

Stephen Appel*, Marcy Bidney*

Geodex 2.0: saving a legacy map series cartobibliography

Keywords: Geodex 2.0, map series, map databases, indexing, web map

Summary: In the early years of library automation many libraries were innovative in their approach to bringing new technologies to use for library search. Some of those technologies have long ago been archived or in some cases lost. Specialized libraries such as map libraries, who were early adopters of new technologies are now finding themselves left with legacy systems that are close to or past their life cycle and we are faced with the decision to either phase the system out of service and archive it, or to look for ways of breathing new life into these systems. The American Geographical Society Library was faced with such a decision recently.

In 1988, the American Geographical Society launched GEODEX: GEOgraphic InDEX System for Map Series. GEODEX was an innovative means for indexing the hundreds of map series held in the collection at the AGSL. At the time of its launch, GEODEX was one of the largest automated cartobibliographies in the world close to 400,000 entries. Changing technology and computer upgrades threatened all of the data held in GEODEX due to operating system incompatibility with the DOS program that was the heart of running GEODEX. That would have meant over 25 years of work and hundreds of thousands of database records would be rendered useless unless there was a way to bring GEODEX into the 21st century. After more than a year of development the conversion of GEODEX is just about complete. A new interface for adding new maps and searching the database bring a fresh look to this impressive database of map series, now including online accessibility for the world. This presentation will talk about the history of this legacy database and the process of converting it to 21st century technologies and what lessons can be learned as we continue to be faced with conversion of early adopted technologies.

Introduction

Developed in the 1980s at the American Geographical Society Library (AGSL), Geodex was a command line based software with a basic graphical user interface (GUI) designed for rapid entering and searching of records for paper maps in large series. Designed by former AGSL Curator Chris Baruth, it was unique and innovative in many ways, but two stand out as particularly innovative for the time. First, Geodex provided a means for geographic searching by coordinates decades before any formal library system could provide that capability. Using bounding box coordinates a user could easily find maps at multiple scales, from multiple publishers etc. that fall within that box. Previously, the only means for conducting this type of search was by using a paper series index that may or may not represent a region at multiple scales. Second, based on principles of shared catalogs where it was believed that “a work should be cataloged once and only once” (Borgman, 1997), Geodex was made to be shared with other libraries with the idea that they would enter their holding data and send those files back to the AGSL to be ‘reconciled’ with the datasets creating a national, and perhaps, international catalog of map series. This enhanced the usefulness of Geodex because after completing a search, the

* American Geographical Society Library, University of Wisconsin-Milwaukee, USA [bidney@uwm.edu]

user could see if a particular map sheet exists and in which library it was available. However, there is little evidence that this system was widely used and implemented in this manner.

Geodex serves as the primary map index database for most of the large national, international and nautical chart sets in the AGSL collection. These maps are not cataloged anywhere else so Geodex serves as the main finding aid for these items. Since its inception in the 1980s we have continuously added our holdings and the database now holds more than 400,000 records. In 2013, after major computer upgrades within the UWM Libraries and the AGSL it was discovered that Geodex, in its DOS format, could no longer run on the upgraded computers and so a decision needed to be made - let the sun set on Geodex and the thousands of records stuck in the database or venture into creating an updated version of Geodex, with the hope of helping to realize its original vision of being a truly shared database among map libraries around the world. So, in looking to the future, the decision to create what has become known as Geodex 2.0 was made. The rest of this paper will give a history and description of the original database and discuss the creation of Geodex 2.0.

We had a number of goals for this project. Geodex 2.0 needed to serve the same purposes as Geodex 1.0 but we understood that we would need to look at the entirety of the software with an open mind and not keep it the same just because that is always how it had been. Breathing new life into the database design of Geodex was a major goal. Another goal was to create a geographic search interface for the staff and public users to utilize when searching for maps. The interface needed to be fluid, fast and seamless to users. We understood that this would be the most challenging aspect of the project but still have high hopes for achievement. The third goal was to be able to realize the original vision of a shared index.

We want to provide means for libraries around the world to enter their series holdings into the database so that when users are searching they will find the sheet they are searching for, regardless of where it is held. While this is an ambitious goal, efforts are already underway to serve this purpose in different ways. The Colorado School of Mines and Michigan State University are both working on their own indexes and making them available on the web. Additionally, the Geodex data has been requested by others who are interested in tackling this project.

The efforts described in this paper are not the first attempts to modernize Geodex. Through the years multiple students and staff have attempted the project. Each time new efforts were attempted they were abandoned. From limited documentation available in the AGSL, it seems that there were two issues that caused the abandonment. It is often the student workers in the AGSL who are attempting the upgrades and student turnover can tend to be high. The other reason we surmise to be technology. It just does not seem that the technological capabilities were available to make a concerted effort until the past four or five years.

The most important and most successful of the attempts prior to this effort was the extraction of the data from the Geodex files into an access database which was then used to produce shapefiles based on the index values in the table. The creation of shapefiles was important because, for the first time, the data locked inside Geodex could be readily shared with other libraries who were interested in adding their holdings into the database. It still did not allow for easy sharing of the data, but these shapefiles have led to efforts in the last few years to begin working toward that vision of a shared database.

Geodex 1.0

Designed at a time when library automation was beginning to become more widespread, it had the potential to provide map libraries with a means for cataloging their collections with information fields relevant to cartographic objects that they were not finding within the newly developed MARC metadata format. In this way, Geodex was designed to be different from a traditional style catalog or index whose primary focus was the cataloging of books. Similar to the developing library automation systems, Geodex was built upon a basis of *fixed* and *open* fields and was designed to maintain a high level of consistency across a map series and set up so that repeating information did not have to be entered for each map. Fixed fields were set at the *File* level and typically included information such as the publisher, format, and catalog location information. Because each file was dedicated to a series, these fields rarely needed to be edited and were therefore *fixed*. Open fields defined the differences between maps in the series and included fields such as *record*, *location*, *scale*, *prime meridian*, *edition*, and various other fields related to years and map units. When editing a record, the open fields would need to be entered each time using codes to help further decipher the information and organize it into a consistent database. For example, when Geodexing a large series such as the USGS US Topographic 24k series, information that is static throughout the whole series such as publisher, scale, format, extents, prime meridian, projection, call number, and other fields could simply be copied over into a new record allowing only the fields requiring a change to be changed. These fields could include the publication year, other dates, map titles, contour or isobath intervals and units, and in the case of sharing data with other institutions, the library at which the sheet is or is not held. A typical process for entering a new sheet was very quick and accurate. On average, about 100-200 of these maps could be entered in a work day. Some maps could take longer, for example, new maps since they would not have had an existing record to copy, changing extents of a map with the same title, and updating existing records to reflect AGSL holdings.

Another innovative feature built into Geodex was the ability to search by geography using the geographic index feature. It could be argued that this feature represents the first computerized capability for geographic search that would accurately search multiple sets of maps at the same time. It works just as geographic search features work today - two sets of coordinate pairs were stored creating a bounding box representing the North, South, East, and West extents of a sheet. Creating and editing records in the geographic index was as simple as entering in these four extents in a degrees, minutes, seconds format. If a map sheet used a prime meridian system other than Greenwich, the original software would convert the index into Greenwich. It wouldn't be until the year 2000 when Endeavor Information Systems announced an enhancement to its Voyager library system that included a geospatial search interface. Endeavor claimed that this enhancement would allow users to search using bounding box coordinates or by an exact point coordinate. (Endeavor Information Systems, 2010) It is unclear if any library using the Voyager system ever implemented this system enhancement.

Originally designed to store data on floppy disks, the database was transferred to a network drive as soon as that capability was accessible at the AGSL. The database functioned by creating one file for each map series, for example file F0063 stores the 1:50,000 scale maps of Germany or F0177 which stores nautical charts with scales larger than 1:399,999. Due to this structure, each of the files that made up the database varied greatly in size depending on how many sheets each set contains.

The nautical chart file itself stores over 10,000 records. In its current state, the database exists simultaneously in different forms. The last extract from Geodex before it became unusable is saved as an access database. The access database can be used to easily create shapefiles. This process was performed frequently and there are a few versions of Geodex shapefiles available. However, these are not very usable because they are static representations of a database that should be *alive*, i.e. they are a snapshot in time of a database that continues to grow.

Envisioning Geodex's potential

Geodex is less like a catalog and more like an index map that has some interactivity and a relational database back end and moving forward we felt it was time and technological capabilities were finally available for it to finally fully function in that manner. Where the original Geodex had combined the search and editing capabilities into one piece of software, we no longer believe that this is the best way to go about meeting the needs of the AGS Library. Instead, a user facing application for searching and visualizing the database should take advantage of the huge advances in web mapping capabilities that have occurred since the design of the software, before the internet was ubiquitous. On the back end, there should be a simple application for adding records to Geodex and for maintenance of the database for GIS staff or, in certain situations, AGSL professional staff. The rest of the paper will discuss the ongoing process of developing a Geodex built for the future.

The Backend

It's often considered ill-advised to reinvent the wheel, but when given the opportunity the option should at least be weighed. For so long, Geodex has been locked in an aging interface with limited access to the inner workings of the application and therefore use has been limited to entering and exporting data from the application. Significant changes to the conceptual model of the database have not really been possible until this point. It is important to consider how Geodex functioned previously because it had a big impact in the design of the data files. For example, each file has an identical field map sometimes with half of the fields being null for a record due to the rest of the records in that file needing that field.

This results in saving large quantities of unnecessary data. Today this would be regarded as poor database design but when Geodex was developed in the 1980s it was necessary to maintain consistency in a simpler piece of software. The backend database structure had always been left alone because the actual Geodex application was still in use and editing it was not a simple task.

There were a number of questions to be considered before moving forward with new database design. Some of the questions we asked were - *How could the database design, specifically the fields, be more efficient with fewer null values in the database overall? What information in the old database is no longer useful? What new information should be added to the database to increase its usefulness? Since the files are no longer stored on floppy disks and data storage availability has increased exponentially, is there a reason for the Geodex files to remain intact or should files be split and/or combined?*

We also took into consideration some best practices for database conceptual models, and were aware of the need to allow for future expansion of the project in ways we may not be thinking about now.

We also wanted to test the use open source database software for this project. While GIS staff and students working in the AGSL have had some exposure to open source GIS software such as GeoDa and QGIS as well as other open source software like PSPP and Gimp, this project would represent a first foray into open source database management software.

After much discussion and learning about the capabilities of a variety of open source database management software programs, we decided to try PostGIS. Essentially, PostGIS is a spatial extension to the popular open source relational database management software PostgreSQL. PostGIS made it easy to migrate the Geodex shapefiles, extracting all of the fields and information from the tables, and thus, created a new database. Using PgAdmin, the software included with PostGIS, we were able to perform both basic and advanced SQL statements on the database tables.

The biggest issue which has impact both on the database and the user interface is whether each file needs to have the same design. Some map series have, for example, multiple dates. A map may have a printed date, a revised date, and overprinted date, an “edition of” date, a photo-revised date, etc. Another series may only have been printed one time and literally every map in the drawer has the same year. Do both files need 6 date fields to accommodate? Or in the case of isobars: currently there are 5 fields: *isobar_m*, *isobar_ft*, *isobar_fa*, *contour_ft*, *contour_m*.

A map is probably not going to have more than one of these fields so they are generally mutually exclusive. In some records there may be a ‘0’ under *isobar_ft* indicating that the soundings on the chart are in ft, but the interval is not known or at least wasn’t entered. But this could easily be reduced to two fields by having one field indicate the type of value (one of the 5 options mentioned above) and then another field with the value. It could even be combined to one field where a value such as “CM6” could be interpreted as “Contours at an interval of 6 Meters.”

Nearly all of the fields in Geodex were critiqued in this way. Now, rather than having 179 different files, there is one feature class in a geodatabase with 179 *subtypes* representing the old Geodex files. It is a big file but it works for our purposes. The geodatabase itself was kept simple and with many changes to the fields resulting in significantly fewer fields 32 now where there had been 58.

Once the new database, which from here on will be referred to as the *back end*, was created, a user interface for adding data was developed in Visual Studio. The first version was a fairly simple windows forms application which essentially replicated the interface capabilities of the original Geodex with a more modern look and more up-to-date capabilities, such as using a mouse. This method was fairly effective but very prone to bugs. For example, if the Caps Lock was activated, it would cause the case-sensitive commands to be incorrect.

When we consulted with other experts on campus regarding the web map application, we realized that we were going to have to move forward with developing Geodex 2.0 with an enterprise GIS software and began development of the database and tools for editing records in ArcGIS.

So, in lieu of a standalone application, we opted for creating an ArcMap Toolbar using ArcObjects and Visual Basic .NET. The toolbar allows for working with a Geodex database located either on a remote server or stored as a file geodatabase on the user’s machine. The toolbar has a tool to add data from the database to the map, a tool to easily filter and query the database, a tool for editing attribute data and another for editing spatial extents easily, as well as a final tool for reconciling holdings information for existing series.

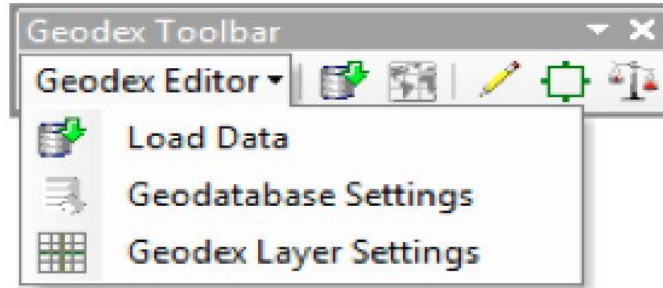


Figure 1. Geodex toolbar in ArcMap.

The Record Editor is by far the most robust of the forms in this application. As you can see in Figure 2, this form is used to edit a record manually, create new records, and other primary functions of the application. In the top Group Box, there is a text entry box, a few button controls, and labels. The text entry box allows the user to load the information from a specified ObjectID. This could be useful if you're using the table and want to make a record with primarily shared information with an existing record.

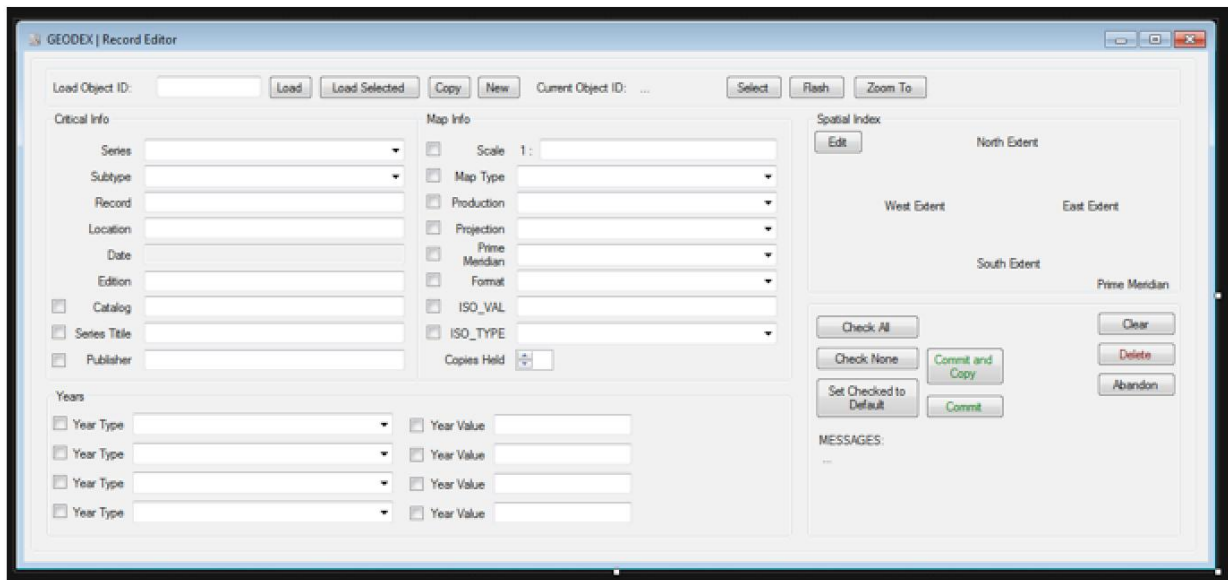


Figure 2 Record editor in ArcMap.

The spatial index editor form is used for directly changing the spatial index extents of a record. This form allows for entry in either Decimal Degrees or DMS. It also can handle entries in multiple Prime Meridians. All index extents in the database are stored in the Greenwich Prime Meridian as it is today the world standard. However, we have many map series with alternate prime meridians and as in Geodex 1.0 all meridians will be transferred to the Greenwich Prime Meridian.

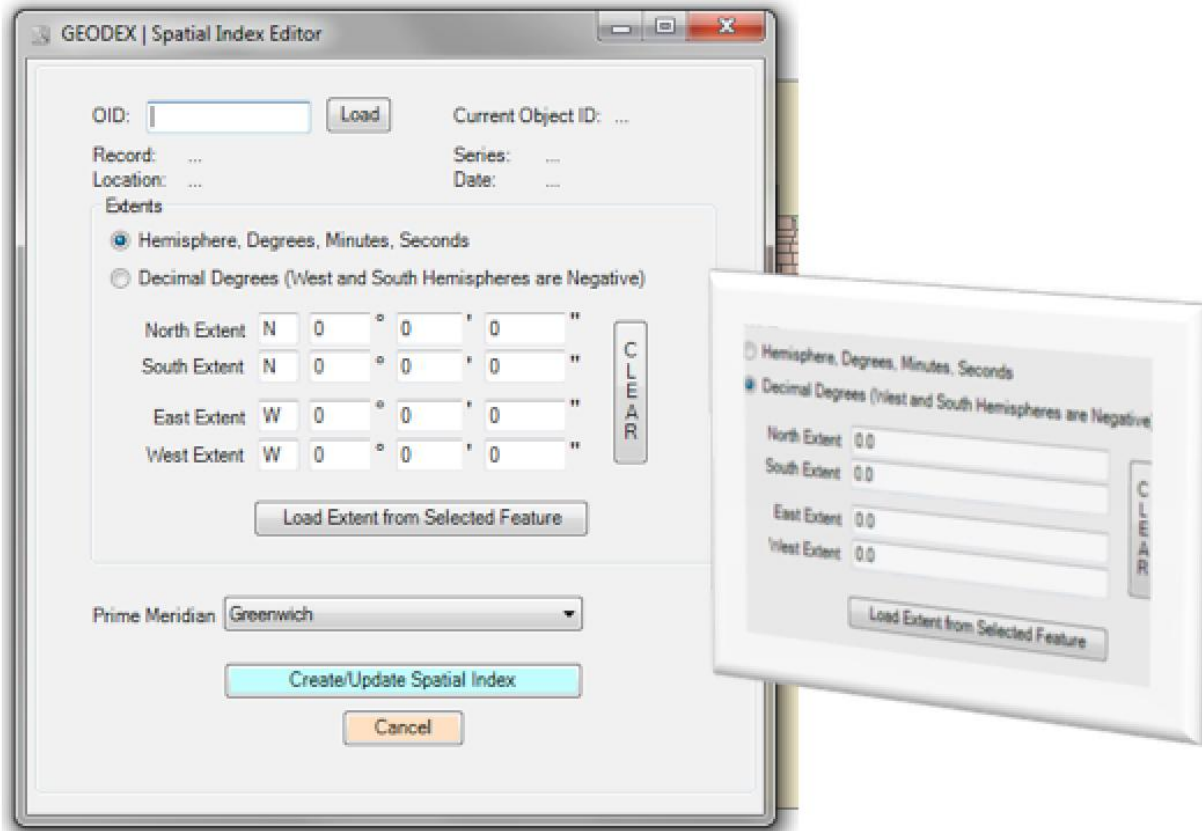


Figure 3. Spatial index editor in ArcMap.

While this new user interface covered the needs for editing and updating the backend database, it was not ideal for searching, even by AGSL staff, and it certainly was not effective for any kind of general user interface, which was one of the primary goals of the project.

Geodex Web Map

We felt it was important to define the capabilities that will be required of an effective user facing web mapping application for using Geodex before we moved forward with the development of the web map interface. A typical example of how the search feature in the original Geodex was used is as follows: A patron wants to see maps from a place such as the location of a village where their ancestors were located. The coordinates of this village are determined using a gazetteer or some other method and then used to search a Geodex file for matching map sheets based on the coverage of the footprint. This capability is absolutely crucial, because it defined the distinctiveness of Geodex, the ability to search for paper maps in the collection based on geographic coordinates. Other simple search capabilities should include the ability to search by title or sheet numbers, query by year of publication, and other relevant fields.

Other, more advanced, queries should be included as well. The ability to view the information in surrounding map sheets would be helpful, particularly if a point falls near the coincident edges of two map sheets. Being able to search by area, such as Europe or South America, would also be useful,

allowing for queries of maps that cover a range larger than a single map sheet. While these are more advanced than the original Geodex search interface, they are not that difficult to accomplish and GIS practitioners will find these processes very similar to geoprocessing operations such as intersect, select by location, select by attribute, buffering, etc. It is therefore not unreasonable to demand these functions from modern web mapping platforms. Another feature that would be helpful, for professional library staff in particular, is the ability to add to and export tables of map sheets for large projects. Perhaps a patron wants a list of *all* the maps that cover a specific area. A selection by location (area) could be done and a table of all resulting map sheets exported easily to be used as a finding aid when navigating the map drawers in the AGS Library or when searching on the digital collections for maps that have been digitized. The ability to query multiple layers at the same time is something that was not possible in Geodex and would be a helpful and, arguably, necessary component of the user-facing web map.

For the geographic search interface it made the most sense to put map *footprints*, which is what is represented by the polygons in the database, on a web map so that users or staff could search for an area or location and see which footprints covered the area and click on the footprints to retrieve the attribute data. In keeping with the open source stack, we initially chose to use Geoserver and connected it to the PostGIS database. Some web map services (WMS) implementations were completed and we attempted to display them in a Leaflet map. Due to the limited experience with javascript by staff and students in the AGSL this effort proved to be a more time consuming than we felt we could afford at this stage of the project. At that point we contacted staff in the Urban Planning department at UWM for some advice on how best to proceed within the confines of available technologies at UWM.

We were encouraged to make use of the ArcGIS Server available on campus. It was recommended we create an enterprise geodatabase on the WebGIS SQL server and then use any web base map we wanted. While this marked a move away from utilizing open source applications, we needed to be able to move forward and this was going to be the easiest way, for now. We were able to successfully copy the data out of the PostGIS database and into the new enterprise geodatabase. So, currently, there are two versions of the database:

- 1) The PostGIS database that is hosted locally and
- 2) an enterprise geodatabase on the campus spatial SQL server.

The Geodex web map could be created in any of the popular web mapping platforms available, but some things are important to consider in the selection. Given the size of the Geodex database, we are able to rule out certain programs because of limitations on data. Map Box and CartoDB fall into this category. Because there are 179 files with almost 400,000 entries collectively, even combining them into one or more files would still far exceed the capabilities of these services without paying for a license.

Open source solutions such as Geomoose or Leaflet could be useful and should not be ruled out completely. Considering that the UWM has an institutional site license to ESRI products, it simply made the most sense to take advantage of the web mapping capabilities offered by ArcGIS.com, the ArcGIS API for Javascript, and even the ArcGIS library for leaflet.js.

Web Map Set-up Challenges

At the time of writing, the web map is not available to the public and is only used by AGSL. To address the slowness of the system and to resolve some search issues we were experience we decided to split the datasets. Splitting the datasets made the most sense for faster loading but in a public instance the division would need to be transparent to the user. Currently, we have divided the database into three sections based mostly on scale – small scale ranging from 1:126,720 to 1:2,500,000, large scale 1:500 to 1:100,000 and because they are used frequently Wisconsin 1:24,000 topographic sheets. Breaking up the database in this manner has slightly increased loading speed but it still can be quite slow.

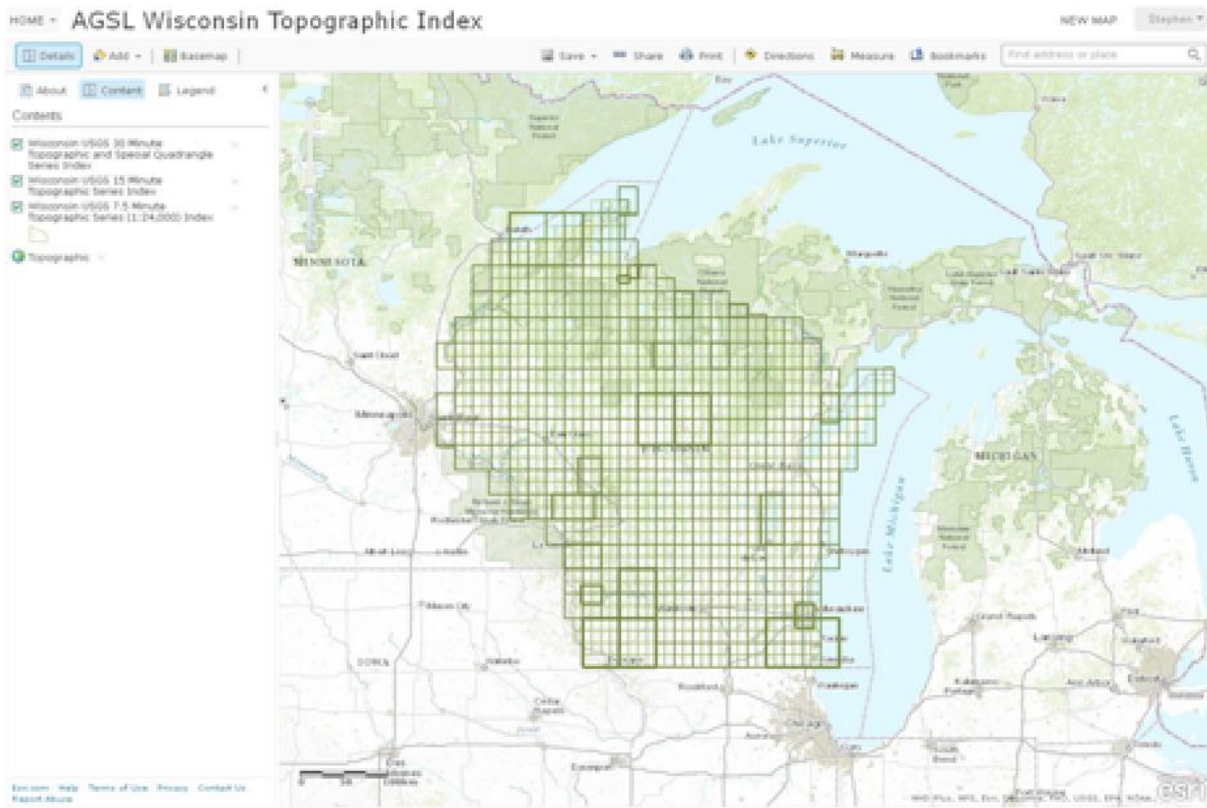


Figure 4. Screenshot of the Wisconsin files displayed in the web map.

Another interesting occurrence we are attempting to fix for the web map interface is that Geodex 1.0 was unable to handle spatial indexes that crossed the 180th meridian. If a map extended over the 180th meridian and was entered into Geodex in the typical way, it would essentially interpret the geography as the inverse, covering all areas in the latitude extent *except* the longitudes that should be covered (think wrapping around the earth rather than straddling the 180th meridian). As a result, sheets that cross over this 180th meridian would need to be entered as two separate records. An excellent example of a file where this occurs commonly is for the New Zealand Oceanographic Series (F0309).

In Figure 5, the two selected features (highlighted in cyan) actually represent just one sheet in the series, but are represented in the Geodex database as two separate records. Their attributes are identical

except for the *map_format* field where one is listed as “Standard quadrangle” and the other as “180 Longitude X-over entry.” In a web map interface, this won’t be an issue so, for series such as this, the issue could be resolved with some simple merge operations should be all that is necessary to make a more manageable database.



Figure 5. An example of the 180 meridian issue.

Conclusion

In the beginning of this project we set out to recreate, revitalize and achieve the original vision of Geodex as a shared map index database. Through much trial and error over the period of more than a year we have made some significant headway and have saved data that represents more than 25 years of work at the AGSL. We are not nearly finished working on this project and are excited to see other efforts at institutions springing up to work on this endeavor. From the many lessons learned along the way we anticipate moving forward with the implementation of a public facing map interface in the coming months with the hopes of adding functionality for other libraries to add their holdings to the database.

References

- Borgman, Christine L. (1997) From Acting Locally to Thinking Globally: A Brief History of Library Automation. *Library Quarterly* vol 67 no 3 pg 215-249
- Endeavor Information Systems. Endeavor Announces Voyager Enhancements, Library Integrations. (2010) *Information Today*. Vol 17, no 3 <http://www.infotoday.com/IT/mar00/news22.htm>