Data Issues Affecting Mapping in the 21st Century

David Clark
National Geophysical Data Center
NOAA/NESDIS
325 Broadway, EGCX1
Boulder, Colorado, 80302
USA

Fax +303-497-6386
David.M.Clark@noaa.gov

As the 21st Century unfolds, we are beginning to fully experience the "Data and Information Explosion." The recognition of this phenomenon began in the late 1960s and early 1970s, but has only now begun to become a reality. One explanation of this may be the catching up of technology with our ability to design and integrate the needs of our society into a technological framework. Of course, one can argue whether technology caught up with our needs or did we catch up with technology? In either case, it has arrived, not only in the Earth sciences, but also in all facets of our society as shown by the integration of the World Wide Web (WWW) into the mainstream of our global society. This unprecedented growth in data and information describing our planet will have profound effect on our study of the Earth and how we describe and map it. For the mapping community, this has significant implications. This data and information from all parts of the Earth system, including the human dimension, and the resulting derived knowledge, will need to be mapped in new and innovative ways, many which have never been even considered before (NRC 1997). This will test our ability to adapt to the new challenges brought on by this unprecedented event. It has been noted that there has been more information produced in the last 30 years then during the previous 5,000 (Pritchett, 1999)

The effects of the data on mapping the Earth System can be divided up into five major issues. They are the size and diversity of spatial databases, the implementation of spatial data infrastructures, the applications of spatial data, the development of national, international and commercial data policies, and emerging technology.

Large and diverse spatial databases

A daunting challenge of the 21st century is the mapping of the large volume and highly diverse data describing the Earth and its environment. This is a result of comprehensive observing and monitoring systems implemented by the scientific community, governmental agencies and private industry. Extensive observing systems are measuring every facet of our environment. The diversity of the types of data, from satellite imagery (i.e. gridded pixels) to global environmental model outputs (gridded points) to survey data (vectors, points) to demographic data (averaged values with a general geographic reference) will make designing a mapping schema very challenging (Tateishi and Hastings, 2000). As result of this diverse observation of the Earth, the volume of data describing the Earth is increasing with the advent of new, high resolution observing systems, including satellite remote sensing, and in-situ monitoring systems. Furthermore, analyses and models using these data are producing even more new data sets from these observed Earth parameters. The size of the data archives is growing faster than we can derive information from it. For example, it is estimated that by 2010, the size of a major US data archive for the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration and the United States Geological Survey (USGS) will be 18,000 terabytes or 18 petabytes (Figure 1). It would take 15.6 million CD-ROMs, each holding 640,000 bytes of data, to accommodate these data. In this particular archive, these highly diverse data range from satellite remotely sensed data to ground based weather radar data to individual point information from a wide variety of in-situ data sources. There are numerous issues which must be addressed regarding these vast and diverse archives, including storage (archival), access, quality, documentation (metadata) and so on. However, these issues are being addressed and will be solved in the years to come, partly by applying technology, partly by new and innovative procedures and methods and partly by the years of experience being gained by the data management experts (Flewelling and Egenhofer 1999). The issue for the mapping community (and for the whole scientific community) is how to effectively and efficiently use these data in our mapping programs (Clark, et al. 1991). The implementation of spatial data infrastructures can be a very effective way to help solve this problem.

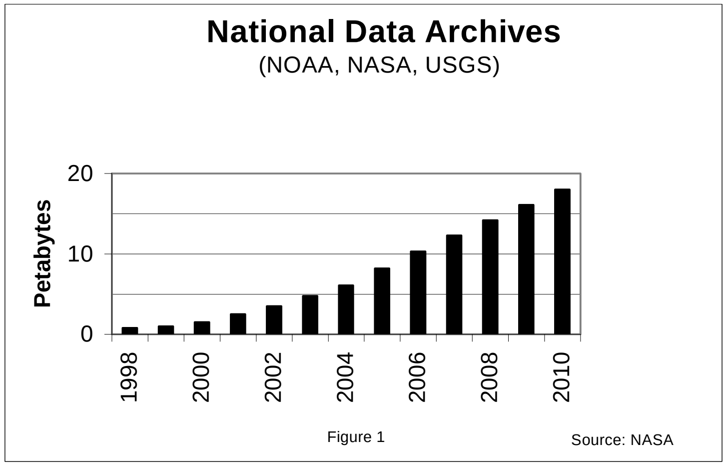

Figure 1

Spatial Data Infrastructures

The data and information society is dependent on the easy and rapid use of a wide variety of data. As noted above, the comprehensive monitoring and measuring of the Earth has lead to a highly diverse set of data which will need to be mapped for scientific, institutional and commercial applications. Spatial data infrastructures can be one way to help address the diversity of the data. In the US, the National Spatial Data Infrastructure (NSDI) is defined as the "means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the Earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, store and distribute such information to meet a wide variety of needs." (NRC 1993)

In the U.S. the Federal government established in 1990 an interagency group, the Federal Geographic Data Committee (FGDC), to coordinate the implementation of the US NSDI (FGDC, 1994). The establishment of the FGDC was as a result of a requirement by the government to coordinate "the development, use, sharing and dissemination of surveying, mapping, and related spatial data" (OMB 1990). President Clinton later strengthened the charge to the FGDC (Executive Office of the President 1994) and today the FGDC, led by the USGS, involves numerous federal, state and local governmental agencies as well as private industry.

Several countries worldwide have adopted spatial data infrastructures, but they differ in type, rationale and application. However, they are all national in scope, refer to spatial data, geographical data or land data, and describe some type of infrastructure for a coordination role in the implementation of the SDI (Masser 1999). As can be seen in Table 1, the countries implementing NSDIs are geographically dispersed and mostly from the developed countries.

National Spatial Data Infrastructures Table 1

Australia	Australian Spatial Data Infrastructure
Canada	Canadian Geospatial Data Infrastructure
Indonesia	National Geographic Information Systems
Japan	National Spatial Data Infrastructure
Korea	National Geographic Information System
Malaysia	National Infrastructure for Land Information Systems
Netherlands	National Geographical Information Infrastructure
Portugal	National System for Geographic Information
Qatar	National Geographic Information System
United Kingdom	National Geospatial Data Framework
United States	National Spatial Data Infrastructure

Source: Masser 1999

Many more countries have begun to realize value of spatial data infrastructures and they have begun to work towards a Global Spatial Data Infrastructure (GSDI). The GSDI is similar in concept to the NSDIs but aims to expand the concept globally (Holland 1999).

The national and global SDIs can be greatly beneficial to solving the data access and usage issues faced by the cartographic community. SDIs can help make a simple task as exchanging spatial data easy. It can also facilitate the implementation of state of the art mapping systems possible (US National Atlas 2001). However, in spite of the great benefits that SDIs will give us, the implementation of SDIs is still a major challenge. In spite of the strong commitment by governments, as noted above in the case of US FGDC, there is still a lagging to the full commitment toward a NSDI. The implementation of the NSDIs and the GSDI will necessitate development of data standards, comprehensive ways to describe the data and involvement of the entire community who is collecting, compiling and distributing spatial data. Changing existing data systems, adding additional human and fiscal resources to implement standards, and implementing new procedures is not popular with data collectors and data managers who already have stretched their resources to the limit. One way to perhaps facilitate the adoption of the SDI is to encourage the private sector to more fully embrace the NSDIs and GSDI.

Data policy

Data policy, which is described as the set of rules, regulations, laws or agreements governing the access and use of data, will have a major impact on mapping in the 21st century. There are numerous data policies currently in effect or are being developed for virtually all types of data and information present in our society today. These policies cover not only the natural and physical sciences but the financial, cultural, and political aspects of our society as well. The United States maintains several types of data policies within its governmental, academic and commercial communities. Much of the scientific data collected by or funded by the US Federal government is in the public domain and can be copied and exchanged freely. The US also has commercial copyright laws and patent protections and is developing new data polices currently through the legislative process (Linn 2000).

Data policies are currently implemented or are being developed in all the countries of the world. In Europe, the European Union has issued the EU Directive on the Legal Protection of Databases, which provides strong property rights in public and private sector databases (European Union 1996). Internationally, the United Nation's World Meteorological Organization has issued several resolutions, which sets out data policy for certain types of data (WMO 1995, WMO 1999). The commercial sector also has its international policy efforts. The World Intellectual Property Organization (WIPO) is an international organization dedicated to helping ensure that the rights of creators and owners of intellectual property are protected worldwide. Currently, there are 175 countries that are members of WIPO.

Each type of data has many data policy issues associated with them. The types of issues may be the same or may be very different. For Earth observation data (i.e., from satellites), some issues are very unique. These include timeliness (data are less commercially valuable as the time from collection progress) and coverage (imagery can be collected across national boundaries without permission of the government). This leads to some interesting data policy considerations (Harris 1997). The European Commission's Earth Observation Data Policy and Europe (EOPOLE) recently published its final report, which recommended initiating European research into the possible development of a comparable data policy for Earth observation and non-Earth observation environmental data (EOPOLE 2000). This may be one way to deal with some of the issues unique to Earth observation data verses other types of Earth systems data.

Interestingly, many of these policies are in conflict with each other and the challenge will be to understand these conflicts and chart a course that benefits all. This will take the close interaction and negotiation of the database rights holder and users to strike the balance between protection and fair use (NRC 1999). The implications for mapping are significant. Some of the "base" information that the mapping community relies on to "baseline" their maps could end up being restricted. The question of data access, copyrights and restrictions could make the "fair use" of map data much more complicated. One new idea that has gained some favor is the "copyleft" concept. Copyleft states that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it. It is a copyright notice that permits unrestricted redistribution and modification, provided that all copies and derivatives retain the same permissions. A notable example of this is the LINUX software system. In summary, if currently public domain data suddenly becomes restricted, many valued mapping efforts could be severely impacted. This issue is potentially one that could be a very significant one for the mapping community.

Diverse Global Applications

As we enter he 21st century, many national and international research, monitoring and observation activities will need data mapped to address the requirements of their programs. Many scientific programs, such as the International Geosphere Biosphere Program (IGBP) and the International Human Dimensions Program, will have very different needs for mapping and displaying the information obtained by these programs. The study of Global Change will necessitate the collection and analysis of a wide range of data (Townshend and Rasool 1993). In contrast, operational programs, such as the Global Observing Programs (Climate, Ocean and Terrestrial) will also have a very different set of mapping requirements (JDIMP 2000), which could include real-time mapping.

The issues for mapping are many. The diversity and volume of the collected data has already been noted. The application of these data will dictate the numerous ways that they will have to be mapped. One simple consideration, the scale of the application, requires a major effort for cartographers. The scale of the Global Circulation Models can range from 5 degrees to less than one degree. Large scale applications such as ecosystem studies may want scales of 1 minute to 1 second. Sometimes, for base and reference data for these types of studies, local and regional data may be used. This example shows that there are major sampling and interpolation considerations that the mapping community will need to address in creating new databases. There are many, many examples of the types of mapping that will need to be done. For example, there is need to map data from observational programs that involves monitoring arrays sending real-time data to central locations via satellite from which predictions are being made. Also, mapping is required to show socio-economic data, such as population characteristics, integrated with ecological, hydrological, climatic and geomorphic data to study health patterns. The examples are almost endless and the new types of mapping which will be needed in the future probably has not been even conceptualized (NRC 1995).

As an example of one way of dealing with the diversity in environmental measurements and as an effort to coordinate observing systems, many of the global programs and organizations have formed a partnership to address environmental issues, research and provision of Earth observations. The Integrated Global Observing Strategy (IGOS) is made up of representatives from the national space agencies, many United Nations organizations, major international scientific organizations such as IGBP and World Climate Research Program, and the international observing programs, GCOS, GTOS and GOOS (Figure 2).

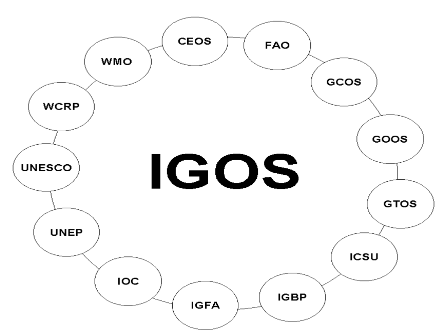

Figure 2

Many of the various environmental initiatives and treaties implemented over the last few decades call for systematic observations of the Earth, and the IGOS will provide the necessary links in support of the international initiatives. IGOS may be a good paradigm to consider for addressing other problems such as data management and mapping.

Rapidly Advancing Technology

Finally, emerging new technology will help the mapping community with these issues. Moore's law says that processing speeds of computers are doubling every 18 months. Disk storage costs decrease 50% each year and storage capacity doubles every 24 months. Fiber optic network speeds double each year. Computer power is now 8,000 times less expensive than it was 30 years ago (Pritchett 1999). What does this mean for mapping? The World Wide Web (WWW) is changing the way we look at data and WWW mapping is a key area of research. Geographic Information Systems (GIS) are evolving to use raster, vector and point data simultaneously. Time varying data will be used in GISs along with spatial data to fully understand our dynamic environment. The impact of technology is evident to virtually all of the mapping community and has changed the way it has been in the past and the way it will done in the future.

However, there is a down side to this technology. Rapidly advancing technology can give rise to a "digital divide." This is a gap between those who are technologically capable and those who cannot, for economic or cultural reasons, implement this technology. This challenge is significant because of the data issues discussed above, but technology will be critical in addressing these problems. In addition to the digital divide, the sheer magnitude of technological advances is making it difficult for efficient assimilation of the technology and some mapping programs are lagging behind. For example, reports indicate that 1.5 million Web pages are added each day, and that the number of websites is doubling every 8 months (Pritchett 1999). It will take an aggressive mapping program to keep its technological edge.

How are we dealing with this technological tidal wave? New and innovative methods for data and information handling are being developed and are becoming freely available. Methods like data mining, data discovery, distributed computing are coming into the mainstream and their principles are becoming well known. Innovative technology like Napster (exchanging data, i.e. music, directly over the WWW without central control) and the public-private partner system, TerraServer (innovative provision of satellite imagery over the WWW), is using technology to provide users with services only dreamed of a few years ago. For the mapping community, innovations like the Open GIS Consortium's Web Mapping Testbed are applying this new technology.

Summary

In summary, these issues illustrate that the ability to manage, display, use and ultimately map data will necessitate development of new and innovative paradigms to address the data and information challenges. Implementation of NSDIs and a GSDI will help in the exchange and use of data for mapping. Certainly, technology will go a long way in helping to address some of these challenges and the mapping community is very in tune with technological advances. But, it needs to be remembered that technology alone will not be a panacea for all of these challenges.

Finally, data policy will play a major role in the data and information society, perhaps the most influential of the challenges faced by the mapping community. It will take all of the creativity of the mapping community to address all of these issues as we move into the next century.

References:

Clark, D. M., Hastings, D., and Kineman, J., (1991) Global databases: their implications for GIS. In: Maguire, J., Goodchild, M. and Rhind, D. (eds.) *Geographical Information Systems: Principles and Applications*. (London, Longman Ltd.), Vol. 2, pp. 217-231

EOPOLE (2000). *Earth Observation Data Policy and Europe - Final Report.*

European Union (1996). *European Union Directive on the Legal Protection of Databases*. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996.

Executive Office of the President, (1994), *Coordinating geographic data acquisition and access the National Spatial Data Infrastructure*, Executive Order 12906. Federal Register 59, 17671-17674.

FGDC, (1994), *The 1994 plan for the National Spatial Data Infrastructure: building the foundations of an information based society* (Reston, VA: Federal Geographic Data Committee, USGS).

Flewelling, D.M., and Egenhofer, M.J., (1999). Using Digital Spatial Archives effectively. In: Fisher, P., Armstrong, M., and Lees, B., *International Journal of Geographical Information Science.* pp. 1-8, Volume 13, Number 1. (Taylor and Francis).

Harris, R. (1997), *Earth Observation Data Policy*. (Chichester, John Wiley and Sons, Ltd.)

Holland, P. (1999). *The Strategic Imperative of a Global Spatial Data Infrastructure.* Presented at Cambridge Conference for National Mapping Organizations held at St. John's College, Cambridge, United Kingdom, 19-23 July 1999.

Joint Data and Information Panel (JDIMP), (2000), *GCOS/GOOS/GTOS Data and Information Management Plan, Version 1.* Report numbers: GCOS-60, GOOS-70 and WMO/TD No. 1004.

Linn, Anne (2000). *History of Database Protection: Legal issues of concern to the scientific community*. Unpublished.

Masser, Ian, (1999). All shapes and sizes: the first generation of national spatial data infrastructures. In: Fisher, P., Armstrong, M., and Lees, B., *International Journal of Geographical Information Science.* pp. 67-84, Volume 13, Number 1. Taylor and Francis.

NRC, (1999). *Question of Balance: Private rights and the public interest in scientific and technical databases*. National Research Council. (Washington DC: National Academy Press).

NRC, (1997). *Bits of Power: Issues in the Global Access to Scientific Data*. U.S. National Committee for CODATA. (Washington DC: National Academy Press).

NRC, (1995), *Finding the forest in the trees: The challenge of combining diverse environmental data*. U.S. National Committee for CODATA (Washington DC: National Academy Press).

NRC, (1993), *Toward a coordinated spatial data infrastructure for the nation*. Mapping Science Committee (Washington DC: National Academy Press).

Pritchett, Price, (1999). *New work habits for a radically changing world.* Pritchett Rummler-Brache

Tateishi, R., and D. Hastings, (2000), *Global Environmental Databases: Present situation; Future Directions*. International Society for Photogrammertry and Remote Sensing. Geocarto, Hong Kong

Townshend, J.R.G., and S.I. Rasool. 1993. Global change. In *Data for Global Change*, P.S. Glaeser and S. Ruttenberg, eds. CODATA, Paris.

United States National Atlas (2001) http://nationalatlas.gov/

World Meteorological Organization (1995). *Resolution 40 (Cg-XII) WMO Policy and Practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities*. World Meteorological Organization.

World Meteorological Organization (1999). *Resolution 25 (Cg-XIII) Exchange of hydrological data and products*. World Meteorological Organization.