Project Assessment

The CLIR-sponsored project **Library of Congress Multi-Sheet Map Collection: Africa** has successfully served as a prototype project for gathering sheet-level metadata for multi-sheet map sets. This is significant because the Library's Geography & Map Division holds some 2 to 3 million maps of this kind. The catalog records resulting from the project provide for the first time an item-level inventory of the individual map sheets held by the Library.

Further, the Library MARC records now include place names (as well as other descriptive elements) for each individual sheet, and also geospatial coordinates for the coverage area of each sheet. This means that the maps will now be discoverable by searching tens of thousands of new place names that previously were not available. It also means the maps will be discoverable by search applications that support geospatial (coordinate-based) search. In this new form of searching the user will create a query by specifying a geographical "region of interest". This will be accomplished by means of an interface that enables the user to draw the region of interest on a map, and then to submit the region (in the form of a point, line, or polygon) as a query. The user will be able to discover maps of interest without prior knowledge of the particular place names used in the catalog records. Geospatial searching will be included in the Library's next generation OPAC. In the meantime other search systems that provide coordinate-based indexing and searching will be able to exploit the data in this way.

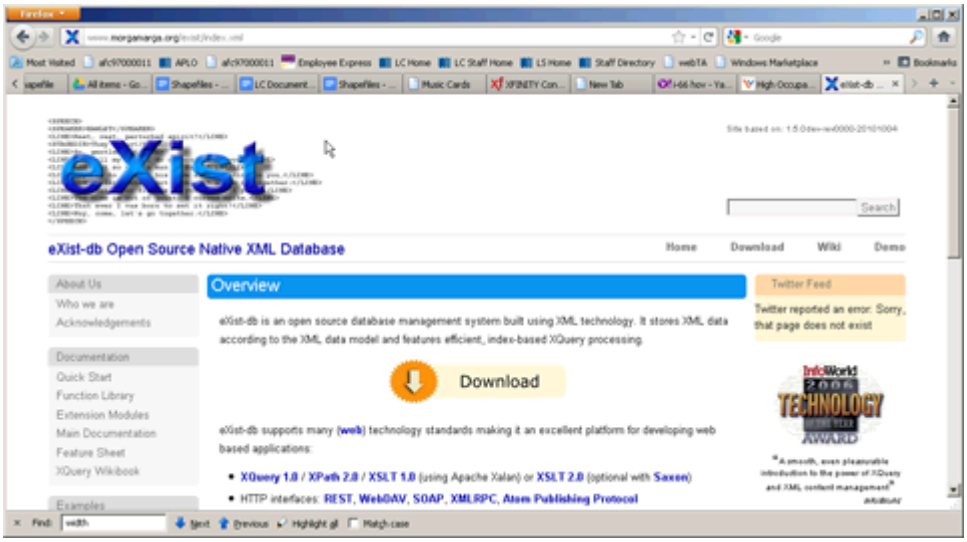
It is also significant that by capturing the coordinates for coverage area of each map it becomes possible to create a visual display of this vector-based data. Viewing the polygons (resulting from capturing the corner coordinates for each sheet) for an entire map set results in a virtual "index map" of the map set. Viewing the data in this way is already a familiar visualization of a map set to librarians and geographers. There are several examples of this visualization in the examples that follow. That this visualization is now closely bound to the bibliographic data is an advantage for both librarian (for the purpose of collection management) and researcher (for identifying and retrieving the particular map of interest).

Finally, it should be noted that the Library gained valuable experience in processing multi-sheet map sets. It was tested and proven that it is possible to combine the output of professional cataloger's (the set-level MARC record) with the data generated by library technicians (the sheet-level data used to enhance the MARC record) in a well-coordinated workflow. It was also tested and proven that XML technologies provide a viable platform for data capture and subsequent data transformations for geospatial data. An XML-based workflow is especially advantageous given that the web-based geospatial community already makes extensive use of XML formats (e.g. KML and GeoRSS). The knowledge gained from this project will inform the planning of subsequent projects to provide item-level control of the Library's large collection of multi-sheet maps.

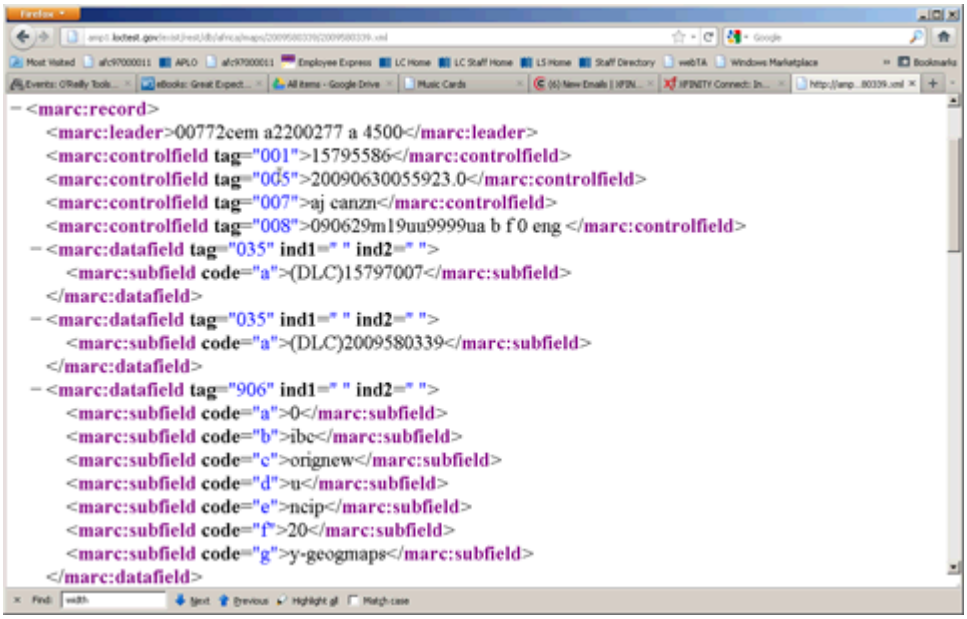
The following notes and screenshots provide an illustrated overview of the project in terms of 1) data capture workflow, 2) distribution of the records via the Library's Voyager OPAC, and 3) examples of how the project data can be viewed and re-used by users.

Data Capture

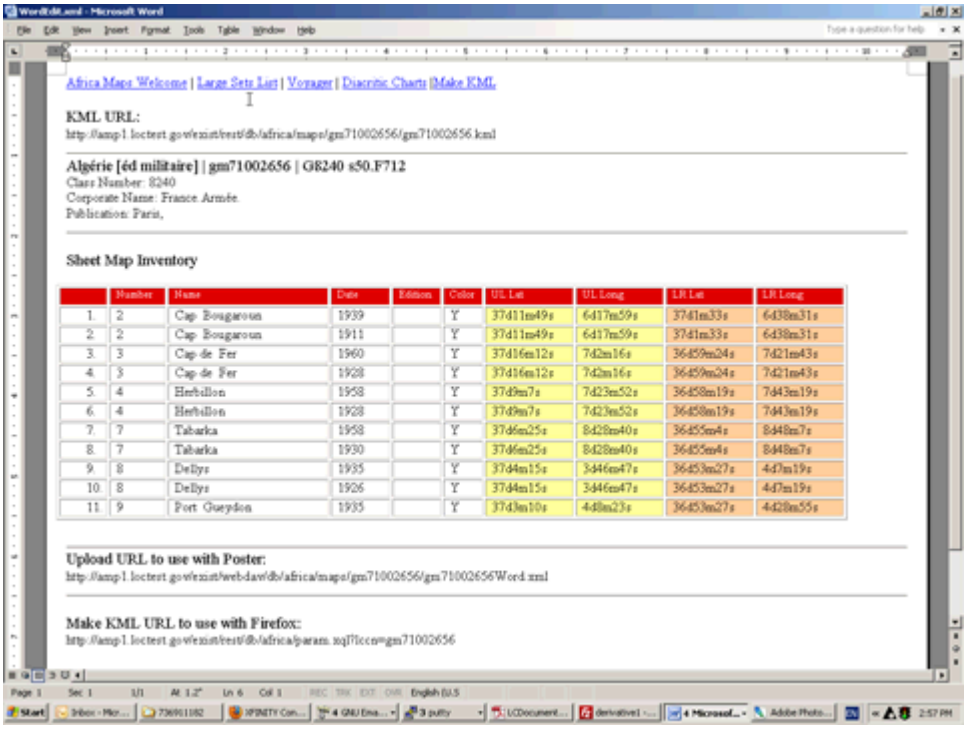
The sheet-level data capture was managed using a native XML database (eXist). Using an XML-based approach made it possible to accomplish both data entry and to also execute the subsequent data transformations.



The set-level MARC records were upgraded in the Voyager database by Geography & Map Division professional catalogers in the usual way. Once upgraded, these MARC records were converted to XML (MARCXML format) and loaded into the eXist project database. The MARC data was then available for use in creating the data entry form, and later, for creating KML files and Shapefiles. The record is also available so the record itself can be updated with the sheet-level data captured in the course of the CLIR project..



1. Sheet-level data entry was accomplished by automatically generating a Microsoft Word document for data entry with the appropriate data fields. Staff responsible for data entry were able to key data directly into this familiar interface. Because Microsoft Word is itself an XML format this document could then be directly loaded to the project XML database.



Africa Maps Welcome | Large Sets List | Voyager | Diacritic Charts |Make KML

KML URL:
http://amp1.loctest.gov/exist/rest/db/africa/maps/gm71002656/gm71002656.kml

Algérie [éd militaire] | gm71002656 | G8240 s50.F712
Class Number: 8240
Corporate Name: France. Armée.
Publication: Paris,

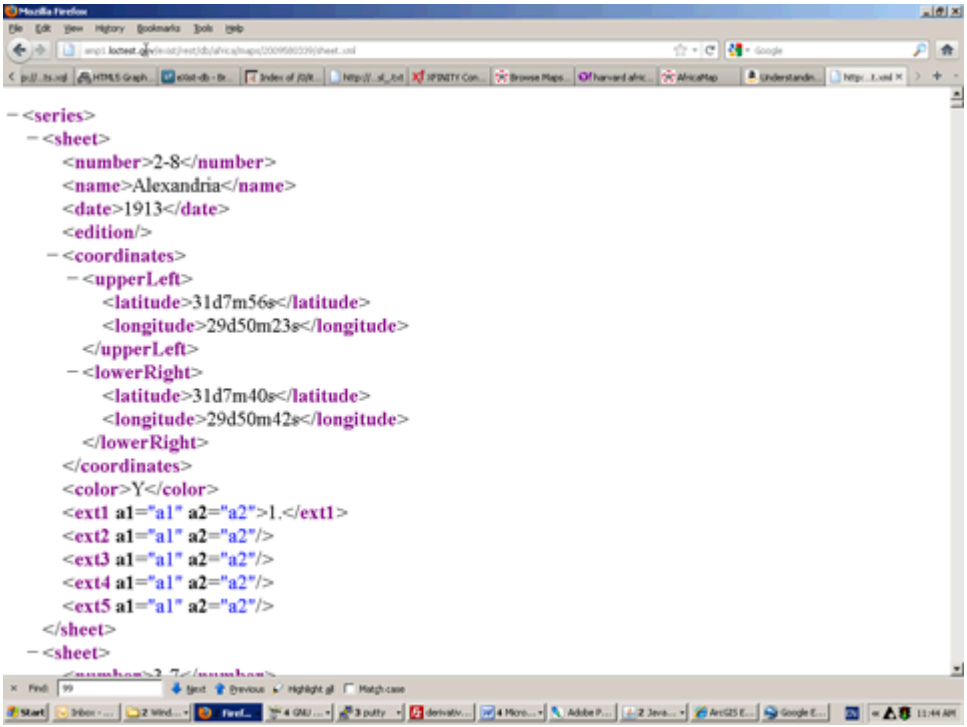
Sheet Map Inventory

	Number	Name	Date	Edition	Color	UL Lat	UL Long	LR Lat	LR Long
1.	2	Cap Bougaroun	1939		Y	37d11m49s	6d17m59s	37d1m33s	6d38m31s
2.	2	Cap Bougaroun	1911		Y	37d11m49s	6d17m59s	37d1m33s	6d38m31s
3.	3	Cap de Fer	1960		Y	37d16m12s	7d2m16s	36d59m24s	7d21m43s
4.	3	Cap de Fer	1928		Y	37d16m12s	7d2m16s	36d59m24s	7d21m43s
5.	4	Herbillon	1958		Y	37d9m7s	7d23m52s	36d58m19s	7d43m19s
6.	4	Herbillon	1928		Y	37d9m7s	7d23m52s	36d58m19s	7d43m19s
7.	7	Tabarka	1958		Y	37d6m25s	8d28m40s	36d55m4s	8d48m7s
8.	7	Tabarka	1930		Y	37d6m25s	8d28m40s	36d55m4s	8d48m7s
9.	8	Dellys	1935		Y	37d4m15s	3d46m47s	36d53m27s	4d7m19s
10.	8	Dellys	1926		Y	37d4m15s	3d46m47s	36d53m27s	4d7m19s
11.	9	Port Gueydon	1935		Y	37d3m10s	4d8m23s	36d53m27s	4d28m55s

Upload URL to use with Poster:
http://amp1.loctest.gov/exist/webdav/db/africa/maps/gm71002656/gm71002656Word.xml

Make KML URL to use with Firefox:
http://amp1.loctest.gov/exist/rest/db/africa/param.xql?iccn=gm71002656

3. When the Word document is loaded to the project database, the sheet-level data is automatically extracted, transformed, and stored in a generic XML format.

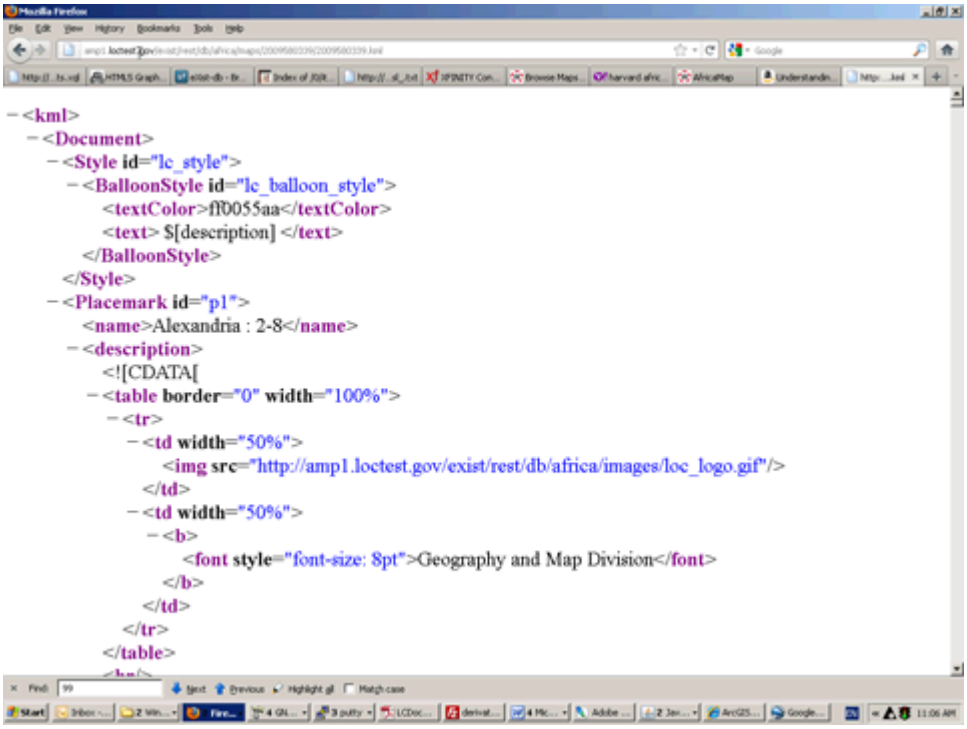


```
- <series>
  - <sheet>
      <number>2-8</number>
      <name>Alexandria</name>
      <date>1913</date>
      <edition/>
    - <coordinates>
      - <upperLeft>
          <latitude>31d7m56s</latitude>
          <longitude>29d50m23s</longitude>
        </upperLeft>
      - <lowerRight>
          <latitude>31d7m40s</latitude>
          <longitude>29d50m42s</longitude>
        </lowerRight>
      </coordinates>
      <color>Y</color>
      <ext1 a1="a1" a2="a2">1.</ext1>
      <ext2 a1="a1" a2="a2"/>
      <ext3 a1="a1" a2="a2"/>
      <ext4 a1="a1" a2="a2"/>
      <ext5 a1="a1" a2="a2"/>
    </sheet>
  - <sheet>
```

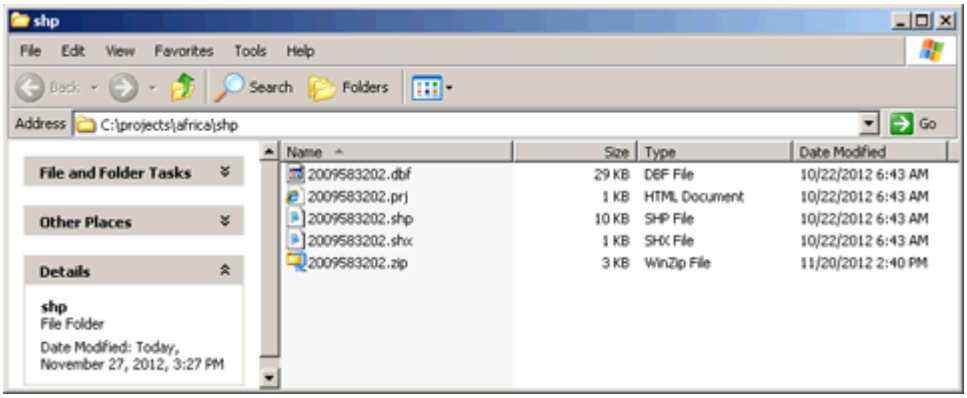
Data Conversion

There are three data conversions that then take place: creation of a KML file, the creation of an Esri Shapefile, and the creation of an updated MARC record that contains the sheet-level metadata. All three make use of both the original MARC data and the sheet-level data.

1. The KML file is an XML format that is widely used in geospatial browsers such as Google Earth.



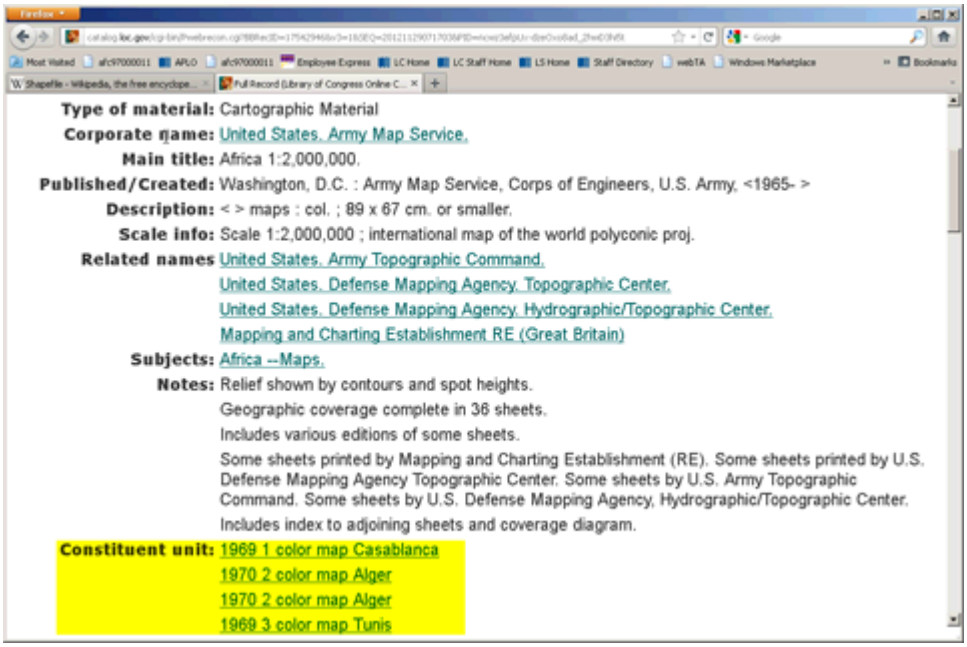
2. The Esri Shapefile is a widely-used geospatial vector data format for geographic information systems software. The Shapefile is a binary format and is comprised of four files that are zipped into one.



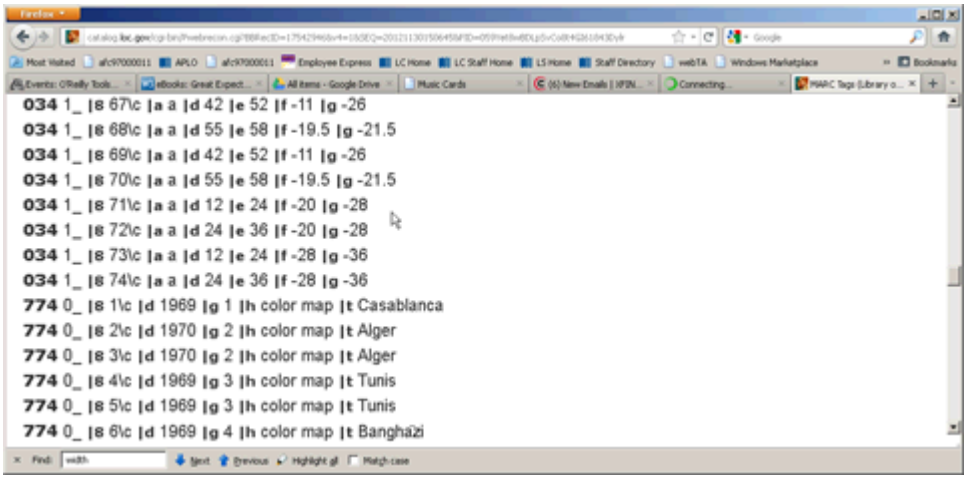
3. Finally, the Library's MARC records are updated with the sheet level metadata and coordinate information, and also with links to the KML file and Shapefile for each set.

Distribution of Project Data

The Library's initial distribution of the data will be via the Library's Voyager OPAC. Here the Labeled view of the MARC record shows sheet-level map descriptions (from the MARC 774 field) as "Constituent items".



The example below shows a section of the Voyager MARC Tag display. The geographic coordinates for the coverage area of each individual sheet map are expressed in the 034 field, and the other descriptive data elements for each sheet are expressed in a 774 field. Note that the 034 field for a given sheet is correlated with the corresponding 774 field for that sheet by means of the matching Subfield 8 values. The presence of the geographic coordinates in the MARC record will enable geospatial searching when that feature becomes available in the Voyager OPAC.

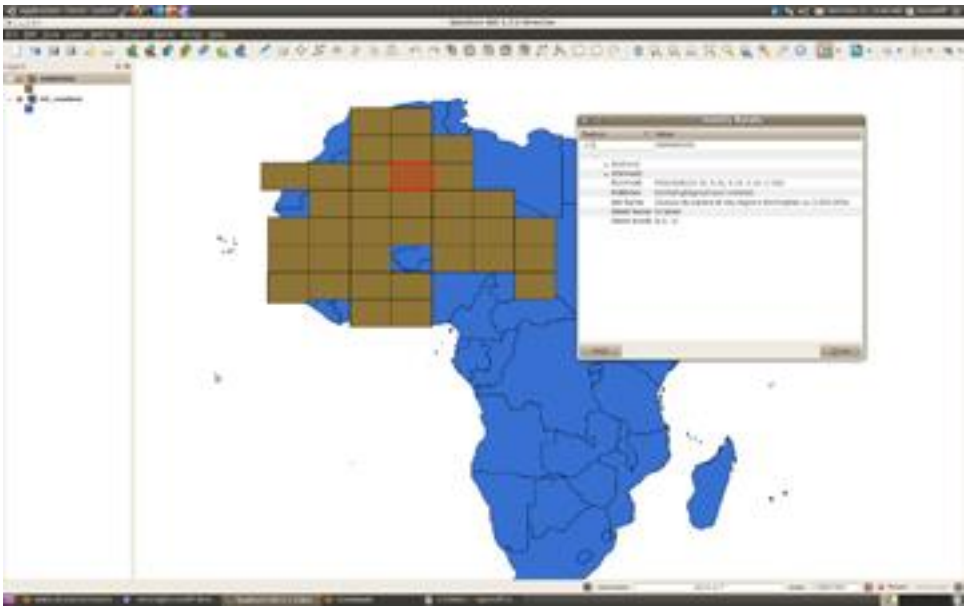

034 1_ |8 67\c |a a |d 42 |e 52 |f -11 |g -26
034 1_ |8 68\c |a a |d 55 |e 58 |f -19.5 |g -21.5
034 1_ |8 69\c |a a |d 42 |e 52 |f -11 |g -26
034 1_ |8 70\c |a a |d 55 |e 58 |f -19.5 |g -21.5
034 1_ |8 71\c |a a |d 12 |e 24 |f -20 |g -28
034 1_ |8 72\c |a a |d 24 |e 36 |f -20 |g -28
034 1_ |8 73\c |a a |d 12 |e 24 |f -28 |g -36
034 1_ |8 74\c |a a |d 24 |e 36 |f -28 |g -36
774 0_ |8 1\c |d 1969 |g 1 |h color map |t Casablanca
774 0_ |8 2\c |d 1970 |g 2 |h color map |t Alger
774 0_ |8 3\c |d 1970 |g 2 |h color map |t Alger
774 0_ |8 4\c |d 1969 |g 3 |h color map |t Tunis
774 0_ |8 5\c |d 1969 |g 3 |h color map |t Tunis
774 0_ |8 6\c |d 1969 |g 4 |h color map |t Banghāzi

The 856 field of the MARC record is used to provide links to the KML file and the zipped Shapefile. These records are stored on the Library system and made available for download. The "Labeled View" of the MARC records displays the URL's as "Links".

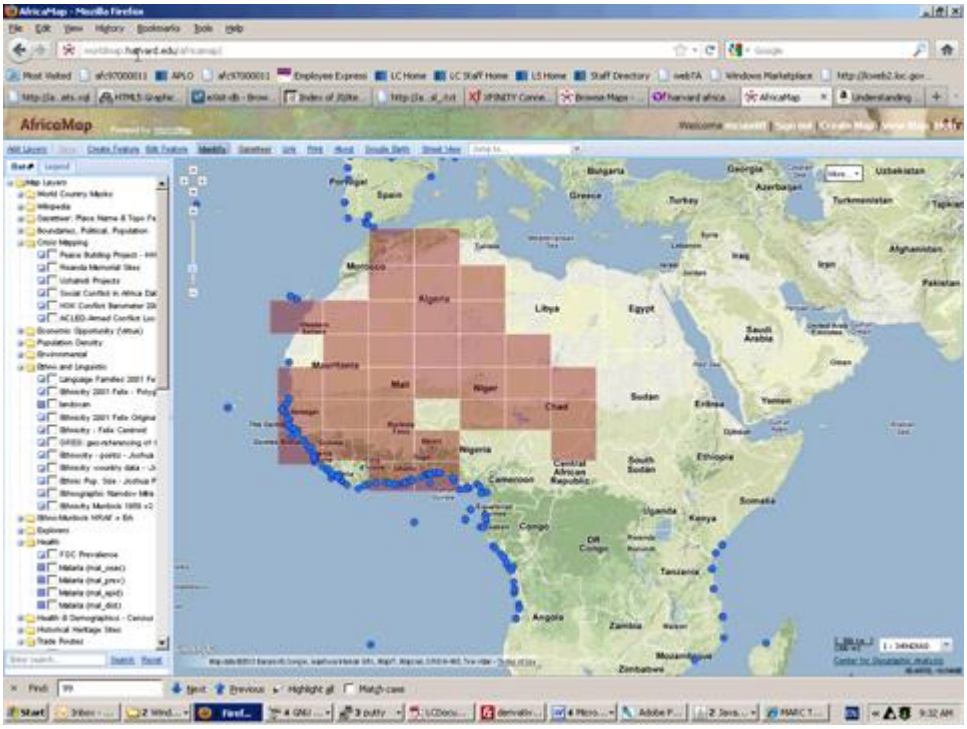
Example Uses of CLIR Project Data

Following are some simple examples of how end users can make use of the Africa Set Maps Project data. The examples below show re-use of the Shapefiles and the KML files.

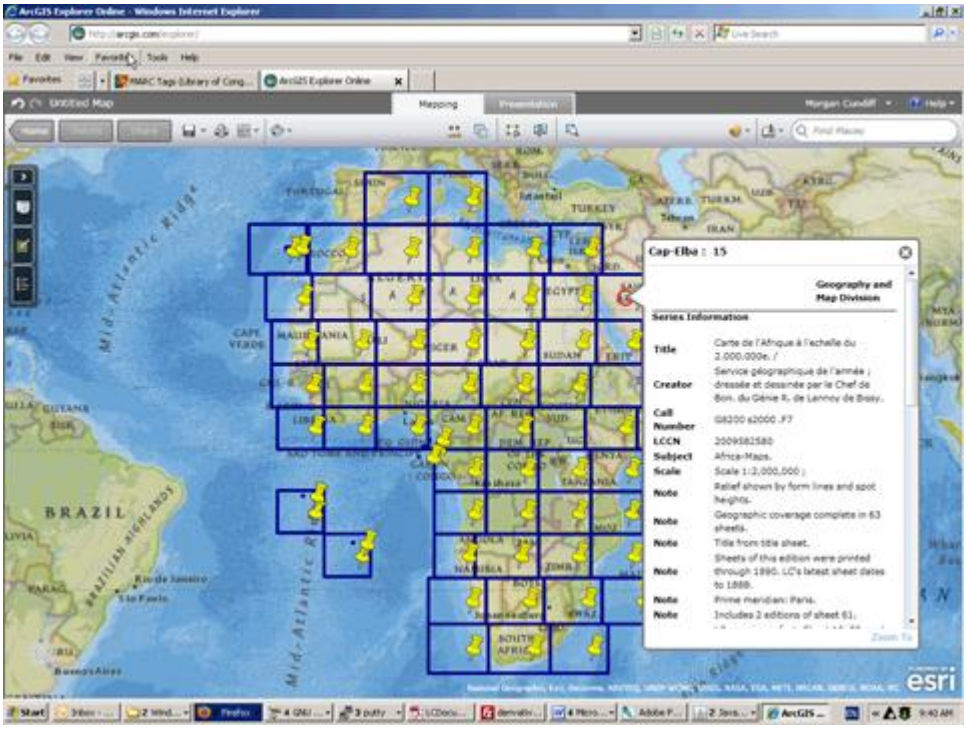
1. Quantum GIS (often abbreviated QGIS) is an open source desktop geographic information system (GIS) application that provides data viewing, editing, and analysis capabilities. The example below shows the Shapefile for a Library set map record overlaid on a base map of the continent of Africa in QGIS.



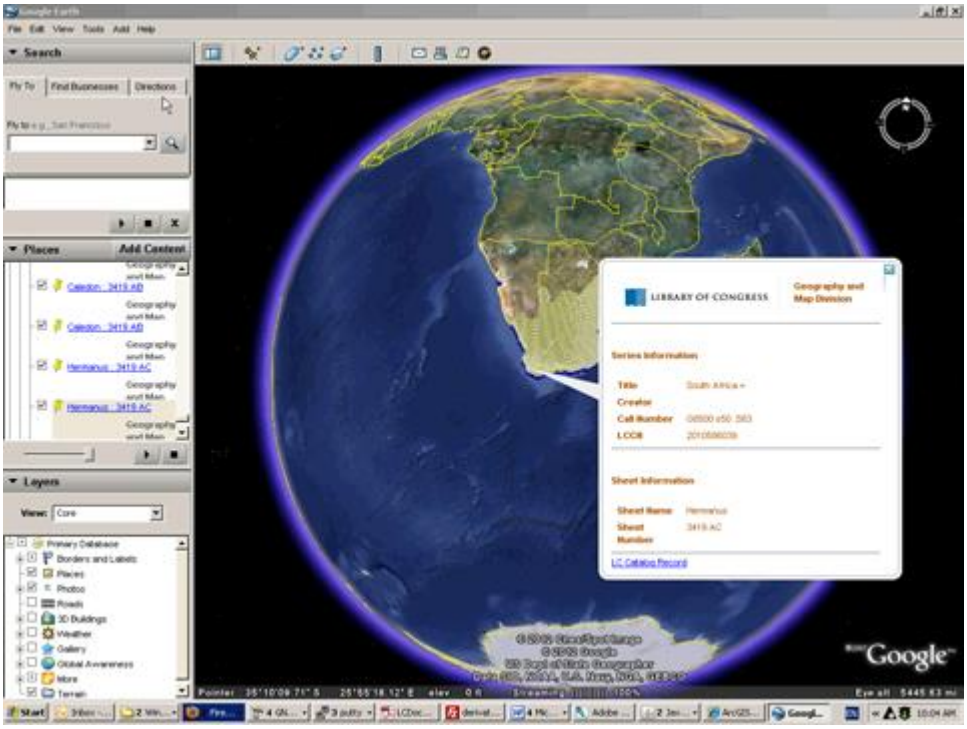
2. The next example shows an LC Shapefile loaded into the Harvard AfricaMap web application (housed at the Center for Geographic Analysis at Harvard University.) This application makes it possible to combine data from other contributed data sets with the LC map information. In the example below the blue dots indicate the location of former slave trade ports.



3. ArcGIS Explorer Online (from Esri) is a platform for viewing, and authoring geospatial information. In the example below the KML file for an LC map record is overlaid on a National Geographic base map.



Google Earth is a 3D geospatial browser that makes it possible to superimpose KML data on a globe created from satellite images.



The KML data can also be searched and displayed in the Microsoft Bing Maps web application.