The Harvard Geospatial Library

http://hgl.harvard.edu

What is HGL?

The Harvard Geospatial Library (HGL) is an on-line repository of digital resources for use in mapping and geospatial inquiry and research. From the Harvard Geospatial Library interface, a user can search for and download geospatial data. The geospatial data held by HGL generally consist of spatial data with associated tabular attributes, raster data, and digitally imaged and georeferenced paper maps. There are approximately 5000 of these digital layers in the repository, and each one of them has an associated descriptive metadata file contained in a searchable catalog. All layers in HGL are also linked with MARC records in the HOLLIS catalog. As a resource primarily designed for the Harvard community, the Harvard Geospatial Library will continue to add data relevant to the research interests of the University as well a wealth of more general base data.

Inception and Direction of HGL

The Harvard Geospatial Library grew from a need to provide the Harvard community with more access to geospatial data from a centralized location. The first proposal for funding came in 1998 from the Harvard College Library which housed one of the University's only centralized/public GIS service locales at the Harvard Map Collection. An LDI (Library Digital Initiative) grant was received in 1999 to get HGL off the ground with a Systems Developer and Project Manager, with HCL providing a GIS Specialist. The V. Kann Rasmussen Foundation provided a key funding boost to add functionality and longer-term personnel funding to the project. The Harvard College Library, the Harvard University Library and LDI funding continue to support the HGL team today.

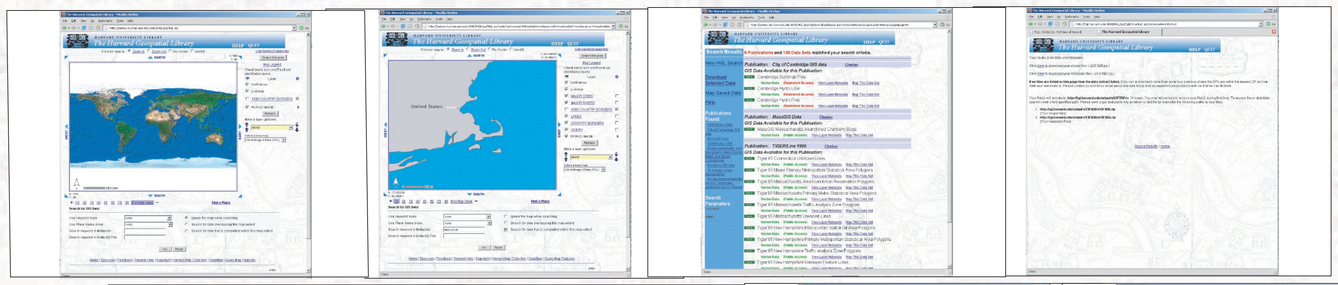
There is an active HGL Steering Committee whose responsibility is to set the mission of HGL and determine how best to meet its goals. The Steering Committee is comprised of individuals from across disciplines who are in-tune with current GIS research needs of core users in the Harvard Community.

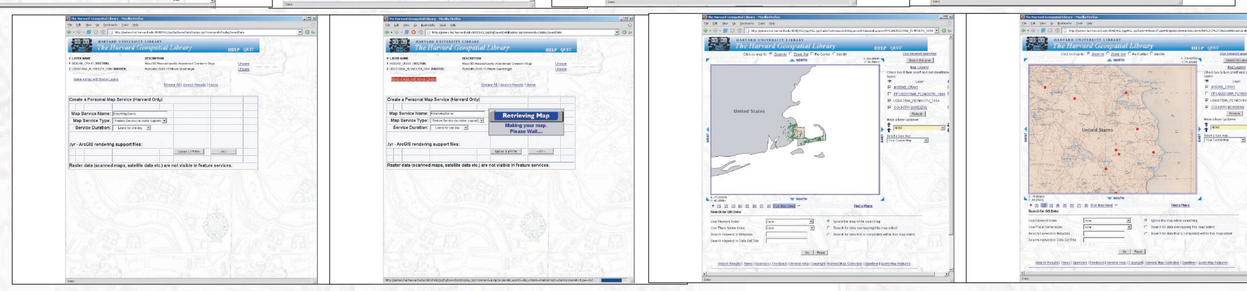
HGL In A Nutshell...

- Data Repository
- Metadata Catalog
- User Interface
- Download and/or Display

Advantages & Impacts of HGL

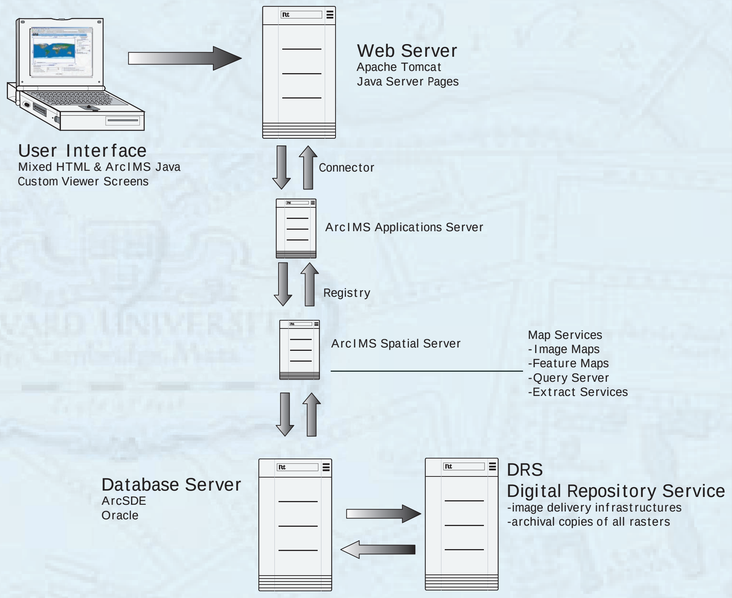
- Data is easier to Find
  - via geographic searching
  - via complete metadata search ability
- Data is easier to Use
  - common formats (e.g. shape files)
  - download only the area of interest
- Data is easier to Share
  - common location
  - HGL relieves the burden of administering servers and databases





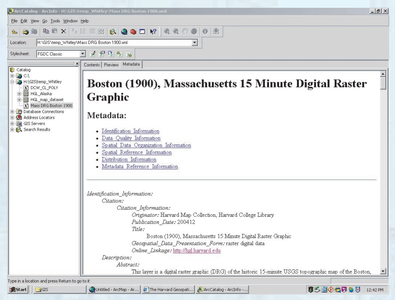
How is Data Stored?

System Architecture

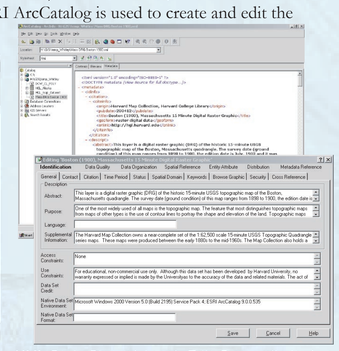


Metadata

A keyword search performed on the HGL home page queries a metadata catalog that is stored in the ArcSDE database. Harvard Geospatial Library utilizes the FGDC CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata) for its metadata, as well as some provisional metadata for data created "in-house". ESRI ArcCatalog is used to create and edit the metadata records.



FGDC Style Sheet displayed in ArcCatalog



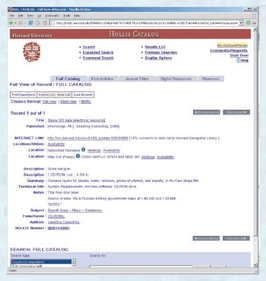
XML Style Sheet (above); Edit Screen (below)

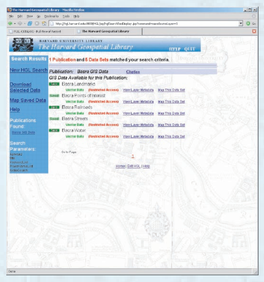
Metadata is stored as XML in an Oracle database, and indexed with Oracle Intermedia. When a search is performed through the HGL home page, it is the XML metadata field that is queried. XML metadata tags are broken down into *groups* of searchable *tags*. These tags are indexed with Oracle Intermedia to provide full text searching capability.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
        <!DOCTYPE metadata SYSTEM "http://www.fgdc.gov/metadata/fgdc-std-001-1998.dtd">
        <metadata>
        <idinfo>
                <citation>
                        <citeinfo>
                                <origin>National Imagery and Mapping Agency</origin>
                                <pubdate>19920</pubdate>
                                <title>Hypsographic, Digital Chart of the World</title>
                                <edition>1</edition>
                                <geoform>vector digital data</geoform>
                                <pubinfo>
                                        <pubplace>Fairfax, VA</pubplace>
                                        <publish>National Imagery and Mapping Agency</publish>
                                </pubinfo>
                                <onlink>geodesy.harvard.edu/.hy</onlink>
                        </citeinfo>
                </citation>
                <descript>
                        <abstract>DCW is a comprehensive 1:1,000,000 scale vector basemap of the world. It consists of
                        cartographic, attribute, and textual data stored on compact disc read only memory (CD-ROM). The
                        primary source for the database is the National Imagery and Mapping Agency's (NIMA) Operational
                        Navigation Chart (ONC) series. This is the largest scale unclassified map series in existence that
                        provides consistent, synoptic coverage of essential basemap features. The database
                        contains more than 1,900 megabytes of vector data and is organized into 10 thematic layers.

                        The HY layer consists of contour lines with elevations expressed in feet above mean sea level, with an
                        interval of 1000 feet.</abstract>
                        <purpose>The Digital chart of the world is a general purpose global database designed to support
                        Geographic Information System applications.</purpose>
                </descript>
```

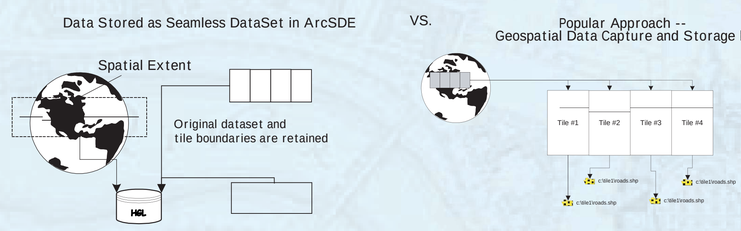
The data contained in HGL is also cataloged using more traditional library tools. As a result, records can be searched for in HOLLIS. For example, a HOLLIS search for "Basra" returns a record for "Basra GIS Data", which in turn contains a link to the digital datasets available through HGL.





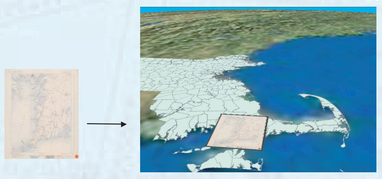
Vector Data Storage in ArcSDE

Vector data layers are loaded into the database via ArcSDE commands and/or ArcCatalog. Many digital datasets are purchased as "tiled" blocks of adjacent data, which can slow down the process of searching, downloading and "clipping" to the geographic area of interest. Vector data that were originally tiled are stored as seamless datasets in ArcSDE. A seamless dataset does not require manual merging of tiles in order to cover one's area of interest. These layers are always distributed in shapefile format for ease of application.

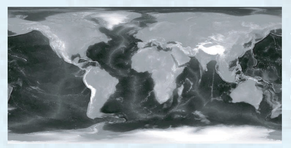


Raster Data Storage in ArcSDE

Raster data layers are approached in two different ways depending on the source file (digitized map vs. traditional raster data sets such as digital elevation models), but both are handled similarly in ArcSDE. Raster data are stored as indexed SDE layers, but SDE utilizes different table types for raster vs. vector.



Historical maps are digitally photographed and georeferenced. An archival TIFF file is stored in the Digital Repository System. The map is georeferenced in ArcMAP and stored in ArcSDE.



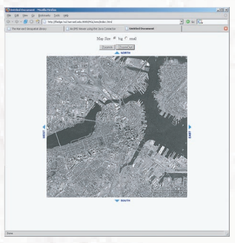
Traditional raster data (e.g. ESRI GRID, DEMs, DRGs) are loaded into the database via ArcSDE commands and/or ArcCatalog. This process is similar to how vector data is loaded. The raster dataset is stored as an indexed SDE layer.

Alternative Access

Accessing the Harvard Geospatial Library is typically done through the User Interface via the URL:
http:\\hgl.harvard.edu
Here are some brief descriptions of *alternative* ways to access the data stored in the HGL database:

WMS Connector:

The Web Mapping Service (WMS) specification is the result of a collaborative effort by the Open Geospatial Consortium (OGC) to broaden access to Map Services. With the ArcIMS WMS connector in place, any OGC WMS "viewer" (a WMS compatable browser or client) can access an ArcIMS Map Service. HGL uses ArcIMS and the WMS connector, so WMS clients can connect to HGL services.



Example of imagery from HGL in a simple WMS client.

Direct Connect:

Direct Connect allows a user to connect to the SDE database directly through ArcMap or ArcCatalog, bypassing the need to raster and vector data through the HGL web-based user interface. Access to this service is granted to users based on their data needs.

A future use of Direct Connect access will include access to ESRI's StreetMap data for use in geocoding.

Web Services [Future Application]:

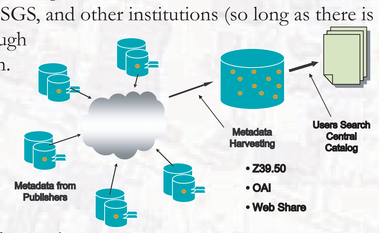
HGL will eventually host many different kinds of geospatial web services, in addition to the map services currently available, such as an address matching service or a census geography service where users can join their own data to census tracts or block groups.

ArcIMS

ArcIMS is software that enables geospatial data to be displayed and delivered via Web services. For instance, the maps that are displayed in HGL are created by ArcIMS, as are maps created at the Geography Network and many other sites. These ArcIMS services can be viewed using many tools, including ArcMap and ArcCatalog, and services from different hosts can be displayed together

Metadata Harvesting [Future Development]

HGL has the potential to tap into other data repositories, such as the Harvard/MIT Data Center, Harvard's Virtual Data Center, USGS, and other institutions (so long as there is agreement between the two parties) through an "Open Archives Initiative" connection. Metadata harvesting would direct metadata from various publishers to be indexed in a central server at Harvard, from which users could search through a central catalog.



An alternative approach to the metadata harvesting concept, would be to use a connector, like Z39.50, to use a protocol to send out search requests to the remote repositories. Instead of piping metadata through a local server, the connector would network out and send "hits" back to the central cataloging system.

Highlighted Datasets

China Historical GIS (Harvard-created Dataset)

MassGIS politcal and environmental data

Historic USGS topos of Mass. (c.1890)

Army Map Service topos of China (c. 1950)

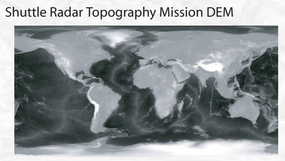
Digital Chart of the World

North Korea Administrative Units

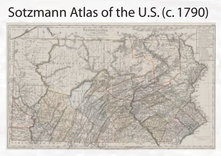
Baghdad Planimetric data

U.S. Exclusive Ecomonic Zone

Shuttle Radar Topography Mission DEM



Sotzmann Atlas of the U.S. (c. 1790)



Future data highlights:

Civil War georeferenced maps
China Population Census data
Massachusetts DRGs
AirSAR Imagery of Central America
Land-Use raster data (world-wide)

Credits:
Poster layout by Whitley Frost, Harvard Map Collection with text and image input by Bonnie Burns, Harvard Map Collection, and David Siegel, Office of Information Services.
May, 2006