

The Preservation and Archiving of Geospatial Digital Data: Some Challenges for National Mapping Agencies

D. R. F. Taylor, T. P. Lauriault and Peter L. Pulsifer
Geomatics and Cartographic Research Centre, Carleton University,
Ottawa, Canada

Presented to the Cambridge Conference, July 2007

Outline

- What is the problem?
- What is the scale of the problem?
- What are the causes of the problem?
- Case studies addressing the problem:
 - InterPARES 2
 - Cybercartographic Atlas of Antarctica
- Potential solutions
- What should national mapping agencies do?

The Problem

- All of the geospatial sciences are making increasing use of Internet technology. Internet mapping, such as Google Maps and Google Earth, is increasingly popular with the general public. National mapping organizations are producing more and more digital data. Much of this digital data is being lost, some of it permanently.

The Scale of the Problem

Cartographers are not alone!

- The entire contents of the 1960 US Census have been lost.
- Twenty percent of the data generated by the 1976 Viking exploration of Mars has been lost.
- Fifty percent of the work of digital composers has been lost.
- Twenty-five percent of the work of digital photographers has been lost.

(John Roeder, InterPARES 2, 2007)

The Scale of the Problem (continued)

Even where the data have not been lost the cost of recovery has been shown to be very high.

- The British Domesday Survey of 1986
- The Canada Land Inventory

The Scale of the Problem (continued)

“In archiving terms the last quarter of the 20th century has some similarities to the dark ages. Only fragments or written descriptions of the digital maps exist. The originals have disappeared or can no longer be accessed.”

(Taylor, 2005)

The Causes of the Problem

- Insufficient attention to the issues being given by data producers. Archiving is often seen as someone else's problem.
- The difficulty of preserving ephemeral and rapidly changing data. Digital technologies themselves give a false sense of security. Existing preservation and backup policies and procedures are often inadequate. At best they preserve, but they do not archive.
- Rapidly changing technologies are leading to obsolescence and lack of backward compatibility.
- Cost and other institutional issues. Archiving is expensive and there is often no money in the budget for it. Databases are often overwritten for operational purposes rather than preserved.

Case studies Addressing the Problem

- **The InterPARES 2 project** (International Research on Permanent Authentic Records in Electronic Systems)
<http://www.interpares.org>
 - A large international, interdisciplinary project supported by the Social Science and Humanities Research Council of Canada that studied the life cycle of records created in complex digital environments from creation to permanent preservation
- **The Cybercartographic Atlas of Antarctica**
 - A complex, multimedia, online atlas which was a scientific case study for InterPARES 2. The production of the Atlas was part of a large research project supported by the Social Sciences and Humanities Research Council of Canada entitled Cybercartography and the New Economy
<http://gcrc.carleton.ca>

Potential Solutions

- The InterPARES 2 project and the experience we have had at the Geomatics and Cartographic Research Centre at Carleton University with the challenges of preserving and archiving complex geospatial information have led to some solutions. Researchers in the InterPARES project suggest a number of these.

Potential Solutions (continued)

1. Select software and hardware that offer the best hope that digital materials will remain accessible over time. Software which meets open specifications and standards is best.
 - Choose software that preserves material as they originally appeared – same look and feel over time. Ensure backward compatibility.
 - Choose software and hardware that allow storing of digital materials easily – interoperability
 - Use software that adheres to standards – ISO and OGC
 - Maintain full specifications and documentation of software, especially if you build it yourself
 - If you customize software make sure you fully document any changes
 - Fully document the construction of your system as a whole to ensure its accessibility. Fully document the context, not just the data.

Potential Solutions (continued)

2. Ensure that digital materials maintained as records are stable and fixed in both content and form. A record is different from a publication or data. Its content is unchanged and unchangeable. Not all cartographic materials are records but some are, e.g. cadastral land records, fundamental datasets.

Potential Solutions (continued)

3. Maintain full identity metadata for all records.

Potential Solutions (continued)

4. Ensure that digital materials carry information that will help verify their integrity – integrity metadata

Potential Solutions (continued)

5. Organize digital materials into logical groupings. Handle them in large sets rather than one by one. Separate records from other digital materials. Organize around the way you work and fully describe the process.

Potential Solutions (continued)

6. Use authentication techniques that foster the maintenance and preservation of records, e.g. cryptology and digital signatures

Potential Solutions (continued)

7. Protect your records from unauthorized action.

Potential Solutions (continued)

8. Protect your records from accidental loss and corruption. Back up regularly and often. Have a rigorous policy for daily backup. Choose a backup technique best suited for your situation.

Potential Solutions (continued)

9. Take steps against hardware and software obsolescence. Transfer hardware functions to software whenever possible and use software which fully meets open standards and specifications

What Should National Mapping Agencies Do?

- Consider issues surrounding long term preservation and archiving as an integral part of your business plan including costs.
- Preservation and archiving issues are best considered at the time the digital records are created, not after the fact.
- As part of your business plan decide what needs to be preserved and why. The major problem in archiving is deciding what not to archive. Decide what your core datasets are and archive them regularly.
- Develop partnerships with your national archives and national libraries.
- Do not be over-dependent on technology as a solution. Storage technology is improving rapidly but storing and preserving data is not the same as archiving it.