[<< issue home]

by Tom Brittnacher
Geospatial Data Curator
University of California Santa Barbara

The UC Santa Barbara Library map collection contains about 450,000 sheets, with an emphasis on medium- and large-scale topographic maps; it is one of the largest map collections in California. However, the Library's map collection is not browsable by users. If someone wants to peruse large map sets, library staff need to transport stacks of maps (in some cases, thousands of sheets) from a staff-only area to a viewing room quite a distance away in the Library's Special Research Collections (SRC) department.

Over the years, the Library has scanned paper maps for users on request, but this is no longer done, although the scans remain on computer servers. As a means to improve access to the map collection, the Library is currently developing a method for sharing those scans by expanding the capabilities of our online digital library: the Alexandria Digital Research Library (ADRL), successor to the Alexandria Digital Library (ADL) of the late 1990s (Figure 1 shows the home page). Users will be able to discover, view, and download (subject to copyright restrictions) the digital map scans through an online interface.



Figure 1. UC Santa Barbara Library's digital collection.

The process of ingesting maps into ADRL involves not only uploading the digital files, but also uploading the metadata that describes the files. The metadata, in CSV (comma-separated values) spreadsheet format, provides the system with information (metadata elements) about each map sheet, such as the title, date, publisher, topic keywords, bounding box, scale, etc. Each row of the spreadsheet

represents a single map sheet, and each column represents one metadata element.

In preparation for ingest, sheet-level metadata describing each individual map scan must be created. In the case of a map that only has a single sheet, this is straightforward. The title, date, and other information have already been gathered and entered into a catalog record. Library staff can extract that information and put it into the metadata spreadsheet. In the case of large map sets, only a set-level catalog record typically exists. As a result, Library staff must additionally collect information about each individual sheet, including titles, sheet numbers, dates, and anything else that varies from sheet to sheet. With large map sets, the process of manually creating sheet-level metadata can be tedious, time-consuming, and error-prone. For our China 1:100,000 map set, which encompasses over 5,000 sheets, it would take two and a half months to compile the metadata spreadsheet.

At UCSB Library, we greatly sped up the metadata creation process for map sets by using GIS software to create digital index maps from which sheet-level metadata can be easily extracted.

What is a digital index map?

Many topographic map sets have an index map, which is an overview map of the geographic region (such as a country) covered by a map set. Typically, a grid system is overlaid onto the geographic region, and each grid cell represents a single map sheet. The grid cells are identified (labeled) using a systematic coding system, usually with a combination of numbers and letters, and that code appears on the individual map corresponding to that grid cell. An index map shows the grid system with the corresponding grid cell codes to help the user find which map sheet covers their area of interest.

Libraries typically make a photocopy of the index map and indicate their holdings by putting a mark on those grid cells for which they have the corresponding paper map (see Figure 2). When patrons inquire about maps available for a certain location, library staff can quickly determine if the library holds a map from that set that covers the desired geographical area.

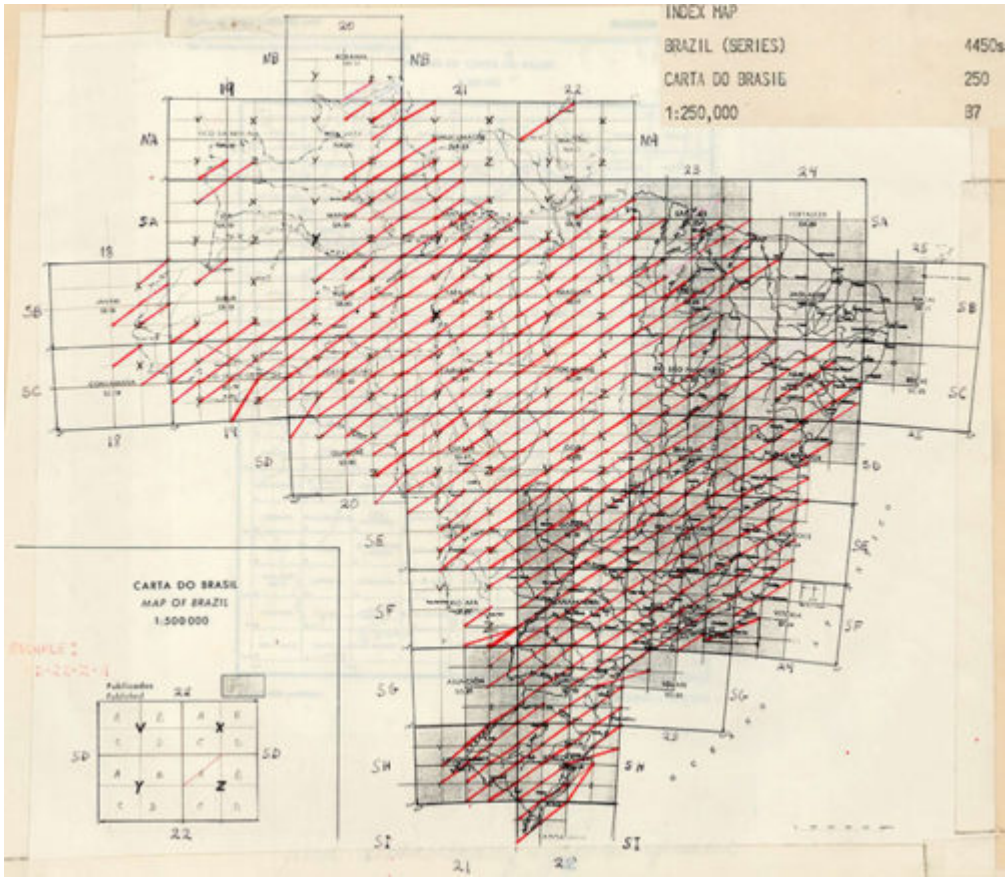


Figure 2. The index map for the Brazil 1:250,000 topographic map set. A red diagonal line through a grid square indicates the paper map is held by the UCSB Library.

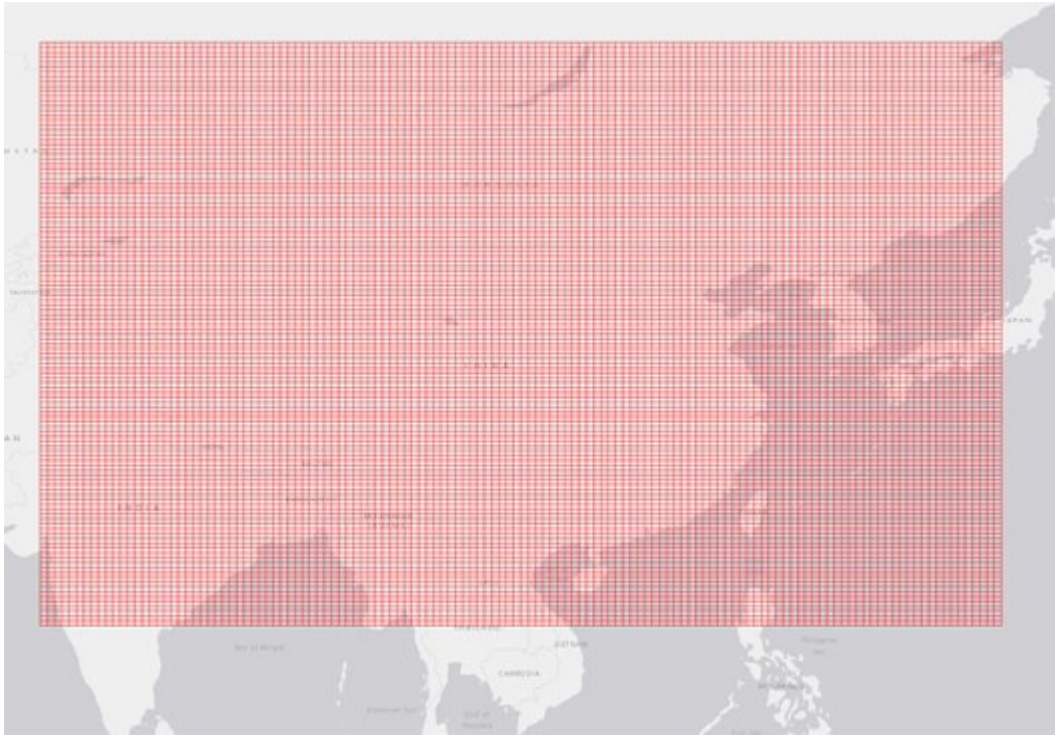
The paper index maps can then be scanned, thus creating a digital version, usually in PDF format. However, to be useful as a tool for creating metadata, the digital version must be in a format used by geographical information systems (GIS) software so that each grid cell can be queried for specific information about the map represented by that grid cell. GIS software tools provide quick and easy ways of generating digital index maps for this purpose.

Creating digital index maps using GIS

Below, I will outline the theory and general steps we followed to create digital index maps for generating metadata. The detailed process of creating digital index maps using GIS software and the background of the technique have been well documented by Christopher J.J. Thiry, Librarian at Colorado School of Mines. Step-by-step instructions are available for several versions of the Esri ArcGIS software

suite at http://libguides.mines.edu/maps/tools. And his article, GIS-based discovery interface to paper map sets, available at http://www.e-perimetron.org/Vol_12_2/Thiry.pdf, provides an overview of the issues around index maps and his methodology for creating online interfaces.

Many topographic map sets use latitude and longitude as its coordinate system, with the index grid based on regular intervals of degrees and minutes. For example, the Soviet topographic map set of China at a scale of 1:100,000 includes sheets that are two degrees wide by one degree, twenty minutes high. The maximum extents of the entire grid system for the China map set are from 73 degrees east to 135 degrees east longitude, and from 18 degrees north to 54 degrees north latitude. We can enter this information into GIS tools to create a digital grid system. In Esri's ArcGIS software, the "Fishnet" tool takes these measurements and generates a grid of rectangles, where each rectangle represents the area of one map sheet (see Figure 3).



Once the grid is generated, the grid cells can be graphically edited to match the actual index. Grid cells in the fishnet that are beyond the extent of the map set can be deleted, and any grid cells that represent irregularly-shaped maps can be edited. Typically, a cartographer will alter map extents to include an adjacent area

too small for its own map sheet. For example, it may be more practical to extend a map sheet by a few inches to accommodate a small corner of land rather than create a separate sheet that mostly contains water. It can also be beneficial to shift a map sheet slightly to better fit the land area along the coast onto a single sheet. These irregularities can easily be represented by editing the grid cells using the GIS software tools (see Figure 4).



In the Brazil 1:250,000 topo set, a map sheet was rotated to more efficiently cover the boundary.

Once the grid cells match the actual sheet areas of the map set, various characteristics of each map sheet can be recorded in the attribute table for its grid cell. The digital index map created in GIS has both a graphic component (in this case, grid cell rectangles) and an attribute table (spreadsheet). Each row in the attribute table represents one graphic feature (one grid cell rectangle), and each column represents one attribute of the feature, such as title or date. Other than a few attributes created by the software for internal use, all attributes are determined and encoded by the creator.

Typically, individual map sheets are identified by applying a systematic coding system that then appears on the individual maps and on the index map. This system allows users to easily determine which map sheet corresponds to their area of interest. For example, the China 1:100,000 map set uses a global grid system developed by the Soviet Union for their topographic maps of various countries around the world. Each row is four degrees wide and is given a letter designation, starting with A for the row from zero degrees to four degrees north. Thus, the band from 28 to 32 degrees north is labeled “H”. Each column is six degrees wide and is

given a number starting with one for the band from 180 degrees to 176 degrees west longitude. Therefore, the band from 120 to 126 degrees east longitude is labeled “51”. Each horizontal and vertical band intersect to make a large grid cell four degrees by six degrees in size. These large grid cells can be labeled by combining the band letter and the band number. As such, the intersection of the H band and the 51 band can be labeled “H-51” (see Figure 5). Each of these large grid cells are further divided into 144 grid cells, 12 cells wide by 12 cells tall. Each of these 144 cells represents a single map sheet. The 144 cells are numbered from the top left corner to the bottom right corner along each horizontal row. Thus, the last map in the first row is map “12”, and the first map in the second row is map “13”. Map “13” is identified by combining the row, column, and cell values to give it the map number “H-51-13” (see Figure 6). As you can see, the index maps are not always easy to read.

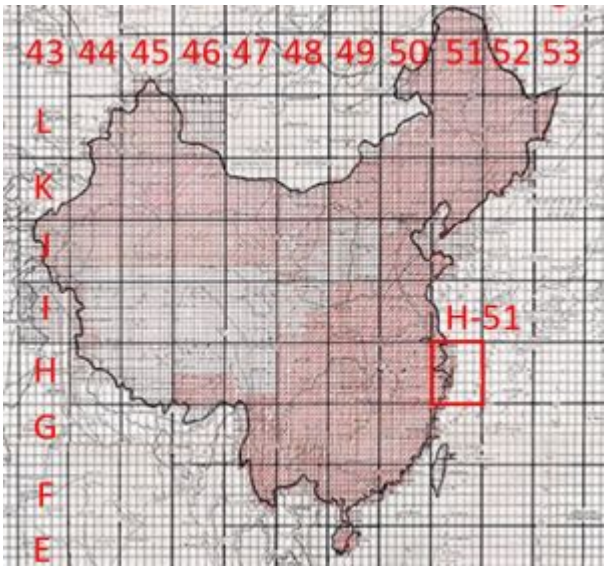


Figure 5: Grid system for China 1:100,000 topo map set.



Figure 6: Detail of index map square H-51.

The grid identification system can quickly be added to the index map attribute table using the GIS software interface by selecting entire rows or columns of grid cells and recording (or “calculating”) their row letters or column numbers in the attribute table. Therefore, unlike the manual process of looking at each individual sheet and recording its map sheet number by hand, many grid cells can be coded at once (see Figure 7). By breaking the process into three separate tasks (coding the rows, coding the columns, and coding the cell numbers), the task can be completed in

minutes for the entire set of over 5,000 map sheets.

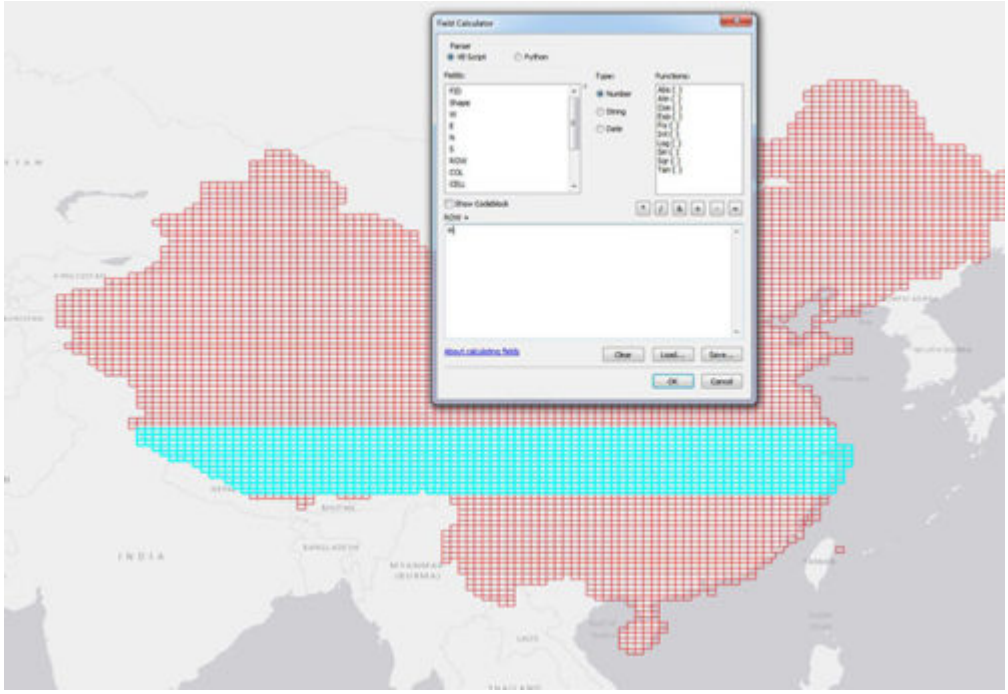


Figure 7: Using ArcGIS tools to record Row H in the attribute table.

The three columns in the attribute table (row, column, and grid cell) can then be concatenated into a fourth column with the completed map sheet number. The concatenation process can also incorporate the dashes between the values. Thus, if the column [ROW] = “H”, the column [COL] = “51” and the column [CELL] = “13”, then the concatenation:

```
SHEET_NO = [ROW]&”-“&[COL]&”-“&[CELL]
```

gives you SHEET_NO = H-51-13. This can be calculated for all rows simultaneously using GIS tools (see Figure 8).

In most cases, the task of coding the grid cell numbers (i.e. column [COL]) usually involves selecting one of perhaps only four or six grid cells, if this level of coding exists at all. The method of selecting grid cells using the graphic interface is then usually quite easy. In the case of the China 1:100,000 map set, with 144 grid cells within the larger grid cell, it is easier to copy and paste repeated patterns in the attribute table rather than use the graphic interface to select every twelfth grid cell. It is best if this is done while the fishnet is still rectangular (prior to removing

grid cells outside the map set extent) so that the repeating pattern is not interrupted by missing cells.

The extents, or bounding box of each individual map, can also be calculated using GIS tools. Rather than viewing each individual paper map sheet to determine the extent, the system can automatically calculate the values. The “Add Geometry Attributes” tool in ArcGIS automatically creates four columns in the attribute table and enters each value (in decimal degrees) for the western-most longitude, eastern-most longitude, northern-most latitude, and southern-most latitude in its respective column. The first four columns in Figure 8 show the four extents, with the columns renamed to designate the extent’s respective cardinal directions.

W	E	N	S	CELL_ID	ROW	COL	CELL	SHEET_NO
115.000122	115.500122	29.333322	29.000122	5235	H	50	99	H-50-99
120.000122	120.500122	32.000087	31.666687	6301	H	51	1	H-51-1
121.500122	122.000122	29.333322	29.000122	5248	H	51	100	H-51-100
122.000122	122.500122	29.333322	29.000122	5249	H	51	101	H-51-101
120.000122	120.500122	29.000087	28.666687	5113	H	51	109	H-51-109
120.500122	121.000122	29.000087	28.666687	5114	H	51	110	H-51-110
121.000122	121.500122	29.000087	28.666687	5115	H	51	111	H-51-111
121.500122	122.000122	29.000087	28.666687	5116	H	51	112	H-51-112
122.000122	122.500122	29.000087	28.666687	5117	H	51	113	H-51-113
120.000122	120.500122	28.666713	28.333313	4981	H	51	121	H-51-121
120.500122	121.000122	28.666713	28.333313	4982	H	51	122	H-51-122
121.000122	121.500122	28.666713	28.333313	4983	H	51	123	H-51-123
121.500122	122.000122	28.666713	28.333313	4984	H	51	124	H-51-124
120.000122	120.500122	31.666713	31.333313	6169	H	51	13	H-51-13
120.000122	120.500122	28.333322	28.000122	4849	H	51	133	H-51-133
120.500122	121.000122	28.333322	28.000122	4850	H	51	134	H-51-134
121.000122	121.500122	28.333322	28.000122	4851	H	51	135	H-51-135
121.500122	122.000122	28.333322	28.000122	4852	H	51	136	H-51-136
120.500122	121.000122	31.666713	31.333313	6170	H	51	14	H-51-14
121.000122	121.500122	31.666713	31.333313	6171	H	51	15	H-51-15
121.500122	122.000122	31.666713	31.333313	6172	H	51	16	H-51-16
120.500122	121.000122	32.000087	31.666687	6302	H	51	2	H-51-2
120.000122	120.500122	31.333322	31.000122	6037	H	51	25	H-51-25
120.500122	121.000122	31.333322	31.000122	6038	H	51	26	H-51-26
121.000122	121.500122	31.333322	31.000122	6039	H	51	27	H-51-27
121.500122	122.000122	31.333322	31.000122	6040	H	51	28	H-51-28
121.000122	121.500122	32.000087	31.666687	6303	H	51	3	H-51-3
120.000122	120.500122	31.000087	30.666687	5905	H	51	37	H-51-37
120.500122	121.000122	31.000087	30.666687	5906	H	51	38	H-51-38
121.000122	121.500122	31.000087	30.666687	5907	H	51	39	H-51-39
121.500122	122.000122	32.000087	31.666687	6304	H	51	4	H-51-4
121.500122	122.000122	31.000087	30.666687	5908	H	51	40	H-51-40

Figure 8: Part of the attribute table for the China 1:100,000 topo map set.

For the China 1:100,000 map set, we were able to create the digital index map, as well as calculate the four extents of the bounding box and the sheet codes in about an hour, much faster than a manual process, which would have taken weeks.



Capturing additional information

Although GIS can be used to quicken the process of collecting sheet-level metadata, it cannot be used to collect non-systematic information, such as sheet names and dates. (Large map sets are rarely published in their entirety within the same year.) For this additional information, we did, however, speed up the process using other tools.

Opening a TIFF file is a time-consuming aspect of collecting sheet-level metadata. For a large map set, the seconds can add up quickly, especially when opening high-resolution images. Handling the large paper maps can also be cumbersome. To combat this problem, we used an automated process in Adobe Photoshop to crop just the corner of the scanned maps that include the sheet number and date, and downsize the resulting image when saving it as a copy so that the image would only take a second to open. We were able to then quickly open several images at once and record the relevant information into a Microsoft Excel spreadsheet. (In the case of the Soviet maps of China, we did not record or transliterate the Cyrillic sheet names.) We also applied this methodology to the Brazil 1:250,000 topo map set, but for those map sheets, we also recorded the sheet names, including diacritics.

Combining metadata for ingest into the digital library

Once we encoded the map extents and sheet numbers using GIS, recorded title and date information using Excel, and gathered map set-level metadata using the catalog record, we combined the information into a single CSV file. Following the addition of some digital library-specific metadata elements, we were then able to ingest the CSV file and the TIFF images into ADRL.

Overall, the China 1:100,000 set of over 5,000 maps took about 14 hours to complete, as opposed to over 400 hours if we had done each sheet by hand. The Brazil set of about 500 maps took about 8 hours. This significantly reduced the effort required to create sheet-level metadata.

Viewing the map sheets in the digital library

Once the map scans and sheet level metadata are ingested into ADRL, there is a

display page for the map set as a whole (see Figure 9), as well as for each individual map (see Figure 10) and the index map (see Figure 11).

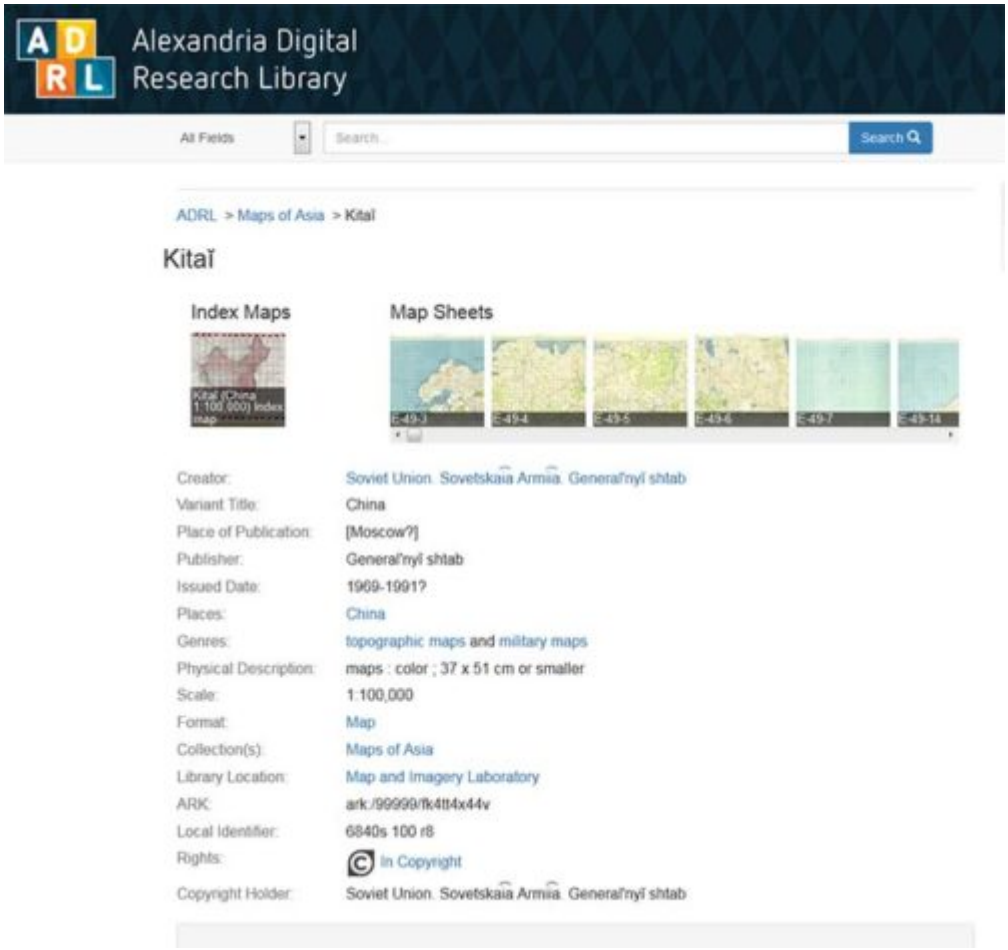


Figure 9: The Map Set display page.

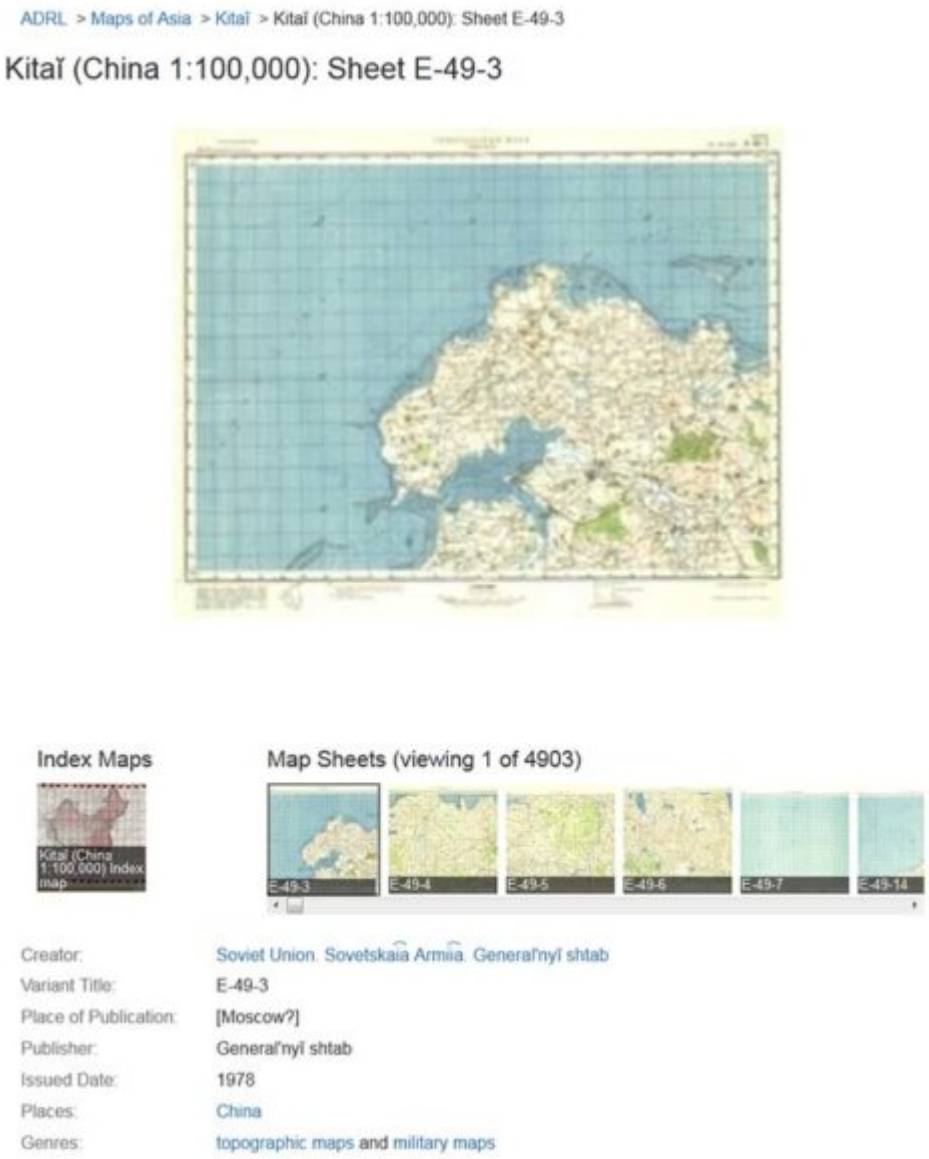


Figure 10: The individual map’s display page.

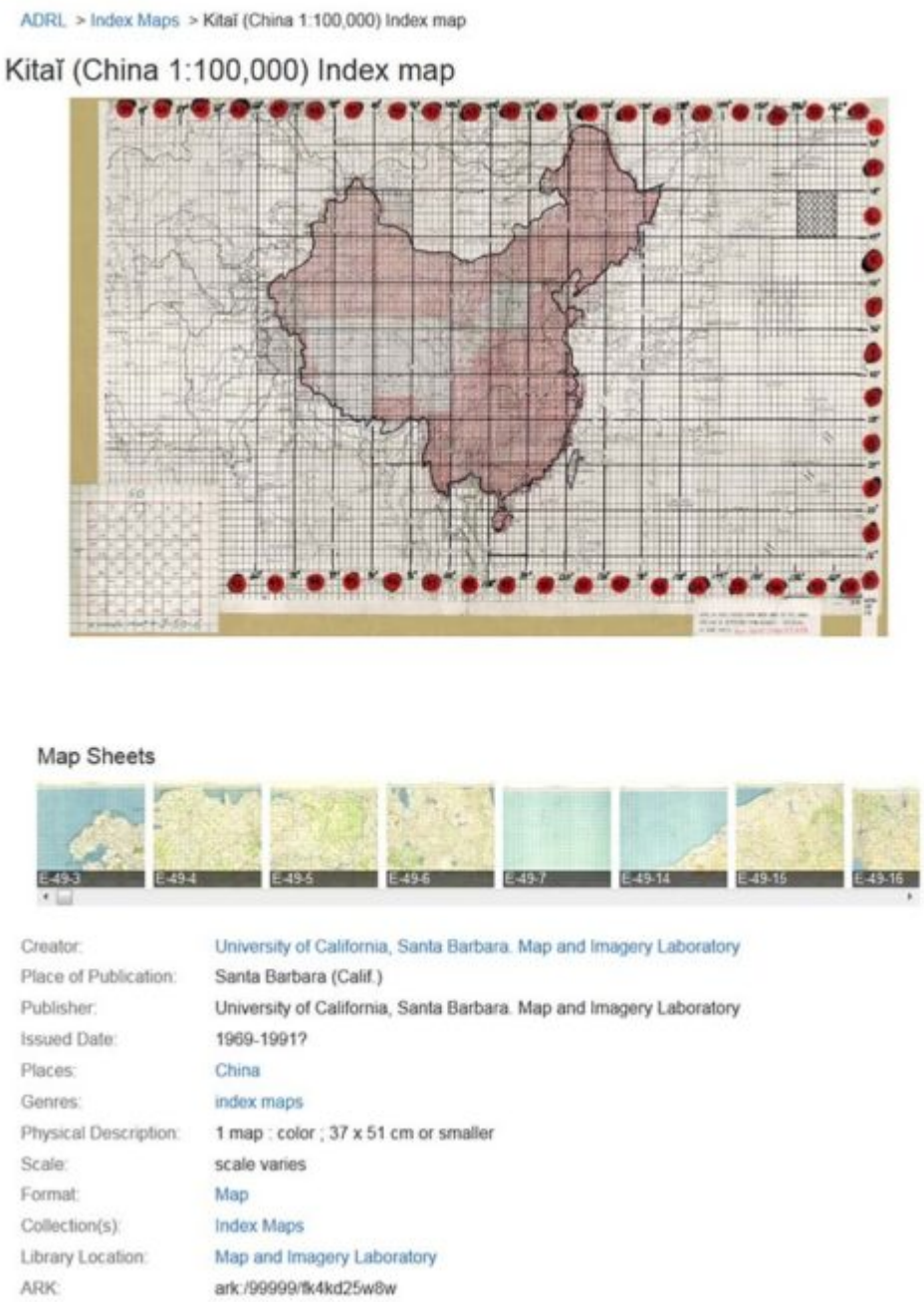


Figure 11: The index map display page.

From any one of these pages, users have access to the index map and all of the sheets. Each sheet's page has information relevant to that sheet only, while the map set page has information relevant to the entire set. Similarly, the index map page has information relevant to the index map itself.

At the moment, only ten test maps and map sets have been ingested as we test the interface. Once the bugs have been ironed out, we will ingest many more of our scanned maps into our digital library.

Reusing and sharing digital index maps

Although we no longer need the digital index map for generating metadata at this point, we joined the Excel spreadsheet to the GIS attribute table and copied over the information to create a more detailed GIS index map that includes the bounding box, sheet number, title, and date. The digital index map can then be used on its own as a search tool (see figure 12). Chris Thiry has used these index maps within his discovery interfaces to identify the map holdings at the Colorado School of Mines. Repositories such as Stanford University's EarthWorks are also developing the capability of interacting with these digital maps to discover and access scanned map images. UC Santa Barbara is hoping to develop this capability as well.

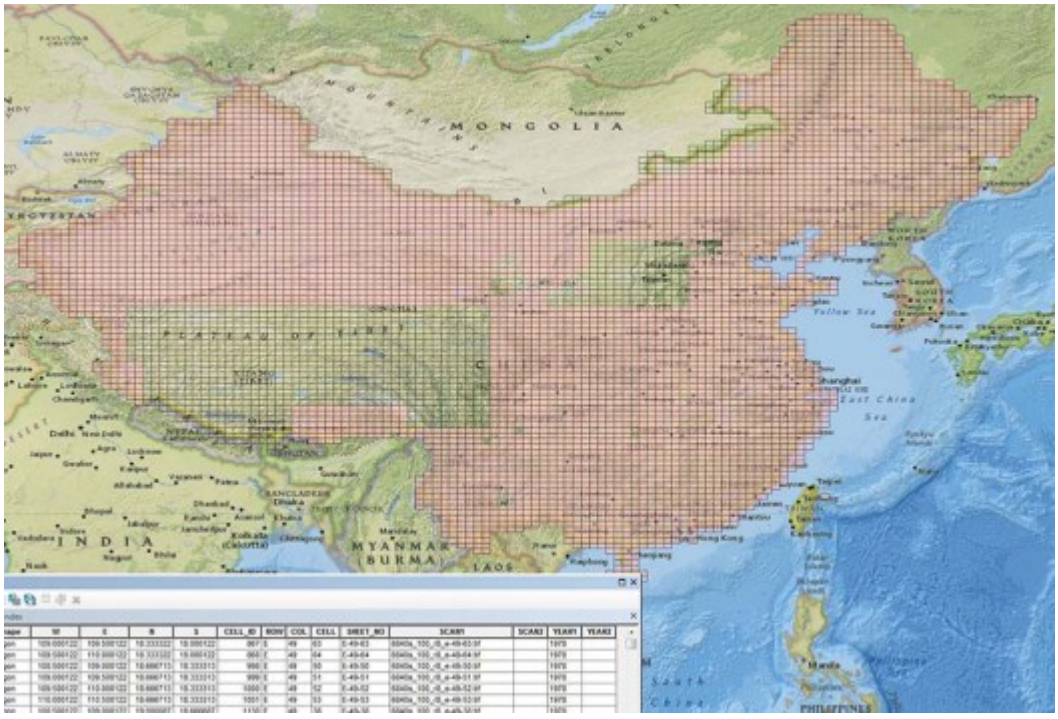


Figure 12: The completed China 1:100,000 digital index map

To reduce the duplication of effort, institutions can share their GIS index map shapefiles by contributing them to the Colorado School of Mines Clearinghouse of indexes to paper map sets. There are instructions on the page for participating. Before creating your own digital index maps, it's best to check the clearinghouse

first to ensure the work has not already been done.

The UC Santa Barbara Library is also partnering with Stanford University and the Colorado School of Mines to explore sharing index maps as geojson files through a GitHub repository. This option provides a method of archiving and sharing non-proprietary, human-readable files. We are currently developing best practices for naming metadata elements so that index maps from multiple institutions can be interoperable.

As more and more institutions struggle to demonstrate the value of their map collections, sharing digital index maps will help us all increase access to this typically-hidden resource. Anecdotal evidence has shown that providing tools for easy discovery of maps significantly increases use of the collection, whether through cataloging or online index maps. This demonstrates that there still is value in maps, as long as users can find them.

[<< issue home]
[next feature >>]

Citation Author: Tom Brittnacher