

Executive Summary

- dates: January 2008 to April 2009 (and beyond)
- participants from Archives: Tom Rosko, Elizabeth Andrews, Mikki Simon Macdonald, Jessica Brody (transcriptionist), Sara Smith (transcriptionist), Krystle Stratum (transcriptionist)

	STAFF TIME (HRS)	TEMP TIME (HRS)
PROCESSING	43	
RESEARCH	20	
SELECTION	28	
DIGITIZATION PREP	34	
DIGITIZATION QC	24	
CATALOGING	86	
CATALOGING QC	10	
DUBLIN CORE MAPPING	17	
TRANSCRIPTION	42	376
TRANSCRIPTION QC & CONVERSION TO PDF	25	110
MEETINGS	195	
IAP SESSION	7	
TOTAL:	531	486
TOTAL HOURS BY ALL:		1,017

- Key Points:
 - 235 items with a total of 1095 pages, including 51 drawings (26 oversize) were digitized from the Kevin Lynch papers (MC 208)
 - MC 208 needed to be organized and described at the collection level before items could be selected
 - Three temps were hired (total of 40 hours/week) to complete transcription
 - The project marks the first time the Archives has, on a large scale, cataloged at the item level
 - All cataloging was completed by one staff member, the Archives Technician
- Suggestions for Future Projects:
 - Use AT for item level cataloging and digital object management
 - Create standard metadata fields for all digital collections
 - Monthly web searches to locate links to our collections (and receive feedback from users)
- Final Steps to Completion (see final section of report for details):
 - QC of DOME (ongoing as updates are made)
 - Archivists' Toolkit
 - Brown Bag
 - Reporting and Documenting
 - Reference Procedure
 - Kevin Lynch Awards
 - Evaluation and Feedback

The following offers a detailed description of how the 1,017 hours were spent throughout the Kepes-Lynch project. Each section represents a necessary step undertaken by the Institute Archives for the successful completion of the project. Each section concludes with how, and why, hours may vary in future projects.

The final section (Ongoing and Future Steps) discusses steps still being taken as the Kepes-Lynch project concludes. In many cases there is no way to truly estimate the hours that will be dedicated to these final steps.

PROCESSING

43 HOURS

- Reviewed *The Image of the City* in order to be able to better understand the perceptual form of the city project, to determine what material belonged in the new series and to write proper collection and item level descriptions
- Processed the material in MC208 pertaining to The Image of the City/Perceptual Form of the City. Reviewed the entire collection and gathered together all of the material related to *The Image of the City*
 - MC 208 was completely unprocessed having come to the Archives in 12 accessions over 12 years. The collection is 16.5 cubic feet
 - Processing included:
 - Creation of new series (& completed basic processing steps) for 2 cubic feet (6 boxes)
 - Determining what material belonged in the new series "Perceptual Form of the City"
 - Compiled contextual information. Wrote series description, scope and contents notes, history note, and notes on The Image of the City
- Reviewed materials in map case to determine relationship (if any) to Perceptual Form of the City

Future Projects

Processing time will vary greatly depending on the collection(s) chosen for a given digitization project. If a collection has previously been processed (described and organized at the collection level) then little work will be needed. Likewise, if a collection has not been processed weeks could be needed to pull it into shape for digitization.

MC208 had not been processed and for this project only a very small portion of the collection needed to be processed [2 cubic feet]. Processing the entire collection (a remaining 14.5 cubic feet in 22 boxes) would have taken significantly more time.

RESEARCH**20 HOURS**

- Further research was necessary to select the material to be digitized. Research Included:
 - review of *The Image of the City*
 - materials in MC 208
 - President's Reports
 - *MIT Directory of Current Research*
- Conducted background research on the Rockefeller grant in order to make decisions regarding copyright
 - Corresponded with Rockefeller Archives
 - Received photocopies of original grant proposal and follow-up documentation

Future Projects

Research will always be necessary in order to provide the proper context for the digital collection and to facilitate material selection. The time needed for this step will vary depending on the scope of the project and scale of the collection.

SELECTION**28 HOURS**

- Reviewed MC208 Perceptual Form of the City series with Jennifer Friedman (5 meetings in total) to determine which documents were to be digitized. *Note:* not all items in the Perceptual Form of the City series were selected for digitization
- Determined selection criteria to be:
 - Documents dealing directly with project planning, inspiration, and documentation
 - If multiple copies of a document – digitize only the copy(ies) with annotations
 - Documentation related to similar projects done by Lynch during the same time period will not be digitized as part of the KL project
- Created a filemaker database (originally excel) to track the selection process and later the overall management of the KL project
 - Documents were assigned tracking numbers (later became KL#)
 - Created titles for each document (originally for basic tracking purposes – later this became the foundation for the descriptive metadata)

Database Fields:

- Document ID
- Document Title
- Box
- Folder
- Description (originally used for condition and author information – not actual content/context description as later needed)
- Date

- page count
- handwritten pages
- annotated pages
- names
- notes on annotations
- map
- dimensions
- related to

Added at time of vendor transfer and qc:

- At vendor(y/n)
- vendor box #
- dated transferred
- date returned
- QC (y/n)
- errors(y/n)

Added at time of transcription:

- transcribed by
- reviewed (one, two, three)

- **235** items were chosen for digitization, with a total of 1095 pages, including 51 drawings (26 oversize), 672 of which are handwritten or include handwritten annotations
- Compiled a list of all names appearing in the selected documents (not those of study participants) for discussion re: privacy and copyright issues
 - 14 authors other than K. Lynch

Future Projects

Time required for selection of materials is dependent on the size of the physical collection and the scope of the digitization project. Time will vary greatly.

Creating a database to manage the selection process is necessary. In the future more time should be spent at this step to collect the metadata necessary to catalog the items. This will not only save time later when cataloging, but will also limit the handling of the original items.

DIGITIZATION PREPARATION

34 HOURS

- Prepared documents for transfer to Boston Photo Imaging
- Prepared test batch of documents and had several meetings with BPI to review/discuss/determine digitization specs
 - Specs based on a few representative samples
- Measured the oversized drawings to determine if a special set up was need to film
- Determined what the preservation/digitization issues were for each of the various types of paper present in the collection
 - Consultation with preservation services
 - Online research

➤ Workflow for prep work:

Preparing Archives materials for vendor:

- Consult filemaker file MC208KL to determine which documents will be sent out for digitization
- Each document will be foldered separately. Staples and paperclips will be removed
- The following will be written on the front of the folder:
 - Unique document number
 - # of sides to be digitized
 - Any special instructions for person digitizing the materials
- When a document is removed from the collection it will be replaced with a document information page that will include the document number, document title, and page numbers. This page will be printed from the MC208KL Filemaker database
- At this time a photocopy will be made of each document. The photocopies will be kept in order and batched according to the folder that houses the original document
- The individually foldered document will then be reboxed into a numbered manuscript box. In each record carton will be an inventory listing of the documents inside
- In MC208KL note the new box number
- Record the date physical possession of the documents transfers to the vendor

Future Projects

The time needed to prepare documents for digitization will vary with the number of items selected. An additional variable will be the types of material being digitized, as well as the condition of the material. Scheduling preservation/conservation review/work needs to be taken into consideration as early as possible.

Less prep work may need to be done if digitization is done in-house as opposed to outsourcing to a vendor.

DIGITIZATION QC

24 HOURS

- Checked physical condition of materials returning from BPI & confirmed digitization (recorded the return in the DB)
- When materials are returned from vendor:
- Check each record carton to be sure that the folders match the folder listing
 - Match each document folder, using the document number, with a document information page
 - Place the documents back in their original order in their original folders
 - In MC208KL record when each document returns from the vendor, and when each document has been relocated back to its original folder and box

- Checked digital files: Opened PDF checked against DB notes and original if needed. Checked TIFFs when problems were found in PDF (and spot checked)
Compared:
 - KL#
 - page count & page order
 - "title"
 - Quality
- Sent xl to Beverly with corrections that may be needed
 - 11 fixes necessary
 - Due to quality and variety of original documents some digital not as crisp as would have liked – using same specs for all easier but not necessarily great - depending on ultimate use and need – fine for what we are doing at this point in time

Future Projects

Digitized materials will always need to be checked for quality and accuracy. The major factor affecting the QC process will be the number of objects digitized. We need to determine who should do the QCing. Should this be done by the collection owner, or a different unit?

CATALOGING

86 HOURS

- This projects marks the Institute Archives first serious foray into item level cataloging
- IRIS training – two training sessions with Jolene
- Worked with Carl to map DB to IRIS
 - While this got the data into IRIS significant clean-up was necessary. Mapping only got the info into IRIS – not into the correct fields
 - DB fields mapped into IRIS:
 - Document Title
 - Date
 - Description
 - Document ID (KL#)
 - Page Count
 - Folder
 - Dimensions
 - Box #
 - Related to (list of KL#s of other items *VERY* related to item)
- created titles and dates based on archival practice and research
- assigned work types to each document
- added subject, and name records to IRIS authority files & attached to item records
- wrote content descriptions for every item – included context of the perceptual form of the city (necessary since items must on some level stand alone)

Future Projects

Cataloging is one step that may take less time in the future (again the actual time will vary based on the number of items being cataloged). On future projects more work should be done during the selection process to facilitate cataloging and minimize handling of the physical material. Much of the metadata can be collected / created at the time of selection (title, author, date, page number).

Cataloging of future projects should be done in Archivists' Toolkit. AT is where the collections will ultimately be managed. AT also provides several different methods of data export (METS, MODS, MARC, and Dublin Core).

Consideration should be given to what other resources (staff) in the Libraries may be able to assist with the cataloging/metadata work, and what particular tasks are suited to which level of staff (professional cataloger, student, etc.).

CATALOGING QC

10 HOURS

- Printed list of works in IRIS
 - checked to be sure all KL numbers were present
 - checked titles and descriptions for misspellings
 - spot checked records for anomalies
- Scrolled through IRIS records to be sure that all "builds" were complete

Future Projects

The time needed to check for quality and accuracy will vary with the number of items cataloged and the depth of cataloging completed.

DUBLIN CORE MAPPING

17 HOURS

- Cataloging was begun in IRIS with the understanding that mapping and ingest pathways to DOME/Dspace were already in place. This was true only for RVC records
 - Archives used additional IRIS fields not used by RVC (eg: description) or used fields in a different way than RVC (eg: extent)
 - Different needs required different mapping (in some cases)
- Corresponded with Carl over the course of several months to determine the best possible mapping options. Main questions were regarding the following data:
 - Author/Research – dc.creator or dc.contributor
 - Page number – dc.format.extent
 - Dates – dc.date.created vs. date.issued
 - Physical location - folder/box - (in IRIS field Repository_No.)
 - Repository information

- The suggestions made by IASC staff were based on significant research into the best practices of other digital repositories; a favorite example being California Digital Library
- Dublin Core fields being used as of March 2009:
(field names marked in red are fields we had suggested using)

dc.creator	author/researcher
dc.date	creation date (specific date or circa date span)
dc.date.issued	year listed in creation date or date span (required field in dspace)
dc.identifier.other	IASC collection number (dc.identifier is being used for the IRIS image number)
dc.description	Description
dc.description.provenance	KL#s for TIFFs and corresponding checksums
dc.format.extent	page count (included " <i>count of pages:</i> " before number) or dimensions (included " <i>dimensions:</i> " before numbers)
dc.publisher (dc.repository.name)	MIT IASC
dc.relation	Indicates other objects that are VERY related to the record being viewed. KL#s used to represent this relationship. While this information is not currently being used, it is our hope that it will be in the future. Included here so that the relationship is not lost.
dc.relation.ispartof (dc.source)	location of original (Collection #, Box #, folder title)
dc.relation.isbasedon	KL#
dc.subject.lcsh (or other)	Subjects
dc.title	Title
dc.type	genre type (text, drawing, interview, etc.)

- Records were reviewed in DOME_test and then later in DOME
 - This process has led to the need for additional research, discussions, and changes to metadata
- At the time of this report time is still being spent check & rechecking metadata in DOME and working with Carl to make the appropriate adjustments

Future Projects

Very minimal time, if any, should be dedicated to Dublin Core mapping. This step should not be needed for future projects. As the Libraries begin to standardize the metadata requirements for dspace/dome, altering the mapping for individual projects should be minimal.

- The decision was made to transcribe all documents (typed and handwritten) to provided full-text searching. This decision was based on the integral nature of the transcriptions (individuals' interpretations and descriptions of the sights, sounds, and smells of the environment) to understanding the research.
- Prep work
 - Performed typing tests with the material in order to calculate time and staff needed to complete the project.
 - Created guidelines for transcription based on LC practices
 - Created a tracking workbook (tab for each week) to track pages completed, time spent, and issues/confusion in documents
 - Doc ID number
 - Date Transcribed
 - # of pages
 - Work type [dropdown containing following choices: interviews, texts, field notes, correspondence, writings, drawings, other]
 - Suggested keywords
 - Problems (y/n)
 - Page(s) [with issues] (to help during QC process)
 - Specific details [dropdown containing following choices: difficult to read main text, difficult to read notations, unsure about word/phrase, unsure of best way to transcribe]
 - Next steps [dropdown containing following choices: need original, need second opinion]
 - Completed date
- Hired students to transcribe the documents
 - Wrote job description and posted with Simmons College
 - Received 20 resumes
 - Interviewed four candidates
 - Hired three candidates (Simmons students in the Archives management concentration)
 - two at 15 hours/week
 - one at 8-10 hours/week
- Training
 - Explained project, transcription guidelines & procedures, and tracking workbook
 - Ran a typing trial to obtain a time estimate (min/page) for both handwritten and typed material
 - This information was used to determine the general work load to prepare each week for each temp
- Transcription dates: September 29, 2008 through December 1, 2008
 - Temps worked onsite 1 day each week (Monday) for approximately 6 hours and remotely for the remainder of week
 - Transcription was done from the digital versions. Originals were provided onsite when needed

- Workflow for transcription
 - Each temp was given a flash drive that contained:
 - Original PDFs (approximately 1 weeks worth of work)
 - Tracking workbook
 - Group meeting each Monday to review the process and ask/answer questions
 - Transcribed documents were backed-up weekly on the R-drive (as were the tracking workbooks)
 - Each Monday the temps were given a new set of files to transcribe
 - Assignments were tracked in the DB

➤ Transcription breakdown

	Temp A	Temp B	Temp C
Documents	57	54	72
Total Pages	351	278	400
Handwritten Pages	190	46	137
Hours	135	96	145

- Transcription was completed at an average of 2.8 pages/hour [time dependent on amount and type of text (handwritten vs. typed) per page]

Future Projects

Transcription may not be necessary for every project. It should be considered on a project by project basis dependent on the needs of the collection and those of the end users.

Even if it is determined that a collection would benefit from transcription we may want to avoid transcribing all documents, and consider only handwritten documents.

The main reason for transcribing the Lynch material was to provide full-text searching within the documents. In the future we may want to consider OCR as an alternative option.

TRANSCRIPTION QC & CONVERSION TO PDF

135 HOURS

110 temp/25 staff

- December 1, 2008 through January 5, 2009
 - Conversion began December 31, 2008 and was completed January 9, 2009
- QC consisted of checking for the following:
 - Misspellings (if words are misspelled in original they will be misspelled in the transcript)
 - Consistency in the use of notations and symbols
 - Difficult to decipher words and phrases and offering second/third opinions on possible translations

- Prep work
 - Created transcription review workbook to track the QC process. Designed to be passed from one temp to the next
 - Document id – (pre filled)
 - # of pages (as a check against info in db)
 - Problems (y/n) filled in by me based upon original transcription tracker
 - Problems (y/n) 2
 - Problems (y/n) 3
 - Pages with issues
 - Specific problems
 - Next steps
 - Comments
- Workflow for transcription QC:
 - Each temp was given a flash drive that contained:
 - Original PDFs
 - Docs transcribed by one of the other temps
 - Transcription review excel
 - Gave A's documents were given to B, B's to C, and C's to A
 - Documents were reviewed, edits made and changes noted in workbook
 - Difficult words were highlighted in red and marked in the workbook to be passed to the next reviewer
 - Group meeting each Monday to review the process and ask/answer questions
 - Files were backed-up weekly on the R-drive
 - After two weeks of review the process was repeated. A's documents were given to C, B's to A and C's to
- Before converting files to PDF
 - Documents were checked to be sure that there were no red highlights and that all questions from the tracking sheet were answered. As well as that the KL numbers matched and had the correct number of pages.
 - A final decision was made on any words or phrases still highlighted in red. If the word was completely indecipherable If red remained staff made the final call on words or used w_____
 - Saved as pdf in the following format KL_002XXX_tr.pdf

Future Projects

This is a necessary step if transcription is undertaken. The time needed will vary with the number of items transcribed.

PROJECT MEETINGS**195 HOURS**

- This is an estimate of time spent in meetings that are not accounted for in other sections of this report. The hours account for the time of all three of the Archives staff members
 - 12 project team meetings (16 hours)
 - 49 hours in meetings
 - 18 hours dedicated to preparing for the meetings and following up on issues raised during the meetings (30 min per meeting per staff)
 - Numerous impromptu meetings of staff
 - Estimate one hour a week (approximately 64 weeks)
 - Two staff members meeting
 - Approximately 128 hours

IAP**7 HOURS**

- It was decided that an IAP session would be a great way to advertise the project
- The project was not yet complete and materials were not available live in DOME
- January 9, 2009
- 5 hours of preparation for presentation
 - Meetings with Jolene and Beverly
 - Creation of presentation slides
- 2 hours for actual presentation
 - Approximately 22 attendees (students, faculty, and staff)

Future Projects

We recommend conducting a similar session for each significant digitization project. The IAP session proved to be an excellent way to drum up excitement for the project and also gave us an opportunity to interact with those in the MIT community interested in our subject.

- Archivists' Toolkit
 - Important step for access and marketing
 - At time of this report 9 hours have been spent entering MC208 finding aid into AT and attaching digital objects
 - Estimate another 10 to 20 hours to get all metadata into AT (this will depend on if we can automate the process or if data will need to be re-keyed into AT) have been discussing this with Carl
- Brown Bag
 - The project team will be conducting a session for the Libraries
 - Will repurpose material created for IAP
- Reporting and Documenting
 - Approximately 15 hours have been dedicated to reviewing the project and reporting
- Creation of Reference Procedures for Digital Objects
- Kevin Lynch Awards
 - Unknown what work will be need to be done for this important event
- Evaluation and Feedback
 - All projects should be followed up with evaluation of the finished product and gathering feedback from users. May include:
 - Official user evaluation surveys
 - Web searches (monthly?) to locate links back to our collections
 - Where do we appear?
 - What are people saying?
 - What can we learn about our users?
 - What can we learn to improve future projects?
 - Web analytics for collection sites
 - What keywords are used to find our collections?
 - What is being searched for within the collection?
 - How is the site being used/searched?