

INSPEL



2002

Vol.36

No.1

Official Organ of the IFLA Division of Special Libraries

INTERNATIONAL JOURNAL OF SPECIAL LIBRARIES
REVUE INTERNATIONALE DES BIBLIOTHEQUES SPECIALISEES
МЕЖДУНАРОДНЫЙ ЖУРНАЛ СПЕЦИАЛЬНЫХ БИБЛИОТЕК
INTERNATIONALE ZEITSCHRIFT FÜR SPEZIALBIBLIOTHEKEN
REVISTA INTERNACIONAL PARA BIBLIOTECAS ESPECIALES

Published in cooperation with the German Special Libraries Association
(Arbeitsgemeinschaft der Spezialbibliotheken - ASpB)

Editorial Board:

Hans-Christoph Hobohm (Editor in chief)

Fachbereich Archiv-Bibliothek-Dokumentation, Fachhochschule Potsdam,

PO-Box 600608, D-14406 Potsdam, Germany.

Tel.: +49 (0)331 580 1514, Fax: +49 (0)331 580 1599; e-mail: Hobohm@FH-Potsdam.de

<http://inspel.ifla.org>

Jakob Andersen, National Library of Education

P.O.Box 840, DK-2400 Copenhagen, Denmark

Kerstin Assarsson-Rizzi, The Library of the Royal Academy of Letters, History and Antiquities, The
National Heritage Board, Box 5405, S-11484 Stockholm, Sweden

Rafael Ball, Forschungszentrum Jülich GmbH, Zentralbibliothek, Leo-Brandt-Strasse
D-52425 Jülich, Germany

Ysabel Bertolucci, Kaiser Permanente Medical Center, Health Sciences Library
280 West MacArthur Blvd., Oakland, CA 94611, United States

Julia Gelfand, Science Library # 228, University of California, The UCI Libraries
P.O.Box 19557, Irvine, CA 92623-9556, United States

Paul Kaegbein, Lehrstuhl für Bibliothekswissenschaft der Universität zu Köln,
Universitätsstr. 33, D-50931 Köln, Germany

David C. McQuillan, Thomas Cooper Library, University of South Carolina
Columbia, SC 29208, United States

John Meriton, National Art Library, Victoria & Albert Museum
Cromwell Road, London SW7 2RL, United Kingdom

Lena Olsson, Stockholm Institute of Education
P.O.Box 34103, S-10026 Stockholm, Sweden

Publisher:

Technische Universität Berlin, Universitätsbibliothek, Abt. Publikationen,

Straße des 17. Juni 135, 10623 Berlin, Germany

Telefon: (030) 314-23980, Fax: (030) 314-24741, email: publikationen@ub.tu-berlin.de

Printer:

Offsetdruckerei Gerhard Weinert GmbH, 12099 Berlin, Germany

INSPEL is published quarterly. The annual subscription rate is EUR 41,00 plus postage. All subscription inquiries and orders should be sent to the Publisher. All other correspondence should be addressed to the Editor.

The contents of this journal are indexed/abstracted in the following information services: ECONIS (Economics Information System), IBZ (Internationale Bibliographie der Zeitschriftenliteratur), ISA (Information Science Abstracts and Fulltext Sources Online), INFODATA (Informationswissenschaft und -praxis), Library Literature & Information Science Full Text, LISA (Library and Information Science Abstracts).

© INSPEL and contributors

ISSN 0019-0217

Contents:

James Boxall: Geolibraries, the Global Spatial Data Infrastructure and Digital Earth: A time for map librarians to reflect upon the Monnshot.	1
Jan Smits Geospatial data access: can we manage to shift?.	22 R
David Cobb: Limits to access. You Can Look But Don't Touch.	49
Betty Kidd: Digitization of Cartographic Materials: National Archives of Canada.	53 R
Pierre-Yves Duchemin: La numérisation des documents cartographiques anciens: supports traditionnels et nouvelles technologies.	67
Scott R. McEathron, Patrick McGlamery et al.: Naming the Landscape: Building the Connecticut Digital Gazetteer.	83 R

GEOLIBRARIES, THE GLOBAL SPATIAL DATA INFRASTRUCTURE AND DIGITAL EARTH: A TIME FOR MAP LIBRARIANS TO REFLECT UPON THE MOONSHOT.*

By James Boxall

Much has been written about the concept of “geolibraries” from different perspectives (Goodchild, 1998, 1999, 2000; Onsrud, 1995; Battenfield, 1998; Boxall 1998; 1999) but little has been done to actually make the case for *libraries*. This may be due, in part, to the reality that geolibraries are defined in a digital-only context. This brings us to the issue of how metaphors become critical, not only to our understanding of the infrastructures we attempt to build, but to the ways in which we try to link the new concepts and structures with the old. Researchers outside of the library community conduct much of the research and development on spatial data infrastructure (SDI). These researchers may hold concepts and ideals of traditional libraries not held or supported by library practitioners. This has both positive and negative effects.

On the positive side, such activity provides for potential collaboration and a new type of shared experience. From the library point of view, geographers and geographic information (GI) scientists who research and work on geolibraries (in all their permutations, such as digital earth) can become allies in an effort to increase access, use and preservation of cartographic materials and geospatial information. The ideals and beliefs held by librarians (curators and archivists as well) however, are not shared by geographers and GI scientists, and the reverse may also be true. They can be shared. They have the potential to be common beliefs. They are not normally connected because of the differing histories of disciplines and the manner in which librarians carry out their professional and academic lives. Simply put, the emphasis in the library is on service, while the emphasis in the academic community is on research. Both communities share the roles of teaching, but even there resides a potential conflict. Librarians have for some time fought the intellectual battle of trying to convince academic colleagues that they too are teachers on campus.

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

Those who have written about geolibraries, as well as associated issues that have always been the domain of libraries such as access to information, archiving and preservation, and collection development, have tended to either over simplify library issues or have focused too much on the engineering and computational aspects of geolibraries. Interestingly, most people involved in geographic information systems (GIS) have assumed that metadata issues and other standards processes related to information description are new and unique, or even highly technical in nature. The work of ISO TC 211 and other such standards organizations tends to involve more non-librarians. This is not to detract from the positive outcomes of that work, but librarians, following the discussions on the margins, find the reinvention of metadata as both amusing and frustrating, since librarians pioneered the development of cataloguing standards long ago.

Librarians tend to talk with, to and among librarians. Even in the broadest communal definition of librarianship and libraries, divisions exist such as those between librarians and curators, or archivists. So divisions and overlap between GIS and libraries should not be viewed as so odd. Why all the fuss? Simply put, there is no need to duplicate effort. Librarians can contribute to geography, and geographers to librarianship. We have common goals and experiences, and we share a common language – the spatial. Key to this argument, of course, is the idea that map and GIS librarians are specialists among specialists, and so working collegially with other specialists in the discipline of geography makes perfect sense.

Geolibraries do provide an excellent and unique opportunity to elevate the work of both GI scientists and librarians and to bridge the gaps between geographers and GI scientists, and geographers and librarians. Map libraries are becoming geospatial information centres, providing access and services; teaching and value added creation/manipulation. Though the level of service varies from campus to campus (and in the US between public libraries), the overlap between the work of geographers and the work of the geolibrary is substantial and growing.

Still, a great deal has been written about access to information without ever thinking or mentioning libraries as being the critical juncture. This is more ironic and troubling because the concept of the library is embedded in the term geolibraries. Goodchild defines the idea of “geolibrary” as a “library filled with georeferenced information” which is based upon the notion that information can have a geographic footprint (Goodchild, 1998). He also explains that the GIS community has been working with geographic information, while georeferenced information is broader in scope to include such things as photographs, videos, music and literature that can be given a locational variable which defines a

footprint. In this way, the idea of the geolibrary immediately extends well beyond the traditional scope of map libraries and archives to include almost all information contained within libraries; he later mentions that it can include information outside of libraries as well. This is the theoretical basis for what we now view as geolibraries. Geolibraries are now seen as components of digital libraries, in large part due to funding of the Alexandria project directed by Goodchild. In practice however, the collection development focus of many geolibraries has been much more narrowly focused.

The most obvious feature of geolibraries, as is true for digital libraries, is the focus on digital information and metadata, as well as the distributed nature of the libraries and 'collections'. But they do not just include technology and information. Buttenfield (1998) wrote: "We could (and should) discuss the technological and cognitive impediments of fully operational geolibraries. It is equally important to consider the institutional, societal, and economic issues, which have not been adequately addressed in current digital library efforts."

Distributed geolibraries provide a useful framework for discussion of the issues of dissemination associated with the National Spatial Data Infrastructure (NSDI). The vision is readily extendible to a global context. (Finding #5, Distributed Geolibraries Workshop, Mapping Science Committee, 1999).

Mike Goodchild, has noted that the Digital Earth, should be viewed as a "moonshot" (Goodchild, 2000). Goodchild et al. touch upon other issues, some of which have been previously explored (Boxall, 1998; 1999; 2000), yet, changes occurring in many infrastructure activities means that it is still useful to review and revamp some of the suggestions for action and reflection. Above all, any discussion of "Digital Earth" (DE), the "Global Spatial Data Infrastructure" (GSDI), and "Distributed Geolibraries" should be framed around the broadest definitions of information and infrastructures; namely to include and focus upon the people, technology and organisations which give rise to and sustain such infrastructures.

This paper begins with an answer. During a presentation at the Association of Canadian Map Libraries and Archives annual meeting in Montreal (June, 2001), I stated that the fundamental missing piece in all this discussion about sub-sections of the Global Information Infrastructure (GII) was the lack of a culture of permanence resulting from the alternative culture of the Internet. We have developed "near ideologies" that promote networks over people. My answer is a shift in dialogue based upon the principle that we should be funding "institutions and intellect" instead of "infrastructure and internet". This view has been called neo-luddite and reactionary. Actually, the phrase is borrowed from e-commerce,

with an alteration. We hear now that total online commerce does not work well, and so we see the rise of the concept ‘clicks and mortar’ as opposed to ‘bricks and mortar’. Sustainable (over 100 years) Internet-based infrastructures require the type of clicks and mortar that gave rise to the greatest libraries and archives ever seen. Digital Earth is wonderful, and technically feasible, given enough resources. We have, however enough time to take a closer look at some of the bigger issues before we launch our moonshot – or before the vision is extended too globally.

The matters for concern are not so ‘neo-luddite’ after all, for even some of the strongest proponents have raised questions about some very difficult and outstanding issues. For example Goodchild mentions the connection or co-existence of two basic trends – the increasing popularity of “things geographic” and the “digital transition”. He suggests we face a basic paradox, “...an increasing marginalization of cartography and an increasing need for good cartographic practice.” Borrowing that phrase we can insert other terms in the place of cartography, such as libraries, archives, education, teachers and teaching and geography. Kofi Annan, Secretary General of the United Nations, addressed the American Association of Geographers meeting in March 2001 and reminded geographers of their critical value to the greater global good. Geography is mainstream again, if it wants to be, and so also may be libraries.

Yet, it is curious that most everything about DE and GSDI is written by non-librarians. Even the recently released Cookbook for GSDI (www.gsdi.org) makes almost no mention of libraries, and those few instances tend to be quaint. Perhaps it is due to the lack of involvement by librarians in GSDI? This should not be surprising when one considers that some of the great thinkers and gurus of this high-tech era give little attention to libraries. For example, a librarian would think that a text called the “Age of Access”, dealing with information access issues, would mention libraries more than a single paragraph (Rifkin, 2000, p.87). Ironically, librarians are not at the forefront of these debates. This is the age of information, a knowledge based economy, and those most concerned with information and knowledge seem to be on the margins. This relates back to Goodchild’s statement in 1999 about the marginalization of certain professions at the a time when they are more needed.

James J. Duderstadt, in Educause Jan/Feb 2001, talked about the coming GII when he said:

We are on the threshold of a revolution that is making the world’s accumulated information and knowledge accessible to individuals everywhere, a technology that will link us together in new communities never before possible or even imaginable (p.56).

This view is underlined by Borgman (2000) although the latter author takes a more academic and critical view of all sides of the issue. Both authors do, however, highlight the reality evident in writings about geolibraries and digital earth – we are well on our way to developing the most widely used, available, fastest, and most complex (yet intuitive) communications and information distribution system ever, with an enormous amount of internationally agreed to *technical* interoperability. The integration of standards at ISO is a testament to this development. Notice I did not say it would be universally available or available to all equally, and I specifically place it in a technical context. This, I fear, will also be the same period of time when the division between haves and have-nots will become even greater. The digital divide will, even under digital earth, grow. This does not mean to suggest that we should ignore progress or fail to advance technology. We can still have a moonshot, but we need to learn from our past moonshots and so-called revolutions. We need to see that one technology or one revolution cannot be the end all. We must move well beyond the hype and the focus on the engineering and computational wonders.

This author would also add that libraries will always exist, and furthermore that libraries have never been more important. But this importance rests on what libraries do, and not on what they are perceived as doing. Libraries are more than storehouses and much more than clearinghouses. We do more than collect and catalogue. Librarians are active participants in the learning process. We are part of a great process of “change”.

In considering possible institutional arrangements for distributed geolibraries, we begin with the assumption that libraries are social institutions that will continue to change but will not be made obsolete by the advent of electronic publishing. Indeed, distributed geolibraries and digital libraries in general will complement the traditional activities of libraries and related institutions. Libraries respond to many complex societal needs. They are used for research, teaching, self-learning, and entertainment. They serve as social and activity centers for many communities, whether these be small towns, neighborhoods, or institutions. (Mapping Science Committee, 1999)

The 1998 workshops hosted by the Mapping Science Committee (under the National Academies in the US) to discuss the ideas of Distributed Geolibraries was one of the few occasions where some of the best and brightest in the fields related to geoinformation (including librarians) were able to meet and review issues and possible developments. Out of that workshop came one of the few substantive reviews and presentations of geolibraries and their potential relationships to other geospatial information infrastructures. In addition to the above quotation, some of

the more salient points raised are summarized here. The entire report is available for purchase, and also distributed by the National Academies Press via the World Wide Web. The report as a whole is very detailed and covers many of the issues that needed to be addressed at the time it was written (1999). However, the US-centric view in the text needs to be addressed soon if any progress is expected at an international scale. Even the authors of the report identify this issue:

The workshop participants were almost entirely from the United States, and this report necessarily adopts a U.S. perspective. Nevertheless it is hoped that it will be read by non-U.S. researchers and developers interested in distributed geolibraries and that it will help to achieve a greater degree of convergence in research and development at the international level (Mapping Science Committee, 1999).

Therefore, much of this essay deals with and uses examples of issues being faced outside of the digital earth “home”. Although many international efforts are not thoroughly discussed, it is recognized that efforts such as the Global Mapping Project and GSDI do take into account significant barriers to access and use from an international perspective.

In 1997, the Washington DC based Association of Research Libraries (ARL) published a special issue of Transforming Libraries titled *Issues and Innovations in Geographic Information Systems* (ARL, 1997). While this exploration of case studies from libraries in the US and Canada highlighted trends and services being provided within libraries, it failed to grapple with the more serious issues. In 1997 the move to implement GIS in libraries was still new enough that serious problems had yet to be recognized and addressed. From this publication however, comes this statement regarding the value of traditional library roles:

Though they are using GIS, libraries rarely focus on the technology itself. Sometimes there is a brief infatuation with the technology, but soon a realization sets in: though the technology is new, traditional skills of librarianship are required to use it effectively. User needs must be evaluated; data must be selected, cataloged and prepared for users; public services must be designed, offered, and managed.

The paper goes on to include a description and summary of interviews with Larry Carver and Mary Larsgaard from the University of California-Santa Barbara project for the Alexandria Digital Geospatial Library. In that section, Carver and Larsgaard make the excellent point: to remember that all the focus on the technology has to be geared to the basic notion of user needs which continues to be the key element of reference services in libraries.

Since the time of the ARL report, and during the whole phase of implementing GIS services within libraries, much has been written about the development of particular aspects of such services, as well as the broader issues associated with geospatial information (Cline and Adler, 1995; Cobb, 1995; Smits, 1999). Also we have seen many developments connected with map library associations and calls for further cooperation, both in and outside the map library community (Boxall, 2000; Wood, 1988). But libraries and archives, especially those working with cartographic materials and geospatial information, are more than just storehouses of media in a variety of forms; we are active learning places and long standing contributors to the economic and social vibrancy of our communities (see Hawkins, 1998).

Now that the hype and fear surrounding Y2K has passed on into the eternal ether graveyard, it should not be surprising that we find new opportunities to really begin to explore ways to extend the valuable uses of computing and telecommunications in all aspects of our work and lives. Libraries and life are being transformed by such revolutionary *and* evolutionary changes. I will not argue that such change has no place. Nor does this discussion assume that the efforts by others to create more, and faster, access to geospatial information is headed in the wrong direction. This paper does assume that the concepts (or metaphors) of ‘geolibraries’, ‘digital earth’ and ‘spatial infrastructures’ do provide the best means to promote cooperation, dissemination and other library qualities but we also must immediately become more clear about our goals and difficulties we must overcome.

Frank Webster among others questions the concept of an information society and why we continue to refer to this as an information age (“why not a computing age?”) (Webster, 1991). How did information become some so sacred? Perhaps we need to accept the potential reality of libraries becoming extinct in order to focus on what libraries have been, are, and should become within Digital Earth. The Distributed Geolibraries text from the Mapping Science Committee presents bluntly:

It is possible that libraries will be the principal means whereby citizens gain access to the services of the distributed geolibraries of the future; it is also possible that libraries *will play no significant part* in that process” (Mapping Science Committee, 1999; emphasis added).

There is time, and then there is “Internet time”. Much has been written about the Internet breaking the stranglehold of geography; distance and time no longer matter when communications and information access can take place anytime and anywhere. Internet time, however, is defined as psychological and social

phenomena. It is visible in several ways, most notably the way in which recent ideas seem very old due to technological changes. This also relates to infrastructure issues. For example, we often have heard that the new spatial data infrastructures are being developed along similar lines as previous major transportation systems. Instead of transporting products and people via trains planes and automobiles, digital networks transport ideas and information. The development of concrete infrastructures for the transport of things took decades, and continues today. The planning process was long and arduous. We must take a similar long view of the digital infrastructures of today, or we may see a breakdown similar to crumbling highways and broken water mains.

We also hear about the digital divide. This is not new, consider developed and developing nations. Some may argue that a western bias suggests the developed world represents progress and should be followed almost as a natural evolution of nations and peoples. Now a new dimension has been added to the already existing divides – a digital divide. Traditional physical infrastructures have been built around the idea of a greater public good, regardless of how the resources for the construction were obtained. A large infusion of capital would be necessary to build a new fully wired and digitally integrated building. Retrofitting an old infrastructure will create a patch-work of wires and screens almost jumbled into the first available space as if the fear of not having the capability outweighed the impact of the retrofit. This is one inherent problem with framing our views around legacy concepts such as infrastructures, as well as focusing on the engineering and computational facets more than the society inequities and impacts.

This is not a nostalgic trip down some false memory path suggesting in any way a return to the good old days – they never existed. Then again, do the good new days exist? It is the *speed of change*, the Internet time equation, which is most disconcerting in relation to the building of these infrastructures, DE, GSDI and Geolibraries for example. The “web” went truly public around 1990 and between 1990 and 1994 the WWW consortium was formed and Mosaic was released. (<http://www.w3.org/People/Berners-Lee/ShortHistory.html>). Now almost every advertisement or notice refers to a www address. Some companies rely exclusively on their web sites for disseminating information about products and services. Try the same experiment in China or Angola. We have seen similar speedy adoption of technology, along with the usual quick and therefore frustrating changes in formats. What came before VHS? What will come after DVD? Ten years of Internet Time is too short a period on which to bet the future of life and work and discourse and culture.

What does all this have to do with map libraries, digital earth and SDI's? Perhaps nothing. Perhaps everything. The Internet, or the web because only the web metaphor seems to be surviving, is only ten years old and has only been a large factor in public consciousness for approximately five years. It is consuming all of our work and discourse in ways that previous technologies have not. Perhaps that is over stating the case, but the research literature and publications in more popular press and electronic discussion lists seem to suggest that the speed of change will only increase. Again, the Distributed Geolibraries offers a glimpse of this:

At the same time there are potential disadvantages to use of the WWW as a mechanism for storing and disseminating geoinformation that will have to be addressed. Little of the information now available via the WWW has been subjected to the mechanisms that ensure quality in traditional publication and library acquisition: peer review, editing, and proofreading. There are no WWW equivalents of the library's collection specialists who monitor library content. But it is easy to be misled into believing that quality control problems of the WWW and distributed geolibraries are somehow different from conventional ones. Users of distributed geolibraries will tend to trust data that come from reputable institutions, with documented assurances of quality, and to mistrust data of uncertain origins, just as they do today.

The quote begins to address several of the critical questions, or 'traditions' that are being evaded by some of those trying to build and gain support for certain infrastructures like DE and GSDI. Aspects of traditional libraries *and* archives are important and should be borrowed/lent to the new infrastructures.

Libraries are, in one sense, part of the infrastructure of learning, discourse and democracy. Libraries and archives preserve and promote our heritage and culture. They are part of the communication process. They also require levels of redundancy and preservation not normally seen as valuable in a delete and overwrite culture that permeates the 'net. Librarians also hold values such as privacy, public access, neutrality, confidentiality, freedom of expression, and equality to be critical to the very nature of our work and lives. Put another way, again quoting from the Distributed Geolibraries workshop:

Whereas the substantive content and focus of geographic infrastructure building have focused on data and information (e.g., the NSDI), the substantive content of traditional libraries has focused on collections of knowledge and to a lesser extent collections of information.

Jankowska and Jankowski (2000) suggest that part of the driving force behind geolibrary developments is the changing nature of the global economy and the

demand for “fast and easy access to information” (p.4). They go on to state that the digital changes in libraries allow us to move beyond traditional roles and services, but ask if we will be able to fulfil our role in the digital period era (p.5). This raises the related matter of the time lag between policy and practice. Whether or not we can adapt and change to meet new roles will somewhat depend upon broader changes in the policy realm. Consider for example the pricing of government geoinformation, a concept not fully appreciated within the US. Imagine if, for example, a geoinformation infrastructure was built around the idea of a take off on the music sharing software Napster, “mapster”. Would “mapster” elicit the same response as Napster? Would the concept prove popular with the user community and threatening to the information creators?

Within all these discussions and developments is the more interesting question of how we attach metaphors to what we are doing. Recall for example the hype that surrounded the “information highway”. Generally, metaphors are short lived; I cannot recall the last time I heard or read “information highway”. Not only is the technological life cycle short, but the metaphorical concepts that we use to describe technology are also short lived. Why is the idea of metaphor critical? Goodchild (2000) suggested that the DE and GSDI, and Moonshot, are metaphors. Without dwelling on a post-modern deconstruction of these terms, nor delving too deeply into the ontology or epistemology of these objects and ideas:

The metaphor of the library is powerful because it immediately suggests a number of important issues. For example, one way to think of a library is as a storehouse of the intellectual works of society, and millions of people from all walks of life have contributed works to our current library system. Can we expect to see a similar diversity of contributors in the distributed geolibraries of our future? However, the metaphor of the library should not be taken too far, and not all aspects of the operation of a library will be useful in envisioning distributed geolibraries. Many of these will be generic and of no specific relevance to the geoinformation that is the focus of distributed geolibraries For example, it is assumed that distributed geolibraries will need to address issues of archiving and preservation (particularly serious issues given the rate of technological change in the digital world), but these are generic to all libraries and are not discussed at length in this report (Mapping Science Committee, 1999).

I would suggest that archiving geospatial information, mentioned above, within the context of geolibraries and/or digital earth will be “killer app”.

The Mapping Science Committee report also raises issues associated with policy. However, this comes from a US based view, and the policy realm outside the US is anything but uniform or stable (MADAME 2000).

The policy challenges presented by distributed geolibraries include the following:

- What are the legal, ethical, and political issues involved in creating distributed geolibraries? What problems must be addressed in the area of intellectual property rights? How will these issues affect the technical development of distributed geolibraries?
- Who will pay for the creation and maintenance of distributed geolibraries? What components might be in the public domain versus those provided by the commercial sector? (<http://www.nap.edu/html/geolibraries/ch3.html>)

The recently released Canadian Geospatial Data Policy Study authored by KPMG Consulting, and funded by GeoConnections makes some key recommendations that could have the effect, if implemented correctly and as promptly as possible, of making more geospatial information available within the public domain. The study makes note of the often quoted 1:4 ratio of economic impact generated by public release of geospatial data. This ratio is now being both questioned and discounted. Ironically, the ratio comes from a report for the Australian SDI authored by the same KPMG (www.geoconnections.org). In addition, in Europe, the development of spatial information infrastructures is fraught with policy barriers (MADAME, 2000).

Can there be compromise? Is there a middle ground in cost recovery/pricing/licensing and crown copyright issues? Branding as opposed to copyright? Crown copyright is currently used as an enforcement mechanism to protect revenue streams, and to impose the concept of Government Information Policy. Certainly, all data users value quality data sources. Canada is very fortunate to have federal agencies and institutions producing useful and high quality data sets and “information” in general. Governments have been and continue to be the single largest producer and collector of geospatial data.

There may not be one “solution” that can meet all users and all requirements. Also, the idea, or misconception, that the policies and implementations (agreements) in other jurisdictions such as the US, Australia, or the EC, are standard and easy to implement across all communities and contingencies is wrong. However, the general tone of the KPMG study is one of creating a more open transmittal of data within the context of the “public good”. The public good is most difficult to define

and depends upon perspective. However, from the point of view of all parties and sides of the debate, educational uses are seen as one clear constituency where improvements are both easy and justifiable. Another area is the government-to-government transmittal of data.

Previously, I argued that the critical issues facing map librarianship are demographics, technology, policy and organisation (Boxall, 1999). First, in looking at demographic issues we must recognize that there have been no recent studies directly related to map librarianship, so it is necessary to infer from more general trends the possible implications for the profession. There is no doubt that, like other professions in the educational sector, we are facing an aging workforce. We have seen over the last few years numerous advertisements for map and/or GIS librarianship positions that have been difficult to fill. It is almost common knowledge that faculty, librarians, teachers, and a number of other professions are facing massive turnover in staff over the next decade. Demographers have been aware of this fact since the 1960's and 1970s when massive employment occurred in these fields. This is true for librarians, and especially so for librarians in specialized areas such as maps, geography and GIS librarianship. 'Map Librarians' are specialists among specialists. The introduction of GIS in libraries (or is that through libraries?) requires applicants to have multiple qualifications in the areas of traditional librarianship as well as qualifications and training in GIS and related technologies. The demand outstrips supply (as anyone who has seen re-postings can recognize). This is not due to a lack of interested people, or persons with adequate training or experience, as much as it is due to the simple fact that the total pool of persons being trained in fields associated with map librarianship (including library and archival studies) is being courted by more lucrative employment and career opportunities. The real issue, in relation to the development of digital earth, is where the librarians will come from in order to help shape the geolibary component of DE. We have all come to this profession via numerous routes, librarians, curators or archivists. However, we must recognize that the competitive job market is having an impact on recruitment and therefore services.

And what of training and professional development? Because of changes in GIS and cartography, not to mention other disciplines, we need to make sure that our skills are current. In Canada a broader study of the human resource issues jointly funded by the Geomatics Industry Association of Canada and the Canadian Institute of Geomatics, and the Canadian Association of Land Surveyors called the HAL suggests that the demand from government and industry for highly qualified GIScientists and technicians will continue to grow. It also highlights the need for more training programs and new methods of delivery for professional and mid-career training. We have an opportunity to become allied with these broader calls

for more solid educational opportunities for GIS and cartography students and for cooperation among practitioners. (Fisher, 1998). In this way, we may begin to realize that we are part of the cartographic visualization process, rather than being the keepers of the output from that process (Kraak, 1998). It is not simply a matter of digital technology changing the nature of work that is of concern to our communities. In fact, digital technologies are merely the physical manifestation of changes in the culture of work. There has been a shift from the notions of public service to public entrepreneurial service. In all these professional qualification issues we need to become more aware of, and more heard among, our colleagues in related disciplines.

Funding issues are another consideration. Libraries do not pay as well as faculty or industry positions. Even Goodchild (1998) suggests that libraries are facing a difficult present and future, even without adding in human resource questions:

At this point in history, libraries are faced with apparently insurmountable problems (Hawkins, 1996). The published corpus of humanity is growing rapidly, and doubling in not much more than ten years. Journal prices continue to rise at well above the rate of inflation. Library budgets are contracting, and libraries are faced with unprecedented problems of security. The pressures to find new approaches, and to take advantage of new technologies, are high.

We must consider another feature of demographics, the changing demographics of our clientele. This is critical when considering new types of services, or in revamping existing services and programs. For example, we find in many universities larger numbers of students drawn from non-resident populations or from groups who have, historically, been marginalized from higher education. Such groups include students whose first language is not English or students with learning disabilities. We also see trends towards more mature, continuing and distance education students. These new mosaics of students and clientele mean adapting services and collection policies to meet their needs. Add into this mix the obvious growth in the overall numbers of persons using cartographic and geographic information in all forms as a result of increased access to and use of digital technologies including web based mapping, and suddenly we can predict increased pressures on our resources, institutions and ourselves in the effort to meet service needs. The vast majority of youth are regular users of the Internet and computing technology and their expectations for digital work and access to information are significantly different from the norm (Weil, 1999).

At the recent Canadian Association of Geographers meeting in Montreal (June 2nd, 2001) presentation the author (forthcoming in Canadian Geographer special issue on GIS) declared standards are dead, but standards are more vital than ever. The

developments related to XML and GML, and ISO standards related to metadata and technical issues are merging with other developments such as the FGDC content standards and MARC21; not to mention such hybrids as Dublin Core. The geospatial and library communities have no shortage of standards, but there is a trend towards a standard with some room for local variations, pricing/access restrictions, and language for example.

With the Open GIS Consortium (OGC) moving forward briskly, there is little doubt that questions related to standards and interoperability are at the fore and will remain there for some time. Standards are recognized as vital and critical to the continued progression towards DE and the GSDI and they do require a greater degree of input from the library community. But standards are not the issue we should be so focused upon. The potential ability to find geospatial information from distributed geolibraries is great; metadata standards will work, so long as we pick one and stick to it. The real question is: who will generate, check, load, monitor, update, and verify the metadata? Who will preserve it? Libraries can do this, but in conjunction with the other demands for services and the budgetary crises, there will be a serious backlog. Even if librarians utilize consorsial arrangements like OCLC or working with individual national library networks, we will still find many serious gaps in the metadata.

But the above ignores the real potential in accessing geospatial information – the potential to aggregate from the local to the global – the true meaning of DE. Some libraries, mostly in the US, are developing excellent collections of data resources and metadata storehouses for local data. But DE is more than the US collections. If standards are somewhat established, then how will we organize the collection and access of the data? Is there an agreement among the map libraries and geospatial information providers to maintain metadata, beyond some well know local cases? Perhaps one immediate action would be to have the library community represented within the DE and GSDI community to develop reasonable processes to deal with the long-term viability of metadata collections. Libraries and archives have a unique perspective that is not normally present in the DE and GSDI community, the long-term view. Librarians deal in decades and centuries and longer. The culture of the digital geospatial community is one based upon the computing and engineering view, which tends to hold that one can re-create anything quickly, delete, overwrite, and download. This is short sighted at best. It is not to suggest that the deeds or views of those in DE and GSDI are wrong. Rather, this is simply a comparison of two cultures. Libraries have the views of openness, neutrality, redundancy and sharing to help make sure the collections we build have some *context and lasting nature*. These views have a place within DE and GSDI, as well as within other local efforts.

Where are the data, really? A modern and well developed nation, 60% of Canada is mapped using 20 year old information at the 1:50,000 scale. We have issues of currency and timeliness as well as barriers to international access to local data, such as cadastral information. Not all US federal information is as easy to obtain as one would think, and the NSDI system is still not seamless. When we add in local data needs to make digital earth a reality, as well as more grey literature and other non-georeferenced information, then we find the reality of digital earth is a long way off. Perhaps efforts at new organisational structures for geospatial information will work (such as the GeoData Alliances)? Perhaps we are merely at the very beginning rather than the middle of the process? If we return to the notion of Internet time, what can we expect in ten years in terms of policy and data access? I would argue that for geolibraries to be truly effective, and for digital earth to really meet the moonshot goal, then local data must become the vast majority of the holdings and collections. Digital earth is going to become the ultimate metaphor for “think global, act local”.

Kate Beard (1995) stated: “Engineers and librarians are two groups that one does not associate as having collaborative interests” and recognized that “fringe areas of library science” are the map librarians, and for engineering are those in spatial information engineering”. With this context the workshop on distributed libraries organized meetings to “frame the discussion by reference to the functions, services, and institutional arrangements of the library, for two major reasons: first, to engage the library community, with its long experience in providing access to information, in the development of a vision for a new kind of library and, second, to provide a familiar and concrete starting point for the discussion” (Mapping Science Committee, 1999).

Boston et al (1998) suggest that in developing new GIS-related services, it is both critical to include internal institutional or agency departments as well as a multitude of external partners who can provide support, guidance and become integral collaborators for creating new products and services. We must begin to recognise that we, under whatever banner or name, are within a specialized group. We are in essence a policy network (Coleman and Skogstad, 1990). In this case, policy does not simply refer to the creation of ‘policies’, but to the broader notion of socio-political associations that seek common goals – networks as communities. I would also suggest that we are part of an emerging network that has a foundation in the best of cartography, geography and librarianship. It does not mean that existing organisations or associations lose their identity. Cartographic societies remain ‘cartographic’, while map library associations remain ‘library’ associations. But we must be honest and recognize that we have too many associations and societies that are carrying out the same functions. Conversely, *because* of our

small size and collegiality, we should be in a better position than most to cooperate and find new, effective means to enhance and sustain our community and our work.

Rosenblatt, S. (1999) clearly states that we “must develop new expertise and build strategic alliances and collaborative relationships with complimentary partners inside and outside” our institutions (p.45). We are facing tremendous technological change that requires inputs of money, time and energy. We must become involved in policy issues related to the access and use of geospatial information in all its forms and will require high-level and substantial political efforts. Organisationally, we need to remove duplication of effort and find ways to coordinate actions, services, training, communication, public and political awareness and research and teaching, providing the services at which we excel. As the move towards a Global Geospatial Data Infrastructure gains acceptance and support, our community of librarians and curators and archivists needs to be well positioned to work within that structure *and* affect its development. We cannot react to such initiatives; we must be part of their development and use. We must also be present with the power and status that comes from representing a broader community of users and creators from the library, cartography, geography and GIS communities i.e. the geomatics community.

Clearly it behoves all associations related to these fields to sit down and openly discuss what it is they want to accomplish and how to be successful. The issues we face –demographic, policy, technological, and organisational – are such that they can only be overcome or dealt with effectively through cooperation and collaboration at levels we have yet to see. Many of these issues, including the need for more structure and collaboration amongst communities of librarians and others is more than adequately explored by Birdsall (2000). Smits (1999) also details the new roles for our professions and adds in analysis and representations. He does make note that this requires new educational opportunities and challenges, including the financial burden many institutions face with regard to technology and data. More importantly he clearly states that map societies must look at cooperating further. He also raises an interesting point of practical concern: which way do we go? Do we federate with library sciences or with cartographic or geographic communities? I think the map and GIS librarians can be even more unique, we can align ourselves with both, and they with us.

As further noted by Goodchild:

Projects like ADL, and the concept of a geolibrary, are bridges between the GIS and library communities..... Thus we find ourselves in the GIS research

community at the beginning of a period of exciting collaboration with the library and information science communities”(Goodchild 1998).

Martha Gorman, President/CEO of GITA (Geographic Information Technology Association) sent an email to the MAPS-L list on May 24, 2001 while attending the GSDI meeting in Colombia. In that note she asked: “where are all the librarians?” Her feeling was that the issues and topics were ideally fitted to the views and traditions of librarians. Many of the issues presented in this essay address the “why not” question. In large part it has more to do with how we have organized ourselves on the margins and less to do with what we have to contribute. We are, for better or worse, on the front-lines providing geospatial information services. We, even within our associations, have few resources with which to support the level of involvement needed or expected. We are also a small group of highly focused and specialized individuals. We need to change that if we are to be effective. And this point brings us full circle to the issue or metaphor of the Moonshot that Goodchild suggested was given to us through the call for the creation of Digital Earth (it was a moonshot speech not heard).

Moonshots like DE require focus and massive inputs of energy, human resources and funds. The original Moonshot (1963-1972) was criticized by the astronauts in the space program. In the beginning the engineers designed the spacecraft without windows and no steering mechanism – an astronaut was merely seen as cargo, along for the ride. We are again building a moonshot with an engineering and computational perspective. Those perspectives are needed and are excellent. But they do not answer all the questions, and they certainly do not emphasize results and impacts. This is a natural progression because we are enamoured with the technical possibilities, but we give short shrift to the human, organisation and socio-economic aspects. In fact, today the technology, not the person, is in vogue.

Why all the hype and why the rush? Perhaps it has to do with new technologies having a culture wrapped in Internet time. Perhaps it has to do with forgetfulness – we forgot that sharing information is still a new idea for most of the world, while the reality is we have always lived in an information age. Digital Earth will not get off the launch pad because the concept is based upon global access to local data, and most of the planet still does not view access as a “public good” and this will exacerbate the digital divide. We also tend to rush with the hope that someone will later clean up the mess. The technical aspects are meaningless because what really needs to be done right away is change focus back to the lasting things; our institutions, our culture our memory, our heritage and its relationship to the future. We need to focus the resources *within* the moonshot on institutions and intellect,

and less on the infrastructure and Internet. These are revolutionary events and times, but this is not a revolution.

References:

ARL (1997) *Transforming Libraries : Issues and Innovations in Geographic Information Systems*, Transforming Libraries, No. 2 February 1997 (ARL SPEC Kit 219), George Soete, Editor Prudence S. Adler (advisor); Washington, D.C.: ARL. <http://www.arl.org/transform/gis/gistrans.html> , Accessed June 11th, 2001).

Beard, Kate (1995). Digital spatial libraries: A context for engineering and library collaboration. Information Technology and Libraries, 14(2),79- 86

Birdsall, William F. (2000) A Political Economy of Librarianship? Hermes : revue critique No.6 (April, 2000) Available from : www.microtec.net/charro/HERMES6/birdsall.htm

Borgman, Christine L. (2000) From Gutenberg to the Global Information Infrastructure: Access to Information in a Networked World (Boston: MIT Press).

Boston, J., Dean, C.W., Phillips, H., and Pope, N.F. (1998). The Public Electronic Library: Integrating GIS Resources and Tools. Library Hi Tech 16(3-4), 100-106.

Boxall, James, (1998). *Spatial Data Infrastructures: Developments, Trends, and Perspectives from Converging Viewpoints – Introductory Remarks* . Special Issue on _Spatial Information Infrastructures in the 1990s (James Boxall, Guest Editor), Cartography and Geographic Information Systems (American Congress on Surveying and Mapping) July1998. 25(3), 129-131.

Boxall, James, (1999). Would you mind showing me a map of amazon.com? Cartographic and Geographic information reference in a digital age, a Canadian view”. Meridian (ALA Roundtable for Geography and Map Librarianship), 16 (1999), 9-16.

Boxall, James, (2000). Spatial Data within the Academy: “Much ado about nothing, or, there is chaos under the heavens and everything is excellent” Bulletin [ACMLA] #109 (2000),13-18.

Cline, N. M. and Adler, P.S. (1995). GIS and Research Libraries: One Perspective. Information Technology and Libraries, 14 (2), 111-115.

Cobb, D. A. (1995). Developing GIS Relationships. Journal of Academic Librarianship, 21 (4), 275-277.

Coleman, W.D. and Skogstad, G. (1990). *Policy Communities and Policy Networks: A Structural Approach*. In W.D. Coleman and G. Skogstad (Eds.), *Policy Communities and Public Policy in Canada* (pp.14-33). Mississauga: Copp Clark Pitman.

Congressional Subcommittee (1999). Congressional Hearing on "Geographical Information Systems Policies and Programs June 9th, 1999. Subcommittee on Government Management, Information, and Technology (Congressman Stephen Horn, Chairman)

<http://www.house.gov/reform/gmit/hearings/testimony/990609h.htm>

Eaton, J.J. and Bawden, D. (1991). What kind of Resource is Information? *International Journal of Information Management*, 11 (156-165).

Executive Office of the President (EOP), (1994). Coordinating geographic data acquisition and access: the National Spatial Data Infrastructure (Presidential Executive Order No. 12906). *Federal Register*, 59(71) (April 13, 1994), 17671-17674.

Fisher, P.F. (1998) Is GIS Hidebound by the Legacy of Cartography? *The Cartographic Journal* 35(1), 5-9.

FGDC Newsletter, 5(1) Spring 2001 www.fgdc.gov

Fleet, C. (1998). The Role of Computer Technology in the Future Map Library. *LIBER Quarterly* 8(2) <http://www.konbib.nl/kb/skd/liber/articles/>

Frew, James, Freeston, Michael, Kemp, Randall B., Simpson, Jason, Smith, Terence, Wells, Alex, and Zheng, Qi, (1996). The Alexandria Digital Library Testbed, *D-Lib Magazine*, <http://www.dlib.org/dlib/july96/alexandria/07frew.html>

Goodchild, Michael F., (2000). New Horizons for the Social Sciences: Geographical Information Systems. *Canadian Journal of Policy Research*, 1(1), 158-161.

Goodchild, Michael F. (2000). Cartographic Perspectives on a Digital Future. *Cartographic Perspectives*, 36 (Spring 2000), 1-9.

Goodchild, Michael F. (1998). *The Geolibary* (Available from: www.geog.ucsb.edu/~good/Geolibary.html accessed June 11th, 2001); see also: *The Geolibary*. In S.Carver (Ed.) *Innovations in GIS 5*. London: Taylor and Francis, 59-68.

Goodchild, Michael F., (1992). Geographic Information Science. *International Journal of Geographic Information Systems*, 6(1), 31-45.

Gorman, Martha (2001) "Global Spatial data infrastructure query" Electronic message posted to MAPS-L@Listserv.uga.edu (May 24th, 2001).

GSDI Secretariat May 2001 Newsletter www.gsdi.org/NL/nl1.jpg

HAL (2000) *Background Papers for Workshops: Geomatics Sectors Human Resource Study*. Hickling, Arthurs and Low Consultants. Study Committed by the Canadian Council of Land Surveyors, the Canadian Institute of Geomatics, and the Geomatics Industry Association of Canada. January 24th, 2000

Hawkins, B.L. (1996) *The unsustainability of the traditional library and the threat to higher education*. Paper presented at the Stanford Forum for Higher Education Futures, The Aspen Institute, October 18.

Hawkins, Brian L. and Patricia Battin. (1998) The Mirage of Continuity: Reconfiguring Academic Information Resources for the 21st Century. Washington, D.C.: Council on Library and Information Resources and Association of American Universities, 1998.

Jankowska, Maria Anna and Jankowski, Piotr (2000). Is this a Geolibrary? A Case of the Idaho Geospatial Data Center. Information Technology and Libraries, March 2000, 4-10

Kraak, M-J. (1998). The Cartographic Visualization Process: From Presentation to Exploration. The Cartographic Journal 35(1), 11-15.

MADAME (2000). *Practitioners Forum on the Economics of Geographic Information and Pricing Strategies: Report of Meeting* (Final) October, 2000 MADAME Project (Methods for Access to Data and Metadata in Europe) Deliverable R4, <http://www.shef.ac.uk/~scgisa/MADAMENew/Deliverables/R4final.htm>.

Mapping Science Committee (1999). *Distributed Geolibraries: Spatial Information Resources*. Summary of a Workshop: Panel on Distributed Geolibraries, Mapping Science Committee (National Research Council [US]). Washington, D.C.: National Academy Press. <http://www.nap.edu/html/geolibraries/>

National Research Council, (1993). *Toward a coordinated spatial information infrastructure for the nation*. Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources. Washington, D.C.: National Academy Press.

National Research Council, (1994). *Promoting the National Spatial Data Infrastructure through partnerships*. Mapping Science Committee, Board on Earth

Sciences and Resources, Commission on Geosciences, Environment, and Resources. Washington, D.C.: National Academy Press.

Obermeyer, Nancy J., (1995). *Reducing inter-organizational conflict to facilitate sharing geographic information*. In Onsrud, Harlan J., and Gerard Rushton (eds.), Sharing Geographic Information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Onsrud, H.J. (1998). *The Tragedy of the Information Commons*. In Taylor F. (ed), Policy Issues in Modern Cartography (Oxford: Elsevier Science), 141-158

Rhind, David W., (1992). Data access, charging, and copyright and their implications for geographical information systems. International Journal of Geographical Information Systems, 6(1),13-30.

Rosenblatt, S. (1999). Information Technology Investments in Research Libraries. Educom Review 34(4) (July/August), 28-32, 44-46.

Smits, J. (1999). The Necessity and nuisance of survival, or how to keep to our senses. LIBER Quarterly 9(2) <http://www.konbib.nl/kb/skd/liber/articles/>

Webster, Frank (1994). What Information Society. The Information Society, 10,1-23.

Weil, N. (1999). Forrester: Youth will change the Net economy. Computerworld Online News August 10, 1999. <http://www.computerworld.com/>

Weiss, Peter N. and Backlund, Peter (1996) *International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization*. Submitted to the Harvard Information Infrastructure Project, June 20, 1996, <http://www.aedue.org/article/weiss.html>

Wood, A.A. (1989). Keynote Address to the Congress of Cartographic Information Specialists Associations (presented November 9th, 1988, Chicago). Bulletin [ACMLA] 70, 9-14.

James Boxall
Killam Library
6225 University Avenue
Halifax, Nova Scotia, Canada B3H 4H8
jcboxall@is.dal.ca
902-494-3757

GEOSPATIAL DATA ACCESS: CAN WE MANAGE TO SHIFT?*

By Jan Smits

Our new earth

In his book *Snow Crash* Neal Stephenson portrays a virtual librarian. This is a digital construct, which looks and acts like a virtual human librarian, as this is easier to communicate with than an icon or any other concentrated digital display. The library he works with is a hyper library, i.e. all the data he has access to are interrelated, either by themselves or because of historical use. The librarian works with Hyper Cards and contains a self-educating algorithm, i.e. he has “the innate ability to learn from experience¹”. When asked who wrote him the librarian answers:

*“I was ... coded ... by a researcher at the Library of Congress who taught himself how to code. ... He devoted himself to the common problem of sifting through vast amounts of irrelevant detail in order to find significant gems of information”.*²

The most relevant source of information for us in this book is a piece of software called ‘Earth’, which “keeps track of every bit of spatial information that it owns – all the maps, weather data, architectural plans, and satellite surveillance stuff”. When we take this one step further we may also add to these geospatial data all locational data, which is part of analogue and digital alphanumeric objects and databases, keeping in mind that a lot of spatial phenomena, including human activity, are constantly monitored³.

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

¹ This may result among other things in the filing of profiles which are constructed of previously posed queries and answers, i.e. FAQ’s and FGA’s, but the profiles may also include topological or contextual intelligence.

² Stephenson, Neal (1992). *Snow Crash*. – London : ROC, p. 101

³ The amount of monitored data with a spatial attribute is overwhelming. To give but a few examples: satellites constantly spew data concerning the physical environment for e.g. weather-forecasts and the Global Change project and keep track of movements in the air and on the water; governments monitor traffic and other movements and keep track of planning and building; censuses are taken at intervals by most countries; biologists recount at intervals biological presence; companies keep large databases of customer-behavior; the world is almost spammed with web/video-cams; etc.

Thus one of the possible futures toward which we are trending or are unconsciously becoming part of is sketched.

What we have to assess is whether we are (going to be) able to work with access tools -whether existing, developing, or probabilistic- which make it possible to find the 'gems of information' our customers are looking for. In this assessment we have to keep in mind that we still mainly deal with Document-like Objects (DLO)⁴, but that in the future we will more and more have to locate first the relevant data and then the software to create a legible file and/or image and possibly a tactile image. This transition from DLOs to databases⁵ will take quite some time and depends partly on the availability of the latter in the public domain, i.e. in libraries, archives and the like accessible for private use. Should the data not be available in large quantities in the public domain than standards for its filing, retrieval and use will be market conforming and hopefully interoperable⁶ and available to the public at large.

Towards a new paradigm

⁴ Though the term is much in use on Internet sites concerning the documentation of Internet sources it is hard to find a definition thereof. After reading through some 30 Internet sources I did not find an adequate definition. I would define a DLO as a unit of information that carries distinct descriptive data, which can be used in a formal description. The following quotes may lead to a better understanding of the nature of DLOs:

Document-like objects which share the characteristic that they can, but need not be, adequately represented in a print format, examples being pure text, textual images, text with printable illustrations, and photographs.

POCKLEY, Simon (1997). *Killing the duck to keep the quack.*

<http://www.cinemedia.net/FOD/FOD0055.html> [Accessed: 11 January 2001]

Non-document-like objects, on the other hand, include such resources as virtual experiences, databases (including ones that generate document-like outputs), business graphics, CAD/CAM or geographic information generated from database values, and interactive applications which might have different content for each user. In the context of image discovery, these sources do not "contain" images as much as they "generate" images. The images they generate may be described as fixed document-like

objects, but the metadata required to describe them (the systems doing the generating) are distinct.

WEIBEL, Stuart, and Eric Miller (1996). *Image Description on the Internet*. A Summary of the CNI/OCLC Image Metadata Workshop, September 24 - 25, 1996, Dublin, Ohio. In: D-Lib Magazine, ISSN 1082-9873, January 1997.

<http://www.dlib.org/dlib/january97/oclc/01weibel.html> [Accessed: 11 January 2001]

⁵ Although R2V (raster-to-vector conversion) and satellite imaging of Land Use/Land cover classification are well established, the latter are mainly in the form of DLOs and not as databases. I would define a database as a collection or BLOB (Binary Large Object) of distinct data which are retrievable through an underlying structured principle, e.g. geographic co-ordinates, the periodic table of elements in chemistry, the genome sequence, etc.

⁶ Interoperability (as defined by the FGDC): A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems. Nerbert, Douglas D. (2000). *Z39.50 application profile for geospatial metadata or "GEO"* <http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm> [Accessed: 11 January 2001]

We are at the moment going through what might be called a paradigm shift. Kuhn used the term paradigm to denote a generally accepted set of assumptions and procedures that serve to define both subjects and methods of scientific inquiry⁷. When the assumptions and the procedures which serve them cannot answer anymore the aim for which they have been formulated or when the aim seems in need of reformulation as the answers are not adequate anymore, new assumptions and procedures have to be formulated which might answer the questions posed. Where Kuhn states that unanswerable questions lead to revolutionary science I would say that at any given moment in time a paradigm meets its paradoxical⁸ frontiers. When the paradoxes accumulate there comes a moment when it seems logical to formulate a new paradigm, which inherently requires new practises⁹. The idea of paradigm shift cannot, however, be formulated too strictly. As knowledge and the data it is based upon are cumulative and assumptions and procedures follow each other consecutively there can be no sharp divide between one paradigm and another. It can sometimes be that some older assumptions are still valid and some procedures still workable within a new paradigm. Only with historic hindsight can one tell when one paradigm has been replaced totally by another. Furthermore a paradigm shift in itself might be an objective occurrence, but it can only be acknowledged as such when there is wide subjective cognizance of the new paradigm.

The problem for us is that we are probably in the middle of a paradigm shift and therefore will have problems defining what assumptions and procedures might be adequate in the future and in what way we can make the transition as smooth as possible, as we cannot imagine future use except in a probabilistic way, presuming we can at least transfer contemporary procedures to future technologies. It feels like falling into a science-fiction story or a fairy tale, where part of the environment (physically and/or mentally) reminds us of our normal or traditional environment, but where there are also many phenomena and modes of thinking, living and communication which are hard to interpret, because we have to imagine them at the vague frontiers of our intellect and emotion. Sometimes we do not

⁷ Kuhn phrases it as the “universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners”. Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, University of Chicago Press, p. X.

⁸ Paradox: a person, thing or phenomenon at variance with normal ideas of what is probable, natural or possible. Hayward, Arthur L. and Sparkes, John J. (1982). *The concise English dictionary*.

⁹ Or as Lyotard phrased it: science comes more and more to the conclusion that her dynamism relies on the paradox, on incompatibility and indecisiveness. Peperstraten, Frans (2000). *Jean-François Lyotard*. In: *Filosofen in deze tijd*, p. 201.

understand them because our frame of reference does not allow for them or because we are afraid to let go of our traditional frame of thinking and imagining. New modes of communication, or in our case defining new kinds of queries and answers, can be especially dumbfounding as we do not yet fully understand the underlying syntax and semantics, nor what they should signify. And most of us are not aware yet that while looking for the golden apples, guarded by the Hesperides, Atlas is going to put the burden of the world on our shoulders. And it might take a long time before we have outwitted him and return the burden.

Kuhn was chided for his paradigm theory, especially the idea of incommensurability of subsequent paradigms, as he could only give a historical review of the phenomenon. This might be a valid criticism were it not for the fact that during his lifetime no total paradigm shift was observable, as is probably also applicable to our own situation¹⁰. I think his theory applies to our situation as the basis on which our assumptions seem to work (the data), the processing of them (software agents), and the way in which we have to manage the production flow (TGIS¹¹), as well as the aims for which these procedures are tools are changing in such radical ways, that the future situation seems incomparable or incompatible with the present or traditional one. Though the assumptions and procedures of the different paradigms are incommensurable they will exist side by side for some time as some of the data within the former paradigm, especially DLOs, cannot be transformed into the new paradigm, as the data from which they are constructed cannot be extracted in separate entities with their accompanying attributes from them and thus cannot function as independent data within the new paradigm. The digitized maps need to be emulated¹². But also certain procedures ought to have low-level processing and retrieval agents, which are emulations of procedures of the former paradigm.

One of the properties of a paradigm shift is that experimental or innovative assumptions and instruments will diverge, i.e. the gap between separate (groups of) collections will widen for some time as, in the first instance, each will define anew its goals and the ways in which to try to achieve them. As with the old paradigm

¹⁰ I think he should have applied his theory more to the totality of sciences or society as a whole than to a specific science. As a result paradigm shifts would take more time than 'revolutionary science' before new paradigms are accepted, as they have to permeate the sciences or society.

¹¹ Temporal GIS. "Although the computer technology for spatio-temporal analysis exists, the GIS community must undergo a paradigm shift to fully appreciate TGIS benefits. It's not just a matter of collecting time-based data within a GIS, but also developing a new way of thinking about time in a spatial sense". Jim Castagneri (1998). Temporal GIS explores new dimensions in time. <http://www.geoplance.com/gw/1998/0998/998tmp.asp> [Accessed: 20 February 2001]

¹² Rothenberg, Jeff (2000). *Using emulation to preserve digital documents*. The Hague, Koninklijke Bibliotheek, 69 p.

(Paris Principles and ISBD/MARC) it will take time to converge upon common goals and instruments. Some collections may have more trouble achieving these than others, depending on feelings of certainty about the goals to be achieved and resources demanded and being available.

In order not to lose hope or get lost we have to assess the available contemporary technology and as there is no other way try to look for ways to emulate, migrate or transform them to newer configurations. At the same time we have to try, however, to innovate technologies or pirate them from other science-fields in order to be able to answer new kind of queries. Unfortunately I'm not practically involved sufficiently in how databases and the relations and topologies of their data are constructed and how these are developing, to totally understand and describe the change. The only thing I can do for the moment is try to imagine how they will develop without going into particulars. On the other hand I can summarize the traditional retrieval mechanisms we use and see how additional and/or enhanced functionality can be developed, and comment on the possible adaptability to our field of new retrieval mechanisms from other fields. When speaking of traditional retrieval mechanisms I think of Dublin Core (DC), ISBD/AACR-MARC (IAM), FGDC/ISO (FI)¹³ standards and the like; when talking about innovative retrieval mechanisms I think of tools like visual geographic interfaces, TextTilling, and Knowledge Discovery and Data Mining.

From DLOs to data

The traditional paradigm we are working within is that of identifying DLOs. What we have to try to cope with is that in the future instead of images constructed of several layers of data printed on top of each other within a single image we will primarily work with single data or datasets. This is a change with more impact on our work than we may imagine. Up till now we have been well versed in reading cartographic representations and translating their contents to our customers who might not be so well versed. We do not have to be cartographers because the image is constructed for us and we mainly have to evaluate whether the sources used in constructing it are up to par for the query posed. We have learned to read the patterns and relations in the single and combined features of the map. We have learned to access the image through its particulars, using its formal descriptive elements and interpreting its cartographic and mathematical properties and subject matter. For interrelations we use geographic and subject thesauri. And when we have a sophisticated retrieval system we can select and combine single features and

¹³ It might be that the FI-standards are mechanisms, which are exemplary for the paradigm shift, while they are innovative in coding nuclear entities and their attributes or properties, and qualifying them.

properties from the descriptions of the items and as an answer to our query we get one or more descriptions of maps that contain the information we are looking for. When necessary we can aid our customer in interpreting the documents he/she gets as a result of the query that was posed.

This method, however, does not always give the desired answer, and usually will only be an approximation to the query posed, as no single item or set of cartographic items fully answer the question(s) posed. This is because we are depending on the selection and interpretation of data by the cartographer who constructed the map. And in the construction from basic to generalised data the cartographer not only uses one to one mechanisms, but also adds or distracts to/from the resulting images in such a way that the image becomes more friendly, artful or usable for the user¹⁴.

The array of products which are produced, use but a fraction of the amount of data available, as physical production of maps is limited. Some data is probably never used as nobody imagines that it can be used for the construction of maps, or because it is economically not viable to use it, or because simply nobody knows that it exists or can be functional for certain purposes.

In his work *Laocoon*¹⁵ the German dramatist and critic G.E. Lessing (1729-1781) “warns poets against empirical descriptions of landscapes, because the consecutive nature of language breaks up into parts what is bodily coexistent. These parts are then difficult or impossible to put together again. What exists side by side becomes transformed into before and after by verbal description. Painting has not this kind of limitation. Nor has the map, a geographer would add.”¹⁶ However, the map is an interpretation of only a small part of the data available to construct it. As soon as it is created it gives the overview which hopefully is an approximation of the reality we seek, meaning we can far more easily interpret the interrelations between its constituent parts.

From evaluating to creating

But the future is probably inversely different. “With an array of analysis tools [researches in future] will work more exhaustively with the ... digital resources than they have so far been able to work with analogue ones. ... Increasingly, research efforts to interpret these data are assisted by data visualization tools. In

¹⁴ The cartographer can even differ from reality when he puts the several layers of a map (e.g. topography, orography, text, etc.) together to make the image more legible or artful.

¹⁵ Lessing, Gotthold Ephraim (1766). *Laokoon oder ueber die Grenzen der Mahlerey und Poesie*.

¹⁶ Hägerstrand, Torsten (1995). *Landscape as overlapping neighbourhoods*. In Benko, Georges B., and Strohmayer, Ulf. *Geography, history and social sciences*. Dordrecht, Kluwer, p. 86.

addition to data mining and visualization tools, future researchers will be aided by intelligent agents that explore the ... [data] looking for information that meets certain user-specified criteria and refine their searches as they accumulate data and knowledge. Digital archives [or repositories] combined with new technologies will liberalize scholarship. They will enable simultaneous access to a range of sources (both local and distant) and facilitate the use of research methods not possible with conventionally printed or hand written records.”¹⁷

Though many of our colleagues understand the words in the premise Andrew Tatham laid down during the LIBER-conference in 1994¹⁸, few of us can yet act it out in reality or understand its full impact.

In the data-age we, or software which acts in lieu of us¹⁹, have to interrogate our customer as to what kind of features, topology, relations, patterns, etc. he/she wants to view in the image to be created. The easiest possibly is to select the area for which the image has to function, together with its mathematical properties, like scale, projection, geographic grid²⁰, etc. Then we have to try to fill in that empty space with point-, line- and polygon-features with their attributes from one or more datasets available to us. And we must keep in mind that some might be stacked one on top of the other as datasets might come in many levels of functionality²¹.

Now we have to try to think what the image we want to create might look like, then retrieve the data with which this image can be constructed, and then use the software that actually constructs this image. This means we have at least to be part-time cartographers to effectuate this process, and keep the critical words of Lessing echoing in our minds. The image to be created should be first formed in our mind's eyes before we select its constituent data to build it, in order that the

¹⁷ Ross, Seamus (2000). *Changing trains at Wigan: digital preservation and the future of scholarship*. London, National Preservation Office, pp. 11-12.

¹⁸ "We shall no longer provide the users with someone else's selection and presentation of data, but with the data itself and with the means by which the user can make their own selection and presentation of this data to inform or to mould their own or other people's image of the world". TATHAM, Andrew (1995). *Can the map curator adapt?* In: *The Liber quarterly*, vol. 5/1995, no. 3, pp. 330-336. [<http://www.kb.nl/infolev/liber/articles/2tatham.htm>]

¹⁹ It is questionable whether idiolect can be coded into software, in order to retain part of our subjective frame of reference, unless this software has the innate ability to learn from interaction with the customer. If so, for regular customers, their own query-profile might evolve, which includes part of their own idiolect.

²⁰ Most of this is 'spatial reference information'. E.g. Phillips, Hugh (2000). *'Colorized' Content Standards for Digital Geospatial Metadata, FGDC-STD-001-1998*.

[http://badger.state.wi.us/agencies/wlib/sco/meta/colorstd/hier_basicsdgm.html#Spatial_Reference_Information. Accessed: 2 January 2001]

²¹ E.g. *TOP10vector*, the 1:10,000 digital topographic map of The Netherlands, consists of 69 layers of information, which might be retrieved per layer or in a combination of layers.

cartographic overview we want to create is not marred by the fact that we fall into the trap of what is before and what is after without knowing what is the whole.

But it also means a radical change in accessing the data we have under our control. Descriptions like DC and IAM are inadequate as they are developed for identifying DLOs. Descriptions like FI are more adequate as they are developed in order to be able to evaluate the inherent quality of digital data or datasets. But they are first-generation standards, which can be used for evaluating and selecting single datasets. Are they also functional for knowledge discovery and data mining, i.e. finding the relevant data-elements or collections of data-elements from disparate datasets, which will be functional in the map-image our customers are requesting²²?

The new methods of processing depend for a large part on newly formulated queries and the belief or acceptance that such queries should be answerable with the data or datasets available. This means that new retrieval technology will develop experimentally and that many developments will turn out to be dead ends. To prevent too many dead ends we have to look at technologies used in those disciplines which work with vast amounts of data, like astronomy or remote sensing, census processing or the actuary field. In these disciplines testing tools try to make out patterns, relations and topology.

Geographic data growth is more than exponential. When we call a one-million map collection large by present day standards in Terabyte measures it might be called not overly large, and in future it might be even called small, also taking into account that a lot of the DLOs are topographical maps. The problem is that our tools with which we manage future data-collections must be more automated and partly independent as the human mind can but cope with a limited amount of data and produce from this a limited set of DLOs. In order to be able to manage vast quantities in the future we have to be prepared to become more and more meta-managers, of the data as well as of the metadata.

And we cannot stick to maps alone, though these are historically the background to our profession. We must be willing to widen our perspective to include pure geographic data, thus bringing geography and cartography back to their

²² N.B. data mining is data-driven, whereas by contrast, traditional statistical, reporting and query tools are user-driven. To be a true knowledge discovery method, a data mining tool should unearth information *automatically*, and extract these in such a way that they can be put to use in areas such as decision support, prediction, forecasting, and estimation. I will not investigate in this article the impact of this technology as this is still too far away from our everyday practices.

dichotomous nature²³. But also we might sift auxiliary data from fields²⁴ which are not geographic in nature but which collect or use much locational data.

Within this frame of mind we have to look at the same time at combining or recombining disparate data of different value and functionality or even better, imagining the data and datasets as an amalgamated whole, in order to come to new insight or answers. However, we have to strive within our competence, while at the same time trying to imagine new, yet unknown, goals in our minds. Some of these processes might be autonomous as TGIS together with data mining can analyse data and their combinatorial values. One of the results might still be a cartographic map in order that the human mind can get an overview and process the data visualised within the image.

Assessment of traditional access-tools

When the documentation community became aware of the fact that electronic data or datasets needed a different approach of documentation then traditional analogue materials a vast amount of initiatives and evaluations took place in the 1990s. Since the early 1990s producer- and user-communities of elaborate digital datasets (DEM's, TIGER-files, remote-sensing images, etc.) started to develop sophisticated standards to document these.

In libraries bibliographic processes have been aimed at the user-tasks: 'find', 'identify', 'select', and 'obtain'. Other metadata-initiatives have added amongst others 'qualify' and 'manage' or 'housekeeping'.

It is understood that information retrieval is measured in terms of recall and precision. If a lot of relevant information is missing, there is poor recall. If we get flooded by a lot of irrelevant information, there is poor precision. This pronouncement is adequate for discerning differences between the relevant metadata systems, which will be described below.

To foster understanding I have created a diagram in which the metadata-systems have been divided into 4 groups. Added is a 5th group which contains innovative tools, which will be treated separately.

²³ And incidentally give more value to the "G" in the name of the Geography & Map Library Section.

²⁴ Like biochemistry, biology, botany, engineering, history, medicine, physics, psychiatry, zoology, etc. See: Millea, Nick (1998). *Delivering digital data into the library: the DIGIMAP project and its impact on the map room - the Bodleian Library experience*. In: *The Liber quarterly*, vol. 9/1998, no. 2, pp. 189-200, table 3 [<http://www.kb.nl/infolev/liber/articles/millea03-11.html>]. Accessed: 2 January 2001];

Morris, Barbara et al. (2001). *EDINA Digimap: new developments in the Internet mapping and data service for the UK Higher Education community*. In: *Liber quarterly*, vol. 10/2000, no. 4, Table 1 and 2, pp. 446-448.

<i>Quality</i>	Simple	Band II	Band III	Rich	Innovative
<i>Quality Level</i>	Band I	Band II	Band III	Band IV	Band V
<i>Diffusion</i>	<----->	<----->	<----->	<----->	<----->
	+++++++>	+++++++>	+++++++>	+++++++>	<+++++++>
<i>Availability</i>	Internet	Internet	Internet / Intranet	Internet / Intranet	Intranet
<i>Purpose</i>	Location	Selection	Evaluation	Analysis	Intelligent analysis
<i>Unit of information</i>	Individual digital information object	Logical set of digital objects; no links between documents	Publication; links between whole and parts	Databases with links between whole and parts on all levels	Databases with geometric links between whole and parts on all levels
<i>Standards</i>	Proprietary	Emerging standards	Generic standards used in information world	Standards used in specialist subject domains	Intelligent software-modules
<i>Form of record</i>	Proprietary simple records	Dublin Core	AACR2, ISBD	FGDC, CEN, ISO, ANZLIC	Not applicable
<i>Format</i>	Unstructured	Attribute value pairs Dublin Core	Subfields, qualifiers MARC USMARC UKMARC UNIMARC MARC21	Highly structured mark up FGDC DTD ANZMETA	Deriving from Knowledge Discovery and Data Mining
<i>Conservation</i>			Nedlib	Nedlib	Not applicable
<i>Input</i>	Robot generated	Robot plus manual input	Manually input	High level of manual input	Intelligent expert system
<i>Conversion</i>		DC/MARC/GI LS	MARC2DC MARC2FGDC	FGDC2MARC	Not applicable
<i>Protocol</i>	http with CGI form interface	directory service protocols (whois++) with query routing (Common Indexing Protocol)	Z39.50	Z39.50 (in future with collection navigation)	Collection navigation

Diagram of access- and retrieval systems

Band I²⁵

These are simple records, created mainly by robot search engines by general search-and-auto-describe missions. As they do not discriminate between the functions of the words they harvest the resulting indexes show a lack of precision for those seeking qualified information. Unaware web surfers who start out to explore the available information in a random way mainly use them.

Band II

On the instigation of OCLC and NCSA workshops have been held since 1995 to try to find a modus in which these metadata can be formulated. The first workshop held in Dublin, Ohio (U.S.A.) found consensus on a set of elements called since then the **Dublin Core** (DC).

²⁵ For a more extensive treatment and developments within Bands I-IV see: Smits, Jan (2001). *Metadata and standards, confusion or convergence?* In: ???

Dublin Core records are specifically created for resource discovery and as such description requirements are generally less precise than for traditional bibliographic records. This, in order that also non-professional documentalists can generate them. Soon the user community (among others the library field) found, however, that more precision could be needed to be better able to query DC-databases. Precision is sought in qualifiers, which can be added to metadata-tags. And when possible using controlled systems²⁶. A problem to be solved is that most controlled systems or vocabularies (like Dewey Decimal Classification 21, Universal Decimal Classification, Library of Congress Subject Headings, etc.) are not freely available on the Internet, in which case those users who have disposal of these can only use their quality. Use of these controlled systems would enhance interoperability with other metadata-systems, like those in Bands III and IV.

Libraries use DC-metadata in an explorative way but also should embed them in the electronic resources they create themselves and make available through the Internet. Too much metadata is created in a post-coordinate process. To create more uniformity of DC-records some organisations provide more or less sophisticated templates²⁷ to enhance the value of the metadata.

As the Dublin Core is maturing a little the user-community is discussing whether the 15 Dublin Core elements might be more coherently expressed if they are related to an underlying logical model, which treats information resources as having logical states (an abstract work or a physical item, for example) that have relationships to each other and to other resources²⁸. The inherent simplicity of DC-cataloguing poses quite some problems to map librarians, mainly due to the fact that there is no consensus yet as to what is 'best practice' and 'minimal adequate' DC-cataloguing. Though the syntax of DC is developing quite well there is a lack of semantics, as there is no rulebook like the ISBD.

²⁶ E.g. <DC:coverage CONTENT="(SCHEME=INTERNATIONAL PRIME MERIDIAN, GMT)"> W123°20'- W121°09' / N49°38'- N48°55' </DC:coverage>

²⁷ E.g. Koch, Traugott and Borell, Mattias (1998). *Dublin Core Metadata Template* <<http://www.lub.lu.se/cgi-bin/nmdc.pl>> [Accessed: 25 January 2001], part of the 'Nordic metadata project'. These and other tools can be found on the website 'Metadata Related Tools'. DCMI <<http://purl.org/dc/tools/index.htm>> [Accessed: 25 January 2001].

²⁸ Weibel, Stuart (1999). The State of the Dublin Core Metadata Initiative April 1999. *D-Lib Magazine*, Volume 5 Number 4. CNRI < <http://www.dlib.org/dlib/april99/04weibel.html>> [Accessed: 25 January 2001]

Band III

In 1961 the IFLA adopted the Paris Principles²⁹ as basis for an international approach to “headings” and “entry words”. They were however solely meant for catalogues of printed books. In 1967 the Anglo-American cataloguing rules (AACR) were published, followed by an interpretation manual for mapcurators in 1982³⁰. The convergence of cataloguing codes started in 1971 when the IFLA published the first of the ISBD’s for monographs and serials soon to be followed in 1977 by the ISBD(CM) for cartographic materials. Almost all ISBD’s were revised in the 1980s and 90s (ISBD(CM) in 1987) and the last to appear in 1997 was the ISBD(ER) for electronic resources. This latter created havoc, because it means that most ISBD’s have to be revised again to align them with the new insights concerning the description of electronic materials.

But there is more to it. As a result of changes in the expression of documents and their transport media IFLA started to evaluate the whole configuration of Paris Principles and ISBD’s³¹. This study evaluated the value of the fields and elements, using a user-oriented conceptual model, where the elements are described as entities, which have attributes (the attributes serve as the means by which users formulate queries and interpret responses when seeking information about a particular entity) and relations to other entities. When they are related to user-services their function is still restricted to find, identify, select and obtain.

The ISBD(CM) under revision, of which a third draft was circulated in May 2000³², tries to amalgamate alterations to the ISBD(CM) and those parts of ISBD(ER) which seem appropriate. The draft is largely adapted to the recent issue of CCQ³³. For the moment there is a choice to extend area 3 (the material specific area) with:

- Statement of accuracy [for scale] (optional)
- Designation and structure of electronic resource
Structure of resource (optional)

²⁹ IFLA Committee on Cataloguing (1971). *Statement of principles, adopted at the International Conference on Cataloguing Principles, Paris, October, 1961*. Annotated edition with commentary and examples by Eva Verona. IFLA.

³⁰ Stibbe, Hugo P. (1982). *Cartographic materials : a manual of interpretation for AACR2*. ALA.

³¹ IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional requirements for Bibliographic Records* (UBCIM publications - new series, Vol. 19). IFLA. Also available through <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>> [Accessed: 25 January 2001]

³² Bäärnhielm, Göran (ed.) (2000). *ISBD(CM): international standard bibliographic description for cartographic materials*. Revision including Electronic Resources. Third draft. (Not published).

³³ Andrews, Paige G. and Larsgaard, Mary Lynette (1999). *Cataloging & Classification Quarterly*, Vol. 27. Haworth. Also published as: *Maps and related cartographic materials: cataloging, classification, and bibliographic control*. New York, Howarth, 487 p.

- Number of files, records, bytes (optional)
- Digital graphic representation method (optional)
 - Object type (optional)
 - File format (optional)
 - Object count (optional)
- Geospatial reference data (optional)
- Designation of electronic resource (when applicable)
- Characteristics (i.e. structure) of electronic resource (optional)
- Digital graphic representation data (optional)

Further there are large changes/additions in/to the ‘edition statement’, and ‘notes on system requirements’. Though the latter seem to be mandatory in all ISBD’s I wonder if they will not serve better under the ‘manage’ part of metadata. Especially after a certain time when most electronic resources need to be converted, emulated, migrated or otherwise changed these data serve mostly internal library functions and not user-needs.

AACR and the ISBD are cataloguing codes, but before the user can search the bibliographic databases and read the descriptions they have to be electronically processed. For this purpose MARC (MACHine Readable Catalogue)-formats are developed. MARC is an implementation of the international standard *Information and documentation - Format for information exchange* (ISO 2709-1996)³⁴.

To alleviate problems with existing MARCs the Library of Congress in consultation with various user communities develops MARC21. This format “is an integrated format defined for the identification and description of different forms of bibliographic material. MARC21 specifications are defined for books, serials, computer files, maps, music, visual materials, and mixed material. With the full integration of the previously discrete bibliographic formats, consistent definition and usage are maintained for different forms of material”³⁵. The format attempts compatibility with e.g. UKMARC and UNIMARC. After consultation with the UK library and information community the BL Executive Committee has decided to adopt the MARC21 format to replace UKMARC in time. One of the reasons for

³⁴ The British Library (1999). *MARC home page*. BL <<http://www.bl.uk/information/marc.html>> [Accessed: 25 January 2001]

³⁵ MARBI (1996). *The MARC 21 Formats: Background and Principles*. LoC <<http://www.loc.gov/marc/96principl.html>> [Accessed: 25 January 2001]

this decision was that “MARC21 represents a more effective route to the eventual adoption in the future of non-MARC metadata standards ...”³⁶.

With the upsurge of Internet applications the Library of Congress started in 1995 the project ‘MARC DTD’³⁷ to enable the conversion of MARC-data to an SGML-environment and vice versa. With this DTD MARC-data can be converted to the Internet, but also organisations can create MARC-like metadata in the Internet and convert it to MARC-databases.

Band IV

Digital spatial databases have been created from the late 1970's onwards and reach nowadays the phase where there is countrywide coverage (on municipal, provincial and state level) and through GIS they can be integrated with other databases. Initially they were a continuation of existing analogue processes, but for the fact that they are more in vector-format and are built up of layers of information, which can be manipulated independently from each other or in concert with each other. To extend operability a lot of producers are digitising their existing analogue data, usually in raster-format. Seeing the benefit of promotion of and the need for a higher return on operating costs producers started to think of ways to make these data more accessible. As the economic stakes are higher then before they sought to create a descriptive system, which incorporates not only data usually associated with ISBD's but also data, which could help users to evaluate and analyse the fitness of use and quality of the digital spatial data offered. Examples of these standards are ANZLIC³⁸ (Australia & New Zealand), FGDC³⁹ (USA), CEN⁴⁰ (Europe) and ISO⁴¹ (global). The latter two probably will come formally into existence in 2001. The ICA Spatial Data Standards Commission will

³⁶ The British Library Executive Committee (2001). *MARC harmonisation* : British Library to adopt the MARC21 format : Survey results. <http://www.bl.uk/services/bsds/nbs/marc/result1.html> [Accessed: 29 January 2001].

³⁷ Network Development and MARC Standard Office (1999). *MARC DTD's, document type definitions, background and development*. LoC <<http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html>> [Accessed: 25 January 2001]

³⁸ ANZLIC (1998). *Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand*. ANZLIC <<http://www.anzlic.org.au/metaelem.htm>> [Accessed: 24 April 2000]

³⁹ FGDC (1998). *Content Standard for Digital Geospatial Metadata* (FGDC-STD-001-1998). FGDC <<http://www.fgdc.gov/metadata/constan.html>> [Accessed: 25 January 2001]

⁴⁰ CEN/TC287 (1999). *The Geographic Information European Prestandards and CEN Reports*. AFNOR <<http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/PUBLIC/WEB/ENGLISH/pren.htm>> [Accessed: 25 January 2001]. To view the full pre-standard you need to contact a CEN-member.

⁴¹ ISO (2001). *ISO/TC 211 Programme of Work*. ISO < <http://www.statkart.no/isotc211/pow.htm>> [Accessed: 25 January 2001]. To get access to the documents and the document catalogue, you need user-id and password. Please contact your national body to get this information.

publish in 2001/2 a book, which will compare all available national and international metadata standards and –systems, alike to their book on transfer standards⁴².

As a lot of data is needed for the evaluation- and analysis-processes the records can run in the hundreds of elements and be anywhere between a few thousand and tens of thousands bytes. As such they can be also used as tools to administrate the datasets they describe.

Though one of the complaints concerning records created with these standards could be that they need a high amount of manual input I guess that better programming of the metadata-module within spatial datasets could help automated processing of numerical values (e.g. object-count) and other automatically generated data (e.g. geometric base values) in the metadata-records.

Towards meta-information systems?

In band IV-descriptions it is already possible to add metadata concerning the administration of the metadata. The metadata itself can provide information to assist in developing migration/emulation processes and to check whether these processes have functioned adequately. Metadata can help in ensuring the level of integrity of the data after necessary manipulations to preserve them for future use. They help to answer the question whether the data have the same information value and quality as before the manipulations took place and at the same time whether unavoidable losses can be qualified⁴³.

This drive is also detectable in the NEDLIB⁴⁴ (Networked European Deposit Library)-project, funded partially by the European Commission's Telematics for Libraries Programme. One of the reports⁴⁵ concerns itself with metadata for long term preservation of electronic publications, based on the OAIS information model⁴⁶. The different kinds of metadata described in this report are:

⁴² Moellering, Harold (ed.) (1997). Spatial database transfer standard 2: characteristics for assessing standards and full descriptions of the national and international standards in the world. Pergamon.

⁴³ Smits, Jan (1999). *Metadata: an introduction*. In: Andrew, Paige G. and Larsgaard, Mary Lynette. *Maps and related cartographic materials: cataloging, classification and bibliographic control*, p. 313.

⁴⁴ While supplies last all reports (up till now some 7 reports have been published) can be requested free of charge from the Koninklijke Bibliotheek by filling in the Nedlib-orderform at www.kb.nl/nedlib/

⁴⁵ Lupovici, Catherine [and] Masanès, Julien (2000). *Metadata for the long term preservation of electronic publications*. Koninklijke Bibliotheek. (Nedlib report series ; 2).

⁴⁶ CCSDS 650 0-R-1: *Reference model for an Open Archival Information System(OAIS)*. Red Book. Issue 1. May 1999. http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html [Accessed: 5 February 2001]

1. Descriptive metadata
2. Administrative metadata recorded for the deposit system management purposes
3. Metadata for preservation

The descriptive metadata can be the same as generated in band II-IV-descriptions, the administrative metadata can be the same as that generated in Band IV-descriptions. The metadata for preservation are a new subset and are used to monitor the actions needed for the long-term preservation of the content information. In the Nedlib report the following elements are given for a “core preservation metadata set”⁴⁷:

- Specific hardware requirements
- Specific microprocessor requirements
- Specific multimedia requirements
- Specific peripheral requirements
- Operating system
- Interpreter and compiler
- Object format
- Application
- Reference information
- Assigned identifier
- URL
- Fixity information
- Checksum
- Change history
- Main metadata concerned
- Tool
- Reverse
- Other metadata concerned

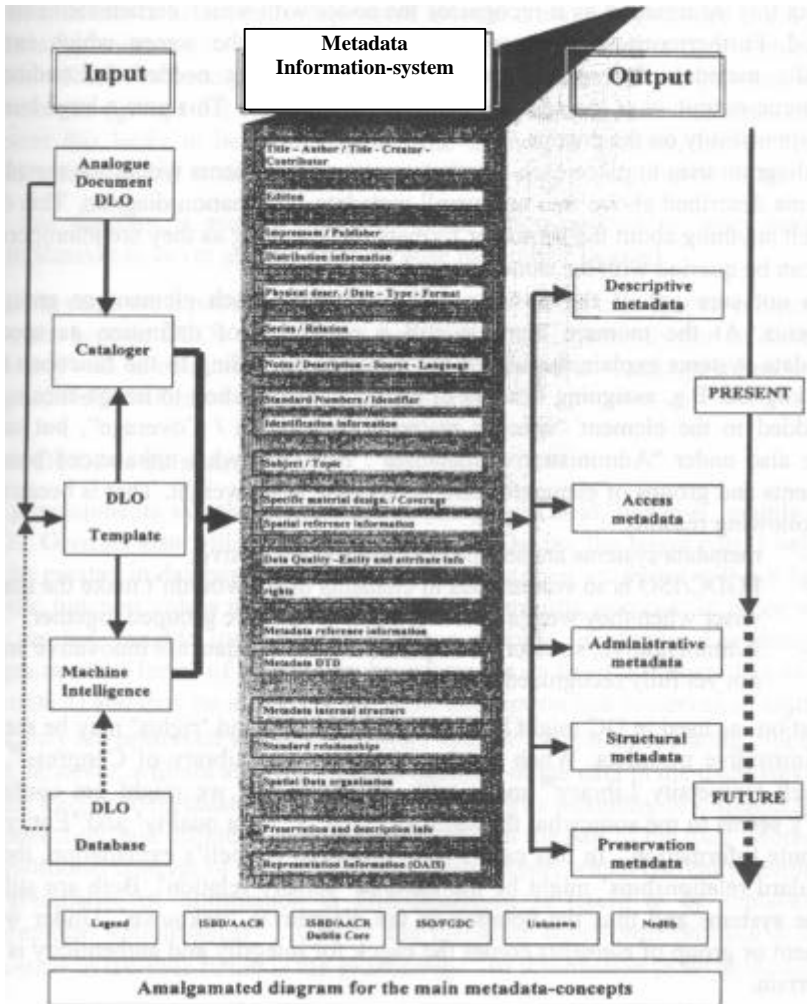
When we take a closer look at some of these elements we see that they have the same function as the data concerning electronic publications in the ISBD(ER), except that in the latter they are relegated to the note-area and are summarized within one element. The Nedlib report is ahead of a possible extensive expansion of the formats used for bibliographic library systems, which eventually might evolve into all-encompassing meta-information systems.

Not all of the necessary metadata needs to be manually processed. When one imagines the repository as a totally integrated system of DLOs and metadata with all relevant links available than an automated processing system can retrieve all the

⁴⁷ Lupovici, Catherine. Ebenda, pp. 17-21.

necessary metadata from the actions implemented for its preservation in the repository and add these to the relevant metadata-record(s). Instead of relying entirely on the prescience of the cataloguer, a digital library might allow direct search of the contents of an information object, particularly for domain-specific information or minor detail that a cataloguer might normally ignore⁴⁸.

⁴⁸ Goodchild, Michael F. (1995). *Alexandria Digital Library. Report on a workshop on metadata held in Santa Barbara, California, November 8, 1995.*
http://www.alexandria.ucsb.edu/publicdocuments/metadata/metadata_ws.html [Accessed: 15 February 2001].



When we take a sharper look at the metadata-systems we can see a lot of similarity, but also a lot of divergence. Not so much in the content of the metadata as well in the way they function at the input- or the output level.

At the Input side one must image that at the moment physical information resources and DLOs are offered to be accessed. Most of these are still manually processed by cataloguers. Some of the data the cataloguer extracts from a template

which is included in DLOs. But one can also image that a computer-program extract this information as it recognizes the codes with which certain elements are tagged. Furthermore scripts may be running behind the screen which extract specific metadata for specific purposes which are not needed for traditional catalogue-output, as is the case e.g. with the Nedlib data. This puts a large burden of responsibility on the creator.

The diagram tries to place each element or group of elements within the metadata-systems described above into an overall metadata-information diagram. This does not tell anything about the format or formats used, as long as they are interoperable and can be queried with the same protocols.

I am not sure yet of the position and function of each element or group of elements. At the moment there is still a confusion of definition as specific metadata-systems explain the use of the metadata according to the functions they get assigned. E.g. assigning headers of technical information to image-files might be added to the element “specific material designation / Coverage”, but might come also under “Administrative metadata”. It is somewhat unbalanced because elements and groups of elements seem to have the same weight. That is because of the following reasons:

- metadata systems are still predominantly descriptive;
- FGDC/ISO is so voluminous in elements that it wouldn't make the reader wiser when they were all included and so they are grouped together;
- Administrative, structural and preservation metadata are innovative and not yet fully recognized.

‘Relation’ as used in DC might be structural metadata; and ‘rights’ may be seen as administrative metadata. When we consider what the Library of Congress⁴⁹ and Cornell University Library⁵⁰ mean with structural data we might get confused. LoC's seems to me somewhat the same as FGDC's ‘Data quality’ and ‘Entity and attribute information’. In this case I have preferred Cornell's explanation, though ‘standard relationships’ might be the same as ‘series / relation’. Both are still in-house systems and thus the notation in the diagram is ‘unknown’. Under which element or group of elements comes the check for integrity and authenticity is also uncertain.

⁴⁹ LoC (1999). Structural Metadata Dictionary for LC Digitized Material. Version 1.03.

<http://memory.loc.gov/ammem/techdocs/repository/structmeta.html> [Accessed: 21 February 2001]

⁵⁰ Cornell University Library (2001). Moving theory into practise : Digital imaging tutorial. Chapter 5: Metadata

<http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-01.html> [Accessed 21 February 2001]

Seen from the output side it might be necessary to extract metadata from different formats, as this is functionally dependent. Should crosswalks between the diverse formats be necessary to effectuate certain output this should be done behind the screen without bothering the user. The five output systems are put tentatively opposite the (groups of) elements from which they extract their metadata. At the moment this looks to be biased against the innovative systems, but in future I think, when seen the amount of metadata used, the opposite will materialize.

But this diagram is mainly meant to foster discussion as to what metadata might/should/can cover and for what purpose they might be used.

In a later stage the input-side will change radically. In the case that all information resources become DLOs or databases metadata will be inherently part of them and obviously will be extracted mainly by computer programs. Catalogers then will be assigned the functions of checking and validating.

Again DLOs vs databases

The developments sketched in the above paragraphs deal, however, mainly with DLOs. Geodata also will be partly created as DLOs, but the larger part, I believe, will be created in databases, where the entities and their attributes will not have a specific function unless they are generated for a specific purpose. In other words they are but variable instantiations within a specific timeframe. The generated images or other forms of output may be referred to as fixed document like objects (see note 4) and may be in need of the same description and preservation metadata when they are preserved within a repository system⁵¹. Such a generating process in itself, however, will not alter the inherent qualities of the data in the databases.

“In the long term metadata ... will evolve to support a hierarchy of different levels of abstraction, and different degrees of expertise on the part of the user. Moreover, the traditional concept of the map sheet or image scene, which currently dominates the granularity of information is GIS-databases, is likely to be replaced by geographical seamlessness”⁵². One of the problems which will produce a lot of discussion in the near future is the *granularity*⁵³ of information to be accessed, i.e.

⁵¹ I believe that examples of these generated fixed DLOs should be preserved in order that future users can obtain knowledge of part of the historical context during the lifetime of the databases.

⁵² Goodchild, Michael F. (1995), Ebenda.

⁵³ A meta-data entity that is associated with the lowest level of granularity of information available to InfoHarness is called the information unit (IU). The IU may be associated with a file (e.g., a man page), a portion of a file (e.g., a C function or a database table), a set of files (e.g., a collection of related bitmaps), or any request for the retrieval of data from an external source (e.g., a database query).

what is the detail of the unit of information which should be considered to merit specific descriptive attention. “In the digital world, granularity of information may take on an entirely new and unfamiliar forms that are no longer linked to the granularity established by the author or publishers”⁵⁴.

Producers of geographical databases will find a descriptive format to describe not only their contents⁵⁵, but what is more important, to describe the potential uses to which the data in their databases can be put to. What really is lacking are search strategies for the data to mold one’s image of a specific or combination of (a) geospatial phenomeno(a)n or how to find the spatial data which can help us with decision support, prediction, forecasting and estimation of a specific spatial phenomenon or combinations of spatial phenomena.

For such processes metadata are inadequate as we are interested then in the inherent qualities of spatial data and their attributes, topology to or combination with other spatial data. This also independent of the geometric and geographic constraints a producer might have put to demarcate its database. And then we haven’t thought yet about the locational data, which may guide us to non-geospatial resources, which can help us to put a possible geospatial image into an overall context.

To imagine how innovative concepts interact with our traditional view of accessing data I have added Band V to the diagram. But a thick line from the other four bands separates this band. It is questionable whether it can be added to the diagram as the functioning which can be described therein might be incommensurate with the other bands and might belong to another paradigm. But for the sake of proposing innovative technologies I have added it anyway.

The next paragraphs will deal with the possible innovative technologies which might help us to find the ‘gems of information’ we are looking for.

Making locational data available

Locational information can be found in many more resources than only cartographic materials or geospatial databases. When we look at documents and databases concerning statistics, history, biology, ecology, travel, etc., many of

Leon Shklar, Leon et al. (1994). “*InfoHarness*” *Information Integration Platform*. <http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/shklar/shklar.html>

⁵⁴ Goodchild, Michael F. Ebenda. One can imagine that automated modules tag every distinct unit of information within a database with metadata which is taken from the structuring principles of this database, e.g. geometric, representational and/or qualitative attributes. When we come to knowledge discovery and datamining this might well be necessary, but it does not need manual input!

⁵⁵ For this purpose they can use the FI-standards.

these are concerned with a certain location on earth. In traditional library-catalogues these are mainly accessed by subject matter, but not according to geographical area. It is my opinion that a researcher is always searching for a body of interrelated information, which will provide knowledge for the subject he/she is working on. And we would like geographic information to be studied in context, which means textual and statistical resources should be available at the same time. Of course we can use the traditional geographical- and subject dictionaries and thesauri for this purpose, but why not use the unique mathematical properties of maps in an age where visual information and the possibility of aggregating information in graphical representations are becoming more and more predominant. When we can translate subject thesauri for locations (like Amsterdam, The Netherlands, U.S.A., etc.) through a conversion table to bounding-box or point co-ordinates the same interface might be used for directing users to these non-map resources.

The problem is best described in a paper by Ray R. Larson⁵⁶, which discusses geographic information retrieval (GIR) within the Alexandria Digital Library project for both map-like spatial objects as well as georeferenced material. "As with traditional print libraries, [...] information can be indexed and retrieved in a variety of ways, ranging from purely descriptive cataloguing of items in the database and topical analysis of content, to more specialized methods of classification and description that exploit the characteristics of digital information". It not only describes the background but also discusses some tools for automated geo-referencing of text.

Larson describes and summarises the GIPSY-model. GIPSY, The Geo-referenced Information Processing System, was developed as a new model of automatic geographic indexing for text documents. In the GIPSY model, words and phrases containing geographic place names or geographic characteristics are extracted from documents and used to provide evidence for probabilistic functions using elementary spatial reasoning and statistical methods to approximate the co-ordinates of the location being referenced in the text. The actual "index terms" assigned to a document are a set of co-ordinate polygons that describe an area on the Earth's surface in a standard geographical projection system. Woodruff and Plaunt describe the GIPSY method for automatic geo-referencing in detail⁵⁷. Later this method was evaluated, together with the POSTGRES method and the

⁵⁶ Larson, R. Ray (1996). *Geographic information retrieval and spatial browsing*. In: GIS and Libraries: Patrons, Maps and Spatial Information (ed. by Linda Smith and Myke Gluck), pp. 81-124 http://sherlock.berkeley.edu/geo_ir/PART1.html (preprint) [Accessed: 27 September 2000].

⁵⁷ Woodruff, A.G. & C. Plaunt (1994). *GIPSY: Geo-referenced Information Processing System*. In: Journal of the American Society for Information Science, 45, pp. 645-655.

TextTilling method in a paper for the Sequoia 2000 project⁵⁸ in the framework of the Global Change programme.

One of the next big projects within ADL is to take all the catalog records from the UCSB Library's online catalog that have geographic subject headings, then apply coordinates to each record, put the records in the ADL Catalog, and see if the resulting searches are as helpful to the users⁵⁹.

On the Internet several (proto)types of Geographic Information browser, which can perform this kind of function we are looking at, are available, e.g.

- The Environmental Resources Information Network [ERIN] Unit of Australia has developed a generic WWW map interface using a collection of simple map images and a standard lookup table to provide visual interactive access to geographically related information⁶⁰.

- MERI (Meadowlands Environmental Research Institute, New Jersey, U.S.A.) built a WWW-based interface to the database, integrating web server, database server and GIS server technologies. Through a map interface, a user can obtain a list of documents that studied a particular area. Conversely, through a text interface, a user can obtain a list of documents that report on, for example, a land use or cover type or a specific water body, with sampling/analysis locations displayed on a map⁶¹.

An atlas as a geographical interface

As an example I would like to sketch a project that was drafted in co-operation with some university departments of cartography in The Netherlands. The idea is to have maps function not only as geographical information sources, but research whether it is possible to have the same maps function as visual interface for metadata-databases. For this project we looked at three possible databases, which could function together within this frame.

⁵⁸ Larson, Ray R. et al. (1995?). *The Sequoia 2000 Electronic Repository*. In: Digital Technical Journal. 7(1995)3, pp. 50-65. <http://www.digital.com/DTJJ04/DTJJ04SC.TXT> (preprint) [Accessed: 28 September 2000].

⁵⁹ E-mail communication by Mary Larsgaard, Map and Imagery Laboratory, Davidson Library, University of California, Santa Barbara (13 February 2001).

⁶⁰ Crossley, David & Tony Boston (1995). *A generic map interface to query geographic information using the World Wide Web*. <http://www.csu.edu.au/special/conference/apwww95/papers95/dcrossle/dcrossle.html> [Accessed: 24 September 2000].

⁶¹ Barrett, Kirk R. , Richard Holowczak & Francisco J. Artigas (1999). *A database of environmental documents about an Urban Estuary, with a WWW-based, geographic interface*. <http://www.awra.org/proceedings/www99/w21/> [Accessed: 27 September 2000].

1. NCGI⁶²: This metadata-database of the National Clearinghouse Geo-Information contains presently more than 1,500 descriptions of digital spatial datasets with visual examples of some 17 producers of geospatial data. The datasets range from several 100s Mb to 10s of Gb.
2. AvN: The second analogue edition of the Scientific Atlas of The Netherlands (AvN, published 1984-1990) contained some 1,000 maps, which give a comprehensive view of the socio-economic, physical and ecological situation of The Netherlands. These maps are now digitised⁶³, but give not up to date information.
3. There is a database of some 30-40,000 descriptions of cartographic documents concerning The Netherlands within the framework of the Dutch Depository collection in the Royal Library. The Royal Library phased out the CCK-system⁶⁴ in 1999 and will convert these descriptions between 2001-2002 to the PICA⁶⁵-database.

The NCGI concerns mainly large-scale datasets with a high economic value. The owners of these datasets, which are usually (semi)governmental or academic bodies, are usually forced to recoup part or most of the costs involved in producing these data. At the same time these producers aggregate these data into the middle- and small-scale data necessary to create the maps used in the Scientific Atlas of The Netherlands. In the age of analogue products they did this service for free. But in the digital age it is hard to get the same services from them. This means we must find a modus to entice them in such a way that they are willing to create and provide these data for free. We think we can reach this goal by offering the maps of the Scientific Atlas of The Netherlands as a visual geographical interface for the NCGI.

The geographical interface would provide several search strategies:

1. When clicking on a map of The Netherlands one can search the underlying metadata-databases for datasets that also cover the whole of The Netherlands.
2. When making a cut-out of a part of The Netherlands one can search the underlying metadata-databases for datasets that also cover that part of The Netherlands.
3. When one zooms in the interface will query the underlying metadata-databases for datasets, which cover that specific part of The Netherlands.

⁶² NCGI (2000). Geo-gids. <http://www.ncgi.nl/Profiel/profiel.html> [Accessed: 24 September 2000].

⁶³ Disciplinegroep Kartografie (2000). *Atlas van Nederland*. <http://avn.geog.uu.nl/> [Accessed: 24 September 2000]

⁶⁴ VELDEN, G.J.K.M., P.J.M. Douma and J.G. Zandstra (1990): *CCK : making cartographic materials accessible*. In: The LIBER Quarterly 2(1992)2, pp. 192-208. <http://www.kb.nl/infolev/liber/articles/cck.htm> [Accessed: 24 September 2000].

⁶⁵ PICA (2000). *About Pica*. <http://www.pica.nl/en/about/> [Accessed: 24 September 2000].

The queries will be effectuated by mainly using bounding-box co-ordinates. There must be, however, also possibilities for searching on point locations, within a radius of a point, a bounding polygon, etc., and when the computer can read the index-map also on named area, administrative sub-division or locality description, when necessary using a thesaurus. But it must be also possible to use traditional geographic and thematic dictionaries to come to the same result.

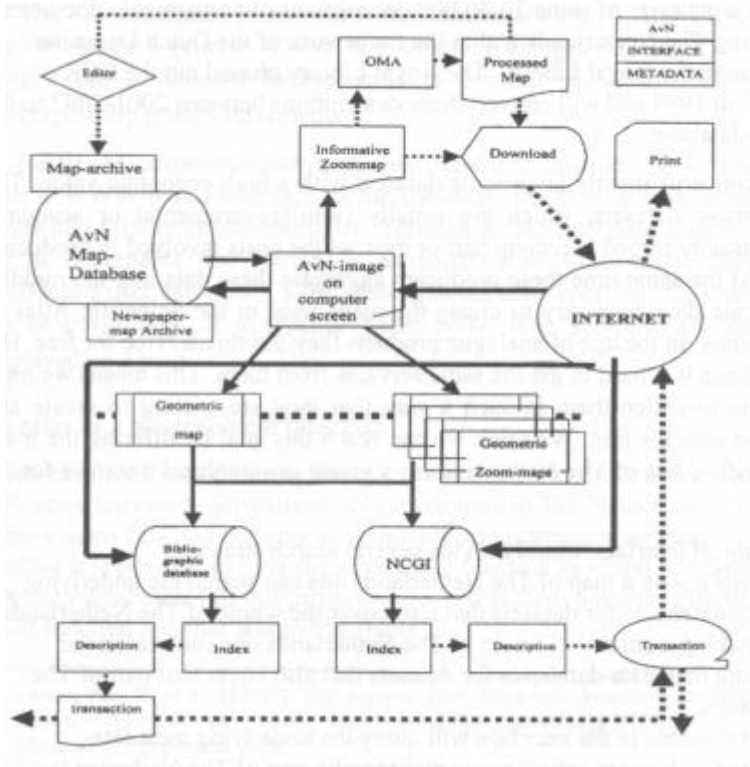


Diagram of a visual geographic interface

We try to sell the idea to the producers as follows. Is there a better way than the one described above to advertise the economically interesting datasets through such an interface and at the same time service education and the general cause! Because the Scientific Atlas of The Netherlands will not only function as a visual geographical interface, but at the same time will be an informational and

educational resource. The educational functions will be enhanced and become interactive when an Online Mapping Application (OMA) will be incorporated with which users can manipulate certain aspects of the maps offered. Maps produced with this application can be selected by an editor and added to an archive to serve as examples for future users.

At the same time some scanned samples of newspaper-maps have been added to the present database to get or give some insights in what way mass media use scientific mapping data to inform the public in general. In future digital maps from newspaper-archives could be downloaded to a sub database of the Scientific Atlas of The Netherlands and have the same function as maps added from the OMA-activities.

But why should we restrict ourselves to spatial metadata-databases for digital materials and not try to include metadata-databases for analogue materials. As long as the descriptions include geographical bounding-box co-ordinates the same kind of queries can be made on bibliographic databases as can be made on the NCGI-database. Because researchers require disparate sources to do their research, and because not all necessary information will be available on the Internet or in ready digital form, it would be a miss when bibliographical databases are left out of this scheme. The 30-40,000 descriptions of maps, which will be loaded into the PICA-database, contain all the necessary geometric or mathematical data necessary to use the query-functions as envisaged with the NCGI.

Furthermore the idea of one-stop-shop information gathering is so prevalent that we must do our utmost to realise this concept with the tools and all the data available.

Conclusion

In my opinion the metadata field will be expanding exponentially in contrast to the wish for basal descriptions as heralded in IFLA's *FRBR*⁶⁶. But the burden of this growth will be put on the shoulders of the creators and the software programmers. Furthermore metadata will be used for a spectrum of functions, ranging from access to preservation. Once a metadata-record has been created it will be constantly added to depending on the need for transformation to keep a DLO or database available for use. Cataloguers will become metadata information system managers and keep themselves busy with checking and validating the input. We also will create innovative user-interfaces which will replace the traditional listings. The most uncertain part of this set up is how the growth of databases will relate to the

⁶⁶ IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional requirements for Bibliographic Records* (UBCIM publications - new series, Vol. 19). Also available through <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>> [Accessed: 24 September 2000]

growth of DLOs, especially in the geospatial field. And in how far knowledge discovery and data mining, which are autonomous processes related to databases, will replace the need for metadata, except as internal checks and validations.

Jan Smits
Koninklijke Bibliotheek, National Library of The
Netherlands
P.O. Box 90.407
NL-2509 LK Den Haag
The Netherlands
tel: +31 70 3140241
E-mail: jan.smits@kb.nl

LIMITS TO ACCESS YOU CAN LOOK BUT DON'T TOUCH*

By David Cobb

Improving access to materials should be one of the goals for all librarians and all libraries. Generally speaking there is little disagreement related to this goal. Many collection development policies eloquently refer to open access and, in the United States, official U.S. government depository libraries are required to provide access to government documents and to provide public access to those collections. And yet collectors, however well meaning, will realize upon further study that there are limits to access and map collections are no exceptions. It is the larger collections that may have more restrictions for it is often their sheer size that creates limits. While most everyone believes that our new digital age and the concept of the shrinking world will open access to all materials, digital access is actually increasing limits to access as well. I will use this essay to outline these limits. Some are obvious and all types and sizes of map collections suffer from some of these limits.

The most obvious of these are library hours. Many map collections are often one person operations, sometimes only part-time, and this will limit access to professional reference service. Many other map collections limit user access to the map file drawers for the purpose of preservation and the safety of users and material and I believe this is a good policy but then requires more staffing. Those map collections that are associated with Government Document units, at least in the United States, usually have more access hours into the evening and on weekends. The more specialized collections often have no evening hours nor weekend access.

A more serious limitation is often the fragile nature of maps. As collections grow, and age, the fragility of the format becomes more apparent. The fact that maps are not protected by bindings or boxes make them an inherently weak format. The ravages of time, a poor environment, and handling are serious threats to a single sheet of paper. In addition, few map collections have either preservation funds or staffing and certain maps may have to be withdrawn from public access. While digital reformatting is a possibility some of our maps are in such a deteriorated

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

state that they require conservation to be readable. There are also several issues related to digital reformatting which I will address later in this paper.

Next is the level of cataloging. According to the 1990 *Guide to U.S. Map Resources*¹ 60% of the collections surveyed were cataloged. In reality, this meant that some collections were being cataloged and not that 60% of the collections had complete cataloged collections. A closer examination reveals that many research libraries only have single digit percentages of maps actually cataloged although several have classified larger percentages of their collections. Again, this classification is almost always internal, not represented in the library's online catalogs, and seldom includes newly acquired digital information. Harvard's Map Collection may be a typical example. Prior to 1990 Harvard's collection was almost completely classified although it should be noted that there were three different classifications: one for books, one for antiquarian maps, and another for current maps. All of these classifications were internal meaning that a potential researcher was required to physically visit the Map Collection, review the card catalog, and then request a specific item. While progress and access has improved I must admit that it remains a part-time effort and it will be sometime before we are able to provide online access to nearly a half million cartographic items. The books and atlases have been reclassified and are now in our online catalog using the Library of Congress classification; new maps added to the collection are cataloged using OCLC and the LC classification; but only a few of the retrospective materials will be found in the online catalog.

Then there is the issue of access and cataloging of digital information. Like many libraries we collect published CDs and various aggregates of unpublished digital geospatial information. For example, the *Atlas of Florida*² was published as a CD in 1994 by Florida State University and the Florida Department of Education. It is included in our online catalog under its appropriate title and author, however, its one subject heading is "Florida—Maps." This CD has numerous data layers showing: camping sites, coral reefs, crime rates and many others but we are limited in our ability to archive and access digital information in library catalogs. The importance of a digital data set is, not the sum of its parts, but the individual parts themselves and knowledge of these "layers" is critical to the use of such geospatial information.

Several libraries are reviewing and experimenting with Federal Geographic Data Committee (FGDC) standards for metadata and their compatibility with library catalogs. A similar project has been underway at Harvard and the Harvard

¹ Guide to U.S. Map Resources. Ed. By David A. Cobb et. Al. Chicago: American Library Association, 1990.

² Atlas of Florida [CD-ROM]. [Atlanta?] EduQuest, [1994?]

Geospatial Library will be launched in 2001. This service will develop separate records for geospatial information: an LC MARC record for the public online catalog and a FGDC metadata record within the Geospatial Library which will provide access to the different layers. A few datalayers will be restricted to Harvard users only, due to licensing agreements, but the goal is to offer access to the system via the Web. Initially, the system will offer access to the Digital Chart of the World, ESRI Maps and Data, Massachusetts GIS Data, Boston Water & Sewer layers, the City of Cambridge GIS data, and parts of the Bureau of the Census Tiger 2000 file. Librarians from the Social Sciences and Earth & Planetary Sciences will receive training on the use of the system thereby creating several points of expertise for faculty and students. Limitations include the fact that not all geospatial digital data will be included in this exploratory environment and there will be a limited number of records relating to paper resources. Types of layers to be considered for future addition to the geospatial library include raster imagery, orthophotoquads, additional state and city data, and foreign census information.

Digital information brings several issues to the forefront that libraries should be carefully considering. The first question is how will copyrighted materials be treated? In the past we have felt comfortable placing signs near our copying machines explaining the copyright laws and the concept of fair use. Today, as map collections begin to offer scanning services, including scanning and processing of large oversized maps, library staff are actually producing the digital files. As a result, the Harvard Map Collection will use a newly acquired large format color scanner to offer only scan to print services for the during 2001. Our intent is to review the issues with our Office of the General Counsel and to develop forms to be signed by the user specifically stating (limiting?) how the digital image will be used.

The final two limitations are points of contention and issues for many researchers – working with digital or paper resources. During this past year a heated discussion took place on the MapHist listserv concerning intellectual property rights and permission fees. While an Ortelius map of the America's from 1570 is no longer copyrighted, a library has the option to invoke its intellectual property rights. In other words, the Ortelius map is owned by the individual library which may, or may not, decide whether the map can be copied and restrict how it may be used. For example, as stated above, we will use our new large format scanner only to make printed copies of the maps we scan. Some scholars have criticized this as holding the individual maps as hostages and that they belong to the public and should be available to anyone. I disagree. Too often, researchers conveniently forget the time, effort, and cost that libraries assume to maintain these national treasures. Those research libraries holding these special collections have often

invested enormous quantities of staff time to care for these materials, cataloging many of them, preserving others, and providing adequate housing. The limitations are often defined to make sure that the item remains safe in the collections for future generations and that the items are copied within appropriate environmental conditions and by professionally trained staff.

These conditions and staff come at a price. It is interesting how everyone expects to be charged by a copying store but resents any charges for a similar service at a library. Most research libraries have a special collections imaging services or a photoreproduction department. Almost always, there are charges associated with this service. I suspect that some libraries charge more than a cost recovery fee and I know that some do so simply to restrict that service because it would overwhelm their staff. In addition, we often research and select materials for imaging via email, regular mail, and telephone. This takes our staff away from other primary duties but, at the same time, makes our collections available to a wider diversity of users around the world. Then the items must be transported to our Imaging Services where they then have to be handled again: placing large maps against the vacuum board, opening folio atlases, and all the time doing light meter tests. The entire process from researching the collection to identify a specific image for the user to creating an 8 x 10 black and white print or a color transparency is completed by professionals paid by the library. Libraries charge permission fees to offset some of these costs and seldom enough to make these services profitable. As we move deeper into our offering of digital services it is probable that libraries will have to charge for customized scanning or for parceling data sets and users should be prepared for such limitations.

Libraries are committed to providing access to their collections but access no longer just means providing the right book for the user. The future will bring more customized services especially as libraries include more geospatial information that is not familiar to a large group of their users. I believe that libraries will increasingly expand their offerings at no charge via the Web but will have to charge nominal fee's to offset the customized reference service that geographical information services may require. I also believe these fees and access restrictions may be limiting but will be offset by the final reward.

David Cobb
Harvard Map Collection
Harvard College Library
Cambridge, MA 02138
cobb@fas.harvard.edu
617-495-241

DIGITIZATION OF CARTOGRAPHIC MATERIALS : NATIONAL ARCHIVES OF CANADA*

By Betty Kidd

Part 1: Context

In the catalogue for the 1972 exhibition, *Archives: Mirror of Canada Past*, which marked the centenary of the Public Archives of Canada, the then Dominion Archivist, Dr. Wilfred I. Smith, noted “The Public Archives of Canada can report some amazing achievements in the last century, but development of public awareness of its role, the encouragement of popular participation in what should be a co-operative mission, and the sharing of the benefits and the increased enjoyment of its treasures must be a challenge for the next century.”¹ In 1972, few could have anticipated the rapid developments in technology which would, within several decades, provide a tool – the World Wide Web – to enable the realization of Dr. Smith’s stated objectives.

On 1 December 1995, the National Archives of Canada launched its World Wide Web site www.archives.ca. Launched at a time that the Canadian public service was undergoing one of the most massive budget cuts in its history – for the National Archives, a decrease of 26% over a three-year period, the site’s content initially featured information concerning the department and descriptions of some holdings but few images or documents.

At the same time within the Canadian federal bureaucracy, there was a growing determination to use technology to offer government services efficiently to the Canadian public. Government programs were being introduced to ensure that this happened and the digitization plans of the National Archives, as a department, were directly impacted. Government On-Line (GOL) (www.gol-ged.gc.ca) is the Government of Canada initiative to enable Canadians to access federal information and services on-line with the expressed purpose to become the world’s most electronically connected government to its citizens by 2004. An integral part of the government’s Connecting Canadians (www.connect.gc.ca) initiative was the Industry Canada program entitled SchoolNet (www.schoolnet.ca) which had the

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

¹ *Archives: Mirror of Canada past/Miroir du passé du Canada*. Toronto: University of Toronto Press for the Public Archives of Canada. p. 2.

mandate to connect Canadian schools and libraries to the Internet by 31 March 1999. In 1996, the Canada Digital Collections (CDC) program (www.collections.ic.gc.ca), also an Industry Canada initiative, was commenced, with a mandate to assist organizations across Canada “to build up a stock of information on Canada for the Information Highway”. SchoolNet and CDC provided Youth Employment Strategy funding for the hiring of young people to undertake digitization contracts. The National Archives quickly became a partner in these initiatives; later in this paper, some of the National Archives’ CDC projects will be discussed.

The actual push to accelerate digitization in the National Archives itself came in the Throne Speech in the House of Commons in October 1999 which promised “to bring culture into the digital age” and noted that the National Archives’ holdings would be digitized and in the Budget Speech of 29 February 2000 which allocated \$75 million to “enhancing Canadian cultural material on the Internet” (note that these monies were to be allocated to numerous federal and other cultural agencies, including the National Archives). Financially, this meant that in the 2000-01 fiscal year, dedicated funds would be available for the first time specifically for digitization and would not need to be “borrowed” from operational funds for other departmental activities. The catch was that the money would not reach departmental coffers until at least mid-year and would have to be expended by the end of the fiscal year – i.e., 31 March 2001. The management decision was to extend a “line of credit”, using operational funds, and to plunge quickly into digitization initiatives. In retrospect and although much was accomplished, certain basic planning steps were skipped, other core departmental programs were negatively impacted and total staff buy-in did not result. In the current year, efforts to alleviate some of these negative aspects are necessary.

In addition, the National Archives, as a federal department, must take into account other federal legislation. Official Languages Act has had a major impact; since Canada is a bilingual nation, this Act must be observed in the presentation of all information, in both French and English, on the Internet. In the planning process, these requirements are an important factor in time estimates and in financial costing. Another is a recently introduced government “Standard for the Common Look and Feel for the Internet” which is necessitating the overhaul of many departmental web sites; undoubtedly, the National Archives site, which has undergone major changes in the past year, will continue to evolve.

Part 2: www.archives.ca

The images on the National Archives Web site reflect the current departmental multi-media approach to exhibitions and other outreach products. Most projects are by subject or theme and encompass all media forms. Unlike the situation for

specific cartographic projects on www.archives.ca, nor will he/she find, except for one virtual exhibition, direct access to documents by media type. Even for this one exception, the term “maps” is unfortunately and mistakenly used for architectural and engineering records as well as for cartographic documents.

The three projects on the NA Web site in which one will find most cartographic images are the virtual version of the exhibition, *Canada at Scale: Maps of our History*, the virtual multi-media *Living Memory* exhibition, and the *Tracing the History of New France* project. As well, a number of “partner” exhibitions are listed under “Canadian Memory Virtual Exhibitions” and are available with a click of the mouse – these include the virtual versions of *Every Name Tells a Story: 100 Years of Official Place Naming in Canada* and *The George Back Collection from the National Archives of Canada*. These projects will be further described as will a project on *Western Settlement* currently underway which will include many maps. The total number of cartographic items available on the National Archives web site at the time this paper was being prepared is less than 100; if one includes the “partner” exhibitions, the number increases to closer to 150.

a) *Canada at Scale: Maps of our History*

This exhibition that was prepared as part of the National Archives’ contribution to ICA’99- the International Cartographic Association conference in Ottawa in August 1999 – was opened in traditional format and subsequently in virtual format. Featuring some 77 archival documents – mostly maps and other cartographic materials, the exhibition concentrated on two thematic areas – the first, the exploration, colonization and development periods and the second, government cartography. Access to the virtual exhibition is provided both through “Exhibitions” and “Places” on the NA web site, and to the individual items, through the “Index” or through the thematic areas noted above. Full descriptive bibliographic entries (AACR2 format) are provided for each item in addition to descriptive text. It is possible to enlarge the small reproduction (the thumbnail) of the item found with the text and bibliographic entry by clicking on it; however, it is not possible to enlarge any portion further nor is any technical information – e.g. number of pixels – indicated.

As would be expected by the thematic titles, many of the “treasures” and significant maps of Canada by well-known explorers and cartographers – including Ptolemy, Zaltieri, James Cook, Champlain (see *Illustration 1*), Joseph Des Barres, Nicolas Bellin, etc. – are featured in this exhibition as are key maps, charts and plans produced by the Canadian government since Confederation in 1867.

Illustration I: From: *Canada at Scale: Maps of Our History Champlain 1632*



b Living Memory

This virtual exhibition of 145 items is based on *Treasured Memories*, a permanent traditional exhibition, the first version of which opened in May 1997 to coincide with the 125th anniversary of the National Archives. The items in the actual exhibition continually change, both for conservation reasons and to make a wider variety of holdings accessible to the public. At the present time, digitized copies of 15 “maps” are featured in the *Living Memory* virtual exhibition.

Access to the documents in *Living Memory* is through theme, time period and media type; this is the only part of the NA web site where the user can actually find maps under the heading “maps”, although the term “maps” is broadened to include architectural and technical drawings as well as cartographic items. When “maps” is clicked, the user is taken to a page with a brief description of the NA’s cartographic holdings – i.e., “more than two million maps, charts, atlases, globes... range from Ptolemy’s atlas of 1490 to the most recent maps published by government and private industry...” and thumbnail sketches of the 15 “maps”. Clicking on the thumbnail sketch will provide a larger image and descriptive text and for several maps, links to the ArchiviaNet bibliographic entry; for an enlargement, one clicks on the image of the magnifying glass. The enlargement includes the number of pixels vertically and horizontally.

To understand this disparate selection of images, it is necessary to know that the exhibition themes are First Peoples (1), New France (2), Newcomers (6), Women (1), War (1), Politics and Government (1) and Arts and Culture (4). The numbers in brackets are the number of “map” images appearing in each thematic area. The same image – the 1822 map by Lligliuk, an Inuk woman (*see Illustration II*) – appears in both “First Peoples” and “Women”.

Illustration II: From: *Living Memory*; Lligliuk, 1822



c) *Tracing the History of New France*

This multi-media project completed in the last year is subdivided into 8 thematic areas and provides access into a bibliographic database for textual records of the New France era. Six of the eight map images in this exhibition appear under “Land”, the first theme; these include maps by Champlain, Bellin, Coronelli (see *Illustration III*), De L’Isle and de Hondt. The other map images are located in the “Seigneurial Regime” and the “Wars” themes. In addition, images of the well-known beaver and cod fishing cartouches or vignettes on the de Fer map of 1698 appear in the “Economy” theme.

Illustration III: From: *Tracing the History of New France* Coronelli, 1688



As for the previous exhibits, the images can be enlarged by clicking on the small image that appears with a brief note. Also, one can click on the “Description of the item” box to read the ArchiviaNet description – this archival data base clearly indicates the link to the provenance of the particular map, - for example, the W. H.

Coverdale collection of Canadiana and the Alexander E. MacDonald fonds. If unknown, the map is designated as a “Single Item.” Linking the descriptions to the database is now the norm for all departmental digitization projects.

For those interested in maps of New France, one should also view the National Archives’ cartographic images in the Virtual Museum of New France on the Canadian Museum of Civilization web site, which will be discussed in the next section of this paper.

d) Every Name Tells a Story: 100 Years of Official Place Naming in Canada

As a federal member of the Geographic Names Board of Canada (formerly the Canadian Permanent Committee on Geographical Names), the National Archives cooperated with Natural Resources Canada and Parks Canada to plan the celebrations in 1997 of the centenary of official place naming in Canada. The actual exhibition, jointly prepared by the National Archives and Parks Canada, was shown at the National Archives although this virtual version resides on the Natural Resources Canada web site (<http://GeoNames.NRCan.GC.Ca/cent/english>)

The exhibition designed for the general viewer, set up as six streets with appropriate signage, used a wide range of archival and current records, including nine maps, as well as “The Carnival” area where questions were located to test the viewer’s knowledge after he/she has visited the other parts of the exhibition. The virtual version replicates the actual exhibition. For quick reference to the digitized items, one can go directly to the “Index of Images” which provides only the location of the original document, a reference number and the number of kilobytes but it is recommended that each of the street areas – e.g. “Names from Explorers” and “International Names” – be visited since much more detailed information is there available.

e) The George Back Collection from the National Archives of Canada

In 1994, the National Archives acquired several sketch books of George Back (1795-1878) with the assistance of Hoechst and Celanese Canada and with a grant from the Department of Canadian Heritage under the Cultural Property Export and Import Act. George Back had created a remarkable record of his expeditions to the Arctic region from 1818 until 1837. His sketches (including sketch maps), drawings and writing document the extraordinary Arctic landscape and topography, its people, flora and fauna. His career was also intertwined with John Franklin’s with whom he served on two expeditions, first from 1819 to 1822 and again from 1825 to 1827. Added to earlier National Archives’ accessions of Back’s graphic materials from 1921, 1955, 1979, 1981 and 1986, these sketch books provided a most attractive body of material when Industry Canada began

identifying material for the Canada Digital Collections (CDC) program. Digitized by the Grade Ten class at Echo Dene School, Fort Liard, Northwest Territories, the collection is located on the Industry Canada website at <http://collections.ic.gc.ca/back/>. The documents are arranged geographically by province and territory and there is a detailed descriptive entry for each. The approximately 30 sketch maps included are to date the largest number of National Archives' manuscript maps available on the Internet (*see Illustration IV*).

Illustration IV: From: *The George Back Collection, National Archives of Canada 1820*



f) *Western Settlement*

Scheduled to be launched in the near future, this exhibition has been in the planning stage throughout the current year. The following quote from the draft project plan provides the context of the current planning:

The settlement of western Canada can be broadly described as the movement of an intrusive European-based population onto an indigenous landscape. This landscape was then transformed to suit the needs of the invading society. Defined as such, western settlement has two broad themes: people and land.²

A multi-media project, it is anticipated that the cartographic component will include numerous maps and plans of Indian Reserves and settlements (based on several publications of the former National Map Collection); examples of fire insurance plans and bird's eye views of western cities and representative sheets of the Three-Mile Sectional Map series. As for other projects to date, existing 105mm fiche will be scanned for this project.

² Jeffery Murray, "Western Settlement: Draft Content Map", December 2000, p. 2.

Part 3: NA Cartographic Material on Other Web Sites

In addition to the presence of cartographic material on the National Archives' web site and to the links to "partner" projects there provided, there are many National Archives maps and plans to be found on a number of other Government of Canada web sites and on university and several private sector sites. A few examples follow:

a) GeoGratis

The largest volume of National Archives cartographic/geomatic records available on the Internet is located on the GeoGratis (<http://geogratias.cgdi.gc.ca>) web site of Natural Resources Canada.

In 1995, the records of the Canada Land Inventory (CLI) program – data and intellectual property – were transferred to the National Archives. Copies of the printed sheets – land capability for agriculture, forestry, wildlife (waterfowl and ungulates) and recreation – had been acquired by the National Archives throughout the program's history starting in 1963. At the time of massive government downsizing in the early 1990s, the CLI with its huge digital data base of over 30 000 map sheet equivalents could not be maintained and was officially discontinued. The National Archives was most eager to acquire these records since the Canadian Geographic Information System (CGIS) associated with the CLI was the world's first fully operational GIS and certainly changed the nature of mapping world-wide. Almost immediately, an ad hoc group representing government and private sector, the "Friends of the CLI", began to explore ways to ensure continuing access to the huge CGIS/CLI data base. Cooperation at the grass roots level in the National Archives, Statistics Canada, Environment Canada, Agriculture and Agri-Food Canada and Natural Resources Canada resulted in the transformation of all the CLI maps into standard data exchange formats. Providing this data on the GeoGratis web site has allowed downloading of digital GIS files for professional and technical audiences. Work continues on similar projects.

b) In Search of Your Canadian Past: The Canadian County Atlas Digital Project

Canada's Digital Collections (CDC) Program, discussed previously, has funded some 400 projects for numerous institutions across Canada, including the National Archives. In addition, in some cases, material from the National Archives has been used to supplement or complete projects undertaken by other institutions. A cartographic example is McGill University's "In Search of Your Canadian Past: The Canadian County Atlas Digital Project (<http://collections.ic.gc.ca/atlas/>) , a searchable site which provides access to land owners' names and to the township maps in the approximately 40 county atlases published in Canada between 1874

and 1881. The Rare Books Division of McGill University holds copies of most of these published atlases but asked the National Archives to provide access to those not in their holdings. The National Archives was very pleased to cooperate since this site is most valuable to genealogists and updates the Archives' effort in 1970 to make these atlases accessible with the publication, *County Atlases of Canada: A Descriptive Catalogue*.

c) *New France Virtual Museum*

As stated on this section of the Canadian Museum of Civilization web site at <http://www.vmnf.civilization.ca>, the virtual museum "brings together as never before traces of New France from wherever they are found around the world." The maps from the National Archives can be accessed through the "Exhibitions" area of the Index. Under the section title "An Overview of Cartography", the maps with brief narrative text, are subdivided under the headings "World", "Americas", "New France" and "Cities" and "illustrate two major elements in the history of geography: the progress of cartography during these centuries and the increasing knowledge of the north american territory, due essentially to the explorations." As well, three additional NA maps of New France can be located by clicking on the word "Maps" or the symbol.

d) *Ottawa in Maps*

On the Carleton University web site (<http://www.library.carleton.ca/madgic/maps/ottawa>), this is the electronic version of the 1974 National Archives publication *Ottawa in Maps: A brief cartographical history of Ottawa 1825-1973/Ottawa par les cartes: Brève histoire cartographique de la ville d'Ottawa 1825-1973* by Thomas Nagy. The project is a joint digital publication of the University of Ottawa Map Collection, the Carleton University Maps, Data and Government Information Centre, and the National Archives of Canada. The images have been created by scanning the 105 mm microfiche of the original maps and utilizing the image compression software, Mr. Sid (Multiresolution Seamless Image Database) that allows zoom-in display. Thumbnails of the maps are shown by date and by clicking, the description of the item is presented and the larger image appears on the computer screen (see *Illustration V*).

Illustration V: From *Ottawa in Maps*; Brosius, 1876



e) *The Atlantic Neptune Online*

The first detailed hydrographic surveying of the eastern seaboard of North America, known as the *Atlantic Neptune*, was the work of Joseph Frederick Wallet DesBarres (1721-1824) and a team of surveyors. The National Archives has extensive holdings of these charts as well as several of the copper plates from which the prints were made.

Atlantic Neptune Online (<http://mercator.cogs.nsc.ns.ca/neptune.html>) owes its existence in large part to the research interests of one of the instructors at the Centre of Geographic Sciences, Nova Scotia Community College (Annapolis Valley Campus). Located on the College's web site, geographic indexes to the charts and a listing of 108 charts with the titles, Stevens numbers and the National Archives' 105mm microfiche numbers (NMC numbers) are provided. Digitized versions of 24 of the charts are available with three zoom levels.

f) *Canadian Heritage Gallery*

Certainly most attractive and well done and despite its name, not to be confused with the federal government department Canadian Heritage, this site is "the most extensive collection of historical Canadiana on the Internet." The Gallery's mission is "to promote the accessibility, knowledge and research of Canadian Heritage by publishing via the Internet an extensive collection of historical photos, original documents, Canadian artwork, maps, and illustrations, fully documented and researched for historical authenticity, and presented in a "user friendly" format; and to advance dissemination of materials and information pertaining to the Heritage of Canada by providing printed or digital reproductions of selected images and caption information, as published on this Site." (<http://www.canadianheritage.org>). The majority of the images on the site are from the National Archives, including

58 maps and plans. The images are offered for sale at prices starting at \$42.00 Cdn for a 5" x 7" print or a 1,000 px X 1,400 px JPEG download or CD-ROM TIFF.

Part 4: The Future

To date, the National Archives digitization program has, in large part, reacted to outside pressures and to the availability of funded initiatives. With more continuity now ensured for an ongoing program, undoubtedly the National Archives will increase the digitization of its holdings in coming years, although only a small percentage – perhaps 2% - will ever actually be available on the Internet. At this time, a more stable organizational structure of the digitization area is being planned. Realization of the impact of digitization on core ongoing programs in recent years, especially in 2000-2001, and corrective action will result in increased staff commitment. Certainly, all staff are eager to see “their” records made accessible to interested parties through the Internet.

The specific requirements of documents in various media forms must continue to be addressed. A non-cartographic example of such thought is the planned philatelic or Canadian Postal Archives site to be launched in September 2001. A cartographic example currently underway is the search for a solution to the lack of appropriate sized scanners for large or “oversized” documents. Although the 105 mm microfiche have been utilized to some extent for cartographic and architectural documents, there are problems in that most of the microfilming to date has been in black and white and consistency of resolution varies depending on the camera used and the timing of the filming. Traditional filming has also been used. This will change in 2002-03 when the department is scheduled to purchase a large scanner for oversized documents. One of the scanners being seriously considered is a KartoScan flatbed scanner for Kongsberg Scanners of Sweden. This equipment was first brought to the attention of National Archives staff at a display of the company’s products at the International Cartographic Association conference in Ottawa in August of 1999. Departmental staff members later met with a company representative in Ottawa and the head of the copying area visited Sweden to discuss technical requirements. The funding for the purchase was confirmed in April 2001 when the department was advised that additional funding would be available for a two year period as part of the “Program Integrity” grants to ensure that the Canadian public service was up-to-date technologically. Identified as a solution for cartographic and architectural material, other large-size documents, such as Indian treaties will also benefit as happened with the 105mm microfilming program introduced in the 1970s.

Cartographic staff have also been identifying the need for a “zoom-in” technology for digitized cartographic materials and in fact, have successfully experimented with Mr. Sid in their portion of the departmental Intranet site. As well, the co-operative *Ottawa in Maps* project employed Mr. Sid. This technology – in particular, the uses made by the Library of Congress – needs to be seriously studied by digitization staff. This type of technology would certainly be beneficial to on-line users of the National Archives’ web site – and incidentally would also be useful for other types of documents in addition to maps.

That maps are scientific documents based on complex mathematical projections needs to be recognized and methodology introduced to ensure that the scale and original dimensions can be reconstituted. A lesson could easily be learned from the 105mm microfilming program where throughout the program’s history, a ruler has been filmed with the map or plan.

In a five-week period in December 2000 – January 2001, users of the NA web site were asked to provide input on digitization projects proposed by NA staff for 2001; some 2900 responses were received. Although the survey was undoubtedly useful, future surveys need to ensure the participation of those groups that have not been regularly using the site; certainly, for cartographic researchers, there are many other sites available with more information on cartography than is found on the National Archives site, and thus, they do not tend to visit www.archives.ca. Unfortunately, not even such groups as the Association of Canadian Map Libraries and Archives and the Canadian Cartographic Association were informed of the survey or asked to participate. Of the 19 projects listed, only two were specifically cartographic in nature and in terms of the voting, “Bird’s Eye Views of Canadian Cities” placed 9th and “Fire Insurance Plans”, 11th. There were also others which would have a high cartographic content – including “Western Settlement” (4th), “Battles of the First World War”(8th) and The North (18th). Not surprisingly, since genealogists constitute a high percentage of the users of the National Archives, the top-ranked projects were those which genealogists would find most useful.

The National Archives needs to look back at its long publishing history of specialized finding aids and catalogues and to seriously consider combining the existing text and descriptions with scanned images of the archival material described. Not only would less staff time be required – although, of course, material received since the publication date could be added – but many useful publications which have been out of print for many years would again be accessible with the added benefits of scanned images and electronic searching capabilities. For cartography, these 1970s publications could be considered: *Index*

to *Township Plans of the Canadian West* (1974); *Winnipeg in Maps* (1975); *County Maps: Land Ownership Maps of Canada in the 19th Century* (1976); *Arctic Images: The Dawn of Arctic Cartography* (1977); *Fire Insurance Plans* (1977) – note extensive additions would need to be added; and *The Riel Rebellions: A Cartographic History* (1979). Several other publications from this period have been or will be used in other ways, including *Ottawa in Maps* (1974) now available on the Carleton University web site, *County Atlases of Canada: A Descriptive Catalogue* (1970), of which much of the information is now available in McGill University’s “In Search of Your Canadian Past: The Canadian County Atlas Digital Project”, and *Maps of Indian Reserves and Settlements* (1980-81) which has been expanded and will be used in the forthcoming *Western Settlement* project. The unpublished listing of trench maps would be of interest to many. And would historians of cartography not be overjoyed with an electronic version of *Sixteenth-century Maps relating to Canada: A check list and bibliography* (1956)?

Continuation of partnering with both federal government and other organizations is vital for the success of the digitization program. As an example, the new Images Canada site (www.imagescanada.ca) is hosted by the National Library of Canada and supported financially by Canadian Heritage as part of the Canadian Digital Cultural Content Initiative. This initiative provides numerous opportunities for increasing the visibility of visual archival materials and the National Archives is committed to participation in the future. As well, there have been and will be joint projects with or participation in projects of other cultural agencies within Canadian Heritage, including the National Library and Canadian Museum of Civilization and with other departments and crown agencies such as Canada Post and Natural Resources Canada.

Closely related to the digitization of images is the necessity to provide accurate and detailed descriptions of and the appropriate contextual information for the digitized items. Thus, the National Archives’ commitment to making finding aids available in electronic form – whether or not the items are digitized, to continue to build ArchiviaNet and to be a major contributor to CAIN (the Canadian Archival Information Network) must be supported and adequately funded. The National Archives’ web site does acknowledge in the ArchiviaNet section that for cartographic and architectural documents, “Most detailed descriptions of this material are not yet online. However, high-level descriptions may be found in the General Inventory, and some digitized maps may be found in the virtual exhibits.” The AG Canada entries created for maps since the 1980s, although electronic, are still not migrated to ArchiviaNet, although some do appear in the National Library’s Canadiana CD-ROMs and most may be migrated to the National Library’s AMICUS database in the near future where they will reside until the

AMICUS based MIKAN system for the National Archives is operational. The “old”card catalogue for maps (1950s to 1980s) is currently being transferred to a digital finding aid format and the first portions of this data base are expected to be available on-line to researchers in 2002. Only additional resources will facilitate access to these invaluable tools and to the other vast number of existing finding aids still available only in paper format.

As noted previously, there are only two cartographic “partner” sites noted under “Canadian Memory Virtual Exhibitions” – the toponymic exhibition and the George Back sketchbooks. There are no references to the other web sites noted earlier in this paper nor to the others not here described; even so, many of these could not be correctly designated as “exhibitions.” There is certainly a need to review these and other web sites in which National Archives holdings figure prominently and to ensure that links are added to various yet undetermined locations on the National Archives web site. Access to web sites should be facilitated at least when the department is named as a partner or as a major contributor (e.g. GeoGratis and *Ottawa in Maps*).

The most recent published *Annual Review* of the National Archives – that is, for 1999-2000 – notes that “With the Internet, the Archives now has a marvellous communication tool to make Canadians aware of its collections. In 1999-2000, some 1.5 million people explored and consulted the National Archives Web site.”³ Certainly, the technology now is available to ensure that the wishes expressed by Dr. Smith in 1972 and reported at the beginning of this paper can now be met – certainly, the future is exciting for the staff responsible for cartographic materials in the National Archives of Canada and indeed for cartographic staff in archives, libraries and museums world wide.

Note: The opinions expressed in this paper are those of the author and do not represent the position of the National Archives of Canada.

Betty Kidd
P.O.Box 370
82 First Avenue
Russell ON K4R 1E2
Canada

³ Annual Review 1999-2000. National Archives of Canada/Revue annuelle 1999 -2000. Archives nationales du Canada, p.10.

LA NUMERISATION DES DOCUMENTS CARTOGRAPHIQUES ANCIENS: SUPPORTS TRADITIONNELS ET NOUVELLES TECHNOLOGIES*

By Pierre-Yves Duchemin

Résumé: La numérisation de documents graphiques, et notamment de documents cartographiques, est depuis longtemps un sujet d'étude à la Bibliothèque nationale. Dès sa création en 1994, la Bibliothèque nationale de France a prévu d'offrir à ses utilisateurs, un grand nombre d'images numérisées à partir de ses collections patrimoniales : elle s'adresse aux chercheurs, aux enseignants et aux étudiants et vise également un public plus large car les collections importantes de documents cartographiques ne sont pas nombreuses en France. Cette communication, centrée sur les documents cartographiques, aborde les divers problèmes qui se sont posés pendant la réalisation du projet.

Par ailleurs, la Bibliothèque nationale de France a produit des documents édités sur CD-Roms : « Mappamondi : une carte du monde au XIVe siècle » qui est une représentation numérique de l'Atlas Catalan, « Les globes de Louis XIV : la terre et le ciel par Vincenzo Coronelli », « La Géographie d'Idrîsî : un atlas du monde au XIIe siècle ».



Mappemonde dans un bonnet de fou (d'après Ortelius, ca 1580)

1. Historique des projets de numérisation d'images à la Bibliothèque nationale de France

La numérisation de documents graphiques, et notamment de documents cartographiques, est depuis longtemps un sujet d'étude à la Bibliothèque

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

nationale : en 1989, fut produit un vidéodisque analogique sur la « Révolution Française » qui contenait des documents cartographiques de la fin du XVIII^e siècle. En 1991, un partenariat avec une société californienne pour numériser la collection de 437 portulans conservés au département des Cartes et plans et les éditer sur CD-Rom a été sur le point d'aboutir mais n'a pu voir le jour en raison de la loi française très stricte en ce qui concerne la propriété intellectuelle et l'usage du patrimoine national : le ministère de la Culture, tutelle administrative de la Bibliothèque nationale, n'a pas accepté cette procédure.



Portulan de Christophe Colomb (1492)

Dès 1994, la Bibliothèque nationale de France a prévu d'offrir à ses utilisateurs, dans un premier temps physiquement présents dans ses salles de lecture, dans un second temps distants, un grand nombre d'images numérisées à partir de ses collections patrimoniales ; en offrant la consultation d'images numériques réalisées à partir d'une partie de ses collections, la Bibliothèque nationale de France s'adresse évidemment à son public naturel, c'est-à-dire aux chercheurs, aux enseignants et aux étudiants de second et troisième cycle. Mais elle vise également un public plus large, bien connu depuis longtemps sur le site Richelieu, où sont conservées toutes les collections spécialisées ouvertes aux chercheurs, aux étudiants... et au « grand public », car les collections importantes de documents spécialisés, et notamment cartographiques, ne sont pas nombreuses en France.

Bien que ce programme ne soit pas destiné à une catégorie précise du « public à distance » de la Bibliothèque nationale de France, il n'en favorise pas moins, de par sa nature et ses caractéristiques, certains usages auxquels l'offre numérique existante n'apporte pas toujours une réponse satisfaisante : curiosité ponctuelle quant à la nature des collections, découverte plus détaillée de leur diversité, connaissance plus approfondie d'une catégorie de documents ou d'une pièce

précise – sans que cela soit exclusif de la satisfaction de demandes plus spécialisées ou plus érudites, qu'autorise la reproduction d'un certain nombre de pièces dans leur intégralité. L'insertion de ce programme dans l'offre numérique doit tenir compte de cette dimension d'« introduction aux collections » par l'établissement de liens avec les différentes composantes de cette offre numérique et certaines parties du site (Présentation des départements).

La variété des documents conservés sur le site Richelieu, ainsi que l'importance de la volumétrie des collections à traiter - on approche les 20 000 000 documents - montre combien il est difficile de concevoir une sélection et de maintenir une politique cohérente à l'intérieur de l'institution quand on doit traiter des documents aussi nombreux et aussi différents dans le cadre d'un projet de numérisation.

Bien qu'à la Bibliothèque nationale de France, la géographie ne soit pas considérée comme un champ spécifique de la connaissance - elle est traitée comme une discipline transverse et les documents cartographiques sont disséminés dans plusieurs départements, même si le département des Cartes et plans est le plus gros réservoir de représentations du monde - les collections retenues montrent une cohérence intellectuelle interne et sont prévues pour avoir une existence propre, avec des buts scientifiques et / ou artistiques, en tant qu'« unités documentaires ». Cette méthode permet d'éviter l'inconvénient qui consiste à offrir des ensembles thématiques incomplets.

2. Critères de sélection

Parmi les critères qui ont présidé au choix des collections, les images numériques ont été créées pour :

- permettre la préservation et la conservation du document original, notamment en ce qui concerne les documents de grand format ou les documents très fréquemment consultés,
- permettre la mise en valeur de collections en les ouvrant à un plus large public, grâce notamment à la transmission à distance,
- favoriser la mise en valeur de collections patrimoniales et prestigieuses de documents rares à valeur documentaire ou esthétique, souvent non publiés, en publiant des disques optiques compacts ou des Cd-webs,
- améliorer les possibilités d'accès au document en proposant un substitut électronique à l'utilisateur final,
- permettre un accès multiple au document,
- autoriser une communication plus importante et plus savante du document grâce à l'utilisation de stations de travail spécifiques (poste d'accès à la

bibliothèque numérique = PABN) qui permettent à l'utilisateur d'effectuer des zooms et d'obtenir une finesse de détails qu'il ne pourrait espérer à partir de l'original,

- offrir à l'utilisateur final une copie électronique du document sur un outil de travail spécifique dans un nouvel environnement technique,
- permettre des sorties payantes sur imprimante de qualité, à l'intérieur de l'établissement,
- un dernier critère a été la complétude et la cohérence de la reproduction photographique, ainsi que la présence de notices bibliographiques informatisées, sur fiches ou sur catalogues imprimés.

Enfin, en France, la situation juridique sur l'utilisation des images numériques n'est pas simple : les lois de propriété et de protection intellectuelle sont très strictes et on doit toujours rechercher d'éventuels ayants-droits. C'est la raison pour laquelle les projets d'images numérisées de documents cartographiques concernent des documents anciens, pour lesquels n'existent pas de problèmes de communication ou de reproduction, puisqu'ils appartiennent à la Bibliothèque nationale de France depuis des années, voire des siècles.

3. Projets d'images numériques de documents cartographiques

La première opération de numérisation d'images a consisté en la réalisation de 260 000 images de « non-livres », parmi lesquelles plusieurs dizaines de milliers de documents cartographiques, de plans d'architecture, de photographies à caractère ethnologique, etc.

Parmi les 11 projets mis en œuvre, 7 sont relatifs à des documents cartographiques ou nécessitent une indexation géographique : cartes de France, d'Europe, du Monde et du ciel des XVII^e et XVIII^e siècles de la collection d'Anville, plans d'architecture du XVIII^e siècle dessinés par Boullée et Lequeu, dessins et gravures de villes et de paysages français du XVI^e au XIX^e siècle de la collection Destailleur, photographies en noir et blanc de la Société de géographie prises en Afrique par les explorateurs français au XIX^e siècle, photographies de Paris en noir et blanc prises par Eugène Atget, photographies en couleurs de villes et de paysages français prises au cours des années 1980 pendant les missions DATAR, etc. Cette première série d'images numérisées est centrée sur la France, mais les nouveaux projets à l'étude possèdent une couverture géographique plus importante.

Les projets réalisés à la Bibliothèque nationale de France concernent des documents appartenant à :

- département des Cartes et plans (1 400 000 documents) qui regroupe cartes manuscrites, gravées et imprimées, atlas, globes, objets géographiques, dessins et notes d'explorateurs, photographies, plaques de verre, etc., de la plus ancienne carte connue à l'imagerie satellitaire, ainsi que les collections de la Société de géographie (livres, périodiques, cartes, photographies, plaques de verre, etc.).

Parmi ces collections, le choix s'est porté sur la prestigieuse Collection d'Anville (1697-1782) :

Parmi les 10 500 documents, dont un grand nombre d'*unica*, qui composent la collection réunie par Jean-Baptiste Bourguignon d'Anville au XVIII^e siècle et qui couvrent la terre entière, le projet de numérisation concerne 2 408 documents cartographiques libres de droits représentant la France, les provinces, les généralités, les diocèses, etc. du XVI^e au XVIII^e siècle, c'est-à-dire globalement l'organisation administrative, politique et religieuse de la France sous l'Ancien Régime, quelques cartes d'Europe, quelques mappemondes et quelques cartes du ciel. Est également incluse la première couverture régulière de la France réalisée à partir de 1747 par les Cassini, dont une version coloriée à la main en toises ayant appartenu à la reine Marie-Antoinette et une version noir et blanc métrique, pour laquelle un taux de résolution plus élevé a été nécessaire (4 000 x 6 000 points par pouce, soit 180 fichiers de 90 Mo, si l'on utilise la version couleur. La carte de Cassini pourrait faire l'objet d'une édition dans la collection « Sources » ; ce document représente en effet une mine toponymique pour la France de l'Ancien Régime. La numérisation des 4 346 images de la collection d'Anville (cartes en plusieurs feuilles) a été effectuée à partir de cibachromes 13 x 18 cm (microfiches monovues). Le catalogage a été effectué à la feuille dans BN-Opaline et certaines notices ont été regroupées en lots cohérents (cartes des diocèses, cartes des provinces, carte de France de Cassini, Gaule, assemblage de 5 à 6 feuilles représentant un cours d'eau, etc.).



Mappemonde cordiforme d'Oronce Fine (1536)



Feuille n°99 (Grenoble) de la carte de Cassini coloriée

- département des Estampes et de la photographie : plusieurs millions de documents, parmi lesquels on dénombre cartes, panoramas, plans d'architecture, etc.

Le choix s'est porté notamment sur les collections de documents produits par les deux architectes visionnaires du XVIII^e siècle Boullée et Lequeu, qui n'ont jamais dépassé le stade du dessin architectural.

- Dessins d'architecture d'Étienne-Louis Boullée (1728-1799)

120 pièces (dessin ou lavis) de grand format (150 x 50 cm) dont quelques documents concernant la Bibliothèque royale, notamment l'installation d'une verrière au dessus de la Cour d'honneur pour en faire la Salle de lecture. Quantitativement peu important, ce projet est néanmoins d'un intérêt considérable : en effet, hormis la collection du département des Estampes, on ne compte qu'une quinzaine d'autres documents de Boullée conservés dans le monde. La numérisation a été effectuée à partir d'ektachromes 24 x 30 cm. Le catalogage a été effectué à la pièce, par numérisation, reconnaissance optique de caractères et



Vue intérieure de la nouvelle salle projetée pour l'agrandissement de la Bibliothèque du Roi

conversion d'un catalogue imprimé en format InterMarc dans BN-Opaline, et les notices ont été regroupées en lots centrés sur un bâtiment ou un projet.

- Dessins d'architecture de Jean-Jacques Lequeu (1757-1825)

713 documents (aquarelle ou lavis) de grand format (jusqu'à 3,5 x 1,5 m). La numérisation a été effectuée à partir d'un microfilm couleur 35 mm. Le catalogage suit exactement les mêmes principes que ceux utilisés pour le projet Boullée.

- Collection Hippolyte Destailleur

4 906 documents (dessin, aquarelle, lavis, gouache, gravure, etc.) de moyen format (40 x 30 cm) représentant des sites de Paris (1 328 documents) ou de province (3 578 documents) du XVI^e au XIX^e siècle. Certains documents étant recto-verso, le total de ce projet atteint plus de 6 000 images. La numérisation a été effectuée à partir de diapositives couleur 24 x 36 cm. Le catalogage a été effectué à la pièce, par numérisation en interne, reconnaissance optique de caractères et conversion en format InterMarc dans BN-Opaline de catalogues imprimés et les notices ont été regroupées par arrondissement pour Paris et par département pour la province.

- département des Manuscrits : quelques documents cartographiques manuscrits.

- Manuscrits médiévaux à peinture

Parmi la sélection de 12 729 pages provenant de 143 manuscrits choisis parmi les plus « représentatifs » des trésors de la Bibliothèque nationale de France, ce « florilège » inclut les documents de la bibliothèque de Charles V, déjà utilisés pour le projet « 1 000 enluminures sur Internet ». La numérisation a été effectuée à partir d'ektachromes 24 x 30 cm ou 4 x 5 pouces pour la majorité des manuscrits et de diapositives couleur 6 x 6 cm.

C'est parmi les manuscrits que se trouve la première application d'imagerie numérique sur l'Internet de la Bibliothèque nationale de France : la page d'accueil du site Web de la BNF contient une base de données graphiques spécifique intitulée « 1 000 enluminures sur Internet ». Cette base de données a été inaugurée en septembre 1995 pour l'inauguration de l'exposition « Culture et pouvoir » à la Bibliothèque du Congrès de Washington. Sept manuscrits à peinture du XIV^e siècle, la période de Charles V, le fondateur de la Bibliothèque nationale, ont été numérisés au taux de 2 000 x 3 000 points par pouce et présentent plus de 850 images zoomables plein écran et accessibles par menus (textes et liens hypertexte vers d'autres textes ou des images, thèmes avec une structure arborescente à 3 niveaux ou feuilletage des manuscrits). Parmi eux, l'« Atlas catalan » par Abraham Cresques, daté de 1375, représente l'état du monde connu à cette époque et constitue le chef-d'œuvre de ce projet. Ce portulan comporte 6 feuilles de vélin, pliées en leur milieu et montées sur des ais de bois, chaque feuille mesurant 64 x 49 cm. L'original a été photographié sur des cibachromes 4 x 5 pouces et numérisé

au taux de 2 000 x 3 000 points par pouce pour les feuilles entières et au taux de 4 000 x 6 000 points par pouce pour les détails. Plus de 50 détails sont disponibles.



Carte de l'Europe, de l'Afrique et du Proche-Orient



Caravane traversant la route de la soie

4. La numérisation des documents cartographiques

Cette communication est centrée sur les collections de documents cartographiques numérisés, parmi lesquelles on trouve des documents parmi les plus intéressants, mais aussi parmi les plus difficiles à numériser et à exploiter correctement, pour trois raisons principales :

- d'une part, les documents les plus intéressants, les plus rares, les moins connus, proviennent souvent des collections spécialisées et ils sont en outre un excellent choix en ce qui concerne la conservation, la communication, la valorisation et les publications,
- d'autre part, les documents cartographiques, en particulier en raison de leur taille et de la précision des informations qu'ils contiennent, soulèvent des problèmes tant pratiques que techniques très difficiles à résoudre pour pouvoir obtenir une bonne qualité et une utilisation efficace des images numériques,
- enfin, à l'inverse d'une page imprimée qui est très « binaire » et ne contient généralement que des points noirs ou blancs, la numérisation de documents

cartographiques relève plus de « l'artisanat », à la fois par la manipulation du document, notamment s'il présente un format qui sort de l'ordinaire, et par l'utilisation de l'image numérique que l'on peut en obtenir.

La Direction des collections, dans laquelle se trouvent les départements qui conservent les collections de la Bibliothèque nationale de France n'est pas seule engagée dans le processus de numérisation : elle partage cette responsabilité avec la Cellule « numérisation » du département de la Bibliothèque numérique et le département de l'Audiovisuel. À l'origine du projet, un des principaux problèmes à résoudre a été de faire admettre aux autres partenaires que les documents cartographiques n'étaient pas tout à fait des documents comme les autres : le marché était prévu pour la numérisation de 300 000 « images » avec un taux de résolution de 2 048 par 3 072 points par pouce. On savait déjà, depuis les tests initiaux, que ce taux était insuffisant pour une bonne utilisation d'images numérisées à partir de documents cartographiques. Ces derniers ont nécessité une procédure particulière afin d'obtenir à l'écran une lecture satisfaisante des toponymes et des détails graphiques : ils ont été numérisés au taux de résolution de 4 000 par 6 000 points par pouce.

Lors de la phase initiale de tests, la question s'est posée de savoir si la numérisation devait être effectuée à partir du document original ou d'un substitut photographique. À l'époque, dans les années 1995-1998, les plus grands numériseurs existant en France ne pouvaient pas traiter de documents à plat de plus de 1,20 m pour les numériseurs à rouleaux ou de plus de 2 m de largeur pour les caméras numériques les plus performantes. Tous les documents sélectionnés dans les projets de numérisation n'atteignent pas de telles dimensions et, dans la plupart des cas, il est possible de numériser un document à partir de l'original, même si les tests montrent que la numérisation d'un inversible couleur de haute qualité donne les mêmes résultats que celle d'un document de taille moyenne. Le problème est différent en ce qui concerne les documents cartographiques dont la taille ne permet pas une numérisation directe : aujourd'hui, il existe des caméras numériques capables de traiter des documents de 4 x 3 m et de restituer l'image en 3D si nécessaire.

En fait, la réponse a plutôt été l'option de la numérisation d'un support photographique : nous étions dans le cadre d'un marché global de numérisation de documents graphiques, parmi lesquels des documents cartographiques et des plans d'architecture. Parmi ces derniers, quelques-uns des plans d'architecture dépassent 3 m de longueur : il est évident qu'ils n'ont pu être numérisés directement et qu'on n'a pas pu faire l'économie d'une campagne photographique. Dans le même ordre d'idées, la carte de Cassini est une mine incomparable en ce qui concerne la

toponymie française du XVIII^e siècle ; il importait donc que l'ensemble des toponymes, quelle que soit la taille de leur typographie, soit distinctement lisible.

La Bibliothèque nationale de France dispose d'une photothèque riche de plusieurs millions de substituts photographiques (microfiches, microfilms, cartes à fenêtre, microfiches monovues, inversibles couleur, inversibles noir et blanc, ektachromes de grand format (jusqu'à 24 x 30 cm), cibachromes, etc.). On aurait pu penser *a priori* que ce trésor allait apporter la solution, mais si l'on en croit les experts en conservation, un inversible couleur réalisé il y a 5 ans n'est plus complètement parfait et un inversible couleur réalisé il y a 20 ans peut être inutilisable, même s'il a été conservé dans un environnement protégé et favorable ; ils soutiennent également que la durée de vie d'une microfiche noir et blanc peut atteindre une centaine d'années, mais il n'était pas question d'utiliser des substituts noir et blanc pour offrir l'image numérique de documents en couleur. On a pu constater que les experts en conservation avaient raison et que nombre de supports photographiques ne présentaient plus un niveau de qualité suffisant. Dans quelques cas, il a été possible d'utiliser des inversibles couleur qui commençaient à virer en leur appliquent une légère correction chromatique selon la charte colorimétrique définie, mais de nombreux documents ont dû faire l'objet d'une nouvelle campagne photographique.

Lors de la sélection des projets, la nature, la couverture et la qualité des reproductions photographiques existantes ont été un important critère de choix : il n'est pas concevable de numériser des cartes à partir de diapositives 35 mm 24 x 36 mm ou de microfiches. Les tests ont montré que les meilleurs résultats étaient obtenus à partir de cibachromes 4 x 5 pouces, d'un piqué plus fin que les ektachromes 24 x 30 cm. Certains projets ont nécessité une campagne photographique complémentaire pour assurer la complétude ou refaire des prises de vues mieux adaptées à une campagne de numérisation que les clichés existants : il a fallu dans certains cas refaire certaines prises de vues de documents dont on pensait la couverture photographique terminée.

Eu égard à la grande quantité d'images à numériser dans le cadre du projet, il est apparu que 300 000 images dépassaient les capacités techniques de la cellule de numérisation et il a été décidé de confier la numérisation proprement dite à des entreprises extérieures dans le cadre d'un marché. Après le lancement d'un appel public à la concurrence et la rédaction d'un cahier des charges, les offres de quatre prestataires ont été retenues, chacun d'entre eux montrant une expertise dans la numérisation de tel ou tel support photographique.

Dans le cas des documents cartographiques, la numérisation a donc été effectuée à partir des cibachromes 4 x 5 pouces au taux de résolution de 4 000 x 6 000 points

par pouce, afin de permettre à l'utilisateur d' « entrer » dans l'image numérique et d'effectuer un zoom sur un détail géographique ou toponymique. Cette procédure, utilisée pour les documents cartographiques, garantit une grande qualité mais coûte nettement plus cher que la numérisation réalisée au taux de 2 000 x 3 000 points par pouce pour les autres documents graphiques.

Dans le cadre particulier des projets de la Bibliothèque nationale de France, le marché passé auprès des 4 prestataires prévoit la numérisation par chacun d'eux de 3 000 images par semaine. En outre, le marché stipule que le prestataire doit numériser en même temps, non seulement l'image elle-même, mais également la « légende ». Celle-ci est généralement composée d'un ensemble de champs extraits de la notice bibliographique, de la référence du cliché au Service de la reproduction qui permet la commande directe d'une reproduction photographique, ainsi que de la clé de la notice bibliographique dans la base de données, utilisée comme lien hypertexte. Le système prévoit, en cas de réponses multiples à une requête, l'organisation des réponses en mosaïques d'images (16 à 32 selon les projets) qu'il est possible, par un simple clic de la souris, d'afficher plein écran et/ou de sélectionner dans un « panier » consultable en fin de requête.

Exemple de légende :

Nolin, Jean-Baptiste (1657-1708). - Roiaume et duché de Septimanie... / dessinée par J.B. Nolin. - [S.l.] : [s.n.], [17e ou 18e siècle]. - 1 carte : 36 x 45,5 cm. - [Cote : BNF C.Pl. Ge DD 2987 (368), microfiche 86/244. - FRANCE-- Sud-Est-- Frontières -- 0561-0752. 16e ou 17e siècle. - SEPTIMANIE. 16e ou 17e siècle. \ Opaline 074046.

Le résultat est tout à fait lisible sur l'écran 17'' du poste audiovisuel (PAV), et est encore meilleur sur l'écran graphique 21'' du poste d'accès à la bibliothèque numérique (PABN). Malheureusement, le poste audiovisuel est doté d'un imageur qui ne permet pas le zoom au-delà de 200 % et une nouvelle solution technique est à l'étude. Si la consultation d'une image numérique sur un écran graphique ne pose pas problème et permet la lecture des toponymes sur une carte, le problème est la taille gigantesque des fichiers à traiter : une carte ancienne, c'est-à-dire un document pas très détaillé, peut dépasser 400 Mo..., même si la compression permet de le réduire à environ 10 Mo en haute définition ou 300 Ko en basse définition. Certains documents particulièrement détaillés se mesurent en Go...

Cette qualité de consultation a pu être obtenue grâce à un sévère contrôle de qualité tout au long de la chaîne de travail et notamment lors du retour des supports de numérisation : la vérification de la conformité à la charte colorimétrique et la vérification de l'adéquation des légendes à l'image font l'objet d'un examen particulièrement attentif. L'utilisation de normes, à la fois pour la

numérisation, la compression des données, l'exploitation et la communication est également un facteur de qualité. Cette normalisation est un pari sur l'avenir : les données seront toujours plus importantes que les supports physiques. Le support peut évoluer, mais les données restent, notamment si elles possèdent une structure normalisée et sont conservées dans un environnement favorable.

En ce qui concerne le support d'exploitation, c'est-à-dire le serveur, des machines spécifiques, de type Sequant possédant un espace de stockage se chiffrant en téra-octets sont utilisées.

5. Catalogage et indexation

Le problème du traitement bibliographique des documents numérisés a dû être résolu avant les opérations de numérisation. Sur ce point encore, de nombreuses questions se sont posées : l'image numérique est-elle un document à part entière ou un simple support de substitution ? Doit-on cataloguer les images numériques à l'unité ou bien par lots ? Dans le cas des documents cartographiques, la réponse a été de cataloguer tous les documents à la pièce, même s'ils appartiennent à un ensemble plus large, artificiel ou non. En outre, l'image numérique d'un document est considérée comme un substitut de ce document, comme un ektachrome ou un microfilm, ce qui permet de lui faire subir le même traitement intellectuel et la même indexation que le document original.

La solution qui consiste à cataloguer chaque document comme une unité documentaire séparée et à créer des liens entre eux afin d'obtenir un ensemble dans la base de données s'avère longue et coûteuse. Les documents originaux ayant été catalogués à l'unité dans la base de données BN-Opaline, il a suffi d'ajouter une adresse logique aux descriptions bibliographiques et de créer un lien réciproque afin d'obtenir aussi bien l'image numérique à partir de la notice de description bibliographique dans la base de données que la description textuelle à partir de l'image. L'utilisation d'un catalogage capable de gérer des notices analytiques ou un catalogage à niveaux a permis de lier précisément une ou plusieurs images numériques à une notice textuelle contenue dans un ensemble plus important. La conversion rétrospective d'un fichier ou d'un catalogue imprimé, par numérisation et reconnaissance optique de caractères, a été utilisée pour obtenir un catalogage à la pièce. Pour un plus grand confort d'utilisation du système de recherche par l'utilisateur, les projets sont organisés en lots, grâce à des notices de regroupement.

Les ISBD, les normes AFNOR, les formats MARC et les listes raisonnées d'indexation matières (LCSH, RVM ou RAMEAU) ont été utilisés alors que la structuration électronique des données, l'hypertexte, l'indexation en « langage naturel » et l'interrogation en texte intégral sont en plein développement. Les bases

de données disponibles à la Bibliothèque nationale de France ne permettaient pas l'expérimentation de solutions plus modernes. Les projets en cours (« Voyages en France », qui associe textes et images, « Mémoires », qui se veut une vitrine de la diversité des collections de la Bibliothèque nationale de France) sont réalisés grâce à des DTD et une structuration en XML.

La base BN-Opaline a bénéficié dès la fin de 1995 d'un accès Telnet, remplacé en mars 2001 par un accès Web, mais ces accès ne concernent que la base de données elle-même et ne fournissent pas les liens à l'image. C'est pourquoi un appel d'offres a été lancé en mars dernier pour la rédaction d'un schéma directeur d'une nouvelle informatisation des collections spécialisées, dans lequel l'utilisation de métadonnées et le lien à l'image numérique dans une structuration XML seront étudiés.

6. Coût de la numérisation

Bien que le coût de la numérisation soit en baisse constante depuis le milieu des années 1990, le coût total d'une campagne de numérisation est élevé, particulièrement si l'on choisit des taux de résolution importants, mais l'on doit garder à l'esprit que le coût de la numérisation ne comprend pas la seule numérisation, mais couvre également les éventuels coûts de restauration, les éventuelles campagnes photographiques, le catalogage, l'indexation, les différents supports physiques, les postes de travail spécifiques munies d'écrans graphiques « haute définition », etc.

7. Les réseaux

L'établissement a dû définir ses besoins:

- combien d'utilisateurs sont susceptibles de se connecter simultanément ?
Actuellement, 1500 ports sont disponibles simultanément,
- quelles images doit-on rendre disponibles pour l'accès à distance ?
La décision a été prise de ne fournir en accès par l'Internet que les images libres de droits, qui ne représentent pas l'intégralité des fonds numériques disponibles à la consultation sur site,
- combien d'images doit-on rendre disponibles pour l'accès à distance ?
La montée en charge progressive du projet fait que chaque mise à jour apporte son lot de nouveautés à la Bibliothèque numérique,
- quelle qualité d'image numérique doit être disponible sur le réseau ?

Il est tout à fait possible de consulter un document à un taux de définition très élevé à l'intérieur de l'établissement et d'offrir une image numérique de qualité inférieure sur le réseau ; cette question est déterminante et structurante en termes de flux de données et de temps de réponse.

Cette volonté de l'établissement de permettre la fourniture d'images à distance a soulevé des questions : les images sélectionnées sont libres de droits et c'est la Bibliothèque nationale qui détient les droits sur les images numériques. À la fois pour offrir à l'utilisateur des temps de chargement acceptables et pour ne pas risquer le piratage, la Bibliothèque nationale de France a décidé de fournir sur l'Internet des images de qualité dégradée. Si cette procédure est envisageable pour beaucoup de documents graphiques, elle n'est pas intéressante pour les documents cartographiques : en effet, la carte de Cassini qui apparaît en image plein écran non zoomable en qualité dégradée est inutilisable. Ce problème va être réglé par la fourniture de l'image complète en qualité dégradée accompagnée d'extraits en haute définition.

Aujourd'hui, la bibliothèque numérique Gallica fournit aux utilisateurs distants de la Bibliothèque nationale de France près de 50 000 documents graphiques.

8. L'édition de CD-Roms

Par ailleurs, à partir de ses collections de documents numériques, la Bibliothèque nationale de France produit des documents édités sur CD-Roms. La collection «Sources» a pour objectifs de présenter sur CD-Rom une sélection des plus importants documents conservés par la Bibliothèque nationale de France et faisant partie du patrimoine de l'humanité. Ce choix d'édition permet de rendre accessibles des documents qui, du fait de leur état ou de leur fragilité, sont difficilement manipulables et donc rarement consultables.

Chaque document fait l'objet d'une reproduction très fidèle par numérisation à haute résolution et offre les mêmes opportunités de consultation que l'original ; la possibilité d'effectuer des zooms à quatre paliers sur des détails améliore même la consultation que l'on pourrait effectuer sur le document original. Ces «fac-similés électroniques» s'adressent à un large public qui peut ainsi déchiffrer pas à pas le document, puis élargir sa recherche grâce aux qui mettent l'accent sur les thèmes abordés, ainsi que sur le contexte historique et artistique de l'époque à laquelle le document a été réalisé. Ces commentaires sont rédigés en trois ou quatre langues selon les CD-Roms.

Enfin, une dernière partie du CD-Rom propose une série d'outils logiciels qui permet à l'utilisateur de «s'approprier» le document grâce à des «carnets» et à

des « calques ». Les carnets et les calques sont stockés sur le disque dur de l'ordinateur de l'utilisateur et gardent la trace des notes, observations et analyses rédigées par l'utilisateur, ainsi que des extraits d'images ou de textes qui lui semblent particulièrement intéressants à conserver.

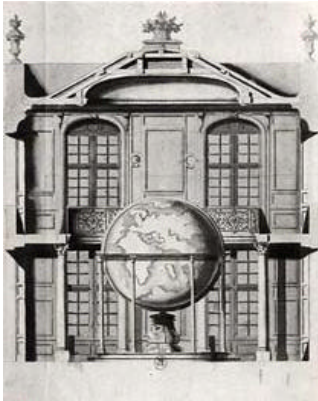
Parmi les documents cartographiques publiés dans la collection « Sources », on peut citer :

- «Mappamondi: une carte du monde au XIVe siècle» qui est une représentation numérique de l'Atlas Catalan de 1375 (Ms Esp 30), avec commentaires en français, anglais, allemand et espagnol, qui reproduit le monde connu à l'époque.



L'atlas catalan (1375)

- «Les globes de Louis XIV: la terre et le ciel par Vincenzo Coronelli», une exploration du globe terrestre et du globe céleste, chacun de quatre mètres de diamètre, réalisés pour Louis XIV à la fin du XVIIe siècle.



Projet de Louis-Étienne Boullée



Détail du globe terrestre



Salon des globes (1782-1900)

«La Géographie d'Idrîsî: un atlas du monde au XIIe siècle» (Manuscrit oriental 2221). En 68 planches et une mappemonde, cet atlas, «divertissement de celui qui est passionné» pour la pérégrination à travers le monde», est replacé dans le contexte du milieu du XIIe siècle et décrit le monde connu en multipliant les itinéraires et les informations aussi bien historiques, économiques que descriptives, voire anecdotiques.

Pierre-Yves Duchemin
National Library of France
Bibliothèque nationale de France
Direction des collections
Mission de coordination
Informatique et numérisation
58 rue de Richelieu
75084 PARIS Cedex 02
FRANCE
Tél. (33 1) 53 79 83 55
pierre-yves.duchemin@bnf.fr.

NAMING THE LANDSCAPE: BUILDING THE CONNECTICUT DIGITAL GAZETTEER*

By Scott R. McEathron, Patrick McGlamery, Dong-Guk Shin
Ben Smith, Yuan Su

Abstract: This article describes the background and processes of the initial work of building the *Connecticut Digital Gazetteer*. The background provides the context of the development of digital gazetteers as a national research initiative within the distributed geolibraries paradigm. The actual technical processes of importing and parsing the data are described and analyzed. To convert point data into data with a geographic “footprint,” spatial extents were identified and added using Geographic Information Systems software and the Land Use / Land Cover (LU/LC) data. The processes of metadata creation for the Connecticut History Online project are then analyzed as well as the actual referencing of this data for use in the cartographic interface. A basic analysis of geographic metadata within the larger context of efforts to merge controlled vocabularies is presented. The article concludes by presenting an outlook on future developments in research on digital gazetteers.

Introduction

Ubi or Where is...?

It is doubtful there is a map librarian alive (or dead for that matter) who has not pondered a scrap of a map and wondered... where... is it from, does it go, or does it show?

When one puts in hours on the university library’s general reference desk, the prevailing question is... *ubi*? For example, where can I find a poetry criticism, a citation, and an article on the national energy policy... *ubi*?

It is reasonable to go to a map when we ask, “*ubi*.” But when we go to the library catalog and ask “where,” we must do it with a textual search. However, a map illustrates topological relationships; relationships of ‘near-ness,’ ‘in-ness’ and ‘far-ness.’ Text does not. Text does not indicate that Chicago and China are worlds apart; in fact, the text-based catalog puts them rather close to each other, alphabetically, on the screen. A map interface would indicate where Chicago and China are in relation to each other, and would indicate that Zinjiang is in China...

* Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

even though it is quite far from it in the catalog. A map interface, a graphical, cartographic interface has the potential of enhancing access to the catalog for many users.

In order to use a cartographic interface to textual place-names in an on-line catalog we must provide the names with spatial awareness. The names must be in a coordinate system. In order to make use of topological relationships; the place-names must have areal geometry from which spatial neighbors can be derived. A digital gazetteer must be built which geographically enables the on-line public access catalog (OPAC). Such a tool has precedence in the Digital Library and Digital mapping research initiatives of the past decade.

In this paper we aim to provide a basic background of building the *Connecticut Digital Gazetteer* by briefly describing some of the primary research initiatives such as distributed geolibraries and the development of digital gazetteers. We will then describe the processes and framework for building the *Connecticut Digital Gazetteer*. We will conclude by providing a vision of digital gazetteers.

Distributed GeoLibraries

In June, 1998 the National Academy of Science's Mapping Science Committee convened a workshop to explore:

- a vision for geospatial data dissemination and access in 2010
- comparisons of different efforts in digital library research, clearinghouse development and other data distribution
- suggestions of short and long term research needs
- identification of policy and institutional issues

“The Mapping Science Committee serves as a focus for external advice to federal agencies on scientific and technical matters related to spatial data handling and analysis. The purpose of the committee is to provide advice on the development of a robust national spatial data infrastructure for making informed decisions at all levels of government and throughout society in general.”¹

The National Academy of Sciences' Mapping Science Committee (MSC) has been directing research initiatives since its 1993 report *Toward a Coordinated Spatial*

¹ Mapping Science Committee.
http://www4.nas.edu/cger/besr.nsf/web/mapping_science?OpenDocument

*Data Infrastructure for the Nation*² set the stage for the National Spatial Data Infrastructure (NSDI). The report established the Federal Geographic Data Committee (FGDC) and the FGDC's Spatial Metadata Content Standards. Subsequent MSC reports have addressed specific components of the NSDI, including: partnerships in *Promoting the National Spatial Data Infrastructure Through Partnerships*,³ in 1994, basic data types in *A Data Foundation for the National Spatial Data Infrastructure*⁴ in 1995, and future trends in *The Future of Spatial Data and Society*⁵ in 1997 and most recently *Distributed Geolibraries; Spatial Information Resources*⁶ in 1999.

The June 1998 workshop met to build a national vision of a GeoLibrary. The participants asked:

- What will it take to build distributed geolibraries?
- What economic incentives can be put in place such that stakeholders in all sectors of the community (business, education, and government) can and will participate?
- What arrangements need to be put in place in the form of institutions, regulations, standards, protocols, committees, and so forth?
- What research needs to be done to address problems and issues for which no methods or solutions currently exist?
- What data sets need to be constructed, and what mechanisms might be used?
- What software needs to be written, and who is likely to write it?

These are the questions that are driving current national and international research agendas in the mapping sciences. The digital gazetteer has emerged as part of that research agenda.

Digital Gazetteers

A preliminary study of digital gazetteers was held at the Smithsonian Institution in Washington, D.C. in October 1999. The goals of the two-day workshop were to

² Mapping Science Committee. *Toward a Coordinated Spatial Data Infrastructure for the Nation*. Washington, D.C.: National Academy Press, 1993.

³ Mapping Science Committee. *Promoting the National Spatial Data Infrastructure Through Partnerships*. Washington, D.C.: National Academy Press, 1994.

⁴ Mapping Science Committee. *A Data Foundation for the National Spatial Data Infrastructure*. Washington, D.C.: National Academy Press, 1995.

⁵ Mapping Science Committee. *The Future of Spatial Data and Society*. Washington, D.C.: National Academy Press, 1997.

⁶ Mapping Science Committee. *Distributed Geolibraries; Spatial Information Resources*. Washington, D.C.: National Academy Press, 1999.

(1) develop an understanding of the potential of indirect spatial referencing of information resources through geographic names and (2) to identify the research and policy issues associated with the development of digital gazetteer information exchange.

The development of interchangeable sets of geographic name data (gazetteers) and interoperable gazetteer services could result in a major improvement in seamless access to and use of a wide variety of information resources through indirect geospatial referencing. The two-day workshop was convened (1) to develop an understanding of the potential of indirect spatial referencing of information resources through geographic names and (2) to identify the research and policy issues associated with the development of digital gazetteer information exchange. -- The vision developed at the workshop is the Digital Earth metaphor for organizing, visualizing, accessing, and communicating information provides a powerful enabling framework for marshalling the resources needed to understand and mediate environmental and social phenomena.⁷

Definition and scope of digital gazetteers

With three key attributes, a gazetteer supports several functions of an information retrieval system:

- It answers the "Where is" question (for example, "Where is Storrs?") by showing the location on a map.
- It translates between geographic names and locations so that a user of the information system can find collection objects through matching the footprint of a geographic name to the footprints of the collection objects. For example, "What aerial photographs cover parts of Tolland County?"
- It allows a user to locate particular types of geographic features in a designated area. For example, the user can draw a box around an area on a map and find the schools, hospitals, lakes, or rivers in the area.

Beyond these basics, a digital gazetteer needs to support:

- the representation of variant names
- information about the names such as authority, etymology, source, and time span for the use of the names geographic footprints (coordinates representing

⁷ Hill, Linda. "Digital Gazetteer Information Exchange (DGIE) Final Report of Workshop Held October 12-14," 1999
http://alexandria.sdc.ucsb.edu/~lhil/dgie/DGIE_website/DGIE%20final%20report.htm

point, bounding box, polygonal, and linear features) information about footprints such as accuracy, measure method, source, and time span

- descriptive text
- data such as population and elevation, and
- relationships between named places (e.g., an 'Is Part Of' relation between a city and a county).

Building the Connecticut Digital Gazetteer

Technical Considerations

The first step was importing Geographic Names Information System (GNIS) data from the Alexandria Digital Library (ADL)⁸ and then storing the data in a local Oracle database. The ADL gazetteer, designed by University of California, Santa Barbara, is a relational database schema⁹ based on the ADL Gazetteer Content Standard¹⁰ and implemented on an Informix relational database management system (RDBMS). Although Informix and Oracle are both relational databases, to import data from Informix database to the Oracle database directly is not convenient. We used the Extensible Markup Language (XML) data format as middleware for importing the GNIS thesaurus (described below). XML has become the standard for information interchange due to its flexibility and simplicity. The designation of both Document Type Definitions (DTD) for the XML documents and the relational database schema are based on ADL Gazetteer Content Standard, however, there were some differences between the relational database schema and the DTD for XML data. So to import XML data into our Oracle database, our solution proceeded in three steps:

1. We parsed the XML documents to machine-recognizable elements. We used Xerces-1_0_3 which is a XML parser from Apache¹¹ to do the parsing.
2. We built the database schema for our local Oracle system. Since the designation of the relational database schema ADL gazetteer use is based on the gazetteer content standard, and the schema provides portability to all RDBMS, we borrowed the schema and use it on our Oracle system.
3. We observed the difference between the DTD and database schema, and developed a mapping method that maps the XML data into our Oracle database.

⁸ Alexandria digital library, <http://www.alexandria.ucsb.edu/adl.html>, 1998

⁹ ADL gazetteer relational database schema, http://alexandria.sdc.ucsb.edu/~zheng/alex-imp/new_gaz/, 1999

¹⁰ ADL Gazetteer Content Standard, http://alexandria.sdc.ucsb.edu/gazetteer/gaz_content_standard.html, 2000

¹¹ apache XML parser, <http://xml.apache.org/>, 1999

To extend our GNIS database, we added data from *Connecticut Place Names* (CPN) gazetteer of historical place names. Here we also proceeded in three steps:

1. We scanned the whole book and used optical character recognition software (OCR) to recognize its characters and saved the data in text files.
2. We built a parser with Java code which parsed the text file and extracted three types of data for each record of place, they are 'place name,' 'geographic type of the place,' and 'citations' for each record.
3. Finally we joined the CHS data to the existing records in GNIS database and stored them in related fields.

Identifying Spatial Extents of Populated Places

The Geographic Name Information System (GNIS) has identified many types of geographic features as points on a map, including Populated Places (ppl) and streams. It is possible to find the spatial extents of many of those named features using Geographic Information Systems (GIS) software and geospatial data.

Of the many themes of geospatial data available, the one that has been extremely useful in identifying the spatial extents has been the Land Use / Land Cover (LU/LC) data classified from the LANDSAT Thematic Mapper satellite imagery. This classification was done by the University of Connecticut's Laboratory for Earth Resource Information Systems (LERIS) within the Department of Natural Resources Management and Engineering. The resulting data set is a 1998 LU/LC grid of 30 by 30-meter cells covering the entire state of Connecticut. The data have been processed into 28 classes, including residential, commercial, farmlands, forests, etc.

Using this data we can identify cells which we consider urban centers or built up areas by aggregating the following four classes of data into polygons: urban residential, medium residential, rural residential and tree and turf complex. Most of these polygons can then be considered populated places. Then, using capabilities available in GIS and database management utilities, such as spatial overlays and text joins, polygons can be selected or excluded and assigned the information from the GNIS ppl points. Thus, we are transferring the named point to an inferred polygon.

Problems may occur when the urban centers are not separated from each other in the LU/LC data due to connected development or urban sprawl. In this case a single polygon may contain several ppl. Points. Clipping the polygon by known administrative boundaries can sometimes solve this problem. In Connecticut, these are town boundaries. Other times a named ppl will occur where there is no

identifiable development in the LU/LC data. In this case the ppl may be a historical place, or a cartographic “locator” with not enough development to be identified. We have found that a significant number of these Populated Places are road intersections that may have been more populous or otherwise significant in the past.

Identifying streams is an easier task since the Connecticut Department of Environmental Protection (ConnDEP) has processed the hydrographic data set (vectors from USGS 1:24,000 Digital Line Graphs) from the GNIS. Similar processes of text joins and spatial overlays can be used to merge the GNIS identification numbers to these streams and to place the GNIS names into unnamed or misnamed streams. Problems arise where there are gaps in the streams due to lakes or marshes that have not been identified by the ConnDEP as part of the stream.

Some problems arise when the urban centers are not separated from each other in the LULC data due to connected development, or urban sprawl. In this case a single polygon may contain several ppl points. Clipping the polygon by known administrative boundaries can sometimes solve this problem. In Connecticut these are town boundaries. Other times a named PPL will occur where there is no identifiable development in the LULC data. In this case the PPL may be a historical place, or a cartographic “locator” with not enough development to be identified. We have found that a significant number of these PPLs are “Corners” which may have been more populous in the past.

Connecticut History Online

In 1999 the Thomas J. Dodd Research Center at the University of Connecticut, the Connecticut Historical Society and the Mystic Seaport Museum partnered in the Connecticut History Online project.¹² The project is partially funded by a National Leadership Grant from the Institute of Museum and Library Services (IMLS). The current collaborative grant will contain a total of about 14,000 historical images from the partners’ historical photographic collections when it finishes in 2002. These images can be searched or browsed in variety of ways, including by keyword, subject, creator, title and date. Geographical sites may be searched using a digital gazetteer developed by the University of Connecticut’s Map and Geographic Information Center. Descriptions of the images are included in detailed cataloging records.

¹² Connecticut History Online. <http://www.lib.uconn.edu/cho/index.htm>

The selection criterion included knowing the geographic place of the photograph. Selectors were asked to select only those photographs whose place was identified. The catalogers, working with the MARC format, noted the place in locally adapted 651 field. The 651 field provides the most 'hooks' or coded geographic data for a digital gazetteer, particularly the 2nd indicator '7' and the sub-field '2' 'Source of heading or term' enables the Digital Gazetteer to process most efficiently.

The catalogers were instructed to use one of four thesauri, *gnis*, the Geographic Name Information System, *tiger*, the Topologically Integrated Geographically Encoded and Referenced, *chs*, the Connecticut Historical Society's *Connecticut Place-names* and a *local* thesaurus in the sub-field 2 of the 651 field. The *gnis* includes all of the words on the USGS topographic map series and is in a standard format. It is available online from the US Geological Survey <http://mapping.usgs.gov/www/gnis/gnisform.html> and from the "Getty Thesaurus of Geographic Names" http://shiva.pub.getty.edu/tgn_browser/. The *chs* is a scholarly, comprehensive study of historical place names, it is in print format only, but will be scanned and entered into the Connecticut Digital Gazetteer. The *tiger* thesaurus is a database of street name and auto address location. The Official TIGER look-up site is <http://www.census.gov/cgi-bin/gazetteer> and MapQuest; a value-added site is at <http://www.mapquest.com/>. Items in the *local* thesaurus are compiled by the catalogers and added to the Connecticut Digital Gazetteer as needed.

Mapping the CHO database

At present the CHO database numbers in excess of 4,500 records. These records list over 6,500 instances of place names. There seemed to be two methods for providing spatial access to the CHO Online Public Access Catalog (OPAC). One, an 'ambiguous search' would search for place-name strings against the 651 of the OPAC. This method may enable the user to search for the place-name 'Mansfield' and all of the named features in 'Mansfield'. This method was determined to be inefficient. The other method, a 'specific search' provided for pre-processing the names in the 651 field. The 6,750± instances of names represent only 627 different names. For example, there are close to 500 instances of the place-name 'Hartford.' By structuring a specific search for Hartford and more efficient use of the Connecticut Digital Gazetteer, the OPAC and the HTML interface is possible.

An SQL process is performed on the CHO database in the Endeavor system. The query process pulls each instance of a 651 field. After the initial SQL process that extracts the place-names from the CHO database as a table, a series of processes are performed. A table for each thesaurus is created for each category. Second,

each instance of a place-name is counted and the count entered in a field. Finally a URL is constructed from the place-name and the category as a 'scripted search.' These tables are then geo-referenced using ArcView; the *gnis*, *local* and the *chs* against the Connecticut Digital Gazetteer, the *tiger* against the TIGER Street file. These coverages are then referenced through an ArcIMS project.

Accessing CHO through a Spatial Interface

We are using ArcIMS on an NT server to link to the CHO OPAC. When the user determines their geographic choice, a 'scripted' query is sent to the OPAC. The script is in the form:

<http://cho.uconn.edu/cgi-bin/Pwebrecon.cgi?DB=local&SAB1=Hartford&BOOL1=as+a+phrase&FLD1=Place+Name+%28651A%29&GRP1=NOT+with+next+set&SAB2=East+Hartford&BOOL2=as+a+phrase&FLD2=Place+Name+%28651A%29&GRP2=NOT+with+next+set&SAB3=West+Hartford&BOOL3as+a+phrase%28651A%29&CNT=25&HIST=1>. This is a particularly difficult example. It indicates the complications of compound place-names like 'Hartford,' 'East Hartford,' and West Hartford.' This search is for 'Hartford' only. We have constructed the ArcIMS CHO Geolocator is a simple search engine with few tools. It is primarily a cartographic interface, not a GIS application.

The Metadata Framework

Jan Smits asks an important question regarding geo-spatial data access: "can we manage to shift?"¹³ To answer Jan's question, we must look to our recent past. We must ask questions to critically evaluate that past and then be prepared to act. We believe digital gazetteers represent a step in making a shift toward increased geo-spatial data access through *indirect geospatial referencing*, a process of using placenames that are spatially aware to query data.

The efforts that are currently underway within the World's community of libraries and librarians to integrate controlled vocabularies are important developments toward a more effective use of names and subjects in searching. These efforts include the International Federation of Library Association's (IFLA) projects of Universal Bibliographic Control International MARC Core Program, Minimal Level Authority Records (MLAR), and Functional Requirements And Numbering

¹³ Smits, Jan. "Geospatial Data Access: Can We Manage to Shift," *Conference Proceedings, 67th Council and General Conference International Federation of Library Associations and Institutions (IFLA)*, Boston, Massachusetts, August 22, 2001, <http://www.ifla.org/IV/ifla67/papers/165-166e.pdf> (1 October 2001).

of Authority Records (FRANAR). Also, in Europe there is the development of the AUTHOR Project: an effort to network the national name authority files. Our contentions are that while these projects are valuable to libraries – they focus primarily on personal and corporate name authorities and do little to enhance access the geographic data.

Libraries have done a much better job in populating authority databases with personal and corporate names than with subject headings and specifically geographic subject headings. For example, in the United States, there are over five million Library of Congress (LC) form Name Authority records. However, there are only 258,822 total Library of Congress form Subject Authority records – most of these are not geographic. In comparison, the GEOnet Names Server (GNS), which provides access to the National Imagery and Mapping Agency's (NIMA) database of foreign geographic feature names, has over 3.7 million place-names. Also, the Geographic Names Information System (GNIS) has approximately 2 million names for U.S. features. These two databases together total about 5.7 million place-names. The difference in comprehensiveness between these two databases and the total number of LC form of Subject Headings is staggering. If one of the virtues of authority control is “precision” in searching: how can this be achieved when so little of the total existing names are represented in the database?

The structures of library authority control are not keeping pace with current intellectual growth nor information production within the World. Nor is the World library community providing comprehensive authority control for the pre-existing legacy of human information and knowledge, especially geographic information. The structures that are being used to provide authority control, such as AACR2R or MARC21 in the U.S. are not in themselves to blame. The failure of authority control centers on our inability to populate the existing structures with data.

Like authority files, gazetteers have tended to be hierarchical – pointing to a preferred name. The notable exemption is the *Alexandria Digital Library Gazetteer*.¹⁴ This project provides a means of exchanging record formats using a flexible metadata content standard (which is currently being revised), providing a common scheme of categorizing places (feature type codes) and search and retrieval protocols. Key to the success of this distributed approach is how effective local efforts, such as the Connecticut History Online project, are in adding the extra value to the metadata and then sharing it with others.

¹⁴ Hill, L. L., Frew, J., & Zheng, Q. “Geographic names: The implementation of a gazetteer in a georeferenced digital library,” *D-Lib* (January 1999).
<http://www.dlib.org/dlib/january99/hill/01hill.html>

The quality, consistency, and of course, the very presence of geographic coding within any given metadata are key variables for the retrieval of relevant data when using a digital cartographic interface to query a database or multiple databases. For example, the MARC21 map format contains several 'mappable' or geographic elements such as the coded cartographic mathematical data within the 034 field. This field holds the geographic coordinates that define the limits of the map being described. However, there are also other fields that can be mapped. These include the 052 geographic classification code and the 651 subject added entry--geographic name fields. Also, beyond the MARC21 maps format, other subject added entries within the cataloging records for other formats often have geographic coding when they have geographic subdivisions assigned. This raises the possibility of being able to reference this data using a digital cartographic interface as well. In other words, a web based map search interface could potentially serve as a portal to large amounts of geographically coded data including bibliographic data that represents information in many varied formats.

The metadata created for the Connecticut History Online project is an example of geographically coded data that may be queried from both an online catalog or from a map interface. The catalogers for the project used Endeavor's *Imageserver* software to input metadata for each image within the MARC21 format. Within each metadata record, there is at least one 651 subject added entry. The following thesauri are being used to determine the subject added entries: gnis, tiger, chs, ipc, lcs, local.

Librarians continue to build a virtual distributed international authority file and map and geography librarians need to assess how geographic name authority data may be incorporated. The Connecticut Digital Gazetteer is an example of a database and tool that incorporates many different vocabularies and their corresponding spatial extents. This will enable indirect spatial referencing of data containing coded geographic data. The development of digital gazetteers is an important development toward the realization of the Digital Earth metaphor and will aid and expand the serendipitous discoveries of geographic information.

Patrick McGlamery
Map and Geographic Information Center
Homer Babbidge Library
369 Fairfield Road
University of Connecticut
Storrs, Ct 06269
Patrick.McGlamery@uconn.edu

Scott R. McEathron
Assistant Map and Geography Librarian
University of Illinois at Urbana-Champaign
418 Main Library, MC-522
1408 West Gregory Dr.
Urbana, IL 61801
Phone: (217) 244-189
macmap@ux1.cso.uiuc.edu