.cgi/GIS/freeourdadata_article.html http://www.journalofmaps.com/cgi-bin/blosxom.cgi/GIS/GRADE_Waelde.html

And the debate continues on….

A third key area within the GRADE project was the consideration of the role of subject-specific data repositories versus institutional repositories (IR). This survey was completed by Pauline Simpson at the National Oceanography Centre, University of Southampton. The final report is on the GRADE website at: http://edina.ac.uk/projects/grade/GRADE_Survey_Report.pdf. The main finding revealed that there were a few IRs interested in absorbing geospatial datasets in the future, but they lacked the existing capability of doing so. Most existing IRs were focussed on Prints materials. An analysis of the positive and negative aspects of subject versus IR for hosting geospatial data indicated that subject-based repositories were more suitable. This is due to the complexities and large volume sizes of geospatial data, compared to other types of scientific data. The overarching message was that the repository environment is rapidly
evolving, and within this there is room for both to have a role. The key point being that it is better for GIS-research data to be archived. If it fits within a Data Centre then it goes there, but if not, it is better that it is deposited in a institutional repository than being ‘lost’ on personal servers or within individual departments.

A fourth central deliverable as part of the GRADE project was investigating informal methods of geospatial data-sharing and sharing practices. This is the first time that there has been a qualitative and quantitative assessment of ‘informal’ sharing patterns and networks between GIS-academics. The results of the Informal Data-sharing questionnaire will be made available on the main project website: http://edina.ac.uk/projects/grade/. The questionnaire showed that the main routes of transfer of geospatial data remain still very traditional, via email and CD/DVD. Newer peer-to-peer sharing technologies have not had the take-up or fan-base expected, and are limiting in terms of GI-specific capabilities.

Evidence shows research geospatial datasets are being shared and presumably reutilised, hence a demand for GI-data. What is indicative is that this sharing is happening between a collective group of users, strongly identifying itself as within a GIS-specific or subject-specific community. What is indicative is that collective groups of users, within GIS or subject-specific communicaties, are communicating and inter-sharing data.

The main, and by far most significant barrier to sharing of geospatial data is issues and confusion over Digital rights, Copyright, and Licensing of GIS data. This can be seen to be one of the major causal factors that has driven more underground ‘informal’ practices of geospatial data-exchange; and, is of utmost priority to the development of any practical formal-sharing mechanism solutions.

Despite this, the top ‘wish list’ factor that the surveyed GIS-academic-community wanted was to have a formal National Geospatial Repository resource, closely followed by having a national central find portal. Clearly, the current informal ad-hoc and personal-level geospatial-sharing is not satisfying a demand.

This work was followed by a workshop considering the use of peer-to-peer technologies to facilitate geospatial data sharing. Investigated by Robin Smith at ICOSS at the University of Sheffield, his efforts entailed looking at social-dynamics and routes of transfer and the importance of networks. This report will be on the main website shortly, and is to be the start of a much wider research on this topic.

A final project deliverable is a report focusing on interoperability issues for geospatial data within repositories. A report addressing this area will be available on the GRADE project web site in the near future.

Although the purpose of the demonstrator repository was never about achieving or targeting the number of users and GIS-datasets deposited, this remains an incredibly positive aspect. And despite the end of funding for the scoping of the GRADE Project, due to its enormous success, the Geospatial Repository will continue.

If you have any further questions about the GRADE project, or would like to know more details about how to register with the GRADE geospatial repository demonstrator or get involved, email the GRADE team on: grade-support@ed.ac.uk. For the latest news see the GRADE project website: http://edina.ac.uk/projects/grade/
Workshops

Geospatial metadata workshops for 2007-08 are being scheduled and the following universities have already offered to host workshops:

University of Sheffield
University of Exeter
University of Manchester
GeoData Institute, University of Southampton

Please contact Tony Mathys (tony.mathys@ed.ac.uk) if you wish to attend one of these workshops, or host one. All that is required is a computer lab with internet access and facilities for giving a presentation. EDINA can provide all the necessary support to publicise, organise and run the workshops.

Upcoming events

Please visit the ‘Selected Conferences and Events’ resource channel of the Go-Geo! portal for current updates at http://www.gogeo.ac.uk/cgi-bin/geoPortal10/Res_AnnualConference.pl

Did you know?

• Go-Geo! portal’s search functionality puts the user in reach of more than 2,400 geospatial metadata records published on numerous UK catalogues.

• In the EU Motive Project (http://www.motive.net) EDINA is working with others on the next generation of data catalogues. These will allow you to not only find data but also link to web services that can deliver the data to your desktop.

• There are now more than 1,300 GI-related resources on the Go-Geo! portal including links to 60+ free software downloads, 75 online services, many with interactive mapping, and 311 book titles from the leading publishing houses including Elsevier, ESRI Press Sage Publications, Springer, Taylor and Francis, Wiley and others.

• The upcoming launch by EDINA of OS MasterMap within Digimap will mean the release to the UK academic community of spatial data with topology. This will set the stage for the creation of numerous derived datasets to support teaching and learning as well as research. Now is the time to become familiar with the Go-Geo! metadata creation resources so you can document the data you create.

• At a recent workshop, Professor Fraser Taylor (Carleton University, Canada) observed that some 80% of the geospatial data collected in Canada over the last 30 years was now lost. Our own research suggests that within UK academia, thousands of datasets are already lost, not really managed or have no official home, what we call ‘orphaned’ data. Now is the time to act so more datasets are not destined to the same fate, Better to preserve them for future generations of academics and students.

• A Department of Energy study in the US estimated that costs incurred for the retrieval of documents ran into millions of dollars. Metadata was seen as the solution for this.

• Plans are afoot to create a lite version of the Go-Geo! metadata editor tool for users wishing to document their spatial datasets for discovery purpose only. The number of elements will be reduced to only those comprising the mandatory UK AGMAP elements. Half of the 30+ elements represented in the lite version require only contact details. A metadata record can be done before the coffee is done.