

Curating Geospatial Data

Geospatial data relate to the location of geographical features and relationships between those features. They are vital for a wide range of business and government functions including defence, transportation, education, engineering, and recreation; in fact, almost all activities rely on geospatial data to some extent. It is widely accepted that the majority of all data held in corporate and government databases include some kind of geospatial characteristics. Such data are important not only for contemporary purposes but are also essential for long-term analyses. Active data management is therefore required to curate, preserve, and provide access to reliable geospatial data over time.

Short-term benefits and long-term value

Geospatial data have typically been produced in analogue form, for example, as paper maps depicting topographic information such as streets, buildings, rivers etc. These data are now more often created, maintained and accessed digitally, increasing the ease with which they can be re-used or re-purposed. Short term advantages of curating geospatial data include:

- Maximising investment: up to 80 percent of the cost of implementing and using a Geographic Information System (GIS) may be tied to the collection and creation of geospatial data. Curation activities help to maximise the initial investment.
- Maximising research benefits: by sharing data from a variety of organisations and countries, comparative analyses of reliable data can be carried out with a speed that would have been impossible to imagine when sharing analogue resources.

Longer-term benefits from geospatial data curation draw on the above and also enable:

- Providing legal evidence: Emerging international agreements such as the Kyoto Protocol may have a huge impact on the value of geospatial data as legal evidence, but the data must be properly curated to be of value.
- Minimising the risk of losing unique data: Even if financial resources are not an issue, it may not be physically possible to recapture the data as it may be temporally dependent.
- Providing historical evidence: Preserving digital geospatial data will provide a record of a given place over time.

HE/FE and eScience perspective

Geospatial data are increasingly used as an interdisciplinary tool and are valuable in a range of academic subjects: 'The use of geospatial data is acquiring new significance as many academic disciplines realise that spatial relationships in the phenomena they study are vital to their understanding. Subject areas range from traditionally spatial disciplines such as geography and geology, through to biology, business, archaeology, epidemiology and history. The relevance of these relationships means that many research projects now collect, analyse and output geospatial data [...] that can be of relevance and use to other researchers and organisations.'

Smith, Mike J in *Use Case Compendium of Derived Geospatial Data*, Dec 2005

Good curation practices are vital for digital geospatial data to be shared between the range of academic disciplines for which they hold value and therefore used to their full potential.

Roles and responsibilities

Data must be created in a reliable and contextually-clear manner if they are to be successfully re-used. This is a particular challenge for geospatial data, which are characteristically complex and vary widely in terms of format, size, portability, quality, accuracy and value. Creators must implement sound creation practices to ensure that their data are re-usable and sustainable over time, and clearly identify any processing the data have undergone. Effective records management practices should also be implemented to ensure that valuable geospatial data are appraised and selected for long-term curation.

The criteria used to appraise geospatial data should reflect both institutional requirements and the potential re-use value of the records for the wider community. Data stewards or curators can establish such criteria according to institutional priorities and financing availability, and facilitate transfer or management of the data for long-term curation. Collaboration between creators and curators will enable sufficient metadata to be preserved with complex data resources so that they can be managed and reliably re-used with authenticity and integrity intact. Due to the often proprietary nature of geospatial data, IT staff must maintain a technology watch to ensure that data formats do not become obsolete and the data remains accessible.

Geospatial data management guidelines and policies help ensure that the necessary tasks are carried out by the relevant personnel, and integrate geospatial data management into the wider organisational activities.

Additional resources

Smith, Mike J in *Use Case Compendium of Derived Geospatial Data*, (Dec 2005)
<http://edina.ac.uk/projects/grade/usecasecompendium.pdf>

CEISIN, *Guide to managing Geospatial electronic records* (June 2005)
<http://www.ciesin.columbia.edu/ger/GuideToManagingGERv1Final.pdf>

Go-Geo website - a tool designed to help users find details about geospatial datasets and related resources within Great Britain tertiary education and beyond. <http://www.gogeo.ac.uk/index.html>

The Open Geospatial Consortium, Inc. (OGC) - <http://www.opengeospatial.org/>

North Carolina Geospatial Data Archiving Project (NCGDAP) <http://www.lib.ncsu.edu/ncgdap/>

NARA Transfer Instructions for Permanent Electronic Records: Digital Geospatial Data Records
<http://www.archives.gov/records-mgmt/initiatives/digital-geospatial-data-records.html>

Federal Geographic Data Committee – a US inter-agency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis.
<http://www.fgdc.gov/>

Scientific formats for geospatial data preservation - A study of suitability and performance
http://www.ncsa.uiuc.edu/NARA/Sci_fmts_and_geodata_HDF.pdf (2004)