Digital Earth as a UCGIS Grand Challenge

Karl E. Grossner
University of California, Santa Barbara
kgrossner@geog.ucsb.edu

Abstract. The 2005 UCGIS research agenda 'challenges' and 'emerging themes' are mapped to the major components of a digital earth system, and the requisite functionality of next-generation digital geolibrary. Large-scale knowledge representation systems are increasingly viewed as essential aspects of the emerging spatial data infrastructure, and could serve as a Grand Challenge for UCGIS with significant benefits to the GIScience discipline, member organizations and the general public.

1 Introduction

At the 2006 Summer Assembly meeting of the UCGIS, the identification of a Grand Challenge as a suitable follow-on to the successful Body of Knowledge project was the topic of multiple workshops in the Research session track. One of suggestions put forward—reprised in a sense—by Keith Clarke, was of "a Digital Earth by 2009." That year has arrived, the question of grand challenges remains and the concept of digital earth systems remains compelling. The term Digital Earth, coined by Al Gore in the 90s, refers to a visionary information system of enormous scope for education and collaborative research; at its furthest extension, arguably a digital "mirror world." The scope is such that it has not yet been comprehensively described, a problem reminiscent of the "defining the elephant" allegory: what it is depends on where you stand.

Several of the fundamental impediments to realizing such a system present in 1998 are either solved or in sufficient retreat that construction of a true *digital earth system* is a feasible undertaking. UCGIS, representing as it does, 84 "observers of the elephant," is uniquely positioned to coordinate the process of defining, then building the first exemplar *digital earth system*. To do so would require substantial work in each of the thirteen research areas (nine "research challenges" and four "emerging themes") described in detail in that organization's *A Research Agenda for Geographic Information Science* (McMaster and Usery 2005). It was generally agreed amongst workshop participants that a Grand Challenge may be defined as a large and difficult but doable undertaking that is of great benefit to society at large and requires the cooperative efforts of specialists in many areas. It should be beyond the scope that a single member or small group of members could manage. It was noted that previous successful Grand Challenges, such as "putting a man on the moon" and the Genome Project, aimed at specific measurable results; it was easy to tell when they were completed.

Since 2006, efforts have been ongoing at UC Santa Barbara and elsewhere to advance the definition of a *digital earth system* from its present somewhat nebulous state to being the basis for an ambitious, worthwhile and doable suite of projects (Grossner, Goodchild and Clark 2008; Craglia, et al 2008; Grossner and Clarke 2007). In this paper I seek to map the emerging preliminary set of requirements for *digital earth systems* to topics in the UCGIS research agenda, in support of the idea that such a system is in fact a suitable Grand Challenge. I also outline the benefits that might accrue from such an effort—to UCGIS, to its member institutions, individual scholars and societies worldwide.

The *digital earth system* is fundamentally a GIS, but merges with other concepts towards what I call a "geographic information and knowledge organization system." As it turns out, a *digital earth system* closely resembles a "distributed geolibrary" as defined at a 1998 National Research Council workshop ("Distributed Geolibraries"). In its reported findings, the case was made that "the library service model that underlies the concept of distributed geolibraries provides a useful way of structuring discussion and of thinking about the resources and research that will be need to make the (Digital Earth) vision a reality" (p.34). Also from the report of that workshop:

> *"A geolibrary is a digital library filled with geoinformation and for which the primary search mechanism is place. Geoinformation is information associated with a distinct area or footprint on the Earth's surface. A geolibrary is distributed if its users, services, metadata and information assets can be integrated among many distinct locations." (p. 7)*

But we find that a *digital earth system* must extend that definition to include many qualities of traditional GIS, of a comprehensive digital atlas, and of open source software development projects. Geolibraries are by and large designed to identify information assets (objects) associated with a place and deliver them in response to queries. To make existing geolibraries more nearly like the educational and "collaboratory" research system in the Digital Earth vision will require further inclusion of a larger body of georeferenced information, expanding the system's role from simply enabling search and delivery of information objects, to breaking open those objects in some degree, and processing what they contain (Goodchild, 2004). This would allow adding to the currently answerable query, "what *do you have* about that place?" another, vastly broader one, "what *is so* about that place?" and even, "what *has been so* there?" A true *digital earth system* should accommodate predictive models as well, able to answer queries about the future with what accuracy their authors' algorithms may provide. An advanced distributed geolibrary—a *digital earth system*—will be able to field all of these questions both from information in its local collection, and seamlessly, from a distributed web of collections worldwide. This would effectively marry the concepts of digital libraries, GIS, and knowledge organization systems (KOS).

Several major elements of the *digital earth system* we undertake to define exist today in commercial systems, in publicly available geospatial data portals and in research applications and prototypes. The tasks before us are to identify missing elements, to map them to key GIScience problems not yet solved, and to propose a plan for undertaking a large-scale collaborative effort at synthesis. Enough elements for such a system are either essentially complete or within sight that the work of designing the framework for that synthesis can begin.

2 Background

The Digital Earth concept first mentioned by US Vice-President Al Gore in 1992 and elaborated upon in his 1998 Los Angeles speech, was that of a geographical computing system for education and research, a “multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data” (Gore, 1998). Gore began the definition of that system appropriately, with a use case scenario:

> *“Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head-mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man-made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a "magic carpet ride" through a 3-D visualization of the terrain. Of course, terrain is only one of the many kinds of data with which she can interact. Using the systems' voice recognition capabilities, she is able to request information on land cover, distribution of plant and animal species, real-time weather, roads, political boundaries, and population.”*
>
> *“A Digital Earth could provide a mechanism for users to navigate and search for geospatial information–and for producers to publish it. The Digital Earth would be composed of both the "user interface"–a browsable, 3D version of the planet available at various levels of resolution, a rapidly growing universe of networked geospatial information, and the mechanisms for integrating and displaying information from multiple sources.”*

The speech motivated a great deal of activity in government, commercial and academic spheres, in the US and globally, and much progress towards achieving such a system has been made. The scope of Gore’s Digital Earth was so large, with discrete technical challenges so difficult, that it was initially productive for individuals and organizations to set about solving particular problems independently of each other, with no *overall* coordination or comprehensive definition of a particular system. The US Government’s Interagency Digital Earth Working Group (IDEW) did coordinate development of interoperability standards for geospatial data sharing and led the development of the Web Map Service (WMS) specification now in common use. A consensus definition arising out of a 1999 IDEW workshop was grand, but quite vague: “Digital Earth will be a virtual representation of our planet that enables a person to explore and interact with the vast amounts of natural and cultural information gathered about the Earth” (1999).

For the period between 1998 and 2001, the US Digital Earth Initiative was a motivational banner raised over various activities related to the sharing and distribution of geospatial data already under way amongst several US Government agencies. However it also included an entirely new undertaking—development of a series of “Digital Earth Prototypes” (2001), including a series of Digital Earth Alpha Version applications on climate and weather for four “scenarios”: museums, education, governance and journalism. That development program was never directly funded and stalled in late 2001 before any specifications had emerged ("Digital Earth Alpha Versions").

Significant related projects in academia included the NSF-funded *Alexandria Digital Earth Prototype* (ADEPT), a "virtual learning system" developed by researchers at the University of California, Santa Barbara between 1999 and 2004. ADEPT followed the Alexandria Digital Library (ADL), an early geolibrary exemplar that remains functional and deserves consideration as one context for developing a next-stage *digital earth system*.

A considerable volume of geographic research related to the Digital Earth concept has occurred world-wide in the last eight years, much of it reported at a series of bi-annual international conferences hosted by the International Society for Digital Earth (ISDE). These Digital Earth Symposia have been held in Beijing (1999), New Brunswick, Canada (2001), Brno, Czech Republic (2003) and Tokyo (2005). The next meeting is scheduled for San Francisco conference in June, 2007, with the tentative theme, "Bringing Digital Earth Down to Earth."

The term, "Digital Earth" remains a motivational umbrella concept and broad goal for geospatial applications with global coverage, but there has been no attempt at a comprehensive definition of a *digital earth system* that could realize the goal of integrating **"**the full range of data about our planet and our history" (Gore, 1998).

3 Digital Earth System Functionality

The specification of a computing system normally begins with a description of what its users can accomplish, and Vice-President Gore's use case scenario is an appropriate starting point. These are the functions both listed and implied in Gore's 1998 speech:

Infrastructure	
Organizational	Grassroots effort of thousands of individuals, companies, researchers and government organizations
	Organic Internet-like growth
	Government-sponsored testbed involving government, industry and academia
	Data from "thousands of different organizations"
Hardware	High-speed (10Gbps) networks
	Huge mass storage requirements
	Satellites providing imagery
	Public access points for highest bandwidth access, e.g. museums
Software standards	("some level of…") interoperability enabling data transfer between disparate systems
	Extensive metadata
Interface; principal application	
Software	3D globe

	Zoom in, out to multiple resolutions; fly through
	Control overlays, including terrain
	Virtual tour of museums
	Personal compilations; email
	Timeline
	Collaboration tools for researchers
	Generate and/or display model results, e.g. land use planning; ecological scenarios
	A path between scientists' results and the public, particularly regarding environmental science
Hardware	Virtual reality helmet, glove
	Voice recognition
Data	
	"Vast quantity"
	Historical, insofar as possible
	Public access and marketplace
	Global " 1 meter imagery"
	Digital Elevation Model ("visualize terrain")
	Land cover and land use
	Plant and animal species' distribution
	Soils, climate
	Real-time weather
	Physically sensed (e.g. GLOBE)
	Roads
	Political boundaries
	Population
Other content	
	Newsreel footage
	Oral histories
	Maps
	Newspapers

4 The Distributed Geolibrary

Five months after Gore's Digital Earth speech, in June, 1998, the Mapping Science Committee of the National Research Council convened a *Panel on Distributed Geolibraries*, chaired by Michael Goodchild of UC Santa Barbara, which met for two days in Washington, D.C. Other members were Prudence Adler, Barbara Buttenfield, Robert Kahn, Annette Krygiel and Harlan Onsrud. The panel produced a summary report that was published by the National Academies Press and remains available online[1]. It delineated the similarities in services and functions between the concepts of a spatial data infrastructure, the distributed geolibrary and the systems in the Digital Earth vision, effectively marrying them in terms of a single system. Any new effort at defining a *digital earth system* represents a resumption of work begun by that group. Excerpts from the workshop report appear in Appendix B of this paper and are referenced extensively in the component mapping exercise below.

5 Digital Earth System Components

As a starting point for defining a digital earth system, I have synthesized elements of the original Gore vision, the framework for digital geolibraries documented in the 1998 NRC report referenced above, ongoing developmental work on national and global spatial data infrastructures and some innovative work on digital cultural atlases undertaken by the Electronic Cultural Atlas Initiate (www.ecai.org) and others. The following outline represents a "for discussion" list of components, or component categories, as a start to a more formal definition.

A. Distributed network of databases

1. Held at participating academic and government institutions; catalogs published to central location in a system-wide standard format
2. Common, integrated three-tiered logical structure
    a. Level I: observational data with global coverage, uniform scale. Maintained centrally and replicated to participant systems.
    b. Level II: observational data and derived knowledge works referencing varying spatial and temporal extents, at any scale. Level II data are in some manner 'certified' as authoritative by an editorial review process
    c. Level III data and knowledge works can be submitted to a participant collection by anyone, registered un-reviewed in a clearinghouse catalog, upon meeting base criteria and format requirements.
3. Novel data model(s) required, to
    a. represent dynamic processes
    b. represent data quality: uncertainty, error
    c. facilitate categorization and integration of field and object data at middleware and application layers
    d. facilitate visualization, knowledge discovery and knowledge representation at application layer
4. All query responses from the system will include Level tags and granular metadata on provenance, quality, suitability

B. Distributed collections of content in various media

1. Georeferenced articles, historical source texts, photographs, video, audio, multimedia educational modules, animations and bibliographic listings.

---

[1] *http://darwin.nap.edu/cart/deliver.cgi?&record_id=9460*.

2. Extending range of information objects beyond the "traditional" geographic material (GIS layers, orthophotos, aerial photos, scanned maps, remotely sensed imagery)
3. Tiered approach, as in Levels II and III for data

C. "Middleware" layer

1. Centrally located, developed collaboratively
2. Search engine and API (application programming interface) that allows the distributed databases and collections to appear as one to application developers
3. Standardized authority lists, including: feature codes, place names, time period and biographical directories
4. An extensible framework for a system of domain ontologies; that is, a robust, flexible means for categorizing knowledge
5. Georeferencing services
6. Bibliographic look-up, e.g. Library of Congress, Cheshire/Melvyl
7. Server-based geoprocessing tools

D. User applications.

1. A web-based digital earth geolibrary
2. Suite of exemplar and prototype web applications and data interfaces for research and education in physical geography, human geography and human-environment relations
3. Wide distribution and high accessibility level, worldwide
4. Accessible to advanced interface technologies, and thick clients, such as virtual globes

E. Organizational infrastructure.

1. Essentially a global collaborative effort, with intellectual leadership and administrative coordination from UCGIS [with others – GSDI, ISDE, ISCGM?]
    a. Defining and designing
    b. Software development
    c. Editorial and review processes and systems
2. Tasks coordinated by a small centralized staff include
    a. Fundraising and grant administration
    b. Organizing meetings

6 Digital Earth, Geolibraries, and The UCGIS Research Agenda

One could, without much difficulty, draw lines describing relationships between nearly all pairs of items in these lists. Only the clearest or strongest of such connections in this observer's perspective are mapped here for purposes of discussion. That said there are 47 mappings in this matrix, and though time and space do not permit discussion of each, an attempt to summarize them follows. The audience for this document is assumed to be familiar with the UCGIS Research Agenda.[2]

Mapping components of a *digital earth system*,

A. Distributed databases	[i, ii]	[8, 7, 1, 4, 3, 10, 2, 6]
B. Distributed content collections	[iii, iv, v, vi]	[4, 11, 13, 10, 9]
C. Middleware services	[i, ii, iii, iv, vi]	[1, 11, 8, 2, 5]
D. User applications	[i, iv, v, vi]	[10, 2, 13, 11, 5, 3, 6, 9, 8, 12]
E. Organizational infrastructure	[i, ii, iv]	[7, 1, 9, 12]

to requisites for a distributed geolibrary,

i. Standards and protocols
ii. Authority datasets
iii. Georeferencing
iv. Cataloging
v. Visualization
vi. Knowledge construction

to UCGIS research challenges:

1. Spatial data acquisition and integration (Jensen, *et al*)
2. Cognition of geographic information (Montello, Freundschuh)
3. Scale (Lam, *et al*)
4. Extensions to Geographic Representations (Yuan *et al*)
5. Spatial Analysis and Modeling in a GIS Environment (Getis, *et al*)
6. Uncertainty (Zhu)
7. Spatial Information Infrastructure (Onsrud, *et al*)
8. Distributed and Mobile Computing (Goodchild *et al*)
9. GIS and Society (Elmes, *et al*)
10. Geographic visualization (Buckley, *et al*)
11. Ontological Foundations for GIScience (Mark, *et al*)
12. Remotely acquired data and information (Hepner, *et al*)
13. Geospatial data mining and knowledge discovery (Yuan, *et al*)

---

[2] McMaster, R.B. & Usery, E.L. (Eds.). (2005). *A Research Agenda for Geographic Information Science*, Boca Raton, FL: CRC Press.

6.1 Distributed databases [i, ii]; [8, 7, 1, 4, 3, 10, 2, 6]

A digital earth system
Collections held at multiple institutions, catalogs published to central location in standard formats; integrated three-tiered logical structure: (I) observational data with global coverage, (II) observational data and derived knowledge works referencing varying spatial and temporal extents, at any scale, (III) un-reviewed data and knowledge works submitted by anyone; novel data model(s) that: represent dynamic processes, data quality, (2) facilitate categorization and integration of field and object data at middleware and application layers, (3) facilitate visualization, knowledge discovery and knowledge representation at application layer; granular metadata on provenance, quality, suitability

Distributed Geolibrary
Libraries have a history of "establishing coordinated, collaborative and multi-institutional relationships" (p. 20). Increasingly, and productively, libraries are providing digital services of their own and referral to those from various providers and locations, all to increasingly distributed constituencies. Robust and flexible standards and protocols are absolutely essential to effective communication and transfer of data between collections.

Library databases typically comprise catalogs of physical holdings and circulation records, but a digital earth geolibrary would include information stores themselves, such as the core (Level I) spatial datasets mentioned above. It would also require several authority datasets to enable the rich connections implied in the Digital Earth vision, namely: place name gazetteers, feature type thesauri, as well as time period and biographical directories. These would be integral to the middleware services discussed below.

UCGIS Research Agenda
There are two key aspects to this requirement—the distributed nature of the system and the requirement for novel data models to serve applications and, ultimately, user requirements. A number of the research issues mentioned specifically in sections 8, 7 and 1 bear closely on the challenges of large-scale distributed information systems. There are calls for examining the status and compatibility of standards, improved models of geographic metadata, and robust system architectures for distributed GIS computing. At a motivational level, there was requested a "…pursuit of explicit conceptual models with substantial potential for providing incentives and overcoming impediments in wide-scale spatial data sharing environments should be a major thrust of the research community." (p.235)

A digital earth system certainly would have many qualities of a geographic information system, but the data models implied by intended functionality are a departure from those of existing GIS, closely resembling those suggested in Peuquet's follow-up to the Extensions to Geographic Representation chapter, which called for a change in paradigm, leading to "truly new representations." Visual and cognitive processes were noted to be "intrinsic to the design (of new data models)" that incorporate true object-orientation. A new approach, "integrated with semantics," is needed to achieve better representation of space-time dynamics and to drive graphics for better visualization capability.

The zooming capability seen in Google Earth would be an important aspect of many digital earth system applications one can imagine, but challenges related to representing multiple scales in databases would need to be met beyond the impressive LOD algorithms that allow speedy maneuvering with orthophoto and elevation data layers.

The Geographic Visualization chapter in the Research Agenda notes, "A disconnect currently exists between visualization and spatial databases," arguing that, "better understanding of cognitive requirements in displays communicating the contents of databases certainly is required." One of the near-term "Priority Areas for Research" calls for "extending representation for visualization to include qualitative, intangible and conceptual information."

The Cognition chapter asks whether "data models (should) match cognitive models of space, place and environment more closely?"

The first item on the sub-agenda for Uncertainty concerns representation, citing a basic unanswered question: "where in the data model is uncertainty information best placed—at the level of objects, geometric entities or cells?"

6.2 Distributed content collections [iii, iv, v, vi]; [4, 11, 13, 10, 9]

A digital earth system
Extending the range of information objects beyond the "traditional" (GIS layers, orthophotos, aerial photos, scanned maps, remotely sensed imagery) to include georeferenced articles, historical source texts, photographs, video, audio, multimedia educational modules, animations and bibliographic listings; tiered approach, as in Levels II and III for data

Distributed Geolibrary
The report clearly delineates how the definition of what is typically considered geographic data, or "geoinformation," might be extended to include "the contents of guidebooks, reports on specific areas, data sets with a geographic dimension, and any other information assets that serve to differentiate one geographic area from another." That is, "information assets having some form of associated geographic footprint, a boundary defining the geographic extent of the information…" All of these materials need to be georeferenced, the automation of which is an active research area.

Beyond that important mechanical step are the larger epistemological, even philosophical issues, raised by mapping raw, non-controversial, observational data to interpretive (read: subjective) knowledge works derived from them. "Traditional libraries collect and catalog primarily knowledge works for good reason. The reading and contemplation of works of knowledge such as books and journals provide context and convey meaning. Currently, such works are one of the best means by which we are able to acquire understanding." (p. 42) Distinctions between information types on the data-information-knowledge continuum must be made clear in any geolibrary, or *digital earth system* for that matter.

UCGIS Research Agenda
The expansion of information types handled by a new category of geographic information systems—such as the *digital earth system* contemplated here—will require work in several research areas, perhaps most notably Extensions to Geographic Representations.

Clear divisions between data, information and knowledge must be incorporated in future in data models; important theoretical constructs for accomplishing this have been put forward in recent years and can be realized and examined in this large-scale exemplar.

A central element in modeling the connection between observational and derived knowledge works is ontology. That a *digital earth system* must have a framework for domain ontologies, with a "top-level common, neutral backbone," is critical here with respect to social phenomena: political, economic and social "knowledge objects" and events in the human geographic realm.

6.3 Middleware services [i, ii, iii, iv, vi]; [1, 11, 8, 2]

A digital earth system
Search; application programming interfaces (API) presenting the distributed databases and content stores as a single virtual entity; standardized authority lists: gazetteer, feature codes, time period and biographical directories, bibliographic look-up; framework for a system of domain ontologies; georeferencing services; server-based geoprocessing tools

Distributed Geolibrary
The middleware layer of a *digital earth system* is the keeper of standards, for metadata, file formats, data transfer and various other specs pertaining to collection catalogs and interoperability generally.

Authority datasets are maintained at this level, to include place name gazetteers and feature-type thesauri. The envisioned *digital earth system* extends this list to include time period and biographical directories.

A geolibrary has georeferencing engines capable of resolving place names to lat/long and potentially, standard and novel hierarchical grids, as well as shielding end-users from having to "understand the complexities of geodetic datums and cartographic projections."

"Cataloging, which serves the critical function of abstracting the information users need to find, examine, assess, and retrieve data," (p. 80) is a service at this layer. "In effect, metadata are the key…automated discovery, indexing, and abstracting tools (that 'perform the functions of abstracting and metadata creation automatically') do not yet exist and will require extensive research and development." (p. 80)

The NRC report proposes that geolibraries extend the range of services they provide to include "tools for analysis, modeling, simulation, decision making, and the creation of new geographic knowledge." Research is required to determine what may be achieved here.

UCGIS Research Agenda
The middleware services in this framework are the common enabling technologies that allow disparate systems to appear and in some respects function as one. The central, institutional roles are discussed below.

Cataloging and search for a *digital earth system*—those elements that furnish access to information—will depend on a robust, flexible framework for developing and managing domain ontologies. It is integral to the semantic, object-oriented data models discussed in Section A above. It is only with ontological commitments that information objects in a database may have meaning; however, "universal, shared taxonomies that are both neutral and comprehensive are an impossibility." Hence, an extensible framework that evolves in "bottom-up" fashion over time is necessary. Such a system is a major research undertaking.

The research issues in distributed computing relevant to middleware layer include: "improved models of metadata," and system architecture problems of client-server processing loads, replication, versioning, and caching.

Wherever categorization and ontology appear, so does cognition research, which informs every aspect of humans' interaction with geographic information: "the construction of ontologies, systems of concepts or classes of what exists in the world, is a cognitive act as well as a reflection of objective reality." (p. 74)

6.4 User applications [v, vi, i, iv]; [10, 2, 13, 11, 5, 3, 6, 9, 8, 12]

A digital earth system
A web-based digital earth geolibrary; suite of exemplar and prototype web applications and data interfaces for research and education in physical geography, human geography and human-environment relations; wide distribution and high accessibility level, worldwide; accessible to advanced interface technologies, and thick clients, such as virtual globes.

Distributed Geolibrary
While a geolibrary (and a *digital earth system*) could make accessible a potentially unlimited volume of text-based information, a major driving purpose and definitive feature of applications built on such a platform will be the enabling of knowledge construction by means of geographic visualization. "This is a novel area with no obvious guideposts, and research will be needed to determine how best to make the user of distributed geolibraries aware of the existence of information and of its important characteristics…in particular, we know almost nothing about how to render *dynamic geospatial data* or how to indicate availability" (p. 81).

The middleware services discussed in Section C. above represent a platform for developers of *digital earth system*-compatible applications for education and research. The use-case scenario from the Gore speech—the schoolgirl with the data helmet and glove exploring scientific and cultural information about the planet—is one possibility. There are countless others.

UCGIS Research Agenda
This aspect of digital earth systems will draw on research in almost every corner of GIScience, perhaps most prominently Visualization, Cognition and Knowledge Discovery/Data Mining. Many of the priority areas for research in GVis are relevant: user interaction tools, representing qualitative, intangible and conceptual information, new interface metaphors, collaboration tools and integration of dimensions (spatial, symbolic, temporal) and modes (dynamism and animation) for visualization.

We presume visualization aids knowledge discovery, and those connected research challenges are summarized nicely thus: "(there is) an urgent need for new methods and tools that can intelligently and automatically transform geographic data into information and, furthermore, synthesize geographic knowledge. It calls for new approaches in geographic representation, query processing, spatial analysis, and data visualization" (p.

365). Likewise, a medium-term objective is to "develop multi-dimensional, interactive visualization techniques with dynamic links to distributed GIS databases to greatly enhance user's capabilities to detect hidden patterns and inspect hidden correlations among geographic variables.

The GIS and Society chapter declares a need to heighten interest and involvement in issues regarding GIS by members of other disciplines, and raising public awareness about the widespread use and usefulness of GIS technologies. Successful, comprehensive, accessible *digital earth system* applications will aid pursuit of that goal immensely.

6.5 Organizational infrastructure [i, ii, iv]; [7, 1, 9, 12]

A digital earth system
Global collaborative effort, UCGIS intellectual leadership and administrative coordination (with others – GSDI, ISDE, ISCGM?); tasks include: defining and designing, software development, editorial and review processes and systems, community, administrative.

Distributed Geolibrary
The central tasks of defining and designing a large-scale geolibrary/*digital earth system* will require consensus decisions about the adoption and/or synthesis of existing standards, protocols, authority datasets (gazetteers, thesauri, etc.) and the "bottom-up" ontology/categorization strategies mentioned. The kinds of educational and research applications envisioned are possible only with the broad acceptance made possible by a global collaborative effort.

The NRC report notes the traditional role of librarians in acquisition decisions and collection development. There is most likely a correspondence between that function and "editorial and review process and systems" referred to as part of a *digital earth system*, but these have yet to be spelled out.

UCGIS Research Agenda
This definition of a spatial data infrastructure would require only minor changes to describe a comprehensive geolibrary as well: "the means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the earth…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 fact, the top suggested "Relevant Project" is the development of "a geolibrary that meets need of all for access to information in public domain and private holdings; public and privately provided services."

The Infrastructure chapter of the Research Agenda notes a key challenge is "providing incentives and overcoming impediments in wide-scale spatial data sharing environments," (p. 235) declaring it needs to be "a major thrust of the research community."

A digital earth system would be a major advance in increasing public access to geospatial datasets globally, and providing innovative approaches to contextualizing them, through participatory tools and the linking of varied information and media types. Its wide-spread use would provide an opportunity to study many of the social issues surrounding GIS use, as delineated in the Research Agenda's GIS and Society chapter.

7 Conclusions

Some common threads appearing in all or nearly all of chapters in the Research Agenda indicate the time may be right for a concerted collaborative effort at a large integrative project—such as the definition and development of a *digital earth system*. There were repeated suggestions that exemplars, demonstrations and prototypes are important on several levels, that it is time to translate recent "theoretical milestones" into real systems, useful for students, researchers and policy-makers (p. 150), and that unified approaches to managing and sharing geospatial data would be beneficial.

A robust *digital earth system* developed in pursuit of many of the UCGIS research objectives would have these positive impacts:

- Improve geographic education and reduce "geographic illiteracy."
- Guide the explosive growth in development and use of geographic computing systems.
- Realize more of the potential for GIS to improve understanding of earth processes and human-environment interaction, leading to better policy-making.
- Develop and encourage gathering of local geographic knowledge and increase public participation in that policy-making.
- Improve access to information in the public domain.
- Dramatically increase the visibility of the GIScience discipline, in academia and the world at large.
- Provide a "factual substrate" for knowledge discovery and framework for knowledge-sharing in the form of an important, new, universally accessible educational and reference resource.

References

Craglia, M., M. Goodchild, A. Annoni, G. Camara, M. Gould, W. Kuhn, D. Mark, I. Masser, D. Maguire, S. Liang, E. Parsons (2008). Next-Generation Digital Earth. *International Journal of Spatial Data Infrastructures Research*, 3: 146-167

Digital Earth Prototypes. (Oct 2001) http://www.digitalearth.gov/prototypes.html

"Digital Earth Alpha Versions" (July, 2000). Retrieved from http://www.digitalearth.gov/alpha/alphaVersionDRAFT03.doc, 10 May 2006.

Distributed Geolibraries: Spatial Information Resources. (1999). Report of June, 1998 Workshop: Panel on Distributed Geolibraries; Mapping Science Committee. Washington D.C.: National Academies Press.

Goodchild, M.F. (1998). *The Geolibrary*. In Carver, S. (Ed.), *Innovations in GIS*. (pp. 59-68). London: Taylor & Francis.

Goodchild M F (2000b) *Discrete Global Grids for Digital Earth*. Proceedings of First International Conference on Discrete Global Grids, Santa Barbara, March 2000.

Goodchild M F (2000a) Cartographic Futures on a Digital Earth. *Cartographic Perspectives* 36: 7-11

Gore, A. (1998). *The Digital Earth: Understanding our Planet in the 21st Century*. Retrieved from http://www.digitalearth.gov, 18 February 2006.

Grossner K., Goodchild M and Clarke K (2008). Defining a Digital Earth System. *Transactions in GIS*, 12(1):145-160

Grossner K. and Clarke K. (2007). Is Google Earth, 'Digital Earth' − Defining a Vision. In *Proceedings of the 5th International Symposium on Digital Earth*, Berkeley, CA

IDEW Workshop Proceedings (1999). [consensus definition]

McMaster, R.B. & Usery, E.L. (Eds.). (2005). *A Research Agenda for Geographic Information Science*, Boca Raton, FL: CRC Press.

Appendix A: UCGIS Research Agenda synopsis

McMaster, R.B. & Usery, E.L. (Eds.). (2005). *A Research Agenda for Geographic Information Science*, Boca Raton, FL: CRC Press.

The analysis provided in the introduction to *A Research Agenda for Geographic Information Science* documented a high degree of interrelatedness between research concepts in the original ten "intertwined and overlapping" (p.13) UCGIS research challenges. It is evident that in this case a simple 13x5 matrix won't tell the complete story, nor tell it most efficiently due to the inherent redundancy. However, it is a place to start. The following excerpts and (intentionally and heavily) paraphrased synopses summarize the research challenges as of 2004—and in seven cases where updates have been submitted, as of 2006.

1. Spatial data acquisition and integration (Jensen, *et al.*) *
2. Cognition of geographic information (Montello, Freundschuh)
3. Scale (Lam, *et al.*) *
4. Extensions to Geographic Representations (Yuan *et al.*) *
5. Spatial Analysis and Modeling in a GIS Environment (Getis, *et al.*)
6. Uncertainty (Zhu)
7. Spatial Information Infrastructure (Onsrud, *et al.*)
8. Distributed and Mobile Computing (Goodchild *et al.*) *
9. GIS and Society (Elmes, *et al.*) *
10. Geographic visualization (Bucley, *et al.*)*
11. Ontological Foundations for GIScience (Mark *et al.*)*
12. Remotely acquired data and information (Hepner, *et al.*)
13. Geospatial data mining and knowledge discovery (Yuan *et al.*)*

* Update reports for these were prepared for 2006 UCGIS Summer Assembly, and are accessible at http://www.ucgis.org/priorities/research/2006ResearchNextSteps.htm

Nine Research Challenges

1. Spatial data acquisition and integration [SD] (Jensen, *et al.*)

This area had the greatest conceptual overlap with other areas in the analysis mentioned above, so a large number of the items listed reappear elsewhere. The important questions and/or problems needing to be addressed include:

Generic integration (conflation) issues

- Accurate integration of recent, "more abundant and precise" spatial data with other current and historical datasets to solve complex problems.
- Are there significant gaps between *in situ* and remotely sensed data? What required data is unavailable, and how can it be obtained?
- A "general theoretical and conceptual framework" for integration of diverse types of spatial data: *in situ* and remotely sensed data, imagery and foundation maps at different spatial, spectral, temporal and radiometric resolutions

- Standards for spatial data collection and representation: remain active in their development and adoption; monitor status and implications if ISO 14000 environmental management standards by private business and other organizations.
- Ability to register a diverse array of spatial databases to a single nationally approved datum.
- Improved methods for estimating and reporting positional accuracy in maps and mapping systems
- Classification systems for levels 3-4 urban land use and land cover information
- Accuracy estimation for change detection maps
- Radiometric correction of remote sensor data: edge-matching of digital orthophotoquad imagery; atmospheric correction algorithms to enable historic comparisons
- Metadata: enabling users' "complete understanding of the content and quality of a digital spatial dataset in order to make maximum use of its information potential."
- Address-matching issues: continued participation in the evolving National Map
- Ethical and moral implications regarding privacy and geospatial data collection and availability

In situ data collection

- Instrument calibration; including physical instruments (thermometers, radiometers) and questionnaires; domain-wide best practices are not well known
- Sampling logic; evaluation varied interpolation methods and efficacy of comparing derived surfaces

Census enumeration logic

- To be resolved: the impact of the geographic database used during field enumeration; minimizing error due to instrument problems; selecting data transformations; assessing the quality and accuracy of a census

Remote sensing data collection

- Assessing the adequacy of currently available datasets and data gathering regimes in place for
    - **urban/suburban socioeconomic characteristics**: land use/land cover, building and cadastral infrastructure, transportation infrastructure, socioeconomic characteristics, energy demand and production potential, disaster emergency response
    - **biophysical characteristics**: vegetation, water, soils and rocks, atmosphere

2006 Update

Much progress was made, and links to many relevant projects were provided, however "many of the challenges remain," and the questions asked are fairly basic. The integration problems faced by the Natiaonal Map project were termed "staggering." Not least of these are challenges posed by the wide variation in resolution between data sets.

2. Cognition of geographic information [CO] (Montello, Freundschuh)

Notes

- "Understanding (cognitive processes) will promote interoperability in distributed systems" (p. 61)
- Cognitive view of geographic information is concerned with depiction, communication and understanding
- We require:
    - Ways to externalize divergent belief systems
    - Improvement to approaches to geographic education
    - Improved user interfaces and query languages
    - Better understanding of individual differences among the many kinds of users of geo-information
    - Better understanding of what and how much information people want and can comprehend, and in what formats it should be presented" (p. 62)
- Modeling the cognitive processes of "spatial experts" would be useful
- "A project like Digital Earth would only reach its optimal effectivenss with research on the cognition of geographic information" (p. 62). It will require research on:
    - Display and visualization of complex geographic information
    - Perception of patterns in space and time
    - Integrating multiple sources of information presented in different sensory and representational modalities
    - Comprehension and communication of scale and scale changes
    - Effective natural language interfaces
    - Individual and group differences in cognition of geographic information
    - Gore: "the hard part of taking advantage of this flood of geospatial information will be making sense of it – turning raw data into understandable information" (p. 63)
- Ontologies – "the construction of ontologies, systems of concepts or classes of what exists in the world, is a cognitive act as well as a reflection of objective reality." (p. 74)

Fundamental Research Questions

1. Should data models match cognitive models of space, place and environment more closely?
2. How well are common GIS operations understood?
3. How does categorization help or hinder understanding?
4. Can natural language queries be incorporated into GIS well?
5. The uses of spatial metaphors in organizing non-spatial information are many – will this work in the same interface with spatial information?
6. Can we evaluate the effectiveness of new approaches to depicting error and uncertainty; temporal change (discussed elsewhere); also virtual environments?
7. How can GI technology be used to improve geographic education?

3. Scale [SC] (Lam, *et al.*)

"Temporal scaling is not well understood" (p. 94)

Research priorities (p. 120)

1. Definitions of scale concepts: clarify the importance of scale and resolution in GIS applications

2. Systematized bases for scale-related decision-making: make explicit the scale-related implications of given analyses; identification of critical scales and of scale-invariant data sets.
3. Practical guidance on data integration and use: guidelines on fitness for use of coarse scaled data, with which to discount results when necessary.
4. Quantifying effects of scale in process and statistical models
5. Intelligent automated generalization methods
6. Cognitive issues: how is scale understood by GIS analysts, by other kinds of users? How do users perceive scale changes?

Potential projects

Generalization: multiple versions of model databases
Metadata to convey scale-related info for potential users of data sets
Design and develop a multi-scale database capable of multi-scale analysis

2006 Update

The authors listed recent scale-related projects relevant to only 3 of their 10 listed research sub-areas. They cited a critical need for further study of how scale and uncertainty affect decision-making.

Some other critical questions still to be solved:

- A clearer understanding of the many perspectives of scale (ecology, sociology, architecture) and how they fit into a spatial approach.
- More work on the effect of scale and resolution on geographical processes.
- A better sense of how the human-social and biophysical perspectives fit together.
- Better approaches for smart generalization.
- A better sense of the relationship between geographical processes and operational scales.
- Work on nongeographical (www) scaling issues and scales and networks

4. Extensions to Geographic Representations [GR] (Yuan *et al.*)

(p. 129 quoting Winston, 1984): "a representation is a set of conventions about how to describe a set of things." In other words, it's the schema, which "defines, and limits, the power of any computational system." It "makes the important things explicit," while exposing the "natural constraints inherent in the problem." (Winston, Ibid)

The representations (data models, essentially) chosen for a geographic phenomenon have a profound effect on interpretation and analysis; "circumscribes what information is accountable, computable and visible in an analysis" (p. 130)
Representation for GIS and analysis is at three levels:
Data models, formalization and visualization
<u>The research challenge objectives</u>: expand representations into volumetric (3-D) and dynamic phenomena and develop analytical approaches that support those extensions, especially in very large, distributed databases. Also, representations of data quality.

The same issues apply to numerous kinds of data: physical phenomena, small-scale human movement (transportation), human activities (disease, migration, cultural diffusion, geopolitical events, economic activities, reporting of news, etc.)

[--These extensions to geographic representations will further enable knowledge discovery and representation of knowledge not typically managed in GIS. There is a factual core to human history and a geographic context to most historical events and circumstances. Better understanding might lead to better policy.
-- The narrative is part of the factual record!
-- It's time to translate the recent "theoretical milestones" (p. 150) into real systems, useful for students, researchers and politicians.] -- KG

2006 Update

Donna Peuquet reported that much work in this area remains "new models in old bottles," that is, the use of existing GIS as a platform, when what is required is a change in paradigm, leading to "truly new representations." Visual and cognitive processes were noted to be "intrinsic to the design (of new data models)" that incorporate true object-orientation. A new approach, "integrated with semantics," is needed to achieve better representation of space-time dynamics and to drive graphics for better visualization capability. Peuquet asks the question "how would gamers design a GIS?"

5. Spatial Analysis and Modeling in a GIS Environment [SA] (Getis *et al.*)

Targets of analysis

Disease
Crime
Water and air quality
Traffic, automotive and otherwise
Physical landscape changes; land cover
"The social, cultural and economic trends that manifest themselves on the landscape"
Accessibility and equality of opportunity

High priority

1. Handling massive data sets, and computationally intensive procedures
2. Analyzing spatial, non-spatial and temporal aspects of problems simultaneously
3. Geostatistics
4. Scales appropriate for the phenomena being studied

Notes

1. The number of space-time studies is unusually low
2. Potential common threads amongst various research domain techniques and traditions should be investigated; theory, models and algorithms may differ, but data requirements may be common
3. Growing number of applications in social sciences (economics, sociology and political science) generating "emerging theoretical constructs." [Krugman, 1991 – economic geography]

4. “may well imply a need for broadening the “G” in GIS to deal with locations…in abstract spaces…” → representations in attribute space; a “comprehensive re-examination of geographic representation in spatial analysis.”
5. → again: “formal and computational models of geographic reality strongly condition the questions we can ask and the answers we can receive from spatial analysis.”
6. (Nyerges, 1991) analytical questions
    a. location and extent
    b. distribution and pattern
    c. spatial association
    d. spatial interaction
    e. spatial change

6. Uncertainty [UN] (Zhu)

Notes

Geographic data are often used under the assumption that they are free of errors. The “beguiling attractiveness, the high aesthetic quality of cartographic products from GIS, the analytic capability and the high precision that GIS can provide further contribute to the undue credibility, at times, of these products.” (Abler, 1987 p. 305) Analyses lead to inappropriate decisions. Current state…falls short of Goodchild (1993, p. 98): all data (to) have accuracy metadata; all processes track and report error; accuracy measures a standard feature of GIS.

“There has been little study of the effects of error on policy-making.”

Attributes susceptible to error: theme, location, relationships, temporal. Discrepancy between geographic data and the reality they are intended to represent = uncertainty. Reality cannot be exhaustively measured. Temporal accuracy includes currency.

A research agenda on uncertainty in GIS data and GIS-based analysis

Representation of uncertainty information

In other words, where in the data model? Object-level, vs. geometric entity-level, vs. cell-level: basically, a big question mark.

Detection of Errors

Needed: efficient and consistent automated error detection processes that every data layer undergoes; hierarchical, rules-based approach to detecting inconsistencies, both internal and with/against the “real world” e.g. boundaries against natural boundaries.

Quantification of Uncertainty

Boolean probability in the case of random error vs. fuzzy possibility in the case of vagueness (something belongs in multiple categories, to varying degrees).

Visualization of uncertainty

Two approaches so far: blurring low quality vs. focusing, by filtering out quality below a threshold, with a toggle perhaps.

Propagation of Uncertainty through GIS-based analysis

Scale incompatibility issues need further study

- Propagation analysis
- Impact on decision-making

- Automated updating of metadata

Comments

Poor communication between academic research community and commercial software industry is noted (Zhu); possible help: “streamlining research findings on uncertainty through complete GIS applications from data collection through decision-making processes,” (in other words, exemplars?) Developed by “a coordinated effort in the form of multi-institutional and multi-disciplinary team and test institute…”

7. Spatial Information Infrastructure [SI] (Onsrud *et al.*)

NSDI: “the means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the earth…includes the materials, technology and people necessary to acquire, process, store and distribute such information to meet a wide variety of needs” (NRC, 1993)
“…the term ‘library’ brings to mind the image of an institution in which the works of the government, the private sector and individuals are made available to the public through a decentralized, yet networked national or global system.”
GSDI, Global Map and Digital Earth distinguish particular perspectives. These authors do not make hard distinction between the infrastructure and the geolibrary institutions it may enable.

Notes

- Government role is key
- Continued improving of access is necessary, for enhanced commercial opportunity, and because the data belongs to the public; safeguard and enhance public access
- Licensing issues abound: e.g. repurposed public data as commercial products (deLorme, etc.); licensing to companies versus public access
- Copyright: traditional models versus open-source “commons”
- “In summary, pursuit of explicit conceptual models with substantial potential for providing incentives and overcoming impediments in wide-scale spatial data sharing environments should be a major thrust of the research community.” (p.235)

Institutional Arrangements

Numerous “lessons learned” from FGDC Cooperative Agreements Program (CAP) of incentives, NSDI, Digital Earth Initiative; Digital Earth should be global – GSDI involved.

Standards

- ANSI/L1 → ISO/TC 211 – what is status? Under way since 1995! www.isotc211.org
- What is driving these standards? Users, industry, software makers? Are they/should they be developed in isolation from other domains?

Priority Research Areas

- Information policy: “what governmental information policies and practices are most addvantageous forpromoting a robust spatial information infrastructure?”

- Access to government spatial information: How do policies affect access to and use of spatial data, and how may needs for access be balanced with privacy and cmmunity security concerns?
- Economics of information: Achieve a better understanding of the economic characteristics of information, especially government information.
- Integration and local generation of spatial information: Develop the "technical and institutional means to support creation of local knowledge;" make such knowledge accessible.

Relevant Projects

1. Assess existing geodata production and maintenance projects
2. Develop a geolibrary that meets need of all for access to information in public domain and private holdings; public and privately provided services
3. Develop alternate strategies for increasing public access to government information
4. Examine pricing and cost recovery practices
5. Compare local, state and federal government policies for data dissemination
6. Guidelines for increasing public participation in particular local-scale project that informs decision-making
7. Collaborative project based on local knowledge
8. Model the components of this expanded view of a spatial information infrastructure, focusing on technology and institutional developments, and how they are embedded in other processes and media. (KG: in other words, design a digital earth system as a geolibrary on an enabling infrastructure)

8. Distributed and Mobile Computing [DC] (Goodchild *et al.*)

Mobile computing

"Goodchild (1998) argues that GIS should be seen as an interaction, not only between human and computer, but between human, computer and geographic reality (HCRI rather than HCI)."
Field computing for data acquisition
Sensor networks
A 'real-time' digital earth – at least with scattered coverage [e.g. James Reserve environmental network in Google Earth]

Distributed computing

3-tiered vision:

- Simultaneous access to multiple servers from a single client
- Virtual databases of tables from various physical locations
- Distributed software – modules residing at various locations

[Driving this is distributed production i.e. the bottom-up generation of geographic data and sharing of geographic knowledge discussed in *Spatial Information Infrastructure* above.] -- KG

"If much geographic information is produced locally, and much of it is also consumed locally, there is little reason to integrate data at a broader scale." [Perhaps this is a question of scale, resolution and generalization. Ways of integrating multiple resolutions within a system; “get off (the zoom in a display) at any scale” to do the work best accomplished at that resolution.] --KG

Libraries and archives
“Much of the data used in GIS analysis is *framework* data…” A great deal of this is accessible from many locations now; search mechanisms remain deficient.
“Extend the notion of a distributed geolibrary…”
“Development of such methods (search based on location) should be a major research priority of the geographic information research community.”

Research Issues

- Examine status and compatibility of standards at national and international levels; identify important gaps and duplications; evaluate level of adaptability to technological change
- Develop distributed services
- Develop economic model of the distributed processing of geographic information
- Transmission efficiencies
- Improved models of geographic metadata
- Theories addressing optimal locations of computing activity
- HCRI (Human-Computer-Reality Interaction)
- Adaptive methods of field sampling directed by real-time analysis in the field
- Study the role of contextual information gathered in the field that informs subsequent analysis
- Examine implications of IDU computing: IP rights, e.g.
- Examine social implications of IDU computing
- Applications of IDU in GIS
- System architectures for distributed GIS computing: database design, client-server processing issues, replication and versioning; caching

Benefits

- Access
- Cost reductions
- Improved decision-making
- Distributed custodianship
- Data integration

2006 Update

- Refinement and adoption of data exchange format standards continues to grow, but there is also a “tension” between the broader standards of W3C SOAP and XML with the specifically spatial OGC GML/Open LS and Google’s KML.
- The emphasis on research related to geospatial computing with mobile devices mirrors the rapid growth in applications and demand.
- Geoportals are maturing, with some earlier related concepts being eclipsed, perhaps: geolibraries, clearinghouses and spatial data archives; ostensibly because geoportals offer GIS functionality in web services.

- Daniel Sui's advocacy of "GIS as a medium for communicating what is known about the planet" is contrasted with GIS as an "intelligent personal assistant."

9. GIS and Society [GS] (Elmes *et al.*)

"…how GIS and its uses alter the construction, perception and experience of time and space, through how social processes affect the form taken by the technology itself, to how the spread of GIS technology affects the geographic political, economic, legal and institutional structures of society."

Major Research Perspectives

Institutional

Extent and economics of GIS adoption by organizations; effect on interaction between agencies and between the public and government agencies

Legal and Ethical

Understanding the nature of geographic information vis-à-vis access, privacy, and "surveillant powers."

Intellectual History

"…technological choices (made and) not made at a given time and by given groups." How technologies come to be dominant; alternatives overlooked?

Critical Social Theory

Focus on the limitations of geographic representation; the absence of a single truth, but the implicit suggestion of such by GIS. Technologies marginalize and empower selectively; how?

[DE is in several respects an answer to these critiques – more factual information in more hands; clarification of sources, etc. (Gahegan, *in press*)] -- KG

Public Participation GIS

How may this be encouraged?

Notes

1. This agenda challenge is pivotal in heightening greater interest and involvement in issues regarding GIS by members of other disciplines. Also raising public awareness about the widespread use and usefulness of GIS technologies.
2. "...what possibilities and limitations are associated with using GIS as a participatory tool for democratic resolution of social and environmental conflicts?" Keith Clarke: it can solve the Kashmiri border dispute, for example.
3. GIScientists have ignored the commercial realm, by and large.
4. [DE will bring many of the societal issues mentioned to greater prominence by virtue of its becoming a significant information resource, at or beyond the level of Wikipedia, for example. It offers a venue for demonstration applications; a testbed for putting theory into practice.] -- KG

2006 Update

"It's a very different world than in 1998;" the advent of mass-market geospatial applications and increased focus on geospatial matters "presents a challenge to the academic community if we want to remain involved in research leadership."

Four Emerging Research Themes

10. Geographic Visualization [GV] (Buckley *et al.*)

There is a big overlap of interests between GVis and several Agenda areas, particularly Cognition of Geographic Information. It is "worthy of our attention" notably for educational applications (curricula, courses, and GVis course materials and software).

Definitions

NSF Panel (McCormick et al.., 1987): "a method of computing…a tool for interpreting image data fed into a computer, and for generating images from complex multi-dimensional data sets."

MacEachren et al.. (1992: 101): "…an act of cognition, a human ability to develop representations that allow us to identify patterns and create or impose order."

Notes

- UCGIS takes a broader view of the term Geographic Visualization than what is relevant for the select group of GIS specialists.
- The extensive work of ICA (International Cartographic Association) is foundational.

Perspectives

There are several conceptual/domain approaches to thinking about GVis, with associated needs and lines of inquiry:

- *Spatial analysis*: "potential misuse (or uniformed use) of visualization to convey deceptive results;" identifying "surrogate artifacts."
- *Knowledge discovery and data mining*: representing "collapsed dimensions" on maps; differences in representation for expert and novice users; relationship of GVis to knowledge creation and developing expertise; "multi-stage approach to KD" (i.e. wizards for setting up data and tools environment for analysis).
- *Spatial data representation*: "A disconnect currently exists between visualization and spatial databases…" Inadequate representation of dynamic processes, complex data, spatio-temporal data.
- *Cognition*: Do spatial databases adequately represent spatial phenomena? Is greater correspondence of database representations with human cognition necessary? Better understanding of cognitive requirements in displays communicating the contents of databases certainly *is* required.
- *Collaborative research*: Special requirements for improved interaction between collaborators and between groups and data being considered need better understanding.

Priority Areas for Research

Near-term

- Tools to allow more user interaction in displays
- Understanding the process of knowledge discovery
- Extending representation to include qualitative, intangible, conceptual information; testing cognitive effectiveness.
- New interface metaphors
- Virtual environments
- Tools for collaboration

- Understanding cognitive issues, e.g. individual differences to inform GVis design
- Integrating the dimensions/modes of GVis: spatial, symbolic, temporal/dynamism and animation.
- Non-conventional graphics: morphing, superimposition, multiple viewpoints

Medium-term

- Adding senses to a GVis interface
- Cognitive effectiveness
- Sensory limitations ("different abilities")
- An "'abstraction control,' interactively varying the realism of (the) scene."
- Displaying increasing volumes of data
- Considerations for larger displays – many more pixels
- Considerations for small displays – fewer pixels (mobile, LBS)

Long-term

- Incorporate graphic designers' awareness of aesthetics and perceptual issues
- Automated GVis systems/services: "show me the relationships between landscape conditions and in-stream habitats…"
- Unify visuals from vector and raster data

2006 Update

Rather than list progress and update research priorities,the authors offered an evaluation of the impact of the UCGIS agenda on the larger domain of visualization generally, and stressed recognition of that relationship as key going forward. They state that "raising the prominence and visibility of UCGIS member institutions' research should be a priority"—necessary in fact to ensure against GIScience "simply being left out of the national visualization research agenda." As noted in other agenda updates, the impact of mass-market geographic visualization products like Google Earth, and the growing standards and open systems approach to geographic computing, is significant to UCGIS members, who are urged to expand their important role in the collaboration between academia, government and marketplace.

11. Ontological Foundations for GIScience [ON] (Mark *et al.*)

Definitions and philosophical foundations:

"To be is to be the value of a bound variable." – Quine (1953)

- Defined: (1) a branch of philosophy seeking to understand and describe the "nature and organization of reality." (Guarino and Giaretta, 1995). (2) "a logical theory which gives an explicit, partial account of a conceptualization." (Ibid), and (3) the psychological study of humans' conceptual systems in relation to to "given domains of objects."
- Theories in the natural sciences express *ontological commitments*, that is, underlying presuppositions. Representing reality in computing systems requires formalizing the taxonomies of knowledge domains—as understood by experts in that domain, but also by users of those systems. That is, "canonical descriptions of knowledge domains and associated classificatory theories."
- Universal, shared taxonomies that are both neutral and comprehensive are impossibility.

- What is required is a framework for domain ontologies, with a top-level common, neutral backbone.

Research Priorities

- Develop a complete upper-level ontology for the geospatial domain, that "define(s) taxonomies of the different types of geographic objects, fields, spatial relations, and processes." (p.341)
- Stepwise, develop detailed ontologies for those subdomains that are the principal application realm of GIS.
- These will provide tools and an integration platform that supports cross-disciplinary communication and collaboration.

Kinds of Research (Approaches)

Philosophical: "clarification of the relations between human knowledge, beliefs, and representation on the one hand, the models and representations embedded in our data systems on the other hand, and the real world of objects beyond." (p.342)
Information science: "methods and tools for describing, accessing, comparing, and integrating geo-ontologies." (p. 342)
Psychological: "eliciting geo-ontologies from human subects (both experts and non-experts)" (p. 342)

Research Objectives

Short-term

"Develop and distribute an upper-level ontology for geospatial phenomena that can be used as a common framework to ensure that independently developed subdomain ontologies are consistent and interoperable"

Medium-term

- "Formalization of an ontology of naïve geographical concepts;" in the context of way-finding systems, e.g.
- Follow-up to ontology research with regard to
    - Vagueness: geographic objects with indeterminate boundaries, for example.
    - Scale: "the interaction between map scale, map resolution, and…level of detail"
    - Change and process, e.g. rivers, lakes, storm fronts
    - "Social ontology": political, economic, legal objects in the geographic realm

Long-term

- "Complete the description and formalization of the ontology of all phenomena at geographic scales, including those phenomena dealt with by other natural sciences such as climatology and oceanography." (p. 344)
- Integration with the Semantic Web

Possible Showcase Demonstrations

"The use of ontologies to facilitate retrieval from spatial databases can provide a visual demonstration of the possible outcomes of the research here described." (p. 345)
Intelligent web geo-services and advanced spatial similarity search engines.

2006 Update

"The popularity of the research topic is highlighted by the share of paper published in competitive GIScience research outlets that deal with research in geo-ontologies. Seven of the 25 accepted full papers (i.e., 28%) at GIScience 2004 addressed explicitly aspects of geo-ontologies. Another two conferences held in 2005 had a strong focus on geo-ontologies: the proceedings of GeoS 2005 included 8 out of 19 papers (i.e., 42%) whose explicit focus was on geo-ontologies, and the proceedings of 2005 Workshop on Semantic-Based GISs (SeBGIS 2005) included 5 out of 15 (i.e., 33%) papers discussing issues related to geo-ontologies. The most pertinent subfields at this moment appear to the design of geoontologies (from upper-level ontologies to domain-specific application ontologies), methods for exploiting geo-ontologies for spatial querying (in particular spatial similarity searchers), the role of geo-ontologies within the Geospatial Semantic Web, and methods to manipulate geo-ontologies, such as aligning and fusing geo-ontologies."

12. Remotely Acquired Data and Information in GIScience [RD] (Hepner, et al.)

The rapidly increasing stream (and stores) of remote sensing data are a case of technology outstripping our ability to make sense of the information available to us. Identifying new applications and providing theoretical foundations, tools and methods for them, is the responsibility of the GIScience community.

"Recent interrelated, fundamental changes in remote sensing policy, basic science, technology transfer, and the private-public mixture of investment and control in remote sensing represents a nexus of activity that has heightened the importance of remote sensing to the national research and economic agendas." (p. 354)

Research Objectives

- Further the integration of remote sensing systems, data sources and analysis procedures with other GIS capabilities (GPS, measurement, visualization, data mining, real-time GIS)
- Act as unbiased evaluator of systems
- Assess social, legal and policy implications of new sensor surveillance capabilities.
- Encourage increased use of remote sensing
- Foster conversion and application of military assets and methods to useful civilian ones
- Investigate methods for using historical data sets to measure environmental change
- Links of remote sensing with social sciences

Short-term research priorities

- Spectral signatures provided in imaging spectroscopy from airborne and satellite platforms
- Calibration of digital cameras and sensors
- Oceanographic multi-platform sensing requires "spatially-referenced data management systems for various streams."
- Legal and social implications of increased surveillance capabilities
- Exploitation of high spatial resolution; better resolution is enabling new applications in urban dynamics, natural hazards and agriculture, e.g.

- Investigation of applications in consumer markets

Medium-term research priorities

- Resurrect archival data sets for applications in global environmental change
- Improvements in autonomous control and automated classification
- Development of artificial intelligence/expert systems to enhance data collection and automated classification.

Long-term research priorities

- Analysis of time series information
- Environmental acoustics for imaging and mapping water column and sea floor

13. Geospatial Data Mining and Knowledge Discovery [KD]

"The wealth of geographic data cannot be fully valued when information implicit in data is difficult to discern. This confronts geographic information (GI) scientists with an urgent need for new methods and tools that can intelligently and automatically transform geographic data into information and, furthermore, synthesize geographic knowledge. It calls for new approaches in geographic representation, query processing, spatial analysis, and data visualization." (p.365)

Definitions

From Fayyad *et al.*. (1996): "…the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data."
Knowledge discovery: "a KDD (knowledge discovery in databases) process includes data warehousing, target data selection, transformation and reduction, data mining, model selection (or combination), evaluation and interpretation, and finally consolidation and use of the extracted knowledge" (Fayyad, 1997, p.5).
"KDD aims to enable an information system that transforms the information to knowledge through hypothesis testing and theory formation." (p.366)
"Data mining aims to discover something new from the facts recorded in a database. It prescribes the steps towards efficient development of knowledge discovery applications." (p. 366)

"Almost without exception, current databases and data management systems are designed without thought to KDD…" (p.368)

Research Frontiers and Priority Areas for Research in GKD (Geographic Knowledge Discovery)

Developing and supporting geographic data warehouses
Better spatiotemporal representations in databases
Diverse data types to include imagery and geo-referenced multimedia
Interfaces and tools usable for diverse group of researchers
Proof of concepts (test cases) and benchmarking of effect of varying data quality on results

Short-term objectives

- Apply DM & KDD techniques to the new generations of geospatial data models; identify analytical and visualization needs.

- Survey exiting spatial analysis methods; evaluate and extend their computational ability in large data sets.
- Develop a "…strong geospatial data foundation," involving distributed databases and distributed processing

Medium-term objectives

- Develop a taxonomy of geographic knowledge; categorize models (methods) for GI computing
- Develop a system for geographic knowledge acquisition and synthesis
- Develop robust spatiotemporal representations and capability for complex geographic queries in large, distributed, heterogeneous and dynamic databases
- Develop robust spatial and temporal reasoning and analytical models to support geographic knowledge formulation through interactive and recursive query processes
- Develop multi-dimensional, interactive visualization techniques with dynamic links to distributed GIS databases to greatly enhance user's capabilities to detect hidden patterns and inspect hidden correlations among geographic variables.

Long-term objectives

- Develop an integrated theory of geographic representation, processing, analysis and representation.
- Enable a full implementation of geographic KDD across distributed databases that allow the general public to inspect climate patterns and regional demographic dynamics, for example, on the Internet.

Appendix B: Excerpts and Notes from 'Distributed Geolibraries'

Distributed Geolibraries: Spatial Information Resources. (1999). Report of June, 1998 Workshop: Panel on Distributed Geolibraries; Mapping Science Committee. Washington D.C.: National Academies Press.

Available at no cost in PDF format from National Academies Press at http://darwin.nap.edu/cart/deliver.cgi?&record_id=9460

Vice-President Gore described a vision of Digital Earth that bears substantial resemblance to distributed geolibraries. (p. 84)

A geolibrary is a digital library filled with geoinformation and for which the primary search mechanism is place. Geoinformation is information associated with a distinct area or footprint on the Earth's surface. A geolibrary is distributed if its users, services, metadata, and information assets can be integrated among many distinct locations.

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? What incentives are needed to motivate people to make their works accessible? If a library exists to serve a community, its first responsibility should be to provide the information needed by the community. (p.19)

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. (p. 20)

"Distributed" (p. 20)

The term distributed refers to the locations of the physical and functional parts of the library and the locations of its users. In today's digital world it is possible for functions to occur in multiple locations, held together and coordinated by communications networks like the Internet. Libraries have responded to this new networked environment by establishing coordinated, collaborative, and multi-institutional relationships. The library building no longer houses all of the services it provides to its users; instead, the institution of the library obtains those services in whatever ways maximize effectiveness and minimize costs, by using resources in the building or from a myriad of sites distributed around the globe.

"Geoinformation" (pp. 22-23)

Geoinformation is information that is specific to some part of the Earth's surface or near surface. It includes maps, of course, which abstract and present information about the locations of phenomena on the surface; it also includes images from the air or space (aerial photos or remotely sensed images) that capture the appearance of the surface using energy (either visible or invisible) radiated from it in some part of the electromagnetic spectrum. Such data were earlier defined as geospatial. In addition, geoinformation includes the contents of guidebooks, reports on specific areas, data sets with a geographic dimension, and any other information assets that serve to differentiate one geographic area from another. Finally, it includes information about the atmosphere above the surface, the geology below the surface, and the oceans that cover two-thirds of the planet's surface. All of these information assets are characterized by having some form of associated geographic footprint, a boundary defining the geographic extent of the information, which is the defining characteristic of geoinformation as the term is used here.

“Characteristics”

In a digital world, however, all [of these] objections disappear, apparently without exception. It is possible to present the digital library user with a picture of a globe; search for locations by name, address, or any other suitable and convenient method; allow repositioning and zooming; search distributed archives for information assets whose footprints match the query, present them to the user in sufficient detail to permit evaluation; and deliver them for further examination and analysis. But although a geolibrary is possible in principle, there are countless technical, practical, economic, and institutional problems that will have to be overcome. Moreover, it is unclear how a geolibrary would deal with issues of intellectual property and how it could be paid for and whether the costs would be outweighed by the benefits.

In other words, a distributed geolibrary would constitute a level of services above those provided by the Internet and the WWW, geared to specific user needs. Distributed geolibrary services offer the potential for more intelligent organization and access, for the creation of new knowledge through analysis of raw data, and for the solution of practical problems. As such, distributed geolibraries are one of a number of new types of Internet services that exploit previously impractical ways of organizing and presenting information.

NSDI (p. 28)

"The National Spatial Data Infrastructure is 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" (National Research Council, 1993, p. 2, emphasis added).

Digital Earth (p. 32)

Distributed geolibraries bear a strong resemblance to certain aspects of the concept of Digital Earth…

Like distributed geolibraries, Digital Earth is about making use of the vast but uncoordinated masses of geoinformation now becoming available via the Internet and about presenting it in a form that is readily accessible to the general user. Like distributed geolibraries, its central metaphor for the organization of information is the surface of the Earth and place as a key to information access.

The library service model that underlies the concept of distributed geolibraries provides a useful way of structuring discussion and of thinking about the resources and research that will be needed to make the vision a reality.

Societal and Institutional Issues

Data, Information and Knowledge (p. 44)

Continuum: data, information, knowledge, understanding, wisdom

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. Traditional libraries collect and catalog primarily knowledge works for good reason. The reading and contemplation of works of knowledge such as books and journals provide context and convey meaning. Currently, such works are one of the best means by which we are able to acquire understanding. "Works of knowledge" are largely synonymous with "intellectual works" and are thus the primary expressions protected by our intellectual property laws.

Both types of information (raw data and derived knowledge works) are valuable, depending on the circumstances and the skills and requirements of their users. To a distributed geolibrary they both look like collections of bits with footprints, and both are retrievable using the same mechanisms; distributed geolibraries would provide a unified means of access. In doing so, however, geolibraries raise issues concerning the relative value of the two types of information.

Both forms of information seem indispensable. There are many questions of a geographic nature that cannot be answered by a right answer but require careful reflection based on both data and prior knowledge works. Providing new data query, search, and display capabilities and services may be important in some distributed geolibraries but providing access to digital works of knowledge is likely to be important in all distributed geolibraries.

In summary, a distinction needs to be drawn between raw data and knowledge works because they appear different from the perspective of the functions and services of a library and with respect to intellectual property rights. Although the NSDI is concerned primarily with the production and dissemination of raw geospatial data, distributed geolibraries could also provide an effective mechanism for the dissemination of knowledge.

Access (p. 45)

The concept of access in an institutional distributed geolibrary environment has two major aspects. One involves technical efficiency and effectiveness in finding desired geoinformation, determining its appropriateness and authenticity, linking to and acquiring it, and electronically processing it if needed. To enable such access, knowledge works and databases must exist somewhere on the network with sufficient metadata and tools available in the system to allow these tasks to be accomplished. The second major aspect of access involves the legal and economic ability of the distributed geolibrary as an institution to provide the geoinformation resources desired by its users, either directly or through the network.

Societal and Infrastructure issues and questions

1. How will local needs for and production of geoinformation be accommodated in a library system that has traditionally emphasized access to books and information with a more general than local focus?
2. Libraries are addressing the need for access to electronic information by developing consortia and networks. How will these new institutional arrangements accommodate and affect the development of distributed geolibraries?
3. Traditional libraries play a significant role in archiving and preserving information. Can this role be accommodated by distributed geolibraries?
4. How can distributed geolibraries deal with inequities of access to electronic systems?
5. Will distributed geolibraries have the effect of enhancing more conventional markets for the information they disseminate?
6. Will distributed geolibraries develop as part of existing library arrangements or complement them?
7. Should the services and functions of distributed geolibraries extend beyond providing users with efficient access to geoinformation to include tools to process and analyze information and create new knowledge?
8. How will distributed geolibraries find an appropriate balance between supplying data and supplying knowledge works?
9. How will each custodian site acquire, give access to, and safeguard the geoinformation in its own collections?
10. How will the distributed geolibrary provide instruction and assistance in the use of digital geographic products and databases? Should users from schoolchild to scientist be expected to be their own reference librarians in the distributed geolibraries of the future?
11. As greater numbers of geographic knowledge works and databases are accumulated in the system over time, will it become increasingly difficult to mine useful information from the available flood?
12. How will the records of humankind be conserved in the distributed geolibrary as an institution?
13. While inclusion of traditional works such as maps in library collections caused few personal information privacy concerns in the past, would the geolibrary's provision for access to detailed databases provide a much greater likelihood for personal information privacy intrusions? What are the principles by which distributed geolibraries would operate in order to protect privacy? How may the principles be enforced and what are the means by which safeguards may be provided in distributed environments?
14. If the generation of knowledge works depends on the resources and intellectual contributions of many persons and institutions, how might intellectual property rights

in these works be appropriately accounted for and how might each custodian manage such rights?

15. How can distributed geolibraries assure that geographic knowledge works and databases are not rewritten or revised by government, private firms, or others to their own benefit? That is, how may one assure that databases are authentic?
16. What incentives other than or in addition to future economic rewards could be effective in convincing individuals, businesses, universities, government agencies, and others to make their geographic knowledge works and databases available over a distributed geolibrary network?
17. Who should decide what is in and what is out of a distributed geolibrary? Should there be a gatekeeper, modeled on the function of a library subject specialist, or should distributed geolibraries operate on the principle of caveat emptor?

Services and Functions (p. 53)

The important library function of collection building, which involves the library staff in making careful decisions about what should or should not appear in the library, has no equivalent on the WWW, where there are no gatekeepers or custodians of quality.

Metadata, or data about data, are likely to become much more important, as libraries seek to refine the services they provide by including more and more tools designed to assist in search, evaluation, and use.

The services of a distributed geolibrary fall into several categories, including services for search and retrieval of items of particular interest, item description and display services, data-processing services, and services for collection maintenance and growth.

While location was handled as one of a number of possible forms of subject in the traditional library, it is the primary basis of search in a distributed geolibrary.

There are three reasons for developing distributed geolibrary services. **The first is economic**…To meet the national mandate to make data collected at public expense available to the public, federal agencies are looking for new ways to disseminate data more widely and effectively…Agencies and companies can also sell data and recover income more effectively using the Internet's growing and increasingly reliable tools for electronic commerce. **The second reason involves the decentralization of geoinformation management**. In a distributed geolibrary there is no need for data to be collected in one place; **A third reason for a distributed framework for geolibrary services is the demand for access**. Public access to geoinformation, particularly by students, can support improvements in national levels of geographic literacy by making it possible for classes to obtain information quickly and easily about any part of the Earth's surface.

Services have been described using broad categories of response to demands. Functions are the actual commands or activities that implement services, and a given function may contribute to more than one service. A function can deliver all or part of a service.

Necessary distributed geolibrary functions
- Search by Geographical Location
- Search by Place Name
- Search by Subject Theme or Time Period
- Item Display and Description
- Collection Creation and Maintenance
- Searching over Distributed Assets (unified, collection-level catalog)
- Integration, Analysis, and Manipulation
- Assisting Users
- Assessment and Feedback

Building Distributed Geolibraries (p. 73)

Many issues can only be resolved by constructing and working with prototypes…Large-scale prototypes are sometimes built in part because it is difficult or impossible to know what is possible without such large-scale experimentation. Without building a distributed geolibrary prototype, it may not be possible to identify exactly what it will do successfully and what it will not do. It may be difficult to know at an early stage how much a distributed geolibrary will cost or whether its costs will be exceeded by its benefits.

The many-to-many paradigm is familiar to librarians, who have traditionally acted as brokers between the publishers and the users of information. Thus, the paradigm shift that is occurring in geospatial data dissemination, in part through a process of technological empowerment, provides a strong reason to look to the library as a metaphor for new dissemination models and suggests that the library is a good place to look for models of distributed geolibraries and for solutions to problems and issues that may arise in building them.

Standards and Protocols
Metadata; file formats; interoperability specs
Spatial Data Transfer Standard (SDTS), the scientific data standards HDF and netCDF, the imagery standards TIFF and GeoTIFF, the military standard DIGEST, and many more.

Authority datasets
The provision for bottom-up contribution essential [KG]
- Gazetteers
- Feature-type thesaurus
- Time-period directory
- Biographical directory
- [KG: domain ontologies]

Georeferencing
If distributed geolibraries are to be useful to people who do not understand the complexities of geodetic datums and cartographic projections, it will be necessary for systems to be developed that are capable of hiding such details or making them fully transparent to the user. Thus, a user ought to be able to access data sets in different

projections and based on different datums and expect the system to handle the differences automatically. Such transparency is not yet available in standard geospatial software products and data sets, and its feasibility has not been demonstrated.

Standard hierarchical grids, [QTM, CSU icosohedral; Pyxis; etc.] Such hierarchical systems may be important internally as indexing schemes for distributed geolibraries (Goodchild and Yang, 1992).

Cataloging (p. 80)

Cataloging, which serves the critical function of abstracting the information users need to find, examine, assess, and retrieve data. In effect, metadata are the key to the many-to-many structure that allows many users to search across many potential suppliers, and its timely creation will be crucial if distributed geolibraries are to function.

The geospatial data community appears to have accepted the notion that metadata creation is largely the responsibility of the producer, whereas the prevailing notion in the library community is that cataloging is the responsibility of the librarian. This reflects a distinct difference in philosophy, since the library practice is based on the notion that ***the librarian may be more skilled in abstracting information*** on behalf of the user than is the producer of the information.

To be successful, a search service designed to help the user of distributed geolibraries find geospatial data and geographic knowledge would have to place heaviest emphasis on the determination of an information object's geographic footprint, either by detecting or inferring coordinates or by identifying an appropriate place name, to be converted to coordinates using a gazetteer. Such tools would perform the functions of abstracting and metadata creation automatically. *Such automated discovery, indexing, and abstracting tools do not yet exist and will require extensive research and development.*

[KG: this leaves aside cataloging with domain-specific ontologies]

Visualization

Some types of geoinformation illustrate close approximations to actual appearance and can be rendered by draping onto a curved surface.

Other information in distributed geolibraries is not rendered so easily. How, for example, would one portray economic information such as average household income using the Digital Earth metaphor? [KG: chloropleth maps] In some cases there may be clever ways of making visible what is normally invisible; in other cases it may be necessary to represent the presence of information using symbols that exploit some other metaphor, such as books or library shelves. *This is a novel area with no obvious guideposts, and research will be needed to determine how best to make the user of distributed geolibraries aware of the existence of information and of its important characteristics*. In particular, we know almost nothing about how to render *dynamic geospatial data* or how to indicate availability, yet we anticipate that such data will be increasingly available to the users of distributed geolibraries. *Current software products are generally incapable of these functions, and much research remains to be done to make them generally available. (p 82)*

Knowledge Construction

Users of distributed geolibraries will need tools for analysis, modeling, simulation, decision making, and the creation of new geographic knowledge. An important component will be the workspace in which the user can process data using many of the functions found in today's GIS…

Research Issues

Many of the topics discussed in this report fall under the heading of "things we do not yet know how to do." In some cases, such as the building of a distributed geolibrary itself, there may be no obviously missing piece of theory or understanding; rather, it may be that we have not yet tried and that given sufficient resources the necessary knowledge will be available. But other items require more focused research. Among them are the following:

- Scalability. We have no experience with building and operating datahandling systems on the massive scales envisioned here.
- Interface design. Most information technologies are designed for skilled users. Distributed geolibraries will be used by everyone, over a wide range of levels of cognitive understanding, and will require new methods of interface design that embody sound principles, some of which have yet to be discovered.
- Merging data. We have very little experience with the massive redundancy anticipated in distributed geolibraries, where many sources of the same data will be available. We do not have techniques formerging data from different sources, across different scales and levels of accuracy, or across different data models or ontologies, or for combining or conflating the desirable properties of sources. Distributed geolibraries will be one of a growing number of applications that depend on the ability to register multiple data sets quickly and easily and to remove obvious discrepancies.
- Indexing. Our methods of indexing data have been developed for the flat two-dimensional world of maps and images. Distributed geolibraries will require comprehensive approaches to indexing that are capable of supporting "drilling down" over a wide range of scales.
- Visualization. While techniques for visualizing static two-dimensional data are well understood, particularly in cartography, we do not have the same level of understanding of appropriate ways to visualize data on the curved surface of the Earth, especially when the data are time dependent. Much more research is needed into appropriate metaphors, techniques, and user responses before these will be as easy as traditional cartographic visualization.

Report Findings

(These are distributed throughout the report)

FINDING 1

A wide variety of human activities could benefit from the services of distributed geolibraries. They include many where the timely pro-vision of information could minimize loss of life or result in more timely and effective use of existing information resources and others where the costs of bad decisions could be avoided.

FINDING 2
Although many projects currently exhibit elements of the vision of distributed geolibraries, the lack of a clear statement of that vision impedes coordination and leads to duplication of effort. A clear statement can provide a sense of common purpose.

FINDING 3
The contents of a distributed geolibrary are not limited to information normally associated with maps or images of the Earth's surface but include any information that can be associated with a geographic location. In this sense the vision extends far beyond the context of the NSDI.

FINDING 4
The vision of the NSDI as expressed by the Mapping Science Committee in 1993 (National Research Council, 1993) did not anticipate the enormous impact and potential of the Internet and WWW. By emphasizing the problems of production of digital geoinformation, it underemphasized the importance of effective processes of dissemination to users of geoinformation. User communities are growing rapidly and are likely to grow even more rapidly if the current difficulties associated with finding geoinformation on the Internet can be addressed.

FINDING 5
Distributed geolibraries provide a useful framework for discussion of the issues of dissemination associated with the NSDI. The vision is readily extendible to a global context.

FINDING 6
Developers of distributed geolibraries will need to consider issues related to intellectual property rights. There are significant differences in both the public access library model and the commercial bookstore model that need to be considered in the broader international debates about the nature of electronic information and databases as intellectual property.

FINDING 7
A distributed geolibrary would support collaborative work, such as multidisciplinary research by teams, decision making by groups of stakeholders, and classroom projects by groups of students. It would provide mechanisms for capturing the knowledge that results from such work and making it accessible to others as appropriate.

FINDING 8
A distributed geolibrary would allow users to specify a requirement, search across the resources of the Internet for suitable geoinformation, assess the fitness of that information for use, retrieve and integrate it with other information, and perform various forms of manipulation and analysis. A distributed geolibrary would thus integrate the functions of browsing the WWW with those of GIS and related technologies.

FINDING 9
Many important applications of distributed geolibraries are best located in the field, using portable systems and wireless communications. Delivery of services to the field is important in emergency management, agriculture, natural resource management, and many other applications.

FINDING 10
There are several alternative architectures for distributed geolibraries, including a single enterprise sponsored by a well-resourced agency, analogous to a national library; a network of enterprises with their own sponsors, analogous to a network or federation of libraries; and a loose network held together by shared protocols, analogous to the WWW.

FINDING 11
New technological initiatives such as the Next-Generation Internet and Internet II are likely to provide extensions to Internet and WWW protocols and orders of magnitude increases in bandwidth. Many of these developments are expected to be relevant to distributed geolibraries.

FINDING 12
A comprehensive gazetteer, linking named places and geographic locations, would be an essential component of a distributed geolibrary. A national gazetteer would be a valuable addition to the framework data sets of the NSDI. These framework data sets are being coordinated by the FGDC, which also has the responsibility for associated standards and protocols. Production and maintenance of the national gazetteer could be through the National Mapping Division of the U.S. Geological Survey (USGS) in collaboration with other agencies and could be an extension of the USGS's Geographic Names Information System.

FINDING 13

The success of a distributed geolibrary will be largely dependent on the ability to integrate information available about a place. That ability is severely impeded today by differences in formats and standards, access mechanisms, and organizational structures. Removal of impediments to integration should become a high priority of government agencies that provide geospatial data.

FINDING 14

Significant research problems will have to be solved to enable the vision of distributed geolibraries. Research is needed on indexing, visualization, scaling, automated search and abstracting, and data conflation. Research on these issues targeted to improve access to integrated geoinformation might be pursued by the National Science Foundation and other agencies sponsoring basic science, as well as by the National Mapping Division of the USGS, and the National Imagery and Mapping Agency.

FINDING 15

While traditional production of geospatial data has been relatively centralized, the vision of distributed geolibraries represents a broadly based restructuring of past institutional arrangements for the dissemination of geospatial data and one that is much more bottom-up, decentralized, and voluntary.

Appendix C. The Gore Digital Earth Speech

Gore, A. (1998). *The Digital Earth: Understanding our Planet in the 21st Century*. Retrieved from http://www.digitalearth.gov, 18 February 2006.

The Digital Earth: Understanding our planet in the 21st Century

by Vice President Al Gore
Given at the California Science Center, Los Angeles, California, on January 31, 1998.

A new wave of technological innovation is allowing us to capture, store, process and display an unprecedented amount of information about our planet and a wide variety of environmental and cultural phenomena. Much of this information will be " georeferenced" - that is, it will refer to some specific place on the Earth's surface.

The hard part of taking advantage of this flood of geospatial information will be making sense of it. - turning raw data into understandable information. Today, we often find that we have more information than we know what to do with. The Landsat program, designed to help us understand the global environment, is a good example. The Landsat satellite is capable of taking a complete photograph of the entire planet every two weeks, and it's been collecting data for more than 20 years. In spite of the great need for that information, the vast majority of those images have never fired a single neuron in a single human brain. Instead, they are stored in electronic silos of data. We used to have an agricultural policy where we stored grain in Midwestern silos and let it rot while millions of people starved to death. Now we have an insatiable hunger for knowledge. Yet a great deal of data remains unused.

Part of the problem has to do with the way information is displayed. Someone once said that if we tried to describe the human brain in computer terms, it looks as if we have a low bit rate, but very high resolution. For example, researchers have long known that we have trouble remembering more than seven pieces of data in our short-term memory. That's a low bit rate. On the other hand, we can absorb billions of bits of information instantly if they are arrayed in a recognizable pattern within which each bit gains meaning in relation to all the others — a human face, or a galaxy of stars.

The tools we have most commonly used to interact with data, such as the "desktop metaphor" employed by the Macintosh and Windows operating systems, are not really suited to this new challenge. I believe we need a "Digital Earth". A multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data.

Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head-mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man-made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a "magic carpet ride" through a 3-D visualization of the terrain. Of course, terrain is only one of the many kinds of data with which she can interact. Using the systems' voice recognition capabilities, she is able to request information on land cover, distribution of plant and animal species, real-time weather, roads, political boundaries, and population. She can also visualize the environmental information that she and other students all over the world have collected as part of the GLOBE project. This information can be seamlessly fused with the digital map or terrain data. She can get more information on many of the objects she sees by using her data glove to click on a hyperlink. To prepare for her family's vacation to Yellowstone National Park, for example, she plans the perfect hike to the geysers, bison, and bighorn sheep that she has just read about. In fact, she can follow the trail visually from start to finish before she ever leaves the museum in her hometown.

She is not limited to moving through space, but can also travel through time. After taking a virtual field-trip to Paris to visit the Louvre, she moves backward in time to learn about French history, perusing digitized maps overlaid on the surface of the Digital Earth, newsreel footage, oral history, newspapers and other primary sources. She sends some of this information to her personal e-mail address to study later. The time-line, which stretches off in the distance, can be set for days, years, centuries, or even geological epochs, for those occasions when she wants to learn more about dinosaurs.

Obviously, no one organization in government, industry or academia could undertake such a project. Like the World Wide Web, it would require the grassroots efforts of hundreds of thousands of individuals, companies, university researchers, and government organizations. Although some of the data for the Digital Earth would be in the public domain, it might also become a digital marketplace for companies selling a vast array of commercial imagery and value-added information services. It could also become a "collaboratory"-- a laboratory without walls — for research scientists seeking to understand the complex interaction between humanity and our environment.

Technologies needed for a Digital Earth

Although this scenario may seem like science fiction, most of the technologies and capabilities that would be required to build a Digital Earth are either here or under development. Of course, the capabilities of a Digital Earth will continue to evolve over time. What we will be able to do in 2005 will look primitive compared to the Digital Earth of the year 2020. Below are just a few of the technologies that are needed:
Computational Science: Until the advent of computers, both experimental and theoretical ways of creating knowledge have been limited. Many of the phenomena that experimental scientists would like to study are too hard to observe - they may be too small or too large, too fast or too slow, occurring in a billionth of a second or over a billion years. Pure theory, on the other hand, cannot predict the outcomes of complex natural phenomena like thunderstorms or air flows over airplanes. But with high-speed computers as a new tool, we can simulate phenomena that are impossible to observe, and simultaneously better understand data from observations. In this way, computational science allows us to overcome the limitations of both experimental and theoretical science. Modeling and simulation will give us new insights into the data that we are collecting about our planet.

- Mass Storage: The Digital Earth will require storing quadrillions of bytes of information. Later this year, NASAs Mission to Planet Earth program will generate a terrabyte of information each day. Fortunately, we are continuing to make dramatic improvements in this area.
- Satellite Imagery: The Administration has licensed commercial satellites systems that will provide 1-meter resolution imagery beginning in early 1998. This provides a level of accuracy sufficient for detailed maps, and that was previously

only available using aerial photography. This technology, originally developed in the U.S. intelligence community, is incredibly accurate. As one company put it, "It's like having a camera capable of looking from London to Paris and knowing where each object in the picture is to within the width of a car headlight."

- Broadband networks: The data needed for a digital globe will be maintained by thousands of different organizations, not in one monolithic database. That means that the servers that are participating in the Digital Earth will need to be connected by high-speed networks. Driven by the explosive growth of Internet traffic, telecommunications carriers are already experimenting with 10 gigabit/second networks, and terrabit networking technology is one of the technical goals of the Next Generation Internet initiative. The bad news is that it will take a while before most of us have this kind of bandwidth to our home, which is why it will be necessary to have Digital Earth access points in public places like children's museums and science museums.
- Interoperability: The Internet and the World Wide Web have succeeded because of the emergence of a few, simple, widely agreed upon protocols, such as the Internet protocol. The Digital Earth will also need some level of interoperability, so that geographical information generated by one kind of application software can be read by another. The GIS industry is seeking to address many of these issues through the Open GIS Consortium.
- Metadata: Metadata is "data about data." For imagery or other georeferenced information to be helpful, it might be necessary to know its name, location, author or source, date, data format, resolution, etc. The Federal Geographic Data Committee is working with industry and state and local government to develop voluntary standards for metadata.

Of course, further technological progress is needed to realize the full potential of the Digital Earth, especially in areas such as automatic interpretation of imagery, the fusion of data from multiple sources, and intelligent agents that could find and link information on the Web about a particular spot on the planet. But enough of the pieces are in place right now to warrant proceeding with this exciting initiative.

Potential Applications

The applications that will be possible with broad, easy to use access to global geospatial information will be limited only by our imagination. We can get a sense of the possibilities by looking at today's applications of GIS and sensor data, some of which have been driven by industry, others by leading-edge public sector users:

Conducting virtual diplomacy: To support the Bosnia peace negotiations, the Pentagon developed a virtual-reality landscape that allowed the negotiators to take a simulated aerial tour of the proposed borders. At one point in the negotiations, the Serbian President agreed to a wider corridor between Sarajevo and the Muslim enclave of Gorazde, after he saw that mountains made a narrow corridor impractical.

Fighting crime: The City of Salinas, California has reduced youth handgun violence by using GIS to detect crime patterns and gang activity. By collecting information on the distribution and frequency of criminal activities, the city has been able to quickly redeploy police resources.

Preserving biodiversity: Planning agencies in the Camp Pendelton, California region predict that population will grow from 1.1 million in 1990 to 1.6 million in 2010. This region contains over 200 plants and animals that are listed by federal or state agencies as endangered, threatened, or rare. By collecting information on terrain, soil type, annual rainfall, vegetation, land use, and ownership, scientists modeled the impact on biodiversity of different regional growth plans.

Predicting climate change: One of the significant unknowns in modeling climate change is the global rate of deforestation . By analyzing satellite imagery, researchers at the University of New Hampshire, working with colleagues in Brazil, are able to monitor changes in land cover and thus determine the rate and location of deforestation in the Amazon. This technique is now being extended to other forested areas in the world.

Increasing agricultural productivity: Farmers are already beginning to use satellite imagery and Global Positioning Systems for early detection of diseases and pests, and to target the application of pesticides, fertilizer and water to those parts of their fields that need it the most. This is known as precision farming, or "farming by the inch."

The Way Forward

We have an unparalleled opportunity to turn a flood of raw data into understandable information about our society and out planet . This data will include not only high-resolution satellite imagery of the planet, digital maps, and economic, social, and demographic information. If we are successful, it will have broad societal and commercial benefits in areas such as education, decision-making for a sustainable future, land-use planning, agricultural, and crisis management. The Digital Earth project could allow us to respond to manmade or natural disasters - or to collaborate on the long-term environmental challenges we face.

A Digital Earth could provide a mechanism for users to navigate and search for geospatial information - and for producers to publish it. The Digital Earth would be composed of both the "user interface" - a browsable, 3D version of the planet available at various levels of resolution, a rapidly growing universe of networked geospatial information, and the mechanisms for integrating and displaying information from multiple sources.

A comparison with the World Wide Web is constructive. [In fact, it might build on several key Web and Internet standards.] Like the Web, the Digital Earth would organically evolve over time, as technology improves and the information available expands. Rather than being maintained by a single organization, it would be composed of both publically available information and commercial products and services from thousands of different organizations. Just as interoperability was the key for the Web, the ability to discover and display data contained in different formats would be essential.

I believe that the way to spark the development of a Digital Earth is to sponsor a testbed, with participation from government, industry, and academia. This testbed would focus on a few applications, such as education and the environment, as well as the tough technical issues associated with interoperability, and policy issues such as privacy. As prototypes became available, it would also be possible to interact with the Digital Earth in multiple places around the country with access to high-speed networks, and get a more limited level of access over the Internet.

Clearly, the Digital Earth will not happen overnight.

In the first stage, we should focus on integrating the data from multiple sources that we already have. We should also connect our leading children's museums and science museums to high-speed networks such as the Next Generation Internet so that children can explore our planet. University researchers would be encouraged to partner with local schools and museums to enrich the Digital Earth project — possibly by concentrating on local geospatial information.

Next, we should endeavor to develop a digital map of the world at 1 meter resolution.

In the long run, we should seek to put the full range of data about our planet and our history at our fingertips.

In the months ahead, I intend to challenge experts in government, industry, academia, and non-profit organizations to help develop a strategy for realizing this vision. Working together, we can help solve many of the most pressing problems facing our society, inspiring our children to learn more about the world around them, and accelerate the growth of a multi-billion dollar industry.