

Research Article

Defining a Digital Earth System

Karl E. Grossner
Department of Geography
University of California, Santa
Barbara

Michael F. Goodchild
Department of Geography
University of California, Santa
Barbara

Keith C. Clarke
Department of Geography,
University of California, Santa
Barbara

Abstract

In a 1998 speech before the California Science Center in Los Angeles, then US Vice-President Al Gore called for a global undertaking to build a multi-faceted computing system for education and research, which he termed “Digital Earth.” The vision was that of a system providing access to what is known about the planet and its inhabitants’ activities – currently and for any time in history – via responses to queries and exploratory tools. Furthermore, it would accommodate modeling extensions for predicting future conditions. Organized efforts towards realizing that vision have diminished significantly since 2001, but progress on key requisites has been made. As the 10 year anniversary of that influential speech approaches, we re-examine it from the perspective of a systematic software design process and find the envisioned system to be in many respects inclusive of concepts of distributed geolibraries and digital atlases. A preliminary definition for a particular *digital earth system* as: “a comprehensive, distributed geographic information and knowledge organization system,” is offered and discussed. We suggest that resumption of earlier design and focused research efforts can and should be undertaken, and may prove a worthwhile “Grand Challenge” for the GIScience community.

1 Introduction

The term Digital Earth, coined by former US Vice-President Al Gore in the 1990s, refers to a visionary information system of enormous scope and with significant potential

Address for correspondence: Karl Grossner, Department of Geography, 4721 Ellison Hall University of California, Santa Barbara, Santa Barbara, CA 93106-4060, USA. E-mail: karlg@geog.ucsb.edu

value for education and collaborative research (Gore 1992, 1998). At its furthest extension, Digital Earth is arguably a digital “mirror world,” storing and managing access to everything that is known about the planet. 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. In a speech at the California Science Center in 1998, Mr. Gore spoke of Digital Earth in broad strokes, but also offered a grab-bag of specific ideas (Gore 1998). If a particular *digital earth system*¹ is to be defined, it is worth examining those ideas in some detail and to answer the first question, “is Digital Earth one thing or many things?” Our aim is not to point out language discrepancies, but to aggregate the various suggestions – ultimately including the goals of existing related initiatives – to arrive at a brief, operational definition of a particular, buildable geographic computing system. The results may seed further discussions in the geographic information science (GIScience) community as to whether a Digital Earth project represents a suitable “Grand Challenge” for the near future, as has been suggested. The Gore speech was a call to action that had immediate impact and remains motivational to many people world-wide. A number of related projects were undertaken and some are ongoing, but the system as described in the speech, one that “puts the full range of data about our planet and our history at our fingertips,” remains largely unrealized.

A frequently quoted passage from that speech begins, “Imagine, for example, a young child going to a Digital Earth exhibit at a local museum . . .” and goes on to describe an extraordinary, publicly available software program with a distinctly educational focus. There is also mention of a “collaboratory – a laboratory without walls” for use by research scientists, and multiple references to sophisticated modeling activities they might perform. A third category of users included most current users of geographic information system (GIS) software in governmental agencies – and importantly, consumers of their output; for example, the technicians, analysts and policy-makers involved in everything from land use, transportation and emergency services planning to global geopolitical strategizing and diplomacy. Presumably, the tasks and requirements of these three groups of users are not the same – are they to launch the same Digital Earth program on their computer desktops? The speech appears to have it several ways. As a single software program, Digital Earth is “a multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data.” That program is “composed of the (browsable, 3-D) ‘user interface’ . . . , a rapidly growing universe of networked geospatial information, and the mechanisms for integrating and displaying information from multiple sources” comprising “both publicly available information and commercial products and services from thousands of different organizations.” Understandably, much is made of the need for interoperability – standard formats and protocols that ensure “geographical information generated by one kind of application software can be read by another.”

Some confusion stems from terms like *a representation*, and *the interface*. If we relax those into their plural forms, *representations* and *interfaces*, we turn what was something of a riddle into something that can be coherently defined. Although the consensus definition that emerged from the NASA-led Interagency Digital Earth Working Group (IDEW) in 1999 read, “Digital Earth will be *a* (italics added) 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” (IDEW n.d.), in fact, IDEW members began development of *Digital Earth Alpha Versions* – multiple applications intended to “advance the Digital Earth (DE) in the near-term, while the DE

community team works to establish a more formal strategy, process, and organization for the long-term program” (IDEW 2001b). Political exigencies interrupted that team’s work in late 2001.

In this article we offer first steps at resuming a systematic development effort like that undertaken for the Alpha Versions. Section 2 is a brief review of the federally coordinated work that followed the Gore speech. In Section 3 we distinguish the current Digital Earth “movement” from the particular large-scale systems advocated herein. In Section 4, the text of the Gore speech is parsed to generate a list of preliminary requirements, and Section 5 briefly surveys some of the progress on them to date. Sections 6 through 8 initiate a renewed design process in these preliminary steps: enumeration of the actors involved in use cases, discussion of organizing metaphors and a restated definition of a *digital earth system*. Finally, in Section 9 we propose some high-level requisites, which point to the remaining research challenges.

2 Federal Follow-up

A US Vice-President can motivate action, and so between 1998 and 2001, a “Digital Earth Initiative,” coordinated by the IDEW, and chaired by the National Aeronautics and Space Administration (NASA), sought to realize the Gore vision according to priorities outlined in the speech: “In the first stage, we should focus on integrating the data from multiple sources that we already have” (Gore 1998).

The federal Digital Earth Initiative was a collaborative grouping of entities and individuals from government, industry, academia, and the public sectors with a stated mission to “accelerate key areas of technology and associated policy infrastructure that are hampering full realization of the Digital Earth vision” (IDEW 2001c). Specifically, it would seek to “to improve the integration of and application of geospatial data for visualization, decision support, and analysis” (IDEW 2001c). As such, IDEW activities focused on interoperability, infrastructure and organizational issues far more than design of a system like the one described in the Gore speech. Government participants included representatives from NOAA, USGS, USACE, EPA, USDA and NSF.² Major standards associations involved included the Open Geospatial Consortium (OGC), the Global Spatial Data Infrastructure Association (GSDI) and the International Standards Organization (ISO).

Results of the three-year IDEW effort included: collaborative development of the current widely accepted Web Mapping Service (WMS) standard, a Digital Earth Alpha Version prototype demonstrating a unified interface for distributed WMS databases, and a Digital Earth Reference Model (DERM) intended to “define the standards and architecture guidelines of Digital Earth” (IDEW 2001a). Additionally, design and development of the aforementioned Alpha Version projects was undertaken, but not directly funded and had stalled by late 2001. These were climate and weather applications based on “user context scenarios” for museums, classroom education, government and journalism (IDEW 2001b). Some current work carries on in spirit; for example, the Linked Environments for Atmospheric Discovery (LEAD) project received five-year NSF funding in 2003 to “make meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves” (Droegemeier et al. 2004).

The Digital Earth Initiative banner raised by the US Government after Gore’s speech flew over a spectrum of activities that had been under way for some years. Many of

them continue to this day and will likely survive changes to working group names and bureaucratic structures. When that banner was lowered in late 2001, the coordination of related activities was taken up by the Geospatial Applications and Interoperability (GAI) working group, a part of the US Federal Geographic Data Committee (FGDC), itself formed in 1990. The GAI charter (FGDC 2004) included language evocative of Digital Earth without mentioning it explicitly:

“(responsibility to) develop and maintain the framework for digital representations of the Earth that enable a person to explore and interact with the vast amounts of natural and cultural information gathered about the Earth. Developments to support this framework should facilitate the integration of multi-dimensional, multi-scale, multi-resolution, seamless data that is readily accessible and enhanced through distributed value-added services.”

GAI was described elsewhere as “an outgrowth of the Digital Earth Initiative,” and remained active until mid-2004 (Evans 2003). Over the next two years, that working group produced a Geospatial Interoperability Reference Model (GIRM)³, described as a tool rather than a set of prescriptive standards. Its authors meant to steer clear of “policies such as human interface guidelines, data content or portrayal requirements, or conventions for data storage or georeferencing,” which *were* the purview of the parent FGDC, but outside the scope of GIRM. The GAI group’s work and responsibilities have since been distributed within the FGDC (FGDC 2004). To all appearances, the “interoperability” part of the GAI acronym remains a key focus; it is unclear whether “applications” are still of interest. Certainly, mention of Digital Earth applications has vanished, at least from government material available on the Web.

We argue IDEW was on the right track in 2001, in terms of its strategy to flesh out requirements with the exemplar “Alpha Versions,” while taking next steps at definition and organization (IDEW 2001b). Several years of progress for interoperability standards in this intervening period mean that exemplars can now include – along with end-user software – a platform home, given some organizing body and project leadership.

3 A Computing Movement

Although the term Digital Earth initially referred to a particular visionary computing system, it has come to represent a largely unorganized global technological initiative, and perhaps an intellectual, or social movement. In that sense, Digital Earth – capital D, capital E – is so far the array, or union, of all networked applications representing some aspect of Earth and its history, as well as an evolving enabling infrastructure of standards and hardware involved in delivering them. However, the underlying public and private data stores belonging to myriad organizations are typically unconnected. It is possible that by virtue of the seemingly organic process clearly under way, a complex of systems will emerge that together “put the full range of data about our planet and our history at our fingertips” (Gore 1998). We can presume Google will approach this goal in some fashion, given its stated mission to “organize the world’s information and make it universally accessible and useful by referencing its location on earth” (Golden 2006). Google Earth director John Hanke has said, “We believe Google Earth is an excellent medium for organizing and sharing the world’s geographic information and

Table 1 Functionality Actions users could perform in the Digital Earth system described in the Gore speech – narrowly stated where offered, in cases quite general

-
- Embed georeferenced data in any quantity (i.e. contribute, publish)
 - View the Earth, i.e. multi-resolution imagery, photographs and other data (to 1 m per pixel) at multiple scales, from multiple viewpoints and able to simulate motion; e.g. animated, still; zoom, pan; orthogonal, oblique
 - Locate information at various levels of granularity by means of *browsing* (maps, lists), *direct queries* and *hyperlinks* to associated data stores
 - Create visualizations of uploaded data
 - Travel through time; display conditions at any place for those time periods the system is aware of, from Mesozoic into the future in the case of predictive models
 - Take “virtual tours” of museums, e.g. Louvre
 - Listen to oral histories and music
 - Collaborate with others in scientific inquiry
 - “Predict the outcomes of complex natural phenomena”
 - “Simulate phenomena that are impossible to observe”
 - Create intelligent software agents that aggregate information found within the system automatically
 - Send content and/or links to content to email recipients
-

we continue to explore opportunities to bring visually compelling and informative content into Google Earth” (Hanke 2006).

Two premises of this article are that: (1) GIScience research should inform both corporate and academic efforts; and (2) a concerted and coordinated design effort for *particular digital earth systems* could accelerate and mold a remarkable undertaking – one possibly deserving of the label, “Grand Challenge.”

4 Parsing the Speech: Software and a Computing Platform

We have examined the text of Gore’s 1998 speech and extracted paraphrased or quoted requirements, grouping them into four “bins”: Functionality, Content, Interfaces and System Architecture (Tables 1 through 4). This step allows us to gauge progress to date (Section 5), to identify use cases and distinct client application domains (Section 6), and to assess what is missing and what is unrealistic. It becomes immediately obvious that the envisioned system can be divided as: (1) multiple application software programs; and (2) a computing platform, i.e. the integrative “middleware” and hardware infrastructure upon which they run.

The term platform has a somewhat flexible usage, referring to many kinds of operating environments, including those comprising hardware, software or both. The requirements discussed in the Gore speech sketch a system architecture and programming platform: thousands of disparate organizations make “quadrillions of bytes of information,” stored on a globally distributed array of networked servers, available to multiple user communities, whose members use disparate software applications. Data from all sources can be merged, even “seamlessly fused” in users’ desktop software, by

Table 2 Content discussed in the 1998 Gore speech

-
- Vast quantities of georeferenced information about environmental and cultural phenomena on and near the Earth's surface
 - Landsat photography
 - "A digital map of the world at 1 meter resolution."
 - A global digital elevation model (DEM)
 - Data layers with global coverage for:
 - land cover
 - distributions of plant and animal species
 - roads
 - political boundaries
 - population
 - real-time weather
 - Directly sensed or observed environmental data with coverage of individual research projects, including "citizen science" efforts like GLOBE
 - Hiking trails and other features in national parks
 - "Value-added information services," e.g. geocoding, routing, processed compilations of census statistics
 - Representations of museum collections
 - Historical data and media content with global coverage for political and cultural topics, e.g. newsreel footage, oral histories, newspaper articles and 'other primary sources.'
 - Pre-historical data, e.g. about dinosaurs
 - Modeled thunderstorms
-

Table 3 User interface elements from the 1998 Gore speech

-
- A "browsable 3D version of the planet"
 - In public exhibits, such as at a museum, head-mounted display and data glove to provide immersive experiences
 - Hyperlink navigation
 - Speech recognition capability
 - Audio capability
-

virtue of a suite of standard formats, protocols and requirements for communication and metadata. For the authors of software intended to access and manipulate the distributed content, this constitutes a development platform. A "Digital Earth-aware," or "DE-capable," program would have the built-in facilities required for either one-way (download) or two-way (upload/contribute as well) communication with one or several clearinghouses, i.e. server hosts offering necessary middleware services. Most of the communications technology involved in this scenario is readily available. However, a number of large issues remain open, including the content and structure of metadata and data models for central databases, depending as they do on the actual and potential requirements of as yet unspecified client software applications. The preliminary requirements from the speech are a starting point for such specification.

Table 4 System architecture described or inferred in the 1998 Gore speech

-
- Databases, content stores, application software are all distributed, i.e. maintained by thousands of organizations worldwide; some are in the public domain, some part of a digital marketplace
 - In aggregate, “quadrillions of bytes of information”
 - Participating servers and access points all on a “high-speed network” (given presumptions of bandwidth limits in 1998)
 - Standard formats, protocols, software and metadata requirements that allow “information generated by one kind of application software to be read by another”
 - Enables the display, integration, and *fusion* of data from multiple sources
 - Individuals are able to “publish” to the system
 - Two levels of functionality – a full level for users on Internet2, and “a more limited level” for consumer-grade Internet access.
-

5 Notable Progress

In ten years, many advances have been made in key aspects of Functionality, Content, User Interfaces and System Architecture listed above. It is beyond the scope of this paper to review them all but we note a few to highlight what we believe is the immediacy and relevance of the Digital Earth concept. The most visible of these involve virtual globes. Keyhole’s *Earthviewer* and the GeoFusion *GeoPlayer* appeared in 2001, and NASA’s own *World Wind* was first released in 2003. These received notice in a still fairly small community of interest. Then in October, 2004, Google acquired Keyhole Corporation, foreshadowing a major development – the June, 2005 release of *Google Earth*, which has captured an enormous interest for a few key reasons: (1) it is free; (2) it is fast; (3) it has its own markup language (KML), which allows anyone to display and easily share their own data; and (4) it is by all accounts fun; this stems from its speed, an easy-to-use interface, high quality imagery and a growing array of interesting content. That *Google Earth* so far falls far short of the Digital Earth vision despite its obvious relation is argued in Grossner and Clarke (2007).

Alongside the meteoric popularity of virtual globes, a “Geospatial Web” has emerged, enabled by newer format standards like KML and older ones that have solidified, e.g. WMS, WFS and the Geographic Markup Language (GML). Open-source software has facilitated geospatial prototyping and development that complements traditional GIS, with spatial databases like PostgreSQL and MySQL, and web mapping tools from the Open Source Geospatial Foundation⁴ such as OpenLayers and MapServer.

One major obstacle to integrating disparate geographic data systems is the difficulty of resolving place name ambiguity. Hill (2006) describes a “unified georeferencing approach” and surveys the significant research efforts aimed at improving and extending digital gazetteers.

The collaborative concept mentioned by Gore was introduced in 1989 by Bill Wulf, came into prominence in the early 1990s (National Research Council 1993) and is an integral aspect of the current cyberinfrastructure focus of US federal funding agencies. Hundreds of such projects are listed in the NSF-funded “Science of Collaboratories”

survey results.⁵ The discussion of geocollaboration for crisis management in MacEachren et al. (2006) exemplifies a research area and methodology brought to greater focus by events of 9/11 and Hurricane Katrina in 2005.

The Gore speech suggested the marketplace potential of Digital Earth might play a part in financing requisite technology. In the intervening years even very small transactions have become feasible (the Geocomm.com portal offers a 1:500,000 Oregon geology layer for \$2.76) but the real potential is more along the lines of value-added services applied to TIGER and US Census datasets, offered for example in the “Marketplace” section of Geospatial One-Stop⁶, a US e-Government initiative.

6 A Renewed Design Process

Such developments suggest that enough components of a potential *digital earth computing platform* are available that it can be deemed feasible. The outstanding “middleware” issues referred to earlier must now be informed by the requirements of particular software applications to run on that platform. We take some preliminary steps here at describing such potential client software. The *unified process* (UP) for software design has been successfully applied to many large projects (Larman 2002, Ambler 2004), and aspects of it are adapted here as a framework for analysis. The UP model describes four project phases: inception, elaboration, construction and transition. We are concerned here with only one or two aspects of the inception phase – the principal goals for which are “to achieve stakeholder consensus regarding the objectives for the project and to obtain funding” (Ambler 2004). In fact, the aim of this article is limited to delineating the project scope implied by the Gore speech with respect to high-level features and use cases, as a starting point for discussion leading to a true scoping exercise. Many aspects of functionality, content, interactivity and system architecture were mentioned. We enumerate elements within those groupings, and presume that at least some in each category are definitional – that is, essential characteristics for the system in question.

Use cases are integral to UP, and an effective method for eliciting requirements for system functionality, content and interactivity in all user-centered approaches to design. In an elaboration phase that follows (typically overlapping somewhat), they will be used to develop more precise functional specifications. From the preliminary high-level requirements listed in Section 4, we derived the following list of the users of Gore’s Digital Earth, alternately termed *actors*; all groups are users, some are potential designers as well. A more complete analysis must follow if evaluation in this inception stage establishes that proceeding further is warranted.

6.1 Actors

Use cases are normally developed in scenarios, with the aim of thinking through the realistic contexts for all user interactions with a system (Larman 2002). Once an instance of a user type has been given life in this way, subsequent questions about functionality can be answered given what is known about that user’s tasks, proclivities, aptitudes and abilities. Vice-President Gore made a wonderful start at one use scenario with “Imagine, for example, a young child . . .” Looking at the entire speech, the users mentioned can be grouped into the following categories:

6.1.1 *Students*

A single “young child” is featured. She participates with other K-12 students in the GLOBE program, an international collaboration supported by several US government agencies.

6.1.2 *Research scientists*

The concept of a “collaboratory” is offered with little elaboration. Potential environmental and biological modeling applications mentioned include those for “the impact on biodiversity of different regional growth plans,” and prediction of climate change based on patterns and rates of deforestation. Agricultural researchers and individual farm operators could perform diagnostics on individual plots of land using more readily available satellite imagery.

6.1.3 *Governmental agency analysts*

Mention is made of applications for municipal land-use planning, “crisis management” (taken to refer to the range of disaster risk analysis and mitigation, preparedness, and response) and crime fighting. At the national and international scale, analysis of geopolitical issues might be enhanced – at least in terms of visualizing geographic factors.

6.1.4 *Governmental policymakers, diplomats and politicians*

Presumably the products of analyses just mentioned could be better and more readily visualized by – and distributed to – the decision-making consumers of the analyses mentioned above. Each is a distinct use case, to be itemized in a subsequent phase.

6.1.5 *Educators*

Use by educators is implied in several respects. Some student users’ activities will be in the form of school assignments, so teachers will be designing curricula to make use of this new resource. Participants in the GLOBE program are being directed or guided in activities designed by educators. Educational requirements are a major consideration in the design of digital earth functionality.

6.1.6 *Museum directors*

In the speech, a schoolchild experiences Digital Earth in a museum, with a head-mounted display and “data glove;” these remain an atypical computing configuration, but costly immersive installations will remain the purview of museums. The bandwidth limitations Gore mentions appear to be vanishing in an exponential descent, making home and school computers the more likely venue, and museum-like interpretive exhibits could prove to be an appropriate interface metaphor for some digital earth systems.

6.1.7 *Commercial marketers*

The speech specifies that Digital Earth should incorporate a “digital marketplace for companies selling a vast array of commercial imagery and value-added information services.” This indicates the involvement of marketing professionals in both the design and ultimate use of the system.

7 Organizing Metaphors: The Geolibrary and the Atlas

A technological development with features and functionality reminiscent of Gore's 1998 vision is the *geolibrary*. The term was introduced coincidentally in the same year, defined as "a library filled with georeferenced information" (Goodchild 1998). Existing geolibraries allow users to retrieve geographic data objects by matching requested locations and thematic attributes with the footprints and metadata of items in one or more collections. Geolibraries are necessarily digital, because making that match entails spatial mathematics essentially impossible in a physical library. A summary report from a 1998 US National Research Council workshop panel on distributed geolibraries (Goodchild et al. 1999), self-described as "a vision and not a blueprint," made the following observation:

"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."

Also, this seemingly prescient suggestion:

"While the prevailing metaphor for human-computer interaction is the office or desktop, that metaphor may not be particularly helpful in organizing information about the Earth. Instead, access to a geolibrary could be through the visual metaphor of the Earth's surface itself; a student interested in Thailand would manipulate a globe on the screen until it centers on Thailand and then zoom in for more detail, as in the Digital Earth vision."

The Alexandria Digital Library (ADL), developed in the 1990s and still operational at the University of California, Santa Barbara⁷, was a pioneering effort at building a geolibrary. Its catalog contains over 15,000 entries for a variety of geographic information objects, including scanned maps, remotely sensed images, digital elevation models and air photographs. Users can identify a place by name or spatial footprint and ask, either broadly or with one or more narrowing filters, "What do you have about there?" The Electronic Cultural Atlas Initiative (ECAI) is an international consortium of humanities and information systems scholars with the goal, "to make virtual collections of scholarly data from around the globe accessible through a common interface (by providing) a means for making data interoperable across formats, disciplines, institutions, and technical paradigms" (ECAI n.d.). The ECAI clearinghouse⁸ offers access to several hundred datasets.

What ADL, ECAI and data portals like Geospatial One-Stop have in common is that their response to queries about a place, or about a theme at a given place and/or time, is effectively, "here are some digital objects with metadata matching your criteria – good luck." Here we draw the distinction between geographic information, as "representation(s) of some part of the Earth's surface" and the many kinds of georeferenced information *referring* to "specific places on the Earth's surface, and yet . . . not normally included in discussions of geographic databases" (Goodchild 1998, p. 59).

To make next-generation geolibraries more nearly like the educational system in Al Gore's vision would require further inclusion of that 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). The currently answerable query, "what *do you have* about that place?" can become another, vastly broader one, "what *is so* about that place?" and even, "what *has been so* there?" An advanced distributed geolibrary, and *digital earth system* for that matter, will be able to field such questions both from information in its local collection, and seamlessly, from a distributed web of collections worldwide. This would effectively and for all practical purposes marry the concept of digital libraries with that of *knowledge organization systems* (cf. Section 8 below). ECAI is moving in that direction by fostering the development of digital cultural atlases, which are GIS-driven geo-historical reference works comprehensive within given knowledge domains.

The Perseus Digital Library at Tufts University⁹ has for several years been at the forefront of research efforts looking to 'open and process' historical materials. Founder and director Gregory Crane, in discussing the potential of enormous digital collections promised by projects such as Google Print, noted that one of the core problems in making best use of such collections includes the two-fold task of first extracting references to people, places, dates and organizations, then automatically generating "atomic propositions" from them, e.g. "PERSON arrived at PLACE" (Crane 2006). Educational digital earth systems able to answer questions from content stores including million-volume georeferenced collections are within sight technologically as advances in natural language processing are increasingly able to generate some very basic but nonetheless useful measures of meaning.

8 Defining a Digital Earth System: A Geographic View

As noted, the term Digital Earth has come to represent a global technological initiative – in a sense, an intellectual movement. We propose here a starting point for defining a (lower-case) *digital earth system*. The Digital Earth concept is inclusive of the next-generation geolibrary, the global digital atlas, and to some extent, geographic information system (GIS) software. A *digital earth system* is then a hybrid of these which does not yet exist, "a distributed digital geolibrary for which the principal user interface is a global atlas, having at least some of the typical functionality of a GIS." Phrased another way, it is "a comprehensive, massively distributed geographic information and knowledge organization system."

It is necessary to parse that definition and define some terms: it is *comprehensive* in that it must contain complete, "blanket" or "Level I" spatial coverage of the globe for a set of base thematic layers at a uniform scale or set of scales (Figure 1). Further, it will contain such additional thematic layers of georeferenced data at any scale, level of detail (LOD) or coverage extent as are made available and accepted for inclusion by expert reviewers (Level II). A third (Level III) tier of content will be un-reviewed material submitted by the global public at large – either explicitly as a candidate for Level II status or simply posted for others to view.

This digital earth system is *distributed* because: (1) there are necessarily multiple, geographically dispersed data stores providing content; and (2) the processing load of server-based query and analytical processes must be shared for performance reasons (Figure 2).

Geographic information refers in this definition to the more inclusive geo-referenced information, "very broadly . . . information about well-defined locations on the Earth's

derived	Level III registered in clearinghouse, but un-reviewed, as with wiki systems; may be nominated for certification and Level II status
observational	
derived	Level II 'certified' data & information/knowledge objects at various scales and coverage extents; central & distributed sources
observational	
derived	Level I 'core' data with global coverage at uniform scale; certified; replicated; e.g. VMap1; .4m photos, imagery; DEM
observational	

Figure 1 Tiered data source structure. Data models must explicitly differentiate observational data and derived information/knowledge objects at all levels

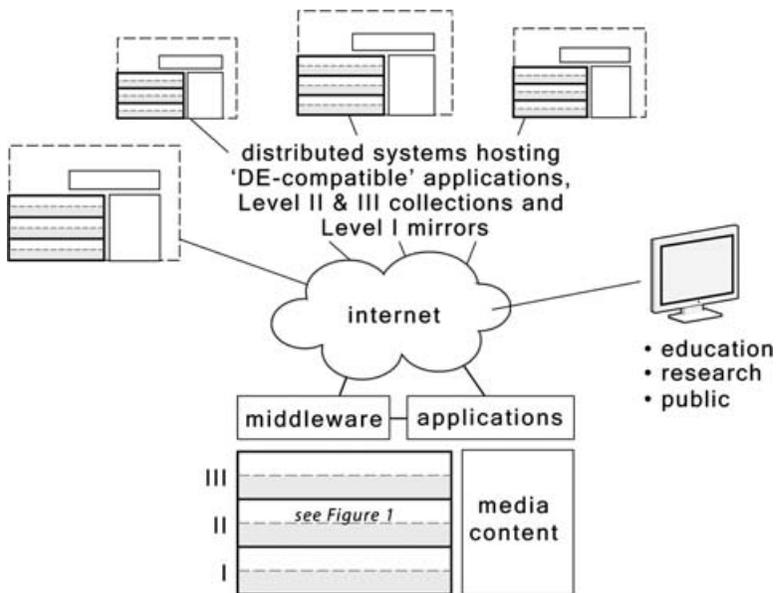


Figure 2 High-level schematic of the distributed, tiered data sources illustrated in Figure 1. Central server(s) differ from distributed DE-compatible collection servers only with respect to the hosting and coordinating of middleware services, broadly defined

surface; in other words, information associated with a geographic footprint” (Goodchild 2000). Since all entities and events have spatial (and temporal) extents, by implication the potential content of a digital earth system is almost infinite. The intent here is not to house all information with a geospatial element, but that any entity, event or process with a particular geographical location *may be* represented in a comprehensive digital earth system; obviously, not all could or should be.

The term *knowledge organization* is explicitly part of this definition for a few reasons. First, distinguishing knowledge from information and data is one element of a general (and admittedly optimistic) statement of epistemological viewpoint. The

formulation of a continuum offered in Longley et al. (2005, pp. 11–12) expresses this well and is widely echoed: *data* as “in some sense neutral and almost context-free . . . raw geographic facts,” *information* as data organized for some purpose, and *knowledge* as information to which interpretation has been added, “based on a particular context, experience and purpose.” Secondly, the vast realm of conceptual knowledge, while not itself intrinsically geographic or spatial, may be entered via spatiotemporal index markers of any geographically located entity or event. Organizing access to that larger realm must therefore be undertaken in this system, at least to the extent of providing reasonable entry points. Finally, while *knowledge organization system* has become an umbrella term “encompass(ing) all types of schemes for organizing information and promoting knowledge management” (Hodge 2000, p. 1), it refers here to a particular neutral, extensible ontological framework, including classification schema informing data model design and authority files, such as gazetteers and time period directories.

9 Essential Components

This “distributed geographic information and knowledge organization system” functions in many respects as a unified entity, comprising participant systems that:

- Adhere to an agreed upon set of protocols and standards for data models, data formats and metadata, allowing it to function as a contributing node in a single, virtual computing system;
- Deliver core imagery and datasets with global coverage extents to any “digital earth-compliant” client software;
- Are comprehensive for one or more additional thematic topics and/or spatiotemporal extents;
- Are universally available.

9.1 Requisites

The features required of a *digital earth system*, and not generally present in existing distributed geospatial data systems, can be grouped as follows:

9.1.1 Extensibility

The initial structure of a core computing platform will be informed by a few exemplar applications, but must allow for extension and adjustment as new requirements surface and the number and type of knowledge domains it serves grows. A consortium of interested participants on the model of successful open-source and standards development projects may be the best approach to coordinating such expansion.

9.1.2 Data model

An approach that is fundamentally different from a typical GIS is required. It must be semantic and ontology-based; that is, structured to allow feature and event attributes to represent meaning in class rules and relationships. Attribute changes over time must be trackable, to permit visualizations of dynamic processes. Furthermore, the model must

enable integration of object and field data sources. The challenges involved are important elements of the GIScience community's research agenda as delineated in Yuan et al. (2005).

9.1.3 *Object-level metadata*

Since both observational data and derived knowledge (concepts which may have contested or simply variant meanings) are to be managed, effective means of distinguishing the two and of representing provenance and quality are essential goals (Peuquet 2002, Gahegan and Pike 2006). Complete and highly granular metadata is required – more accessible and visible than is currently the norm. Comber et al. (2005) show how metadata standards can be expanded to capture ontologies operational at the time of data production.

9.1.4 *Multi-tiered distributed database*

The volume of information required means contributions must be facilitated, but a high standard of authenticity is necessary for the system's core data layers. It is also important that the distinction between observational data and derived knowledge be fundamentally clear. These requirements can be met with a 3-tiered database system, as discussed above and illustrated schematically in Figures 1 and 2.

9.1.5 *Integrated authority lists and middleware*

Existing clearinghouse and portal systems can present unified listings of distributed GIS data layers, but the types of queries to be served by a digital earth system require a central, integrated set of authority lists, including place name gazetteers, and directories of time periods and biographical information. Above all, a centralized and extensible framework for domain ontologies is the key to data integration across collections.

10 Conclusions

We have reviewed the genesis and evolution of the Digital Earth vision, and enumerated the component parts of its initial expression in Al Gore's 1998 speech. We find it presently comprises a nascent software development platform, multiple application software programs and a loosely organized intellectual movement. We have begun adapting methods of the inception phase of the Unified Process for software design towards defining and designing a particular realization of that vision, a *digital earth system*. The goals in this phase should be: (1) achieving consensus by interested parties regarding the broad objectives for the project; (2) assessment of its feasibility; and (3) a preliminary plan for proceeding.

We have made the case that sufficient progress has been made on platform-specific issues and challenges (chiefly with respect to some interoperability standards) that an iterative rapid-prototype design process, such as the one interrupted in 2000–2001, may profitably resume. We propose that next steps include the identification of at least two specific software applications and the initiation of their design; documentation for the 2001 Digital Earth Alpha Versions (IDEW 2001b) might be a worthwhile starting point for that discussion. The important remaining challenges at the platform level can be identified more precisely by gathering requirements derived from specific use-case scenarios for prospective client application software.

10.1 A Grand Challenge?

For these next steps to occur, a key issue becomes “who will fund this?” One possibility is that organizations like Google or Microsoft, who are racing towards their own conceptions of a Digital Earth (Jones 2005, Butler 2006), will find it useful to fund related research. What is missing from that scenario is a community voice helping them define targets, and a research community focused on solving the issues. The breadth and depth of the remaining research challenges are considerable, but may ultimately be met given a broad-based global collaborative effort, requiring strong leadership. Depending on how ambitiously the initial design criteria are framed, this effort can reasonably be termed a “Grand Challenge,” such as might be undertaken by the University Consortium for Geographic Information Science (UCGIS), in partnership with one or more other organizations from academic, government and commercial domains. Certainly, seeking to “put the full range of data about our planet and our history at our fingertips” (Gore 1998) may qualify as grand and would require significant progress on many difficult but surmountable challenges.

Acknowledgements

Karl Grossner’s doctoral studies are supported by the National Science Foundation’s IGERT in Interactive Digital Multimedia, Grant # DGE-0221713.

Notes

- 1 We capitalize Digital Earth in referring to the system discussed in the speech and the broad movement exemplified by the biennial International Symposia on Digital Earth; the lower-case digital earth system refers to a prospective particular computing system.
- 2 National Oceanographic and Atmospheric Administration (NOAA); US Geological Survey (USGS); US Army Corps of Engineers (USACE); Environmental Protection Agency (EPA); US Department of Agriculture (USDA); and National Science Foundation (NSF).
- 3 The most recent version (1.1) was released in December, 2003.
- 4 OSGeo, a non-profit organization with the goal to “promote the collaborative development of open geospatial technologies and data.” See <http://www.osgeo.org/> for additional details.
- 5 See <http://www.scienceofcollaboratories.org/> for additional details.
- 6 See <http://www.gos2.geodata.gov/> for additional details.
- 7 See <http://www.alexandria.ucsb.edu/> for additional details.
- 8 See <http://ecaimaps.berkeley.edu/clearinghouse/> for additional details.
- 9 See <http://www.perseus.tufts.edu/> for additional details.

References

- Amblar S W 2004 *The Object Primer: Agile Modeling-driven Development with UML 2.0*. New York, Cambridge University Press
- Butler D 2006 Virtual globes: The web-wide world. *Nature* 439: 776–8
- Comber A, Fisher P, and Wadsworth R 2005 You know what land cover is but does anyone else? . . . an investigation into semantic and ontological confusion. *International Journal of Remote Sensing* 26: 1143–61
- Crane G 2006 What do you do with a million books? *D-Lib Magazine* 12(3): (available at <http://www.dlib.org/dlib/march06/crane/03crane.html>)

- Droegemeier K K, Chandrasekar V, Clark R, Gannon D, Graves S, Joseph E, Ramamurthy M, Wilhelmson R, Brewster K, Domenico B, Leyton T, Morris V, Murray D, Plale B, Ramachandran R, Reed D, Rushing J, Weber D, Wilson A, Xue M, and Yalda S 2004 Linked environments for atmospheric discovery (LEAD): A cyberinfrastructure for mesoscale meteorology research and education. In *Proceedings of the Twentieth Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, Seattle, Washington
- ECAI (Electronic Cultural Atlas Initiative) (2007) ECAI Research Goals. WWW document, <http://ecai.org/tech/researchgoals.html>
- Evans J D (ed) 2003 Geospatial Interoperability Reference Model. WWW document, <http://gai.fgdc.gov/girm/v1.1/girm1.1.pdf>
- FGDC (Federal Geospatial Data Committee) 2004 GAI Strategic Plan. WWW document, <http://gai.fgdc.gov/library/StrategicPlan20040312.html>
- Gahegan M and Pike W 2006 A situated knowledge representation of geographic information. *Transactions in GIS* 10: 727–49
- Golden K 2006 Google Earth: Organizing the world's information geographically. WWW document, http://www.mbari.org/seminars/2006/winter2006/google_golden.htm
- Goodchild M F 1998 The geolibary. In Carver S (ed) *Innovations in GIS*. London, Taylor and Francis: 59–68
- Goodchild M F 2000 Cartographic futures on a Digital Earth. *Cartographic Perspectives* 36: 7–11
- Goodchild M F 2004 The Alexandria Digital Library Project: Review, assessment, and prospects. *D-Lib Magazine* 10(4): (available at <http://www.dlib.org/dlib/may04/goodchild/05goodchild.html>)
- Goodchild M F, Adler P S, Buttenfield B P, Kahn R E, Krygiel A J, and Onsrud H J (Panel on Distributed Geolibraries, Mapping Science Committee, National Research Council) 1999 *Distributed Geolibraries: Spatial Information Resources* Washington, D.C., National Academies Press
- Gore A 1992 *Earth in the Balance*. Boston, MA, Houghton Mifflin
- Gore A 1998 *The Digital Earth: Understanding our Planet in the 21st Century*. WWW document, http://www.isde5.org/al_gore_speech.htm
- Grossner K and Clarke K 2007 Is Google Earth, “Digital Earth?”: Defining a vision. In *Proceedings of the Fifth International Symposium on Digital Earth*, Berkeley, California
- Hanke J 2006 Google press release, 9/13/2006. WWW document, http://www.google.com/intl/en/press/pressrel/earth_featured_content.html
- Hill L 2006 *Georeferencing*. Cambridge, MA, MIT Press
- Hodge G 2000 *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington D.C., Council on Library and Information Resources Publication No. 91 (available at <http://www.clir.org/pubs/reports/pub91/pub91.pdf>)
- IDEW 2001a Digital Earth Prototypes. WWW document, <http://www.digitalearth.gov/transition.html>
- IDEW 2001b Digital Earth Alpha Versions. WWW document, <http://www.digitalearth.gov/alpha/alphaVersionDRAFT03.doc>
- IDEW 2001c The Big Picture: Digital Earth and the Power of Applied Geography in the 21st Century. WWW document, <http://www.digitalearth.gov/BigPicture.doc>
- IDEW (no date) What is Digital Earth? WWW document, <http://www.digitalearth.gov/vision/WhatIsDE.doc>
- Jones M 2005 Speech at University of California San Diego November 16, 2005. WWW document, <http://ceoa.ucsd.edu/Improved-NTSC.mp4>
- Larman C 2002 *Applying UML and Patterns: An Introduction to Object-oriented Analysis and Design and the Unified Process*. Upper Saddle River, NJ, Prentice Hall
- Longley P A, Goodchild M F, Maguire D J, and Rhind D W 2005 *Geographic Information Systems and Science*. New York, John Wiley and Sons
- MacEachren A M, Cai G, McNeese M, Sharma R, and Fuhrmann S 2006 GeoCollaborative crisis management: Designing technologies to meet real-world needs. In *Proceedings of the Seventh National Conference on Digital Government Research*, San Diego, California
- National Research Council 1993 *National Collaboratories: Applying Information Technology for Scientific Research*. Washington, D.C., National Academy Press
- Peuquet D J 2002 *Representations of Space and Time*. New York, The Guilford Press
- Yuan M, Mark D M, Egenhofer M J, and Peuquet D J 2005 Extensions to geographic representation. In McMaster R B and Usery E I (eds) *A Research Agenda for Geographic Information Science*. Boca Raton, FL, CRC Press: 129–56