

Some thoughts on Digital Archiving

Christine Earl and Tracey P. Lauriault

As creators of cartographic data and maps, cartographers are in the business of producing a specialized historical record for future generations, a record of our environment for our period in history, a record of our society's values, and a record of our technical achievements. It is a record which is worth preserving for future generations, generations which will have different technologies, methodologies and knowledge assets to assist them in uncovering social and environmental phenomena we can not uncover today, and who may wish to study our data. It is also something we would surely like to be able to look back on in our own lifetime to see where we've come from and how we got here.

But consider the range of physical forms of this record over the last thirty years. How enduring will these forms prove to be? Paper has its limitations as a storage medium but nevertheless, paper documents have survived many hundreds of years. What of electronic media? What will happen to today's representations which, in many cases, exist only as digital files coded in proprietary formats and requiring currently existing hardware and software to interpret them? What about the ephemeral representations of interactive Internet mapping? Will researchers in 2033 really be able to access our data and look at them in light of new ideas?

These issues have only recently begun to be addressed and, so far, it is safe to say that they have been dealt with in scattered and ad hoc ways and not in terms of any comprehensive set of policies, standards and strategies (Brown, 2003). Many CCA members will remember the Canada Land Inventory, an initiative which motivated the development of the first GIS – these data on land use, a valuable historical record, have been rescued from oblivion only through a massive and costly joint restoration project conducted by four federal ministries in Ottawa (Bleakly, 2002).

There are now the beginnings of a groundswell of research and interest in the issues of preservation of digital geospatial data. The social science data and library communities are actively discussing them and now some early Canadian geospatial data archiving policy and research initiatives are gaining momentum. The Ad Hoc Committee on Archiving and Preserving of Geospatial Data under the Policy Advisory Node of GeoConnections is investigating an information management infrastructure that will enhance the Canadian Geospatial Data Infrastructure and as well, the International Research on Permanent Authentic Records Electronic Systems II (InterPARES II), a collaborative research project funded by the Social Sciences and Humanities Research Council (SSHRC), has under its Scientific Focus, two case studies investigating archival issues related to geospatial data; they are archaeological GIS records and the Cybercartographic Atlas of Antarctica. Research into and interest in the preservation of digital geospatial data is beginning to increase and as creators of these data, cartographers and geographers ought to pay attention to these initiatives and think about what best practices we can evolve to enhance the durability of our work.

Preservation Issues

Among the issues which relate to the preservation of digital data, technological obsolescence is paramount and several strategies have been proposed for its mitigation. These are a) the deliberate maintenance of old hardware and software; b) the incorporation of the capacity to emulate old technology in the development of new; and c) periodic migration of information or data from one storage medium or format to another (see <http://www.nla.gov.au/padi/topics/18.html> for a good summary of preservation topics).

No less essential than the data themselves is information about them in the form of complete metadata, including information about the archival history of the data and a measure of data accuracy to accompany archival authenticity. The metadata can be stored with the data themselves, or separately but linked to them, and must always be migrated or refreshed with the stored data. The metadata should provide contextual information for the data archive so that the circumstances of its creation and use may be correctly understood. This is to ensure future users will have due regard for the original data context. To date there are some excellent (albeit not universally adhered to) geospatial metadata standards, and some emerging standards for information objects such as video, animation, and photographs, although these need to be extended for scientific uses. However, there are yet no standards for multimedia information objects such as sonification, visualizations, scent, real time webcams, heightened reality, flight simulations, etc. The emergence of new forms of digital media being georeferenced and hyperlinked to Internet maps and atlases and the growth of distributed Internet mapping systems are presenting significant archival challenges.

The costs of establishing and maintaining good and useful archives of digital geospatial data will be considerable, especially in the initial stages while policies and strategies are hammered out, but if our work is to have any long-term value, such costs will need to be borne. In future, perhaps archival and information management practices will be included routinely in project estimates and become standard practice.

While there is obviously no single, simple solution to the problem of preservation, it is encouraging that social scientists, librarians and archivists are already giving some thought to geospatial data. These data have unique value. Are those of us in the cartographic field thinking about the problems of their preservation? We invite you to consider the issues raised in this essay and to send your comments to the Editor for inclusion in a future issue of *Cartouche*. Please tell us what you do and what you think should be done. All contributions will be welcomed.

References

Bleakly, Denise R.(2002), "Long-term Spatial Data Preservation and Archiving: What Are the Issues?", Sandia National Laboratories (SAND) Report 2002-0107, Albuquerque, New Mexico.

Brown, David L., Welch, Grace and Cullingworth, Christine (2003), "Management and Preservation of Geospatial Data" (Draft), Ad-Hoc Committee on Archiving and Preserving Geospatial Data, Policy Advisory Network Node, GeoConnections, Ottawa.