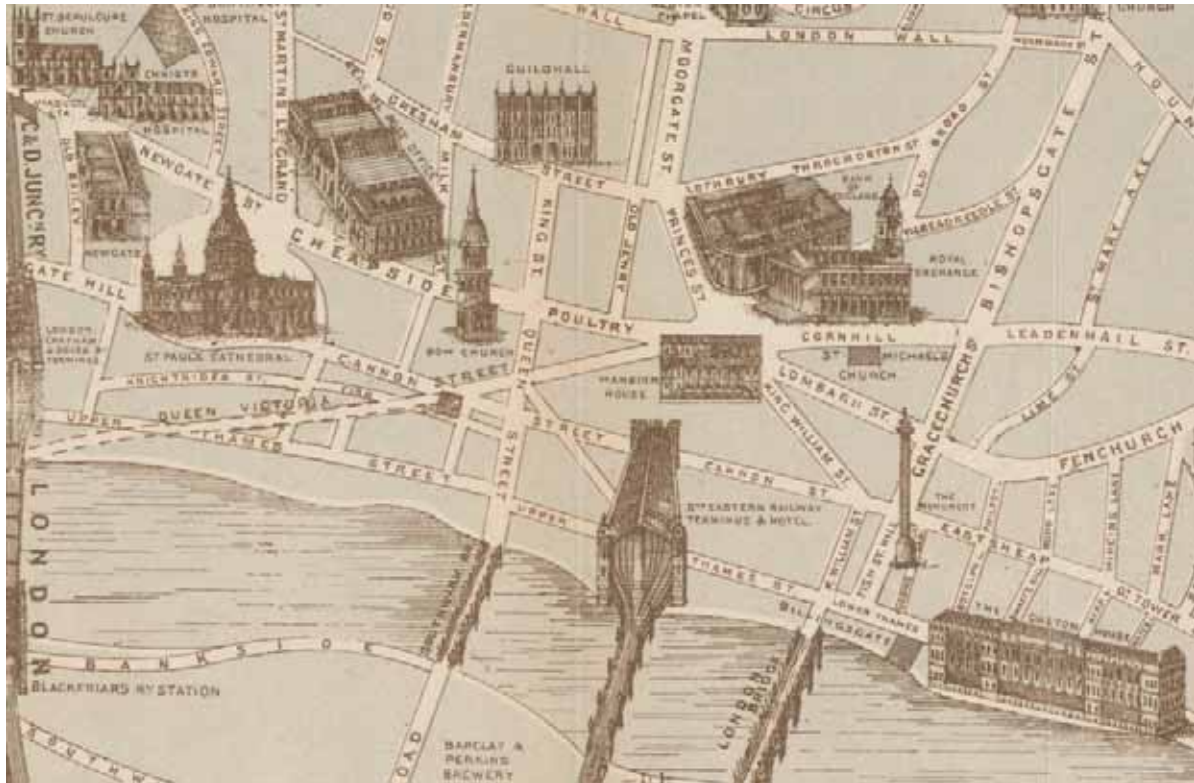


LDI Round 8  
Harvard College Library  
Harvard Map Collection  
Imaging the Urban Environment



Final Report  
May 5, 2009

Submitted by  
Bonnie Burns

## TABLE OF CONTENTS

<b>Introduction .....</b>	<b>2</b>
<b>Material Selection Process.....</b>	<b>3</b>
<b>Project Staff.....</b>	<b>4</b>
<b>Implementation Methods .....</b>	<b>5</b>
<b>Project Schedule .....</b>	<b>7</b>
<b>Budget.....</b>	<b>7</b>
<b>Lessons Learned.....</b>	<b>7</b>
<b>Appendix A: Inventory of Completed Work .....</b>	<b>9</b>
<b>Appendix B: Project Workflow.....</b>	<b>11</b>

## Introduction

The Harvard Map Collection is the oldest map collection in America, and has grown to be among the largest with ca. 400,000 maps, more than 6,000 atlases, and several thousand reference books. Topographic maps, city plans, nautical charts, and thematic maps comprise this excellent research collection representing all chronological periods and significant map makers.

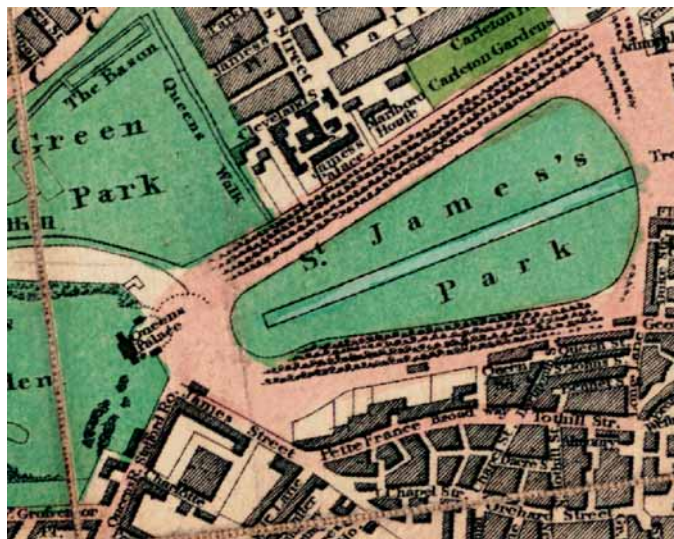
Scholars from around the world visit Harvard, drawn in part by the excellence of the library collections. While our world class resources include maps and geospatial data, these particular resources, so important and relevant across all disciplines, remained largely undiscovered.

The *Imaging the Urban Environment* project was designed to improve access to the resources of the Map Collection and to increase awareness of the global scope of the Collection's holdings. This project addresses the need expressed by a diverse group of faculty for quality images of historical urban maps to be used in the classroom, as well as for graduate and undergraduate research. In addition, the project makes these maps available for use in Geographic Information Systems format, allowing for in-depth time series analysis of the urban landscape.

*Imaging the Urban Environment* combined the knowledge gained from two previous LDI funded efforts: *Cataloging of Antiquarian Cartographic Materials* and the *Geospatial Data Access Project*. The former provided valuable experience in cataloging the maps and atlases in the collection, while the latter led to the development of an efficient workflow for digitization and georeferencing of materials. The experiences of the two projects were instrumental to the success of *Imaging the Urban Environment*.

The goal of IUE was to catalog, preserve, digitize and georeference cartographic materials related to urban areas around the world. The original proposal only included a preliminary list of cities for inclusion, and set a target of 520 maps cataloged and digitized, with 400 maps georeferenced. The difference in the target numbers is accounted for by the fact that not all maps are suitable for georeferencing, for example bird's eye views are not planimetric and cannot be assigned to an area on the globe with any semblance of accuracy.

At the end of the project timeline, July 1, 2008, approximately 700 items had been cataloged, about 545 were digitized and just over 300 had been georeferenced. As the result of an addendum to the grant made in June 2007, 8 fire insurance atlases of the Metro Boston area were also included, which are available through the Page Delivery Service. These consisted of 13 volumes and an additional 333 images.



Map 1: London, 1806

## Material Selection

The original selection criteria involved fewer than 20 cities, with the goal of selecting a map every 50 years for each city. Further examination of the collection revealed that this was not a realistic plan, since some of the cities were so new that selecting a map every 50 years would only result in 3 or 4 maps total being included. For other older cities, the collection is strong but only for certain time periods, so for example we have many maps of London but the majority of them date to the 19<sup>th</sup> century.

It was decided to extend the list of cities and to shorten the time interval when appropriate. In addition, maps of a selected city that were deemed of particular importance were included without regard to the date of publication. Only maps in the public domain, either due to age or publisher, were included in the project.

The final list of cities selected for inclusion:

Amsterdam	Hague	Nuremburg
Antwerp	Halifax	Paris
Athens	Hamburg	Philadelphia
Barcelona	Havana	Pittsburgh
Beijing	Istanbul	Pompeii
Berlin	Jerusalem	Portland, Oregon
Bordeaux	Kansas City	Prague
Brussels	Leipzig	Providence
Calcutta	Lisbon	Quebec
Cambridge, England	Liverpool	Rome
Capetown	London	Rotterdam
Chicago	Los Angeles	Rouen
Cincinnati	Louisburg, Nova Scotia	San Francisco
Cleveland	Louisville	Seattle
Cologne	Lyon	St. Louis
Copenhagen	Madison	St. Petersburg
Denver	Madrid	Stockholm
Dresden	Mexico City	Strausberg
Dublin	Milan	Tokyo
Edinburgh	Milwaukee	Venice
Florence	Moscow	Vienna
Frankfurt	Munich	Warsaw
Gdansk	Naples	Washington
Geneva	New Orleans	Zurich
Genoa	New York	

In addition to the maps on individual sheets, fire insurance atlases of Boston, Cambridge and the surrounding area were included.

## Project Staff

### Project Manager:

Bonnie Burns, GIS Specialist,  
Harvard Map Collection,  
[bburns@fas.harvard.edu](mailto:bburns@fas.harvard.edu), 5-2417,  
6-8125

### Project Associates:

Wendy Gogel, Digital Library  
Projects Liaison, HUL Office  
for Information Systems,  
[wendy\\_gogel@harvard.edu](mailto:wendy_gogel@harvard.edu), 5-  
3724

### Project Team:

David Cobb, Curator, Harvard  
Map Collection,  
[cobb@fas.harvard.edu](mailto:cobb@fas.harvard.edu), 5-2417,  
and 6-3670



Map 2: Lissbona, 1740

Joseph Garver, Reference  
Librarian, Harvard Map Collection, [garver@fas.harvard.edu](mailto:garver@fas.harvard.edu) 6-8717

David Siegel, GIS Applications Engineer, HUL Office for Information Systems,  
[dave\\_siegel@harvard.edu](mailto:dave_siegel@harvard.edu), 5-3724

Marc McGee, Geospatial Resources Cataloger, HCL/FAS Cataloging Support Services  
[mmcgee@fas.harvard.edu](mailto:mmcgee@fas.harvard.edu), 5-2409

Amy Phillips, Maps Cataloger, HCL/FAS Cataloging Support Services (Project hire)

Whitley Frost, GIS Technician, Harvard Map Collection, (Project hire) [wfrost@fas.harvard.edu](mailto:wfrost@fas.harvard.edu) 5-  
2417

Louise Baptiste, Conservator, Harvard Map Collection, [lbaptist@fas.harvard.edu](mailto:lbaptist@fas.harvard.edu), 5-2417

### Project Advisors

Robin Wendler, Metadata Analyst, Office for Information Systems, Harvard University Library

Stephen Chapman, Preservation Librarian for Digital Initiative, Weissman Preservation Center,  
Harvard University Library

## Implementation Methods

This project followed the well-established workflows of previous LDI projects. The position of Map Cataloger from the *Cataloging of Antiquarian Cartographic Materials* project was extended using funds from IUE, as was the GIS Technician position from the *Geospatial Data Access Project*.

Maps were selected by HMC staff based on the selection criteria listed above. Some of the selected maps, specifically those of New York, Philadelphia and Washington D.C. had been cataloged during previous projects. All other maps were sent to the Maps Cataloger and HOLLIS records were created. Each item was given the series title "Harvard Map Collection Digital Maps: Imaging the Urban Environment" in order to facilitate discovery of digitized items in the HOLLIS OPAC. For items that had been cataloged previously, the series title was added to the existing record.

Once cataloged, the maps were assessed by staff of the Weissman Preservation Center and placed into one of four categories: No treatment needed, treatment in HMC before digitizing, treatment at WPC before digitizing, treatment at WPC after digitizing. Treatment in the Map Collection was limited to surface cleaning and repair of minor tears, and this work was completed by the HMC Conservator. Items requiring treatment at WPC before they could be photographed were prioritized and transferred to WPC in batches as requested by the staff there. Many items did not make it to WPC at all within the time frame of the project, and therefore were not digitized.



Map 3: Washington, DC 1936

Maps that required no treatment, or that could wait until after digitizing, were batched and transferred to the Digital Imaging Group in Widener Library, as were maps that had been through treatment at WPC and returned. These were generally sent in groups of 10-15 sheets, which could be anywhere from 2 to 15 maps, as many maps are on multiple physical sheets.

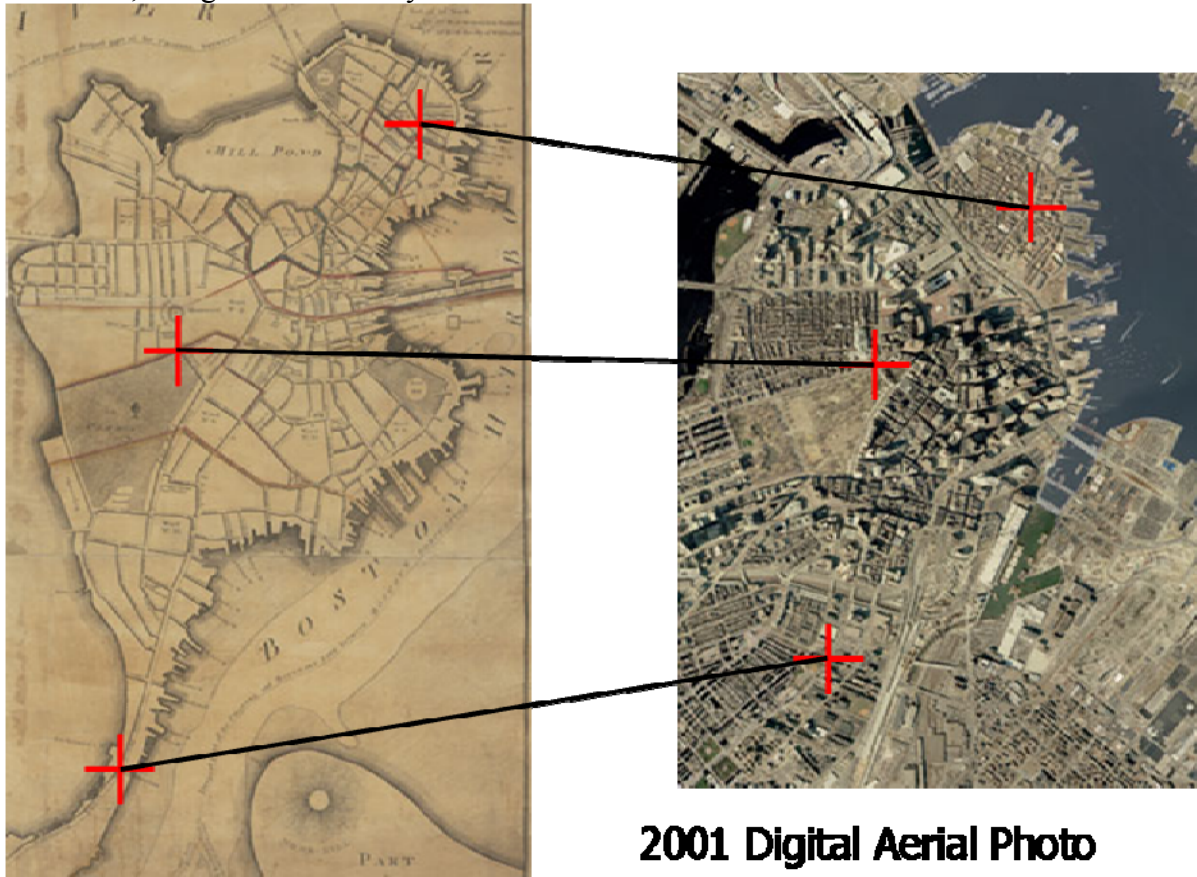
There was no formal standard resolution for the photography, instead a more subjective standard was used. The maps had to be at a resolution sufficient to read the smallest text or geographical detail on the map. This generally worked out to a resolution of 250 – 300 dpi. If possible, this was done in a single photo, however in some cases the maps were too large and were photographed in multiple shots that were subsequently stitched together to form a single image.

Maps on multiple sheets were generally photographed separately but the separate tiffs were also stitched together.

Once photographed, the resulting TIFF files were deposited into the DRS. In addition, a JPEG2000 file was created and deposited. A URN was created for each JP2. In the case of multisheet maps, a TIFF and JP2 were deposited for each sheet, and a URN was created for each sheet and for the stitched version. In the case of an oversize single sheet map, only the stitched together version was deposited and named. The URNs were inserted into the HOLLIS records by the Maps Cataloger.

Once photographed, the maps were returned to the stacks or set aside for further treatment at WPC.

The next step in the workflow was georeferencing. Maps were downloaded from the DRS as TIFF files, and georeferenced by the GIS Technician.



**1797 Street Map of Boston**

**2001 Digital Aerial Photo**

Georeferencing is the process of assigning real-world coordinates to an image. In order to do this, the GIS Technician first found a GIS (geographic information systems) dataset to use as a

source for coordinate information. She then identified points on the map image that could also be seen in the GIS data. This was often a historic building or a road intersection. In some cases, the historic map was already labeled with latitude and longitude, which simplified the process. In other cases, a GIS data set of sufficient detail was difficult to find. If a suitable dataset was not available, the GIS Technician would utilize the paper map resources of the collection. She would find an appropriate paper map that included coordinates – these were usually created by national mapping agencies. This map was scanned, georeferenced using the known coordinates and used as a source to georeference the historic map.

The georeferencing was done using ArcGIS software from ESRI. The process creates a new text file, called a world file. The world file was deposited to the DRS and a relationship ("IS\_WORLD\_FILE\_OF") was established between each world file and its associated TIFF.

Once the georeferencing was completed, the image and world file were downloaded from the DRS by the Geospatial Resources Cataloger, who created metadata describing the georeferenced image using the FGDC Standard for Digital Geospatial Metadata. This metadata and the image were then deposited into the Harvard Geospatial Library and made available to the general public. New URNs were created pointing to the item in HGL and those links were also added to the HOLLIS record.

### **Project Schedule:**

The grant was awarded in September, 2005, but due to previous commitments, we were unable to start work on the project until July 1, 2006. An addendum was made in June 2007 for inclusion of fire insurance atlases in the workflow. The project concluded on June 30, 2008. There were significant stops and starts in the course of the project due to leaves of absence and staff schedule changes, and this contributed to the length of time that the project went on.

### **Budget:**

Cataloging and data processing salaries and fringe: \$145,215  
HCL-DIG imaging services: \$34,520

### **Total Expenditures: \$ 179,735**

Project Award: \$192,000 -- remaining money in the fund was returned to the LDI.  
With the grant addendum, the estimated costs for salaries and fringe were \$146,000 and the costs for imaging were \$46,000, an underrun of approximately \$12,000. Many items that were anticipated to be part of the project, including many of the fire insurance atlases from the addendum, were not digitized because we could not get everything through the workflow in a timely fashion. Preservation treatments took longer than anticipated, and many items could not be treated before the project ended.

### **Lessons Learned:**

*Schedules are made to be broken.*

The Map Collection selected a large number of maps for this project, many of which required treatment at WPC. This worked out well for the first half of the project, however staffing issues and changing priorities at WPC led to a slowdown in preservation treatments. At the end of the project schedule, over 100 sheets were still awaiting treatment. Condition was taken into account to a certain extent during the selection process, but it perhaps should have been a more important selection criteria.

*If it ain't broke, don't fix it*

The Imaging the Urban Environment Project built on the experiences of two previous cataloging/digitization projects with great results. We had learned our lessons in the *Geospatial Data Access* and *Cataloging of Antiquarian Cartographic Materials* projects. We stuck with a workflow that functioned well and used the same methods for tracking materials. Items moved through smoothly, and we could always find items if they were needed for a patron request, which happened frequently. The workflow incorporated multiple checks by different staff members, so mistakes were caught and corrected before they could cause problems. Overall, the previous experiences are what made the IUE project so efficient and successful.



Map 4: Chicago, 1853

## Appendix A: Inventory of completed work

Below is a table describing the progress made on the project. The number of selected and cataloged items refers to specific maps – things that are given a HOLLIS number. However, many of these items are on multiple individual sheets. Billing and georeferencing are generally tracked by sheet, therefore the numbers in the "Digitized" and "Georeferenced" columns refer to the number of sheets, not necessarily maps.

<b>CITY</b>	<b>Selected</b>	<b>Cataloged</b>	<b>Total Sheets</b>	<b>Digitized</b>	<b>GeoReferenced</b>
<b>Amsterdam</b>	12	12	14	11	11
<b>Antwerp</b>	7	7	7	3	0
<b>Athens</b>	7	7	10	10	10
<b>Barcelona</b>	6	6	6	3	3
<b>Beijing</b>	6	6	8	6	6
<b>Berlin</b>	9	9	12	6	6
<b>Bordeaux</b>	9	9	20	18	17
<b>Brussels</b>	9	9	9	5	5
<b>Calcutta</b>	8	8	8	6	0
<b>Cambridge, UK</b>	3	3	4	4	4
<b>Capetown</b>	9	9	9	8	0
<b>Chicago</b>	13	13	18	16	16
<b>Cincinnati</b>	7	7	11	9	9
<b>Cleveland</b>	6	6	7	3	3
<b>Cologne</b>	8	8	9	8	0
<b>Copenhagen</b>	10	10	14	11	11
<b>Denver</b>	7	7	7	4	4
<b>Dresden</b>	8	8	7	7	7
<b>Dublin</b>	8	8	10	8	8
<b>Edinburgh</b>	7	7	7	5	5
<b>Florence</b>	10	10	18	7	7
<b>Frankfurt</b>	7	7	7	4	0
<b>Gdansk</b>	5	5	5	4	0
<b>Geneva</b>	5	5	6	4	0
<b>Genoa</b>	9	9	9	7	0
<b>Gibraltar</b>	8	8	8	6	6
<b>Hague</b>	7	7	7	4	0
<b>Halifax</b>	5	5	5	5	5
<b>Hamburg</b>	14	14	18	11	11
<b>Havana</b>	7	7	7	7	7
<b>Istanbul</b>	13	13	13	6	0
<b>Jerusalem</b>	9	9	14	10	10
<b>Kansas City</b>	6	6	6	6	6
<b>Leipzig</b>	8	8	8	6	6

<b>Lisbon</b>	11	11	11	9	9
<b>Liverpool</b>	7	7	9	4	0
<b>London</b>	29	29	38	37	35
<b>CITY</b>	<b>Selected</b>	<b>Cataloged</b>	<b>Total Sheets</b>	<b>Digitized</b>	<b>GeoReferenced</b>
<b>Los Angeles</b>	7	6	4	4	4
<b>Louisburg, NS</b>	2	2	2	2	2
<b>Louisville</b>	4	4	4	4	3
<b>Lyon</b>	7	7	7	2	0
<b>Madison</b>	4	4	4	3	3
<b>Madrid</b>	7	7	12	12	12
<b>Mexico City</b>	6	6	9	4	4
<b>Milan</b>	9	9	9	4	0
<b>Milwaukee</b>	6	6	10	9	9
<b>Moscow</b>	8	8	8	4	4
<b>Munich</b>	10	10	10	5	0
<b>Naples</b>	9	9	10	5	0
<b>New Orleans</b>	9	9	13	13	13
<b>New York City</b>	23	23	71	70	70
<b>Newport</b>	3	3	3	3	3
<b>Nuremburg</b>	8	8	8	5	4
<b>Paris</b>	25	25	45	35	25
<b>Philadelphia</b>	22	22	27	26	25
<b>Pittsburgh</b>	6	6	6	5	5
<b>Pompeii</b>	5	5	5	4	0
<b>Portland, OR</b>	5	5	9	8	8
<b>Prague</b>	5	5	5	4	0
<b>Providence</b>	5	5	5	5	5
<b>Quebec City</b>	10	10	11	11	11
<b>Rome</b>	17	17	26	22	22
<b>Rotterdam</b>	7	7	9	7	0
<b>Rouen</b>	8	8	8	5	0
<b>San Francisco</b>	14	14	15	14	14
<b>Seattle</b>	7	7	7	7	7
<b>St. Louis</b>	7	7	7	6	6
<b>St Petersburg</b>	12	12	12	9	9
<b>Stockholm</b>	10	10	10	7	7
<b>Strausburg</b>	9	9	9	5	0
<b>Tokyo</b>	7	7	7	4	4
<b>Venice</b>	13	13	20	9	9
<b>Vienna</b>	18	17	19	13	13
<b>Warsaw</b>	4	4	7	5	0
<b>Washington, DC</b>	16	16	16	16	16
<b>Zurich</b>	6	6	6	4	0

## Appendix B

### Project Workflow.

Select maps that will be digitized

Set start schedule and batch size with the DIG

- Up to 15 items a batch
- 2 to 3 batches per month

Set cataloging priorities

- Maps will be cataloged in HOLLIS before they are reformatted by the DIG

Catalog each item in HOLLIS (if not already cataloged)

- Create a MARC record for the individual map or map series
- Add additional series title if needed

Confirm there is an inventory control number on each individual item

- Hand written call number or
- Stamped call number or
- Pre-printed number on the map

Conservation treatment

- Attain conservation assessment for maps in order of cataloging priorities
- Classify items according to conservation needs
  - No treatment needed
  - Minor treatment needed (surface cleaning, tears) that can performed at HMC
  - Treatment needed at WPC before reformatting at DIG
  - Treatment needed after reformatting at DIG

Create the inventory for DIG

- Call number
  - Image file name: NRS path urn-3:FHCL:n
  - Title or description: used to identify the piece
  - Owner/collection/batch: repository name/collection name/local batch identification (
  - Total # of originals: # of original items in this batch, used for tracking purposes
- Note: Annotations that are meant to clarify instructions about individual items for the digital photographer should be added by hand to a hard copy version of the spreadsheet and included as a packing slip with the original material when it is transferred for imaging.

Bring batches to DIG along with paper and electronic copies of inventory sheets

- DIG will create
  - TIFF files
    - File name is
      - call number,
      - map sheet number
      - followed by .TIFF extension
  - JP2 files
    - File name is
      - call number,
      - map sheet number
      - followed by .JP2 extension
      - NRS path urn-3:FHCL:n
- The DIG created TIFF files and a JPEG2000 files are deposited in the DRS. The JP2 file is given the “IS\_DERIVATIVE\_OF” relationship with the TIFF.

Download and georeference TIFF files and deposit the TFW files to DRS

- Download the TIFF file from the DRS, and use ArcMap to georeference the image. The coordinate system of the georeferenced image should match that of the paper map if possible, but it is more important that the selected projection is a standard (non-custom) projection in ArcIMS and ArcSDE. Update the georeferencing so that a TFW file is written.
- Calculate the MD5 signature for the TFW and deposit in the DRS.
  - Use the following settings on page one:
    - Owner code: FHCL.MAPS
    - Billing Code: FHCL.MAPS.GEO\_0001
    - Metadata type: TEXT
    - Mime Type: text/plain
    - Role, purpose and quality: NA
    - Access Flag: P
    - Usage Class: HIGHUSE
    - Create Date: enter the date the files were created
  - Use the following settings on page two:

- Character repertoire: Unicode
- Character map: UTF-8
- Descriptor type: ESRI\_WORLD\_FILE
- Use the following settings on page three
  - The owner supplied name is the same as the owner supplied name of the original TIFF file.
  - Enter the file signature from step 5
  - Signature type: MD5
  - Use the browse button to select the file.
- Add the "IS\_WORLD\_FILE\_OF" relationship to the file. Enter the DRS number of the TFW file in the first box, and the DRS number of the original TIFF in the second box.

Load the georeferenced image and metadata into HGL

- Create a complete, FGDC compliant metadata record for the TIFF image
- Export the metadata from ArcCatalog using the FGDC XML option
- Load the georeferenced image into the production SDE using ArcCatalog
- Calculate 4 pyramid layers
- Set the projection id
- Load the metadata into the GS\_NEW and GP\_NEW tables using the MetadataImporter program.
- Add records to the global\_raster table recording the DRS IDs and type of file, using SQLPlus or Toad. For SQLPlus, the command is:  
 INSERT INTO GLOBAL\_RASTER VALUES (<sequential#>, <hgl\_table\_name>  
 <hgl\_layer\_id>, <DRSID#>, <file\_type>);  
 The file type must be either "TIFF File" or "TFW File"
- Reindex
- If necessary, create a URN for the publication using the NRS Web Admin tool, and send the URN to the appropriate person to add to the HOLLIS record.

Required components for depositing scanned maps:

- DRS Web Admin (with depositor rights)
- ArcMap
- ArcCatalog
- shp2sde
- md5.jar
- SQLPlus
- mdLoader.java

#### HGL interaction with DRS

Each HOLLIS record for an electronic resource will contain a URN that will run a query to HGL

- Query to HGL will contain the HOLLIS ID
- A intermediate look-up table exists between HOLLIS-ID and HGL-ID
  - The look-up table has both the HOLLIS ID and the HGL Publication ID
- A publication will be identified by the query from HOLLIS

Result set

- The URN will link to a result set in HGL
- The result set will be a list of layers for a publication
- The user will be able to select a specific layer for viewing

Downloading results

- HGL will send a query to DRS
- The query will:
  - Include the DRS Object IDs for the related files
    - TIFF
    - TFW
  - The objects will be extracted from the DRS using Asynchronous delivery