Who's Saving the Files? Towards a New Role for Local Map Collections?

by ROBERT B. PARRY

INTRODUCTION

The theme of LIBER's 2002 conference of the Groupe des Cartothécaires, "Strategies for survival", reflects the increasingly widespread fear of map librarians that the future of their collections (and therefore also their profession) is threatened from numerous sides: by changing perceptions of the value of their predominantly paper collections by their paymasters (not so much perhaps by their users), through commodification of data (making it increasingly expensive to obtain and restrictive to use), by the perpetual squeeze on public service provision, and not the least by the direct distribution of data digitally via the web. This fear gave rise recently to a heated debate on MAPS-L, the American map librarians' list-server, of a paper on "The map library's future" (Keller, 2001). Although Keller's views were not popular with many of his respondents, at first sight his arguments seemed relatively uncontentious. Keller saw the role of the future map librarian as a facilitator who would assist in finding, understanding and using geographical information, and perhaps provide specialized hardware facilities, e.g. for printing large format maps from digital data. The data, however, would not necessarily be stored in a map library, nor would it be necessary physically to access the data sets from a map library. So the curator becomes less of a curator, but more an expert advisor and source of knowledge.

Increasingly government agencies are constructing their own National spatial data infrastructures and providing web-facilitated access to both metadata and data, thus subverting the role of map collections. But to preserve superseded data has perhaps been of less concern to such government institutions. Campbell has suggested that a major role of the 21st century map library should be to archive digital data, a role in his view for the legal deposit libraries (Campbell, 2000). This suggests that the curatorial role of the map curator should in fact continue into the digital age. But why, and how?

This paper presents arguments in support of such a role, and suggests that it should be shared beyond the domain of national libraries to include also smaller collections, such as those in universities and, in the UK context, in county record libraries. Although major national collections, with their legal deposit status, are pre-eminent in archiving and preserving their nations' mapping, university collections have had an important role to play in the tertiary spatial education of students and in research in their host institutions. The future of these collections is perhaps even more threatened than that of

major national ones, since university collections are perceived as being less concerned with the archiving of historical material and more concerned with contemporary data. Yet I shall contend that all map curators have an interest, and perhaps a responsibility, to look to the preservation of contemporary data sets, which are now in a more fragile digital form than their hard-copy predecessors. There is a role for collections, both small and large, regional and national, to become pro-active in this area.

This paper does no more than to ask a few questions and present a few ideas, based largely on its author's rather parochial, British-based experience and impressions. It would be interesting to learn however whether it creates any resonance with the experiences of LIBER members in other parts of Europe. The author plans to carry out a more extensive trawl of UK university and county record collections, and data providers, in order to investigate whether there is more evidence to underpin the arguments presented here.

LOOKING BACK

First, let's look back. A column in the periodical *Geotime,* recently reported the comment of a contributor to a National Science Foundation conference that believed a digital library should only contain material "born digital". We live in an age when instant and current are good, and the past may seem of little consequence. Yet we find all too often that the past catches up on us. And here the traditional map collection has often been a help. Two examples from the UK will illustrate this. Firstly, there is contemporary concern about the pollution of groundwater by nitrates: does this come from the application of inorganic fertilizers, or has the release of nitrogen as a result of the plough-up of grassland fifty years ago to feed the nation with cereals been a factor? ("Now, where can I find copies of those old land use map from the Dudley Stamp survey of the 1930s?"). Or another problem: my house is subsiding! ("Are there any old maps which indicate that the developer should have known that this was a former landfill site?"). Both are questions asked – and answered – in the author's university collection.

Map libraries have been important repositories of these kinds of data, relating to the relatively recent past. But until recently the data were not digital: they were captured, stored, and distributed on paper. That realization has led to a flood of activity in the UK to scan old editions of British Ordnance Survey maps, and to "rubber sheet" them to fit contemporary grids. The appearance of many of these on the web (http://www.old-maps.co.uk/), or via purchasable CDs, has benefited a wide range of users. The most comprehensive and ambitious UK project has been the creation of an online, time-layered electronic map library by Sitescope. So much for "born digital".

In Britain, at the University of Portsmouth, Humphrey Southall and colleagues have been carrying out a rather more difficult exercise, capturing digitally the historic boundaries of Britain as part of the Great Britain Historical GIS Project, while at Exeter University, Richard Oliver and Roger Kain recently completed Historic Parishes of England & Wales. An Electronic Map of Boundaries before 1850, held and distributed by the History Data Service at the University of Essex. Such work is difficult because of inconsistently preserved records, but at least, in most cases, the records *have* been preserved somewhere and, with some effort, can usually be retrieved.

Please note that map libraries did not initiate most of these digitising projects, and this means that map collections are being by-passed in terms of historical as well as contemporary digital data.

LOOKING FORWARD

However, that is all about yesterday's history. The concern in this paper is with tomorrow's history, and the starting point is two recent occurrences. The first immediately set alarm bells ringing. It was a message posted on the CARTO-SOC list-server (1 February, 2002) from a researcher at Imperial College who wanted to build an historical database of UK postcodes. While there may not be much intrinsic interest in postcodes themselves, they are important as spatial units for storing and mapping certain kinds of thematic data such as geodemographic and lifestyle data (Raper *et al*., 1992). The Chorley Report (Department of the Environment, 1987) recommended that postcodes should be used as basic spatial data units in geographical information systems. However, one of their limitations is that they are not a stable boundary set: many thousands of changes are made annually to UK postcode geography at the unit postcode level. Therefore for mapping purposes, historical thematic data needs historical postcode boundaries. The researcher who posted the message had tried to obtain the quarterly update files for previous years, but found that the Post Office, and associated companies, failed to keep data older than two years.

The second occurrence stems from a contribution by Pip Forer, a geography professor in New Zealand, to the book *The map library in the new millennium* (Parry and Perkins, 2001). Forer was invited to write about "the business" of mapping, and in particular the way some government surveys had been effectively if not actually privatised. However, his chapter took a somewhat different turn. In it he makes a distinction between the traditional mapping industry as exemplified by national mapping agencies, and a sector "that is based on more purpose-specific … spatial data capture for mapping and analysis" (Forer, 2001, p 189). He describes a new map-making industry "out there", based on geocoded data and the ability of almost anyone to make maps, leading to a new, diffuse "industry". What, he asks, are its implications for map libraries?

Now the contention in this paper is that in our scramble to come to terms with new digital and communications technology, we have indeed neglected the need to capture and archive some kinds of spatial, or spatially encoded data which may be of critical importance in the future. Certainly, we have put much effort into tackling questions of how to capture and preserve the *topographic* digital archive for the future (Fairbairn, 1993; Fleet, 1999), but in doing so have tended to concentrate on national data sets and national libraries. Meanwhile there has been a tendency to neglect the similar importance of both local and national thematic data sets (including, especially, boundary data to which those thematic sets are mapped) and in particular to neglect what to some may seem expendable data, but which previously, being in paper form, would have been less readily destroyed.

WHAT KIND OF DATA?

Because increasingly data are being spatially referenced, GIS and computer mapping systems make it possible to combine data into a limitless number of thematic maps. Therefore it is presumed that a digital archive will be primarily concerned not with the maps themselves, but with preserving the "mappable" data, and the boundary polygons where these are needed to give the data spatiality.

Boundary data: Postcodes, zip codes, NUTS boundaries, administrative and judicial boundaries, planning regions, health areas, police districts, counties, communes, enumeration districts, street gazetteers, all these spatial units, or "area footprints", are receptacles for information. Without these, the thematic data cannot be mapped. Therefore their preservation is crucial.

Spatially encoded data: It has been suggested that between 60 and 80% of all UK government data can be classed as "geospatial" (Nanson, Smith and Davey, 1995). Anything with an address can be mapped, and it doesn't all have to be government data: house prices, crime rates, traffic flows, noise contours, health statistics, lifestyle data, valued landscapes, heritage sites, are examples of the range of themes.

Ephemeral data: Data collected for a specific purpose: e.g. to investigate traffic congestion in relation a new road development or to map the spread of a disease outbreak (such as the Foot and Mouth outbreak in Britain). Often, such data are also local rather than national in scope and importance.

WHO SHOULD BE RESPONSIBLE FOR PRESERVING SUCH DATA?

A report by the Royal Statistical Society and the UK Data Archive (2002) on *Preserving and sharing statistical material* made no direct reference to maps or indeed to spatial data, but it was concerned with the same issues of recognizing the long-term value of certain data sets, and to make provision to share these data. It raised the question of who is responsible for (a) deciding what should be conserved for the future, and (b) holding and making available the archive.

The report clearly placed the custodial responsibility with the data collectors. But is that sufficient? We have already seen that in the case of postcodes this hasn't been occurring. And we know from past experience that major producers of mapping have not necessary been thorough in maintaining a record collection. This does not mean necessarily that the producers should be relieved of this role. With so much data becoming available, the question also has to be asked: "Is it worth preserving?". To which the answer often can only be: "Only time will tell!". But there needs to be some guidelines and some means of encouraging data providers to consider the value of the data beyond their own specific requirements.

"It would be impractical to preserve all material of a statistical nature but it is important that the potential value of all such material is considered as part of the selection process." (Royal Statistical Society, 2002)

In the case of personal data there may be a conflict between the requirements of the 1998 UK Data Protection Act (e.g. that data may be kept for no longer that necessary) and the longer-term value of the data, though anonymized data may be preserved.

WHAT ARE THE REQUIREMENTS?

Of course, it is one thing to accept donations of data stored on CD-ROM and lock them away in a cupboard and forget about them. To ensure their use when required has cost implications both financially and in staff time. The data sets need to be provided with metadata, and their existence made known, perhaps through the map library's web site or via some centralized source, such as a metadata warehouse. More significant in the long term perhaps, is the need to preserve the data in a format, which ensures they can continue to be used. Many major data sets from the early days of GIS and computer mapping have required, or potentially require, a huge input of effort and funding to render them operable with contemporary computer systems. Examples are the Canada Land Information System and the British Domesday project ("Digital Domesday Book lasts 15 years, not 1,000" – heading in *The Observer*, 3 March 2002). Perhaps this will be less of a problem in the future with the development of OpenGIS.

So here are some of the things that would need attention:

- The data need to be adequately catalogued, e.g. to Dublin Core standard.
- The metadata need to be shared over the web.
- Spatially encoded data need to be linked to appropriate boundary data sets.
- The data should be available to a wider user group than the local.
- Confidentiality and IPR may need to be protected.
- Custodial responsibility needs to be clear: are the data donated or loaned?
- Permissions for use and legal constraints need to be granted and understood.
- Data need to be technology-proof.

IS THERE A ROLE FOR THE LOCAL COLLECTION?

"When material is georeferenced, there is a clear advantage to locating its digital version on a server close to, or within, the geographic footprint of the material, since that is where interest in it is likely to be highest." *(Goodchild, 1998, p. 62)*

This paper sets out to stress the importance of the local dimension in spatial data collection and accessibility, and this quote from Goodchild may point the way. Certainly his model applies well to America, where most notably at the University of Connecticut, McGlamery has used generously distributed local data sets to build a GIS and mapping system, making these data accessible and usable to a wide range of consumers, both locally and beyond (McGlamery, 1995). In the UK, government data, including local government data, are not so generously distributed. The same opportunity has not arisen, and it would be difficult to develop such a fully integrated provision of local data.

But as in the USA, the preservation of such data does not have to be the exclusive preserve of a centralized archive, particularly where the data apply to a local region.

"… a digital geolibrary can consist of a collection of servers, each specializing in materials about its local region". *(Goodchild, 1998, p. 61)*

In the UK context the location might be the county record office, or an academic library such as a university map library. The county record office seems an obvious candidate, since its remit is to archive local material, and its clientele has local interests. But first enquiries indicate that little progress has been made in this area.

Perhaps a better contender at present is the academic map library in so far that this, or at least its host institution, has the hardware, expertise and Internet links for storing and sharing and using such data. One has to admit, that in terms of *use*, the principal demand of one's academic clients is not for local mapping, but for national and international data:

topographic, geological, soil, climatic, demographic, the traditional mapping themes, continue to be of predominant interest. Some of these are now accessible, free, via the web, or by subscription to online services such as Digimap, UK Borders and the Census. So here too, we are losing a great deal of our function to centralized providers (such as EDINA and MIMAS).

However, looking back again, we find that our map library at the University of Reading has over the years been the recipient of donations of local hard-copy material. Often this has been the local component of a national map series, but value-added in some way. The most recent example is of many hundreds of large-scale (1:2500) OS maps of the local area, a disposal of the Reading Valuation Office because, yes, they had migrated their records to a digital system. These maps were shared between Berkshire Record Office, the University's map collection and Reading Public Library as a gap-filling exercise. But they were not just OS maps: they had added value. Hand-drawn annotations provided historical property boundary and land valuation information. Reading Public Library was particularly excited about this added information, and it could well be argued that this collection should not have been split up! But at least the records have been preserved.

Such donations have not been actively sought. However, in the future such value-added data will be held in a computer file, where it may be continuously updated, or if not, more easily disposed of when no longer needed by its owners. Ten years ago, the county of Berkshire established a GIS to handle local transport and environmental data. Some 40 data sets on themes ranging from highway networks to wildlife heritage sites, to accident black spots to Second World War pill boxes were mapped to a common reference base. Will that material be retrievable in another ten years? Perhaps then we need to take a more active interest both locally and nationally about what is being collected digitally and whether it is being conserved for the future.

"A serious collection ... needs to work with various data providers in identifying the form that a data archive might take and what should be held locally. Observation suggests this is not an issue that current data providers are looking closely at." (Forer, 2001, p. 200)

Apart from the important function of assisting spatial data users alluded to by Keller and by many of the contributors to *The map library in the new millennium*, there is therefore scope for map libraries to become specialized in the acquisition and conservation of local, and in some cases national data sets, and then to share them, over the web with other collections. Cooperation is also factored in through persuading local government and commercial organizations to donate data sets that they no longer regard as important for their own purposes. In the past, many collections have benefited from hard copy donations of this type. But is the same thing happening with digital data? Not yet it seems. Perhaps map curators could actively promote a culture, which encourages this.

IN CONCLUSION

This paper has done little more than raise some questions. But the issue could be important: digital data are too easily lost. If it is important, is anyone doing anything about it – in other European countries besides the UK? Is it something in which major national collections should play a leading role, or should it be left to the data providers? What are the funding and custodial implications? And, finally, the local component: is this an opportunity for local collections to build bridges with local data collectors and to specialize in saving local digital data?

REFERENCES

1. Campbell, A.: "Where are map libraries heading? Some route maps for the digital future." *LIBER Quarterly*, 10 (2000) 4, 489-498

2. Department of the Environment. *Handling geographic information*. London|: HMSO, 1987.

3. Edinburgh Data and Information Access: http://edina.ac.uk

4. Fairbairn, D.: "The future history of our landscape". *The Cartographic Journal*, 30 (1993), 62-67.

5. Fleet, C.: "Ordnance Survey digital data in UK legal deposit libraries". *LIBER Quarterly*, 9 (1999), 235-243. http://www.kb.nl/infolev/liber/articles/fleet11.htm

6. Forer, P.: "Taking care of business: map libraries and the new "mapping" industry". Chapter 13 in: Parry, R.B. & C.R. Perkins (eds.). *The Map Library in the New Millennium*. London: LA Publishing, 2001. Pp. 188-206.

7. Goodchild, M.F.: "The geolibrary". In: Carver, S. & S. Openshaw (eds.). *Innovations in GIS 5*. London : Taylor and Francis, 1998. Pp. 59-68.

8. Great Britain Historical GIS Project: http://www.gbhgis.org/

9. Kain, Roger J.P. & Richard R. Oliver.: *Historic Parishes of England & Wales.* An Electronic Map of Boundaries before 1850 with a Gazetteer and Metadata: http://www.ex.ac.uk/geography/research/boundaries.html

10. Keller, C.P.: "The map library's future". *Cartographic Perspectives*, 38 (2001), 73-77.

11. McGlamery, P.: "MAGIC at the University of Connecticut: building the virtual map library". *Bulletin, Special Libraries Association Geography and Maps Division*, 179 (1995), 14-30.

12. Manchester Information & Associated Services: http://www.mimas.ac.uk

13. Nanson, B., N. Smith & A. Davey: *What is the British National Spatial Database?* Paper presented to the 1995 Association for Geographic Information conference, 1995.

14. Raper, J.D., W. Rhind & J. Shepherd. *Postcodes: the new geography*, London : Longman, 1992.

15. Royal Statistical Society and UK Data Archive. *Preserving and sharing statistical material*. Colchester : UK Data Archive, 2002, 48 pp

16. Sitescope: http://www.sitescope.co.uk/

WEB SITES REFERRED TO IN THE TEXT

Old editions of British Ordnance Survey maps: http://www.old-maps.co.uk/

Edinburgh Data and Information Access: http://edina.ac.uk

Great Britain Historical GIS Project: http://www.gbhgis.org/

Manchester Information & Associated Services: http://www.mimas.ac.uk

Sitescope: http://www.sitescope.co.uk/