

The Global Public Commons and Marketplace in Geographic Data: A Conceptual Model for Meeting the Needs of Government, Commercial, Scientific and Nonprofit Sectors

Professor Harlan J. Onsrud
Department of Spatial Information Science and Engineering
5711 Boardman Hall, Rm. 340
University of Maine
Orono, ME 04469-5711
Tel: (207) 581-2175
Fax: (207) 581-2206
Email: onsrud@spatial.maine.edu

Abstract

Nations across the globe are rapidly developing and expanding their national spatial data infrastructures. Compatibility among these efforts in terms of policy, technical, and operational arrangements that would better allow sharing and exchange of spatial data among nations has become a major challenge. One means to deal with knowledge sharing and discovery is to promote standardization in the establishment of spatial data infrastructures throughout the globe in order to greatly simplify search, information extraction and management practices. The Global Spatial Data Infrastructure (GSDI) is a concept by which personnel from across the globe in government agencies, private companies, and academic institutions are developing common technical, policy, and administrative concepts in order to enhance the more effective use and sharing of geographic information. Yet, what might an ideal global spatial data infrastructure look like? Is it possible to develop a practical infrastructure that would allow vastly increased numbers of individuals and organizations to both draw from, contribute to and benefit from such a global resource? This presentation outlines a conceptual model for a GSDI that incorporates both a public commons and marketplace in geographic data and services that would allow anyone with access to the Internet to both draw from and contribute to GSDI development.

Introduction

One dream for the global spatial data infrastructure is an operational environment that would efficiently support an active *information commons* integrated with a thriving *marketplace* in spatial data and online services. Currently spatial data is still very difficult for the average person to find, use and share. Society is poorer when its members can't find and use existing information resources nor contribute their own information work products in an effective manner.

Current GIS can be categorized operationally under three primary architectures as illustrated in Figure 1. The vision addressed here is focused on the first of these three architectures. That is, we start with the assumption

that potentially hundreds of thousands of individuals in a nation want to draw spatial data from and contribute back their own spatial data and work products to a common shared pool or marketplace. In an economically privileged nation, one can envision that such sharing and exchange relations could be supported across the existing Internet. In an underdeveloped nation, one might think in terms of high bandwidth Internet access occurring over the airwaves with inexpensive *smart* radio receivers and transmitters. We are approaching an era where the airwaves will soon function like the Internet with the ability to send and receive smart radio packets without interference.¹ Regardless of the geographic location, the assumption is that at least tens of thousands of individuals working at the local level within a nation might desire to share, exchange, sell or barter spatial information with each other. What type of organized legal and technological infrastructure could allow this to happen more effectively?²

1. Desktop - Software and data reside on the users computer but the spatial data may be drawn from anywhere on the Internet or elsewhere.
2. Database – (a) Many users access and use a common centrally maintained remote database for multiple purposes. Users may apply many different software applications to the database with the software residing typically on the users desktop. (b) In some instances, many users are each able to contribute updates to a standardized database or perhaps networks of distributed but standardized databases
3. Web Services - Systems where both data and software typically reside on a host server for a narrowly focused GIS application or well-defined task that the user may run from a remote location. Alternatively the software and data sets for the tightly integrated web services may be distributed.

Figure 1. Common Spatial Data Use Environments or GIS Architectures

A prominent Internet-based geographic information *commons* connecting users and contributors and an affiliated efficient *market* connecting sellers and buyers would benefit everyone by better supporting sharing and trade. Facilities under development in the U.S. such as the National Map³ and

¹ Adaptive radio is an emerging wireless technology that scans the spectrum wherever the device is located and avoids interference by tuning transmissions to available gaps. Such devices modify their timing, frequencies and power constantly adapting to other airwave traffic around them. Smart receivers track and respond to the changes in order to receive the transmissions

² For present purposes, the vision outlined does not address the second and third categories of Figure 1 because those architectures for the near term future are most likely to be supported by larger commercial and governmental enterprises. Further on in time in a web environment where Moore's Law holds true, it is easy to envision that even small businesses or an individual will be able to host their own major databases or powerful web services.

³ <http://nationalmap.usgs.gov>

Geospatial One-Stop⁴ as well as similar efforts across the globe are solid first steps in this direction. However, most spatial data infrastructure development efforts to date have fallen short in addressing many of the current impediments to making sharing and sale transactions efficient for the common everyday user of geographic information. Such users include not only those within the mapping professions but all those within the commercial, scientific, civic and non-profit sectors that are regularly applying mobile, imaging, GIS, and GPS technologies to accomplish myriad daily tasks.

Below we describe the concept of a global *commons* and *market* in geographic information and services. Although discussed separately, we suggest that the two concepts should function seamlessly from an operational perspective. The following descriptions are meant to be descriptive rather than prescriptive. Numerous technological and legal alternatives exist by which the same ends might be achieved so details are provided in the discussion below primarily for illustrative purposes. The goal of this short thought piece is to convey a practical technological and legal *vision* by which those from across the globe might more readily draw from, contribute to and benefit from a common global information resource.

GSDI: The Commons Component

A key premise in most organized spatial data infrastructure efforts is that national governments will be unable to gather and maintain more than a small percentage of the geographic data that users in their nations want and desire. Thus, the national initiatives are depending typically on the cooperation of those already gathering spatial data and those using GIS to meet specific needs to help construct and maintain the spatial data infrastructures for their nations. Some of the impediments to widespread spatial data sharing are well known from asking GIS users why they are not currently creating metadata or making their datasets more readily available to others. Most of these impediments are unrelated to a need for increased funds. For many organizations, if their budgets were doubled they still would not use the increased funds to make their geographic data sets more accessible to their own communities or the rest of the world. They are inhibited by further impediments that money alone is unable to address.

Common wisdom suggests that intellectual property laws and the markets they protect create the only practical environment for producing and sharing useful information. That is, profit motivations drive all major resource development. Yet the history of the web shows us otherwise. In the Internet age, we now have numerous examples of massive voluntary resource production and sharing. In some instances, tens of thousands of individuals have worked collaboratively or as independent contributors in creating new knowledge resources or producing new software.⁵ Further, when the

⁴ <http://www.geo-one-stop.gov>

⁵ Well known examples of collaborative information resources created and maintained through open source or open access principles include Linux (operating system software), OpenOffice (application

production of important “public goods” is involved, government funding for direct production may be the most practical and efficient means for producing information goods.⁶

The approach we suggest for the commons component of the GSDI has at its core an alternative economic model for the production of information resources.⁷ That is, the model we present in this section derives from the open source movement and that phenomenon highlights the fact that there are numerous ways to generate substantial revenues from the production of spatial data and products other than claiming direct royalties or rents in the uses of those products.

Further, a general incentive premise of our emerging spatial data sharing model is that, as individuals, most of our conduct in daily life is not driven primarily by profit motives. Many creators of spatial data, although not all, would be more than willing to share all or some of their spatial data sets if, among other reasons,

- (1) it was much easier to do,
- (2) creators could reliably ensure and retain credit for their contributions to the public commons,
- (3) creators could acquire substantially increased liability protection from use of the data they make available to the public, and
- (4) creators could obtain other non-monetary benefits, such as long-term archiving, elimination of the need to provide web access to their files, increased credibility for their organization, and explicit recognition for their contributions.

Following is a short list of operational characteristics for a global commons in geographic information that a future system should be able to support.⁸

1. A non-expert user creates a GIS data set that he or she wants preserved and accessible to the rest of the world.

Examples might be a geology professor who has created numerous GIS data sets related to a research project, a community group that has tracked vehicle and pedestrian accidents or crime incidents in their

software), open access journals (e.g. see <http://www.doaj.org>), encyclopedias (e.g. see <http://www.wikipedia.org>) and text books (e.g. see <http://www.otp.inlimine.org>). Information work products directly germane to the spatial data infrastructure community produced by following similar open approaches are available as well (e.g. see <http://freeGIS.org>).

⁶ Government funding has been argued as economically rational from a *public goods* analysis perspective for such information goods as initial research and development of the Internet and creation of core spatial data sets upon which a large proportion of citizens and industry benefit directly or indirectly.

⁷ See Benkler, Yochai, Coases’s Penguin, or Linux and the Nature of the Firm, 112 Yale L.J. (Winter 2002-03) V.04.3, <http://www.benkler.org/CoasesPenguin.PDF>.

⁸ A web mock-up demonstrating in greater detail the envisioned capabilities of the system may be linked through <http://www.spatial.maine.edu/~onsrud/PublicCommonsofGIScience.htm>

neighborhood over the past year, and a high school class that has classified and mapped all the tree species in their community.

2. The non-expert GIS user accesses a web site and, through a series of responses to a transcript, creates automatically license provisions to be permanently identified with his or her submitted data set(s). The contributor agrees to (1) apply one of a limited selection of "open access" licenses to the data set or (2) dedicate the file to the public domain.

The concept here is to create a broad and continually growing set of freely usable and accessible data at local jurisdictional level scales similar in effect to the public domain data sets created by federal agencies in the United States. One can assume by default that federal data sets in the U.S. are usable with no intellectual property limitations attached to them whereas this is not true for most other geographic datasets. The intent with this web licensing approach is to provide an easy legal mechanism by which any individual may affirmatively and permanently mark their dataset such that the world knows where the dataset came from and that the dataset is available for use without the law presuming that the user must first acquire permission.

The basic concept of an "open access" license is that any subsequent user may freely use the data file. The advantage of an "open access" license over placing the file in the public domain is that the license can ensure that 1. the originator and all value-adders have a legally enforceable right to credit for their work, 2. liability exposure may be substantially reduced through the license provisions, and 3. the license can prevent the efforts of the originator and value-adders from being captured by others and removed from the commons.⁹

3. In continuing through the web transcript, the user creates a metadata record with deep meaning through minimal effort. The user is walked through a series of questions with limited choice responses and the transcript automatically changes depending on responses to previous questions. From the responses, the metadata record is automatically created.

Non-expert users will never take a meta-data course nor will they ever have familiarity with many technical terms. Therefore, open-ended questions with free-form responses will need to be minimized for lay producers of geographic data sets. The system guides the user in providing deeper meaning by using pull down menu selections and then asking the user to select from among definitions for the terms selected. Those definitions along with ontologies are used to predict and simplify future menu choices in the metadata creation process. Commercial software companies are pursuing similar interview approaches and automated population of some metadata fields internal to their software

⁹ Linux is perhaps the most well-known cooperative information resource development venture utilizing such a licensing approach.

but a non-proprietary capability able to stand alone or be incorporated by any private company should allow more extensive creation of metadata with greater uniformity imbued with deeper meaning.

4. The transcript responses and the actual data file to which the responses apply are submitted to an automated processing facility. The processing facility adds to the spatial data file an encrypted identifier that is embedded in the file but does not interfere with the file nor is it stripped from the file through typical GIS operations.

5. The facility automatically returns a copy of the "marked" geographic data file back to the user and, at the discretion of the submitter, retains a copy in the long-term archive. Any person gaining access to a copy of the file in the future is able to readily determine the status of the legal rights in the data and the full metadata history of all value-adding done to the file. This might typically be accomplished through a simple "drag and drop" to freely downloadable software that links back to the full license and metadata history.

6. Whether maintained on the open web or maintained in a long-term electronic archive, anyone would be able to search for, access and download the file and freely use it. Use of the file would be limited only by minimal legal restrictions that have been imposed to enhance rather than restrict the ability of the file to be maintained in the open.

By going to the extra time and effort of creating a license and metadata, creators get something in return under this model. Through this approach they obtain visible credit when their data set is used in the products or services of others. This not only builds prestige but allows a creator to far better demonstrate to politicians and funding authorities the value of supporting their organization. Those sharing also obtain a level of liability protection through the user license never acquired when data is simply released. They also obtain a potential archiving service. Further, the numerous alternative means for generating revenue as being witnessed in the open source and open content communities remain available. Thus, the most frequently given reasons by local governments and scientists for not making data sets available to others are addressed by the model.

Not all levels of government, private citizens or private companies will want to make any or all of their datasets available through this legal framework. However, those choosing to contribute are likely to discover economic and social advantages in doing so. The current public domain in geographic data provides valuable "raw materials" for innumerable value-adding activities by the commercial, non-profit and government sectors. This important economic resource can be made even more vital by vastly increasing the numbers of individuals and organizations able to readily contribute to and draw from the resource.

GSDI: The Marketplace Component

The Internet has greatly expanded the ability of spatial data and service sellers to reach customers and has allowed an expansion in data offerings and suppliers. Yet considerable obstacles still exist in connecting consumers and suppliers. Efficiency for the consumer in finding, accessing, evaluating and acquiring geographic data can yet be greatly improved.

A networked global marketplace in geographic information should support a capability in which any seller or licensor, no matter how small or large, has the ready ability to:

- post to the system standard meaningful metadata about their spatial data offerings,
- define detailed yet standard license or sale conditions,
- tie data offerings to seller defined pricing formulas that can accommodate automatically variations in quantity, quality, legal conditions of use, and similar data product variables, and
- participate in automated financial transactions and delivery of product.

A seller should be able to alter their product offerings and conditions of use quickly and efficiently. Any purchaser, no matter how small or large, should be able to find from among thousands of offerings for their geographic area those offerings that explicitly meet their licensing and technical needs, do efficient comparison shopping among offerings, and buy or license and acquire the data they desire within minutes of finding it. The seller's financial account would be automatically credited whenever a sale or license is accomplished through the system.

Such a system might be maintained by either a non-profit corporation explicitly established for the purpose or by a for-profit enterprise, either of which might apply a small fee against all financial transactions in order to support the system. Alternatively, such a system might be developed and maintained by government as core infrastructure.

Following is a short list of operational characteristics for a global *marketplace* in geographic information that a future system should be able to support.

1. A small commercial firm in a small city in Russia creates one or many geographic data sets that the company would like to make available by sale or license to any interested consumers throughout the rest of the world.
2. The commercial firm accesses a web site that allows the company to efficiently establish license options to be offered to consumers and to establish prices. The web site automatically generates a standard list of commercial license options. The company checks off those data use conditions it is willing to make available to consumers. The firm is then led through a process of developing one or more pricing formulas that increase or decrease depending on combinations of variables such as volume and constraints on use. In this manner, automated comparison-shopping for potential consumers is greatly facilitated. The license or sale conditions are

permanently tied to the data set such that all downstream users are put on notice of the legal uses that may be made of the file or derivatives.

3. In continuing through the web transcript, the company creates the metadata record. The data supplier is walked through a series of questions with limited choice responses and the transcript automatically changes depending on responses to previous questions. From the responses, the metadata record is automatically created.

4. The transcript responses and the actual data file to which the responses apply are submitted to an automated processing facility. The processing facility adds to the spatial data file an encrypted identifier that is embedded in the file but does not interfere with the file nor is it stripped from the file through typical GIS operations.

5. The facility automatically returns a copy of the "marked" geographic data file back to the company. The marked file embeds by reference the license and metadata information. The facility maintains the metadata for all files enhancing the efficiency of searches but the locations of actual files might be distributed. Automated searches become much easier for those seeking data that meets their specific conditions of use, technical conditions and price.

A coordinated global *marketplace* in geographic information would be supported by standard metadata creation processes, flexible but consistent licensing and sale condition options, and pricing and transaction support structures. Through the availability of freely downloadable "file identifier" software, any person gaining access to a copy of a file through the system is able to readily determine the status of the legal rights in the data and the full metadata for the file. Appearance of the identifier embedded in the file would be strong evidence that the file is controlled by the license provisions attached to it and that users should be held responsible for adhering to those provisions.

Summary

Many individuals and organizations throughout our individual nations are creating GIS datasets with few incentives or little ability to effectively share their works with the rest of the world. An integrated global *commons* and *marketplace* that allows for expanded and more effective sharing and sale of geographic data among vastly increased numbers of both specialist and non-specialist geographic data users would result in a resource of immeasurable value to society. As was true with the Internet itself, the GSDI will begin to have its greatest impact when it is used daily by millions and those millions use the infrastructure not to just gain access to the geographic data and services of others but to offer up and make available their own geographic data, work products and online services.