U.S. Department of the Interior
U.S. Geological Survey

The National Map Catalog Technical Discussion Paper

Tutorial Introduction to *The National Map* Catalog

Version 0.2.0 Unreviewed draft. Comments invited.

Larry Moore

Mid-Continent Mapping Center
September 2004

Table of Contents

1 Introduction

The National Map has two characteristics that make it somewhat different from most other cartographic data registries and clearinghouses:

***The National Map* implementation is based on open standards**. In the early days of map services (only a few years ago), every implementation defined its own interface. This of course caused interoperability problems, which in turn led to efforts to define standard interfaces for map services. The Open GIS Consortium, Inc. (OGC) is a non-profit standards organization that is leading the development of standards for geospatial and location based services. The USGS is committed to using OGC standards to make *The National Map* open and non-proprietary.

***The National Map* is database driven**. A typical WMS application does its own service data management, often by hardcoding WMS addresses in the application. This works fine for cases such as a researcher publishing a project on the Web. In such a case, one organization owns the application and most or all of the data, and the data are probably in only one or two services. But this solution does not scale to many data sources. As the number of contributing services grows, the effort needed to maintain the service links within the application code becomes very expensive. If more than one application wants to use the same collection of services, the cost of this data management becomes prohibitive.

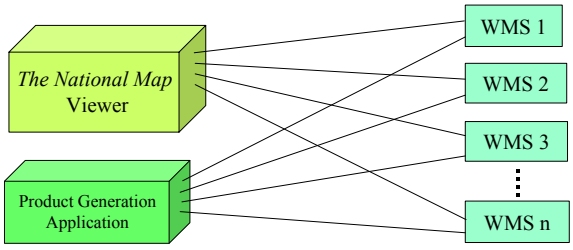

Figure 1. Multiple applications trying to keep track of multiple data sources. This many-to-many relationship can be managed directly only if the number of applications and sources is very small. As the number of data sources and applications grows, it becomes prohibitively difficult to keep the overall system consistent.

A more scalable solution is to have a shared inventory of WMSs maintained in a database. Such a database is called a **catalog** in the OGC literature. Applications need only hardcode one address, that of the catalog.[1] From the catalog, applications retrieve the URL addresses of map services, as well as metadata about the services to help the application decide if a particular service is useful. The database allows inventorying, monitoring, and evaluating WMSs to be centralized, saving individual projects the effort of doing these things on their own. **The catalog does not store geospatial data. It stores pointers to, and information about, geospatial data.**

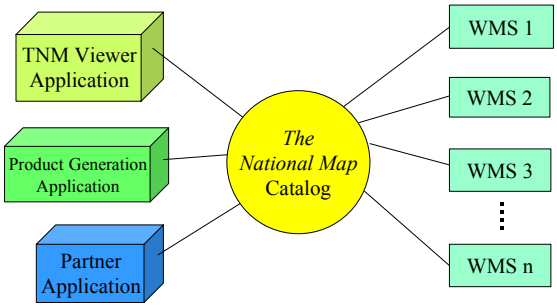

Figure 2. A central database resolves the many-to-many relationship. The database tracks information about the locations of data sources, so the individual applications don't have to. This solution scales to arbitrarily large numbers of data sources and applications.

[1] Of course, it is easy to see that maybe this isn't enough, that multiple catalogs are likely to be created, so we will eventually need a catalog of catalogs. This is a reasonable idea, and we may actually see such designs in the near future.

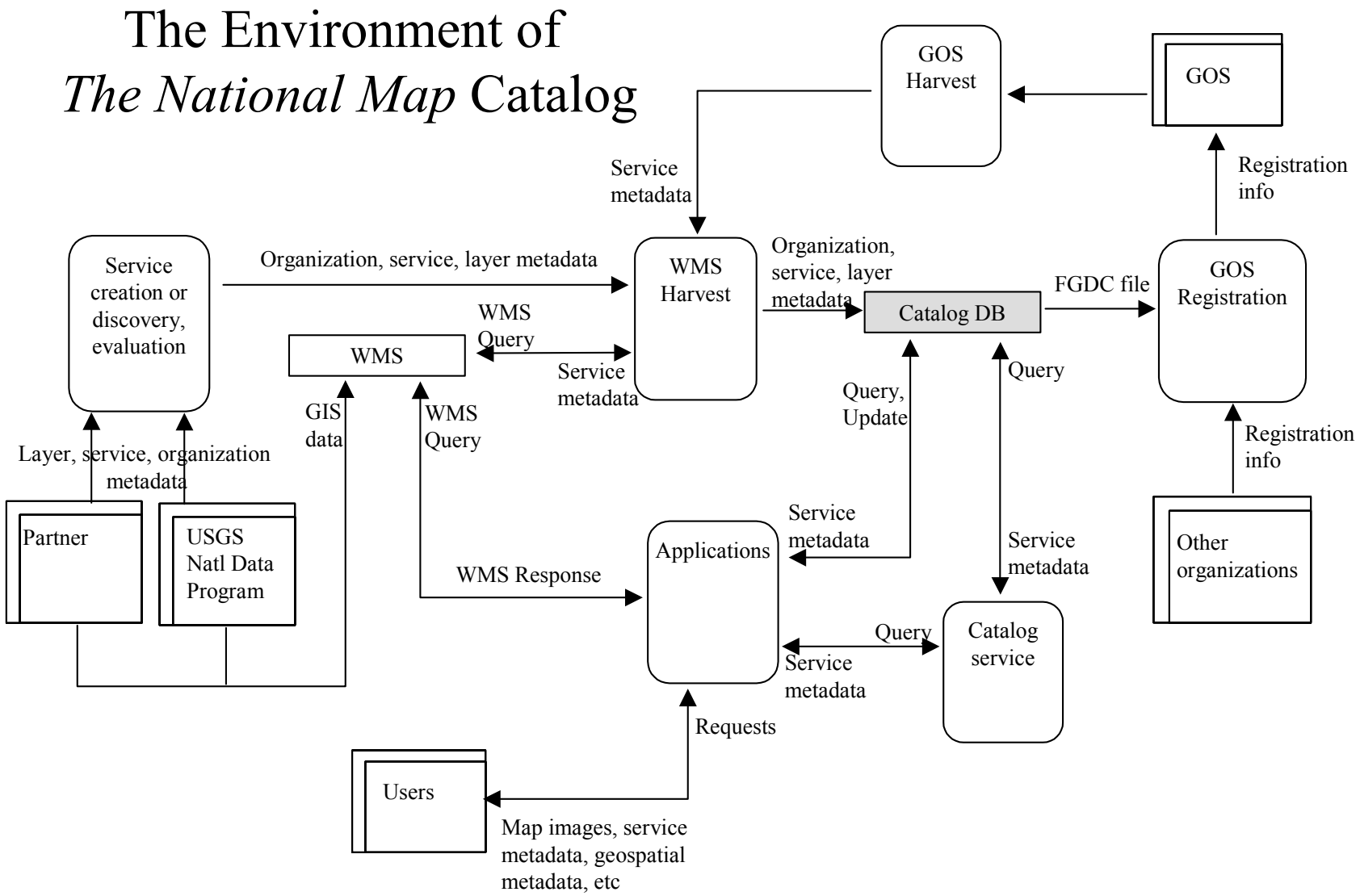

Figure 3. The diagram is a logical description, not a picture of any physical implementation. Rounded boxes are processes, rectangles are data stores, boxes with boarders are external systems, and arrows are data flows. The extent of this "system" is arbitrary, selected to help explain the role of the Catalog. Attachment A is a data dictionary for this diagram.

2 The Catalog Environment

A frequently asked question is "what is the process for getting data into *The National Map*?"

The question, unfortunately, does not have one simple answer. Like all complex systems, *The National Map* cannot be described using the waterfall model: there is no single linear process with clean start and end points that describes how to put data into *The National Map*. This is not a shortcoming of the system or its documentation, it simply reflects the reality of modern information systems.

Figure 3 is an abstract representation of processes and data in the "near vicinity" of the catalog database. It represents the logical environment surrounding the catalog, not the physical implementation of *The National Map* system. The diagram's entities do not necessarily correspond to servers, network connections, or specific pieces of software.

The figure is basically a data flow diagram (DFD). Rectangles represent data stores, boxes with rounded edges represent processes, and rectangles with a double boarder on the top and left sides represent external systems. The arrows represent data flows. The diagram illustrates that there is no single start point and no single end point, nor any unique path for putting data into *The National Map*.

The centerpiece of the diagram is the catalog database. The internal structure of the database is discussed later in this paper, but in this section it is a black box.

This is not a complete or especially detailed description of *The National Map* system, yet is still complex enough to be confusing. To analyze the diagram, we will consider pieces of it from the perspectives of different categories of people. In the following sections, different roles are treated as figuratively "occupying" the process boxes of Figure 3. Occupying a process means understanding the purpose of that process, the data flows into and out of the process, and knowing at least a little about adjacent processes.

2.1 External User Perspective

End users occupy the box labeled **Users** at the bottom of Figure 3. Users communicate only with **Applications** – the rest of the entities in the diagram are hidden from view. In a sense, the main purpose of *The National Map* is to create the illusion of relative simplicity for the benefit of users. Anyone who has tried to assemble (for example) a collection of county datasets into a State-wide dataset has a feel for how complex and tedious GIS work can be. *The National Map* applications do this kind of data assembly constantly and automatically, but the details are mostly hidden from view.

This does not mean that users are unskilled. Users interact with applications, and GIS applications tend to be complex in their own right, albeit for different reasons and in different ways than the internal complexities of *The National Map*.

The number of *The National Map* applications is still small, and only one (the public viewer) is used by significant numbers of people.[2] If *The National Map* is successful, the number of

[2] Other applications are internal to the USGS, and are used by only dozens of people. Quantifying the use of the public viewer is difficult, but we know that it responds to around 20,000 requests for map images per day. This number refers essentially to screen refreshes, so it is much larger than the number of people accessing the viewer. Still, it seems

applications that use the catalog will certainly grow. Having multiple applications is a good thing, and does not imply duplication of effort. It is analogous to the use of multiple clerical tools (editors, spreadsheets, databases) for normal office work.

2.2 Internal User Perspective

"Internal users" are responsible for populating and maintaining the catalog database, as well as monitoring the catalog service interface and the behavior of *The National Map* applications[3]. These users must understand all of Figure 3, though not necessarily the internal workings of all individual boxes.

Internal users spend much of their time at the **WMS Harvest** box, a process that includes many functions, some manual and some automated. They work with the owners of WMSs (or their representatives) to add new data to *The National Map*. The harvest process receives information about WMSs from partner organizations or MPOs, or from other data registries such as Geospatial One Stop (GOS). Some information comes from the WMS itself, through the WMS GetCapabilities request. Information about the service is organized and stored in the Catalog database. For a detailed discussion of these activities, see "Registering Web Map Services in *The National Map* Catalog."

Internal users continuously monitor the catalog and the WMSs registered in the catalog. Some of this monitoring is automatic and some is semi-manual. But even automatic processes eventually lead to some human action. For example, software continuously checks all the registered WMSs to see if they are awake – that is, responding to queries with legal answers. When a service is down, a USGS employee contacts the service owner and investigates the cause of the problem.

Monitoring activities use internal applications. The **Applications** box has two communication paths to the catalog, one through the **catalog service**, and one directly to the database. Internal applications sometimes use direct database connections to the Oracle instances of the catalog, while external applications (such as *The National Map* viewer) should get their information through the catalog service.

2.3 Data Provider and MPO Perspective

A data provider is an organization that owns a WMS that contributes data to *The National Map.* Data providers can be USGS Geography data programs, programs of other USGS Disciplines, or non-USGS organizations. Non-USGS providers are commonly called **Partners,** and communicate with the USGS through MPOs (Figure 3 omits some of this detail, showing only that **Partners** and **USGS National Data Programs** are distinct).

Data providers normally do not maintain WMSs just for the benefit of the USGS; they usually have their own customers and applications. Providers must give the USGS harvest process information about their service(s), but need not alter their service configuration or do anything that disrupts existing applications.

---

likely that the daily users of the viewer number in the hundreds or low thousands. Whether or not this is impressive depends on your perspective. Compared to many USGS Web sites, it is fairly large. But it is roughly 100 times lower than the number of daily users of Terraserver, and many thousands of times lower than the daily users of Google.

[3] Currently, all internal users are USGS employees, but this is not a technical requirement of any of the systems involved.

The box labeled **service creation or discovery, and evaluation** includes many diverse functions. Other GIS organizations create WMSs, often for reasons that have nothing to do with the USGS or *The National Map*. MPOs search for existing services, or sometimes persuade other organizations to create services, but do not create services of their own. MPOs often do some limited evaluation of other organizations' services before passing information on to the **WMS harvest** process. Partners monitor and maintain their own services, and may need to transmit certain types of change information to the USGS even after a service is registered in the catalog.

2.4 GOS Perspective

In Figure 3, Geospatial One-Stop (GOS) and the related data flows represent a class of geospatial registries. At this writing, GOS is the most prominent of such registries, but others exist at all levels of government. The objective of coordination with GOS is to reduce duplication of effort, especially the duplication (from the Federal point of view) and annoyance (from the State point of view) of having multiple Federal agencies contact the same State agency to request the same GIS information.

The problem of "one stop registration" has a conceptually simple solution: we need only agree on a single central data registry. The reality is more complex, because different registries have different purposes and store different sets of information. The organizations that own them therefore are not likely to agree on one particular "master registry." *The National Map* and GOS are duplicate registries only at extremely high levels of abstraction; at working levels the two systems clearly use different data for different purposes.

Nevertheless, any two geospatial registries will almost certainly have ***some*** commonalities, and it is clearly desirable to find and exploit them. The loop in Figure 3 (**GOS Registration -> GOS -> GOS harvest -> WMS harvest -> Catalog DB)** illustrates a working example of two registries using each other's data.

GOS requires an FGDC metadata file as input to its registration process. *The National Map* can create such a metadata file from data in the catalog. These files are automatically created and made available to GOS. Therefore, anything that is registered in *The National Map* is automatically also registered in GOS. This mechanism was put in place in mid-2004, but at this writing has not been fully incorporated into the routine operations of either organization.

Conversely, the GOS registry contains the basic information needed to register a WMS in *The National Map*. A mechanism for partially automating this is under development. The process cannot be fully automatic, because *The National Map* requires some information from a data provider that is not required for GOS registration. Therefore, a separate contact with the provider may be necessary, but at least different questions will be asked.

2.5 The Catalog Service Perspective

The catalog database is an inventory of WMSs that serve base-category cartographic data. By June 2004 the inventory contained more than 150 services managing 2800 layers. As this inventory grows, its value to the general GIS community will also grow. Other organizations will want access to the database for their own purposes.

The value of the database is as a source of data for GIS applications, as opposed to a source for human-readable reports. The accepted way to make a database visible on the Internet is through an

information service. A service typically takes requests in the form of URLs and returns responses in XML documents. This allows any computer on the Internet to access the database through a standard interface, without compromising the security of the source database or requiring any proprietary software on the client side.

Internal applications may access the catalog through a direct Oracle connection, but external applications such as *The National Map* viewer should always get their information through the catalog service.

The catalog service API (application programmer interface) is documented in a separate report, "*The National Map* Catalog Service: A Guide for Application Developers."

3 The Catalog Data Model

Figure 3 shows the catalog database as a black box. This is the only perspective needed by most users, but at least three groups need some understanding of the inside of this black box:

1. Internal users who must populate and maintain the database.
2. Application programmers who want to use the data from the catalog for their own applications.
3. USGS managers and MPO representatives (and possibly some of their State counterparts) who must plan budgets, negotiate partnerships, write reports, and make policy decisions.

3.1 Partners, Services, and Layers

At the core of *The National Map* business model are these two assertions:

1. *The National Map* is based on partnerships.
2. *The National Map* will use open standards implementations of Internet information service technology. In the very near term, Web Map Services (WMS) will be preeminent.

These ideas can be expressed as a simple entity relationship diagram (ERD), as shown in Figure 4.

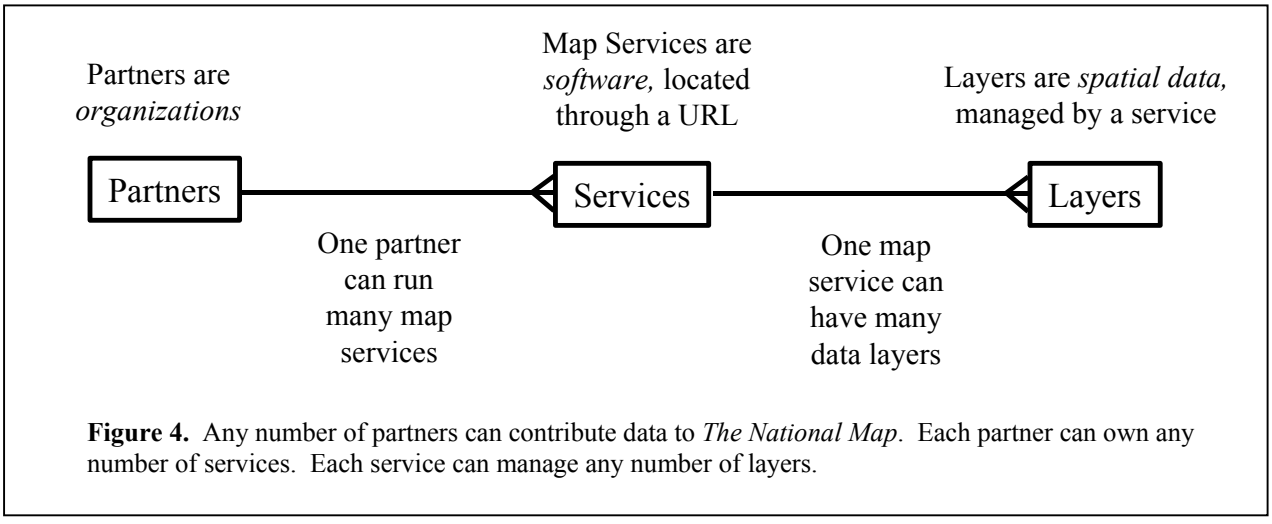

Figure 4. Any number of partners can contribute data to *The National Map*. Each partner can own any number of services. Each service can manage any number of layers.

As of June 2004, the database tracked 70 partners, 155 services, and 2894 layers. 794 of these layers are available for viewing in the public viewer.[4]

Figure 4 is both a simple version of *The National Map* business model and a first approximation of the internals of the catalog black box of Figure 3. The two one-to-many relationships illustrate why the partner-service-layer data of *The National Map* cannot be managed with simple lists and software hardcoding. These business data have too many dimensions, and the numbers – even at this early stage of the program – are already big enough to preclude simple forms of data management.

The complete data model is of course much more complicated. Layers are associated with themes and subthemes; layers have one or more spatial reference systems; layers belong to one or more

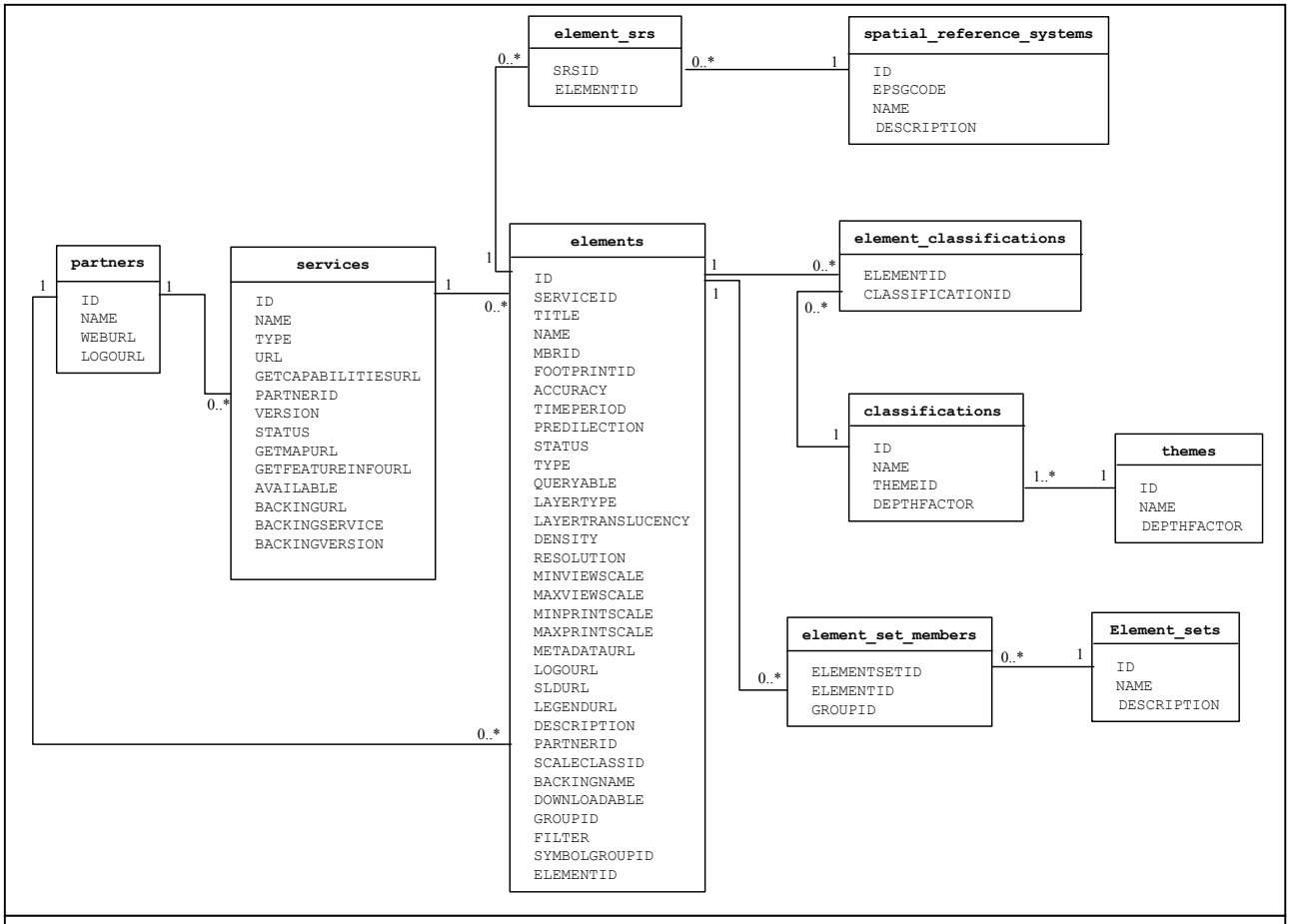

Figure 5. A more detailed view of the catalog data model. This diagram shows the table names and attributes for the more important tables of the actual database, but is still not complete. This diagram uses UML, rather than traditional ERD, notation.

[4] The large difference between the total number of layers and the number of public layers is caused by several things. Many organizations' WMS include layers that are simply outside the scope of *The National Map*, such as ecological data or weather data. Many State and local WMSs also serve data that duplicates coverages served by the USGS (for example, hydrography data derived directly from USGS DLG or NHD data is common). There are both technical and programmatic reasons for registering such layers in the catalog anyway. Selecting which layers to show in the viewer makes the application more coherent and is one of the things that separates *The National Map* from a data clearinghouse.

sets. Partners can be part of consortiums or groups. Partners have one or more points of contact, each with one or more phone numbers and email addresses. Information about attribute-to-symbol mappings occupies a separate and extensive set of tables (see Styled Layer Descriptors below).

3.2 Other Metadata and Catalog Maintenance

The catalog database contains a variety of other metadata: spatial reference system, nominal data scale, suggested ranges for display scales, association of each layer to USGS data themes and subthemes, links to FGDC geospatial metadata, links to backup sites for the service or layer, geographic footprint and minimum bounding rectangle, and so on.

The catalog brings together various levels of metadata (organization, service, layer, feature), from various sources (services' self-descriptions, organization points of contact, created by manual data inspection, created by monitoring software). The overall purpose of this information is to improve the quality of applications that depend on the distributed data sources – the catalog gathers together and organizes information about services and layers so individual applications don't need to. The work of keeping track of all this information has been centralized.

The benefits of this centralization may not be immediately obvious. The GIS state-of-the-practice is still to acquire static data files, either through download or fixed-media purchase, then organize these files on a local computer to serve a specialized application. Internet services make it possible to do GIS work without making local copies of each and every dataset of interest. Instead of downloading a file and loading it into your GIS, you instruct your GIS to connect to a remote service. This reduces or eliminates many tedious retrieval, formatting, and data management tasks.

But as the number of remote sources grows, different data management problems emerge. Many of these are variations of the "dead link" problem common on normal Web pages. Services get moved and renamed; layers can be added, removed, renamed, or have other attributes changed; organizations lose interest and drop their services from the Internet. Most of these kinds of changes are detected by the catalog monitoring systems. When an unexpected change happens, USGS personnel contact the service owner, find out why the change occurred, and update the catalog information as appropriate. This does not mean that the catalog never holds any dead links, or that all services that contribute to *The National Map* are always up. It does mean that dead links usually don't stay dead very long, and that if a service is down, it isn't because nobody is aware of that fact.

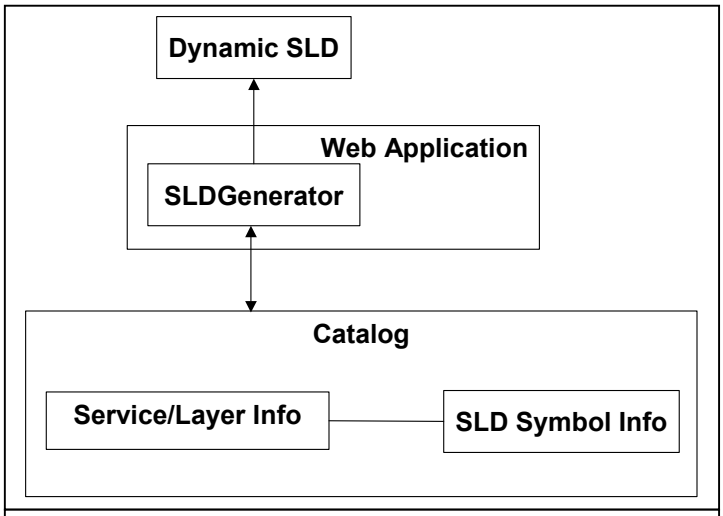

Figure 5. Storing the attribute-symbol relationship in the catalog permits software to create SLDs at runtime.

3.3 Styled Layer Descriptors

An obvious problem with building a national application from State and local data sources is inconsistent attributes and symbols. Local organizations choose attribute models and symbols for their own reasons, and are not necessarily able or willing to change either for the benefit of other organizations.

Styled Layer Descriptors (SLD) are a partial solution to this problem. SLDs are an extension of the WMS specification. They are implemented with XML documents that contain instructions such as (figuratively speaking) "for features with attribute A that are part of layer X in service Y, override the default symbol and use the following combination of colors, line widths, etc.."

Another way to say this is that SLDs implement a conceptual mapping from a particular GIS attribute model to a particular display symbol set.

The application to *The National Map* is obvious: it is desirable to display (for example) all Federal highways with the same symbol. The target symbol set is, in theory, a constant national symbol set. But the source GIS attribute models are many in number. There is no assurance that any two

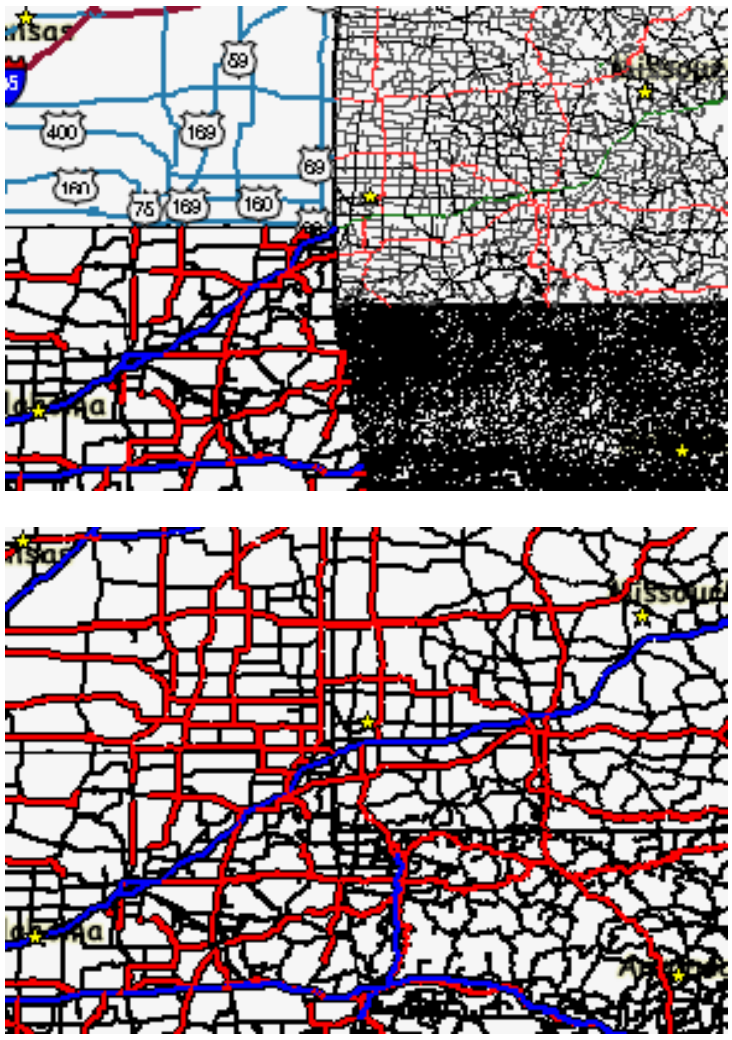

Figure 6. The top image displays the road symbols provided by the State services at the corner of KS-OK-AR-MO. The bottom image is the same area with styled layer descriptors (SLD) applied.

Using SLDs degrades the original data by collapsing the States' attribute models to a kind of lowest common denominator. Nevertheless, the enhancement to visual clarity is so great that the application of SLDs must be regarded as a quality improvement. One way to look at this is that the **signal** of road information is reduced slightly by collapsing the attribute models, but the visual **noise** is reduced a huge amount by standardizing the symbols. The signal-to-noise ratio is therefore much more favorable in the bottom image.

The SLDs are applied only in *The National Map* viewer. The original data are not disturbed, nor are any other applications that use either the default symbols or different SLDs.

data sources will use the same attribute model. This means that the attribute model of each new data source must be mapped to *The National Map's* symbol set. This mapping is necessarily manual. But it need only be done once for each data source, if the source attribute model and the target symbol set are both stable.

This mapping between attribute model and symbol set is stored in the catalog. The technology of relational databases makes the attribute-symbol relationships maintainable, just as it makes partner-service-layer relationships maintainable.

SLDs allow national applications to be created from local data sources without altering the source data or disturbing any existing applications that use either the default symbols or a different set of SLDs.

4 Database-Driven Applications

The most common method for publishing GIS data on the Web is to organize data of interest in one map service, then build an application (a viewer is the most common example) to present the data from that service. This is the limiting case of Figure 1: one application and one WMS. This works fine for researchers or small organizations that want to publish small numbers of data sets,[5] but as Figures 1 and 2 illustrate, this solution does not scale.

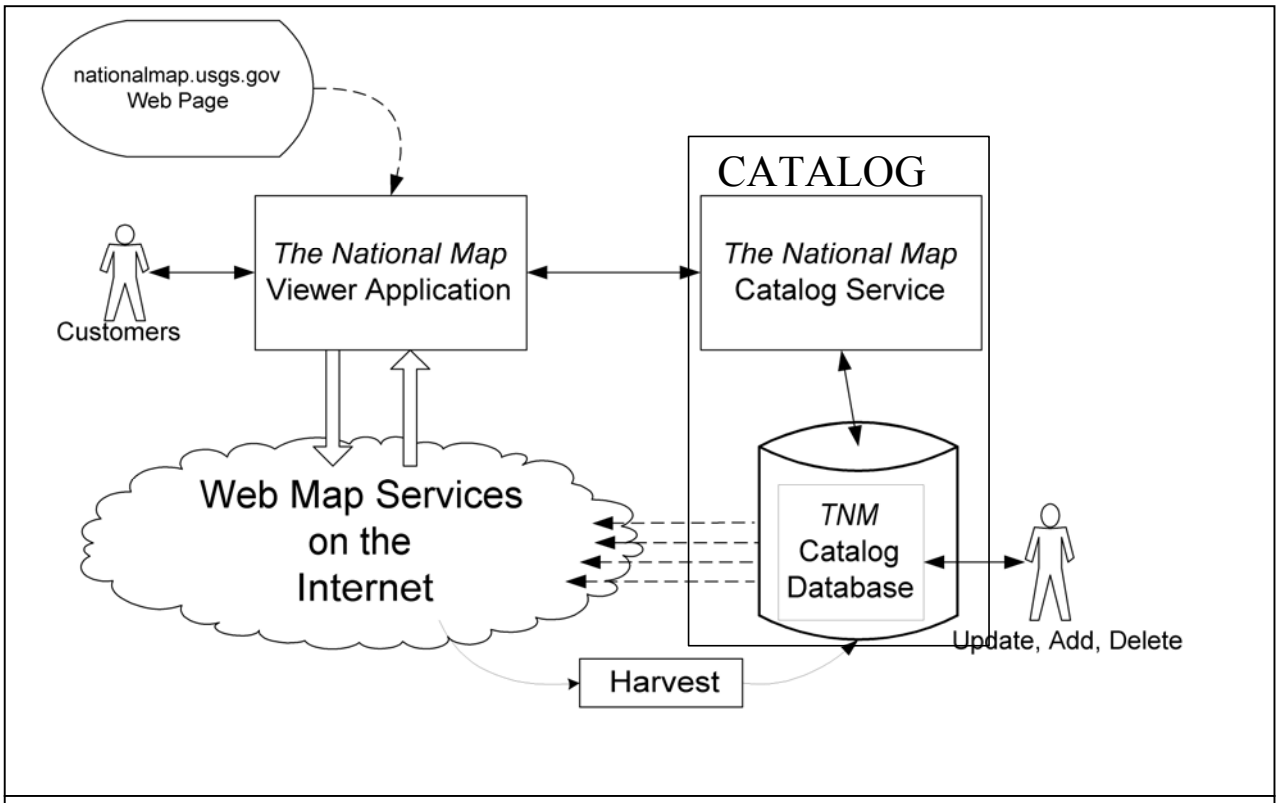

Figure 7. Catalog-driven applications get information about where to find WMS data from the catalog service. The application then connects directly to WMSs out on the Internet to view or retrieve the actual geospatial data. The work of keeping track of relevant WMSs has not been eliminated, but it has been centralized. The overhead of maintaining this information is assumed by the USGS, represented by the human figure in the lower right. The "viewer application" box is just one representative of a potentially large family of applications.

Figure 7 is a different way to view the system represented in Figure 2. Figure 7 shows only one application (*The National Map* Viewer Application), but the same design applies to any number of applications. Each application must "know" (hard-code) only one service: the catalog service. By querying the catalog service, the application can find the addresses of WMSs that may have data of interest. Once relevant addresses have been found, the application then queries the source WMSs directly.

The work of keeping track of WMS locations and characteristics is not eliminated in this model, but it is centralized. The human figure in the lower right corner represents USGS labor and overhead for keeping the inventory of WMS locations current. This maintenance involves some direct interaction with WMSs and their owners, first to find out what they contain, and later to monitor them for changes and continued availability.

5 The Open GIS Consortium (OGC)

The following is from written testimony of David Schell, President of the OGC, Inc., before the House Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census:

> There are two kinds of geospatial standards: data content standards and software interoperability standards.
>
> The OGC is the only organization that develops and promotes geoprocessing software interoperability standards. Interoperability involves different systems exchanging data and instructions in real time through open, consensus-defined interfaces. The OGC's geospatial software interoperability standards help software users and developers answer questions such as: How can my different Geographic Information Systems (GIS) … exchange geospatial data and geoprocessing instructions in real time? What open interfaces do I need to build a spatial data catalog, or "spatial search engine", that works as well as Google and Yahoo work with text data? These software interface and encoding standards are called OpenGIS® Specifications. OpenGIS Specifications are free and publicly available software engineering specifications similar to those that underlie the Web. The OGC does the same kind of work the World Wide Web Consortium does, but our efforts are focused on complex and diverse geospatial technologies. The OGC creates the spatial dimension of the Web. The OGC's standards make it possible for spatial information of all kinds to be easily communicated over the Web.

As a matter of policy, the USGS is committed using open standards, to the fullest extent practical, to implement *The National Map*. "To the fullest extent practical" is an important caveat, because the relevant standards, their commercial implementations, and *The National Map* itself are all evolving. Like other instances of technological evolution, it is often difficult to coordinate all activities.

At this writing, the policy basically means two things:

1. All Web Map Services (WMS) and Web Feature Services (WFS) that serve as data sources for *The National Map* must be OGC-compliant, or at least reasonably close to the state of the practice (in this early stage of the game, few OGC interfaces are bug-free, and there is not yet any generally accepted way to certify OGC compliance for a service).

[5] Not necessarily small volumes of data. The key is the number of different layers, not the quantity of data.

2. The catalog information service conforms to an OGC discussion paper, a precursor to a specification. As the OGC Catalog specification becomes formal, the catalog service will be modified as needed to conform to the specification.

These requirements do **not** mean that non-OGC interfaces and applications must be abandoned to participate in *The National Map*. A WMS may have multiple interfaces. The most common example is ArcIMS, which has a native interface based on ArcXML, but a second interface that is OGC-compliant. The OGC interface is a piece of software that essentially translates OGC queries into ArcXML queries, and ArcXML responses into OGC responses.

6 References and Further Reading

Schell, David. Written Testimony before the U.S. House Committee on Government Reform, Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census. Hearing on "Geospatial Information", June 23, 2004
http://reform.house.gov/UploadedFiles/Schell%20Testimony.pdf

For pointers to more technical information about *The National Map* Catalog, see http://mcmcweb.er.usgs.gov/catalog. Other papers referenced in this document are posted here.

Attachment A. Data Dictionary for Figure 3

Data Stores

WMS = Web Map Service. A map service that conforms to the OGC Web Map Service Specification. A map service serves raster map images (is opposed to GIS spatial objects). This is a data store only from the perspective of applications that use it. From the perspective of the WMS itself, it is more like a process.

Catalog. USGS-owned Oracle database that tracks information about WMSs (and, in the future, other types of GIS information services) that provide data to *The National Map*, including the relationships between partners, services, layers, and geographic extents.

Processes

Service creation or discovery, evaluation. A large collection of organizational procedures, data management systems, and software tools that collectively contribute to negotiation of partnerships. In most cases the goal of partnerships is to gain access to a WMS that serves data of interest to *The National Map*.

WMS harvest. A collection of technical processes and software tools for populating the catalog database. Inputs are information from MPOs and data providers, which come in a wide variety of forms and various levels of completeness. These organizational inputs allow additional semi-automated input from source WMSs.

Applications. Any software that uses information from the catalog to find and query WMSs. *The National Map* viewer is currently our best-known application, but is not the only one. Included in this box are "internal" applications that constantly evaluate and update the catalog. The diagram does not show that internal applications may be able to alter data in the catalog, but external applications can only query the catalog.

Catalog Service. An Internet information service that provides a public API (application programmers interface) to the catalog database. Like other information services, the catalog service takes queries in URL and returns responses in XML. The API spec is published in a separate document, posted on the catalog Web site at http://mcmcweb.er.usgs.gov/catalog

GOS registration. Process for publishing a dataset in Geospatial One-Stop. Datasets that are registered with *The National Map* in the catalog are registered in GOS through a semi-automated transfer of metadata. This costs the USGS a small amount of effort, but relieves *The National Map* partner organizations from having to register with GOS.

GOS harvest. This process is the "reverse" of GOS registration. In this case, a partner that has already published data in GOS is mostly relieved of the responsibility to reregister with *The National Map*. At this writing, this is the only process in Figure 3 that is not yet implemented.

Terminators

Partner. Any organization that is not a USGS national data program, but owns a WMS that provides data to *The National Map*.

USGS National Data Program. One of the official USGS data programs, such as hydrography, elevation, or transportation. These programs may own data production systems and WMSs, or may

be intermediaries between Partners and the rest of the catalog system, or both. The distinction between Partners and Data Programs is important for policy and administrative activities, but is not important for the technical processes of populating and maintaining the catalog.

Users. Any person or external process that uses an application. Most users access the viewer to look at map images created by the overall *The National Map* system, but the term includes many other types of people and processes.

Other organizations. Any person or organization that registers data directly in GOS without being involved with any process of *The National Map*.

GOS. Geospatial One-Stop.

Data Flows

Layer, service, organization metadata. Information about an organization (such as its home Web page), the information services owned by that organization (such as their URLs), and the layers managed by those WMSs (such as their names, spatial reference systems, etc).

GIS data. Geospatial data that are the primary inputs to a WMS. These data are, in fact, the reason for having a WMS. The WMS is just a way to make delivery of the data more standard and convenient.

WMS query. Any one of the three kinds of queries specified by the OGS WMS Specification (GetCapabilities, GetMap, and GetFeatureInfo). Queries are in URL syntax, and are tightly defined by the OGC WMS specification.

WMS response. A response to a WMS query. Depending on the query, the response may be a raster image or an XML document. In either case, the characteristics of the response are defined by the OGC WMS Specification.

Service metadata. Information about a WMS, most importantly its URL address, but may also include other information such as the names of its layers.

Geospatial metadata. Metadata about the sources, currency, and content of layers, typically as defined by the FGDC Geospatial Metata Content Standard.

Request. Something a user asks an application to do. In most cases, the application defines the domain of possible requests by the interface of buttons and forms presented to the user.

Query. A question or request for data from an application to the catalog or the catalog service. The domain of these queries is defined by the content and structure of the catalog. The form of the query is defined by the interface being used (the catalog service, SQL, another application…)

Update. An instruction to change or add data to the catalog database, from an application to the database. Requires write privileges to the database.

FGDC file. An XML file of geospatial metadata that conforms to the Federal Geographic Data Committee's Geospatial Metadata Content Standard. The GOS registration process has specific format requirements for this file.

Registration info. Information required to register a GIS dataset in GOS. Basically a minimal set of FGDC metadata, combined with some organization metadata about the owner of the dataset. Supplied to GOS through an interactive, online form.