

Book Info

Title: Northern antiquities
Author: Mallet, Paul Henri
Publisher: H. G. Bohn
Date: 1847

Search inside this Book

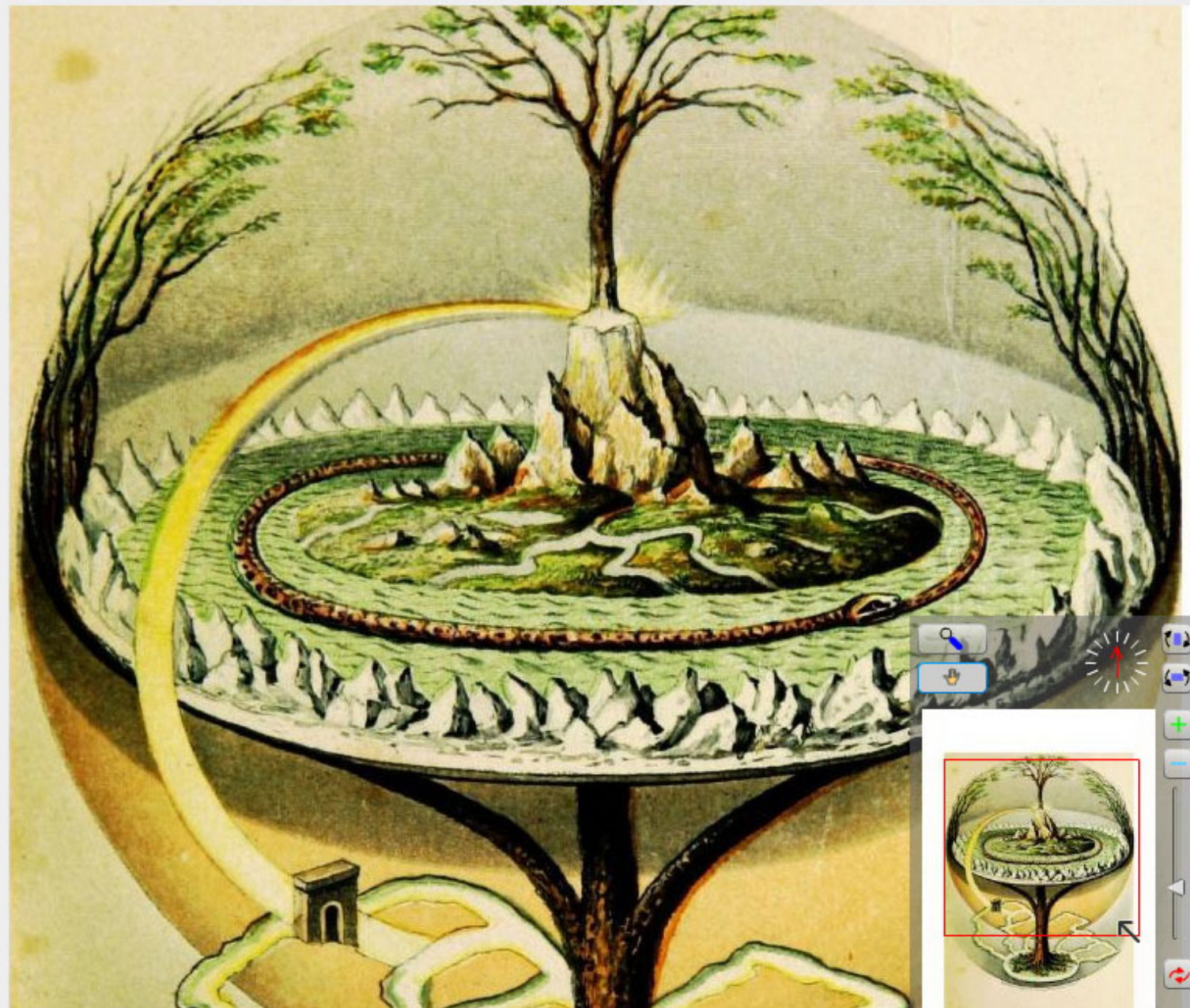
Search

Explore this Book

Link to this Book

Download/Print this Book

Pages Thumbnails Raw Text Zoom In Zoom Out Reset Zoom Use Flash



Page 4 of 588 100 Items per page

Book Info

Title: Northern antiquities
Author: Mallet, Paul Henri
Publisher: H. G. Bohn
Date: 1847

Search inside this Book

 Search

Web-based workflow software to support book digitization and dissemination

The Mounting Books project

books.northwestern.edu

Open Repositories 2009 Meeting, May 19, 2009

Explore this Book

Link to this Book

Download/Print this Book

Page 4 of 588 100 Items per page

What we will cover today

- History of Northwestern's Kirtas program
 - Scope of local book digitization
 - Workflow
- Goals of the Mounting Books project
- Details of the workflow system
- About Fedora architecture for Northwestern Books
- Future directions

Kirtas @ Northwestern University Library

2005 Acquisition and initial testing (in Preservation)

2006 Production of print facsimiles of Brittle books





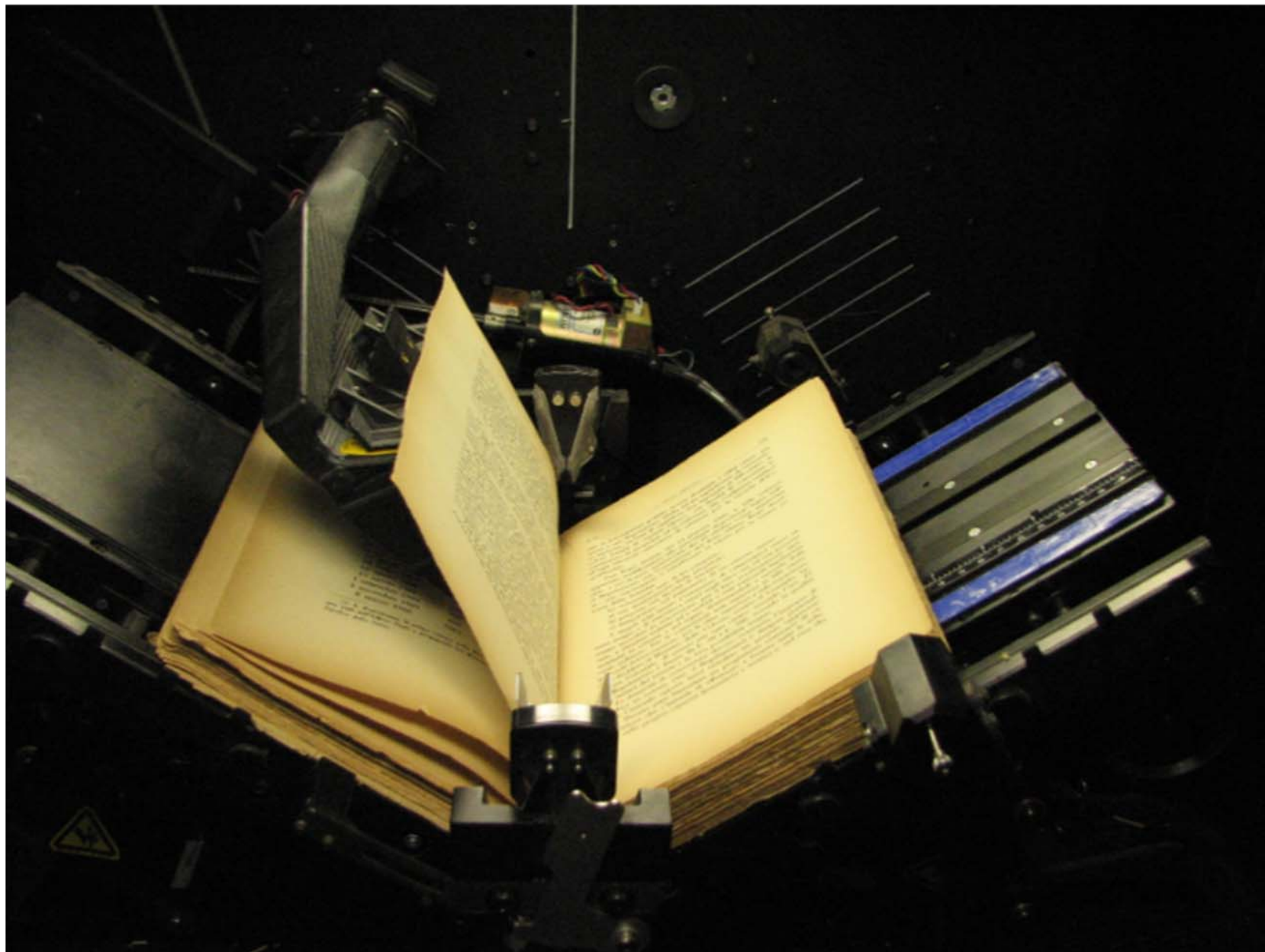
Problem

We had a book scanner.

We weren't doing anything with the scans other than printing them, binding, returning to shelf.

We didn't have a system in place to publish them to the Library's Fedora repository.

The book digitization workflow



Production / Staffing

Kirtas:

APT 1200 rated @ 1200 pages/hour (20 pages/minute)

Assumes no setup time, 1200 page book

Produces raw images

Real-world:

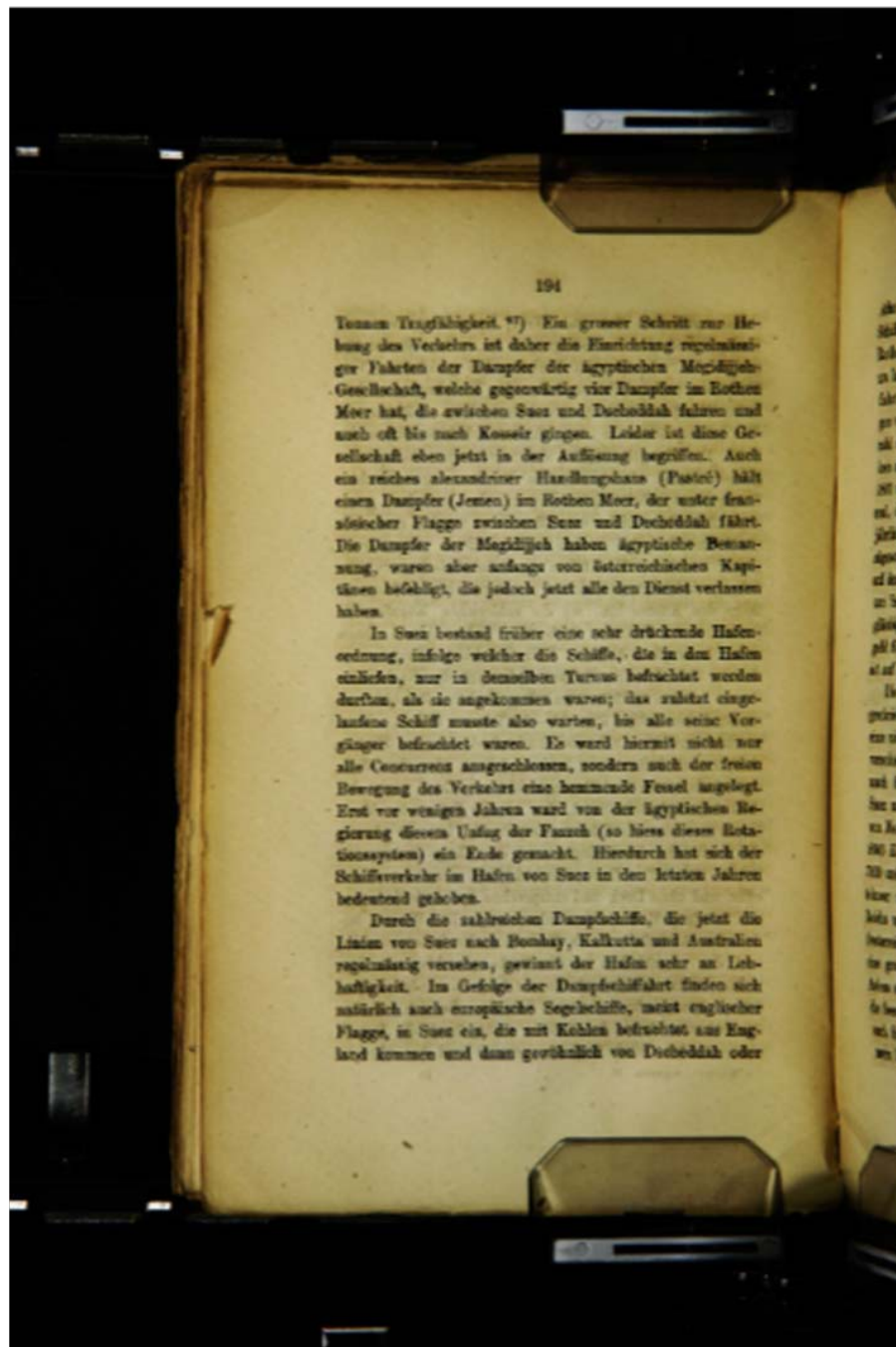
APT 1200 produces 6-900 pages/hour

Each student processes 1-200 pages/hour

Currently 6 students: 2 scanning, 4 processing

Before Book Workflow Interface software:
120 books/quarter

Brittle Image Processing



Tonnen Tragfähigkeit.²⁷⁾ Ein grosser Schritt zur Hebung des Verkehrs ist daher die Einrichtung regelmässiger Fahrten der Dampfer der ägyptischen Megidijeh-Gesellschaft, welche gegenwärtig vier Dampfer im Rothen Meer hat, die zwischen Suez und Dscheddah fahren und auch oft bis nach Kosseir gingen. Leider ist diese Gesellschaft eben jetzt in der Auflösung begriffen. Auch ein reiches alexandrinischer Handlungshaus (Pastré) hält einen Dampfer (Jemen) im Rothen Meer, der unter französischer Flagge zwischen Suez und Dscheddah fährt. Die Dampfer der Megidijeh haben ägyptische Bemannung, waren aber anfangs von österreichischen Kapitänen befehligt, die jedoch jetzt alle den Dienst verlassen haben.

In Suez bestand früher eine sehr drückende Hafenordnung, infolge welcher die Schiffe, die in den Hafen einliefen, nur in demselben Turnus befrachtet werden durften, als sie angekommen waren; das zuletzt eingelaufene Schiff musste also warten, bis alle seine Vorgänger befrachtet waren. Es ward hiermit nicht nur alle Concurrenz ausgeschlossen, sondern auch der freien Bewegung des Verkehrs eine hemmende Fessel angelegt. Erst vor wenigen Jahren ward von der ägyptischen Regierung diesem Unfug der Fazzeh (so hiess dieses Rotationssystem) ein Ende gemacht. Hierdurch hat sich der Schiffsverkehr im Hafen von Suez in den letzten Jahren bedeutend gehoben.

Durch die zahlreichen Dampfschiffe, die jetzt die Linien von Suez nach Bombay, Kalkutta und Australien regelmässig versehen, gewinnt der Hafen sehr an Lebhaftigkeit. Im Gefolge der Dampfschiffahrt finden sich natürlich auch europäische Segelschiffe, meist englischer Flagge, in Suez ein, die mit Kohlen befrachtet aus England kommen und dann gewöhnlich von Dscheddah oder

Tonnen Tragfähigkeit.²⁷⁾ Ein grosser Schritt zur Hebung des Verkehrs ist daher die Einrichtung regelmässiger Fahrten der Dampfer der ägyptischen Megidijeh-Gesellschaft, welche gegenwärtig vier Dampfer im Rothen Meer hat, die zwischen Suez und Dscheddah fahren und auch oft bis nach Kosseir gingen. Leider ist diese Gesellschaft eben jetzt in der Auflösung begriffen. Auch ein reiches alexandrinischer Handlungshaus (Pastré) hält einen Dampfer (Jemen) im Rothen Meer, der unter französischer Flagge zwischen Suez und Dscheddah fährt. Die Dampfer der Megidijeh haben ägyptische Bemannung, waren aber anfangs von österreichischen Kapitänen befehligt, die jedoch jetzt alle den Dienst verlassen haben.

In Suez bestand früher eine sehr drückende Hafenordnung, infolge welcher die Schiffe, die in den Hafen einliefen, nur in demselben Turnus befrachtet werden durften, als sie angekommen waren; das zuletzt eingelaufene Schiff musste also warten, bis alle seine Vorgänger befrachtet waren. Es ward hiermit nicht nur alle Concurrenz ausgeschlossen, sondern auch der freien Bewegung des Verkehrs eine hemmende Fessel angelegt. Erst vor wenigen Jahren ward von der ägyptischen Regierung diesem Unfug der Fazzeh (so hiess dieses Rotationssystem) ein Ende gemacht. Hierdurch hat sich der Schiffsverkehr im Hafen von Suez in den letzten Jahren bedeutend gehoben.

Durch die zahlreichen Dampfschiffe, die jetzt die Linien von Suez nach Bombay, Kalkutta und Australien regelmässig versehen, gewinnt der Hafen sehr an Lebhaftigkeit. Im Gefolge der Dampfschiffahrt finden sich natürlich auch europäische Segelschiffe, meist englischer Flagge, in Suez ein, die mit Kohlen befrachtet aus England kommen und dann gewöhnlich von Dscheddah oder

it. 27) Ein grosser Sc

ist daher die
ampfer der
gegenwärtig

igkeit. 27) Ein grosser S
hrs ist daher die Einricht
r Dampfer der ägyptisch
lche gegenwärtig vier Dan
zwischen Suez und Dsched

Patron/Collection Image Processing

ANALYSIS OF 1973 "ON BOARD" BUS SURVEY

In comparing the results obtained in 1972 and 1973, three major conclusions seem apparent:

First, there appears to be a 30% turnover in bus users as evidenced by the response to question #1 (Did you respond to the bus survey last year? 70.2% yes; 29.8% no). This conclusion is based on the high degree of correlation between 1973 responses and those obtained in 1972.

Secondly, there appears to be greater bus usage by approximately 10% of the bus patrons. This conclusion is derived by subtracting those who did not respond to last year's survey (29.8%) from those who indicated they used the bus more often than last year (43.9%). This conclusion is predicated on the assumption that the difference indicates those persons who were bus patrons in both 1972 and 1973.

Thirdly, the responses indicate increasing utilization of bus service for those persons exceeding the median age of the local population (26.9 years). This conclusion is derived by comparing the relative percent of the total population by age group as enumerated by the 1970 census, and the percent of bus patrons by age group as a function of the total bus patron population respectively.

Characteristics of Bus Riders

The typical bus patron, responding to the 1973 study, has similar characteristics to those of 1972. The overwhelming majority are women (89.0%) and are between the ages of 35-64 years (51.4%) who do not have an automobile available for the trip (83.6%) and who use the bus for transportation to work (48.7%).

Desirable Attributes for Bus Service

Several questions were asked to determine the consistency of responses and to solicit various aspects of those attributes considered desirable in bus service. Eleven of the questions utilized in 1973 were identical to those used in 1972 to provide a basis for comparison.

Two of the primary concerns in 1973 remain the same as those expressed in 1972. Most patrons in 1972 (68.6%) wished half hour service while patrons in 1973 reaffirmed that desire (74.7%). Frequency of service is also a major concern of the bus patron. In 1972, 36.6% would use the bus more often if service was more frequent while in 1973, 29.4% responded affirmatively.

In 1972 the primary criticism of bus service was that buses did not run often enough (25.1%). In 1973 the major concern has shifted to bus service not being on schedule (28.2%). The shift in concern appears significant in that frequency of service as a problem declined by only 2.4% while an inability to maintain schedules doubled (13.8% to 28.2%).

ANALYSIS OF 1973 "ON BOARD" BUS SURVEY

In comparing the results obtained in 1972 and 1973, three major conclusions seem apparent:

First, there appears to be a 30% turnover in bus users as evidenced by the response to question #1 (Did you respond to the bus survey last year? 70.2% yes; 29.8% no). This conclusion is based on the high degree of correlation between 1973 responses and those obtained in 1972.

Secondly, there appears to be greater bus usage by approximately 10% of the bus patrons. This conclusion is derived by subtracting those who did not respond to last year's survey (29.8%) from those who indicated they used the bus more often than last year (43.9%). This conclusion is predicated on the assumption that the difference indicates those persons who were bus patrons in both 1972 and 1973.

Thirdly, the responses indicate increasing utilization of bus service for those persons exceeding the median age of the local population (26.9 years). This conclusion is derived by comparing the relative percent of the total population by age group as enumerated by the 1970 census, and the percent of bus patrons by age group as a function of the total bus patron population respectively.

Characteristics of Bus Riders

The typical bus patron, responding to the 1973 study, has similar characteristics to those of 1972. The overwhelming majority are women (89.0%) and are between the ages of 35-64 years (51.4%) who do not have an automobile available for the trip (83.6%) and who use the bus for transportation to work (48.7%).

Desirable Attributes for Bus Service

Several questions were asked to determine the consistency of responses and to solicit various aspects of those attributes considered desirable in bus service. Eleven of the questions utilized in 1973 were identical to those used in 1972 to provide a basis for comparison.

Two of the primary concerns in 1973 remain the same as those expressed in 1972. Most patrons in 1972 (68.6%) wished half hour service while patrons in 1973 reaffirmed that desire (74.7%). Frequency of service is also a major concern of the bus patron. In 1972, 36.6% would use the bus more often if service was more frequent while in 1973, 29.4% responded affirmatively.

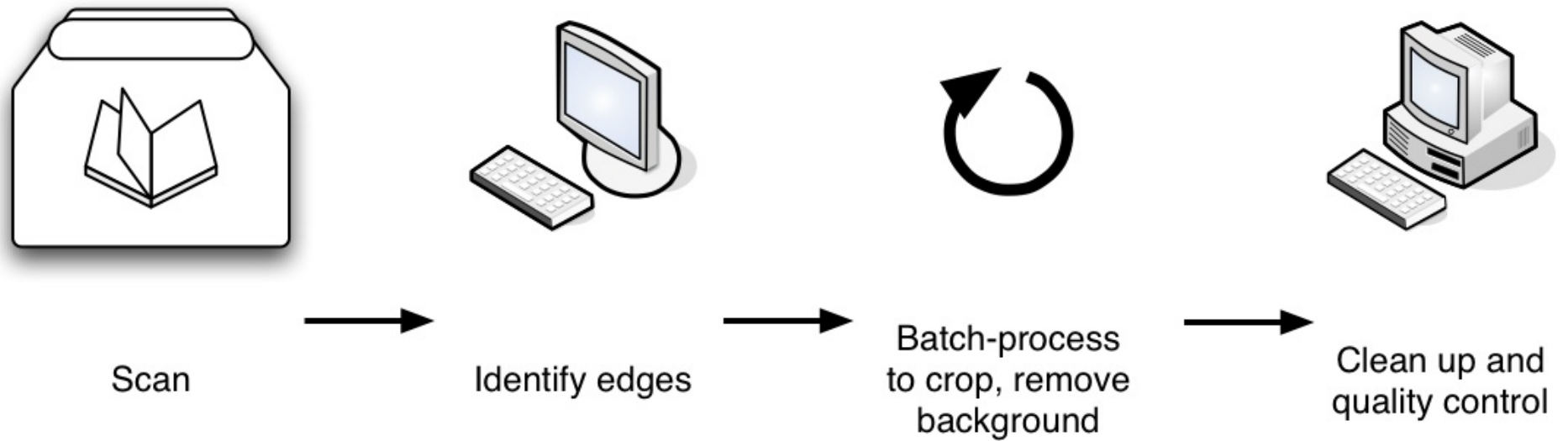
In 1972 the primary criticism of bus service was that buses did not run often enough (25.1%). In 1973 the major concern has shifted to bus service not being on schedule (28.2%). The shift in concern appears significant in that frequency of service as a problem declined by only 2.4% while an inability to maintain schedules doubled (13.8% to 28.2%).

Kirtas @ Northwestern University Library

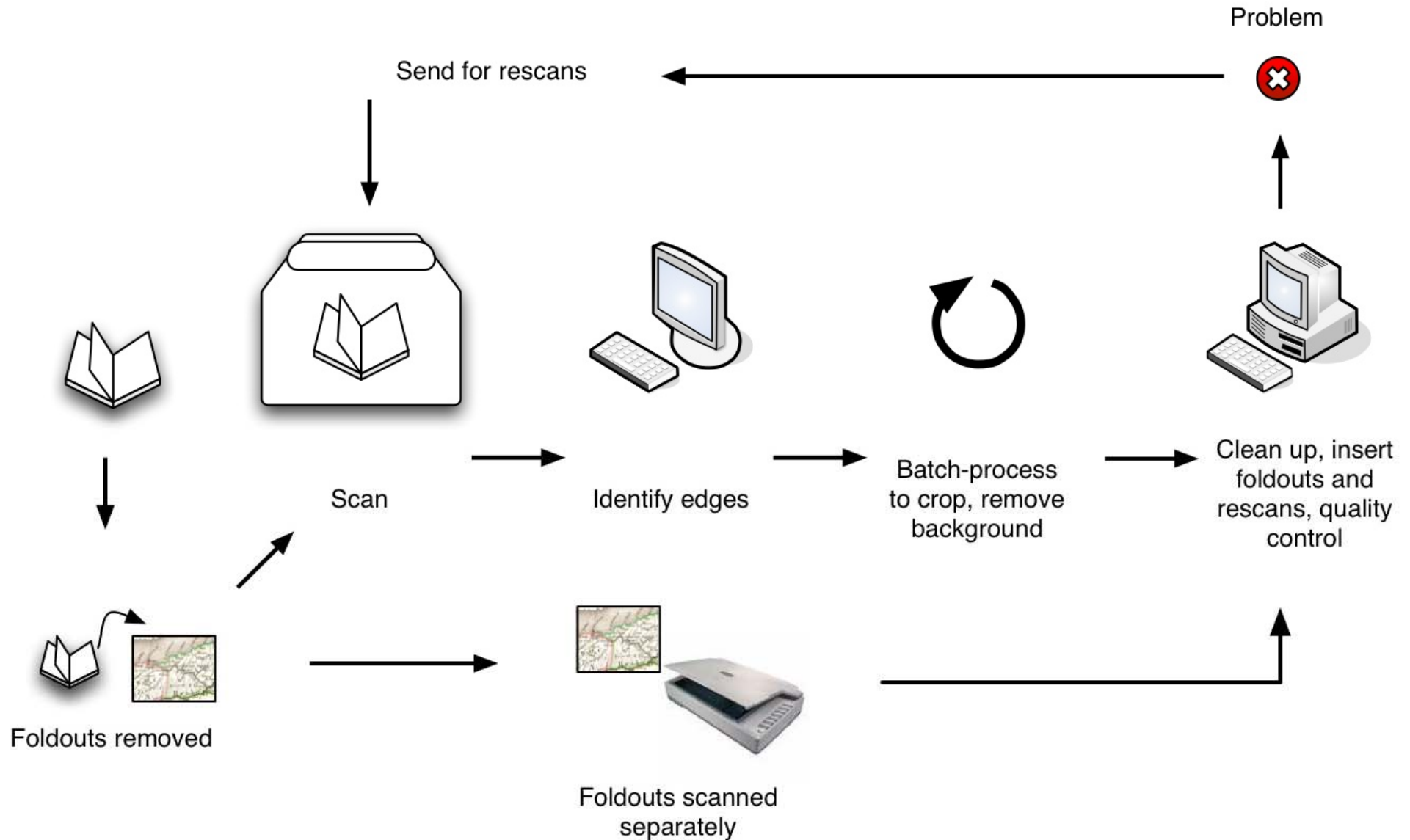
- 2007 Shift away from Brittle reformatting on Kirtas
New Digital Collections department formed
Applied to Andrew W. Mellon Foundation
- 2008 Collection/Patron/Brittle, quarterly queues
New focus on digital-only delivery

Mellon-funded Mounting Books project
- 2009 Book Workflow Interface (BWI)
books.northwestern.edu

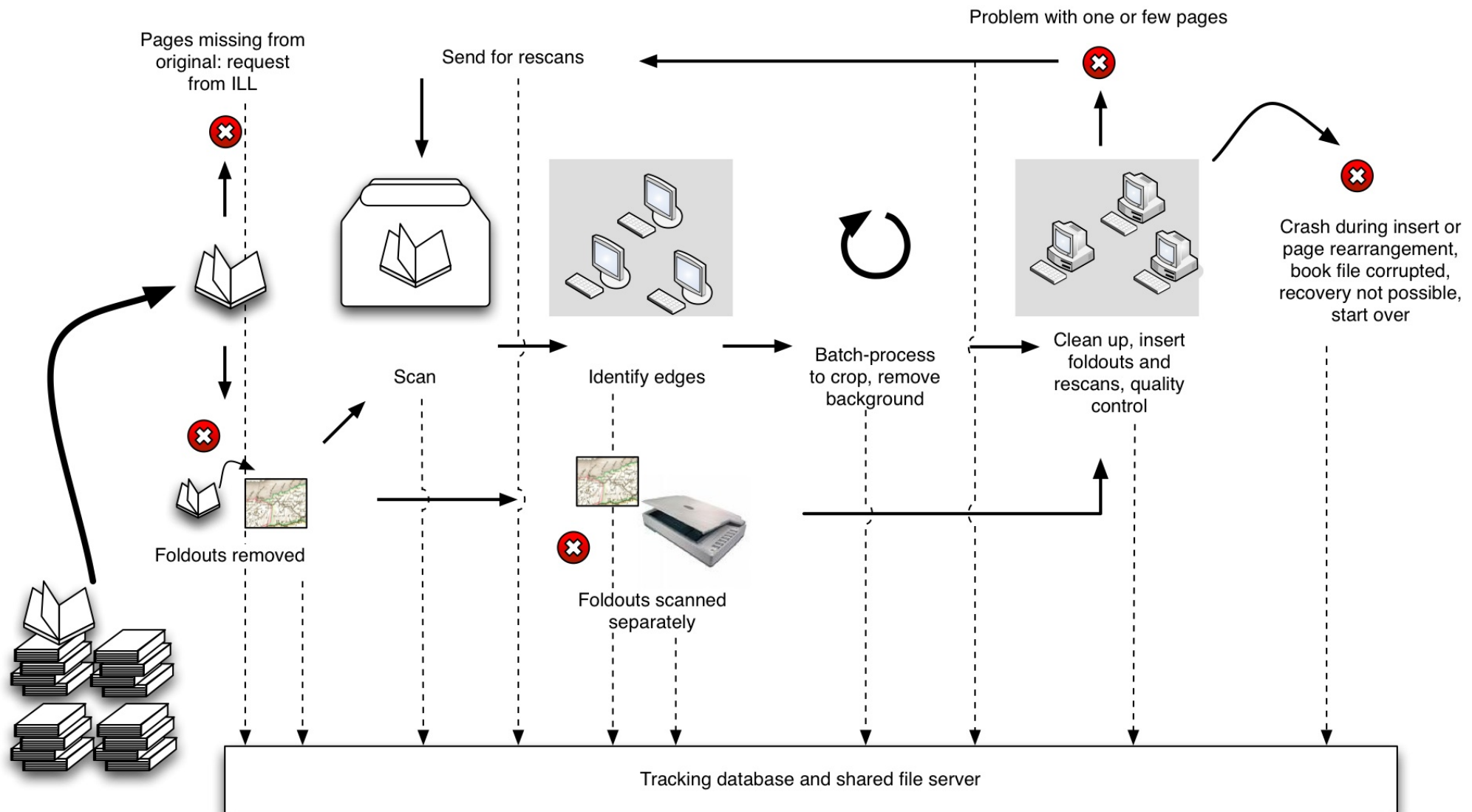
Simple workflow



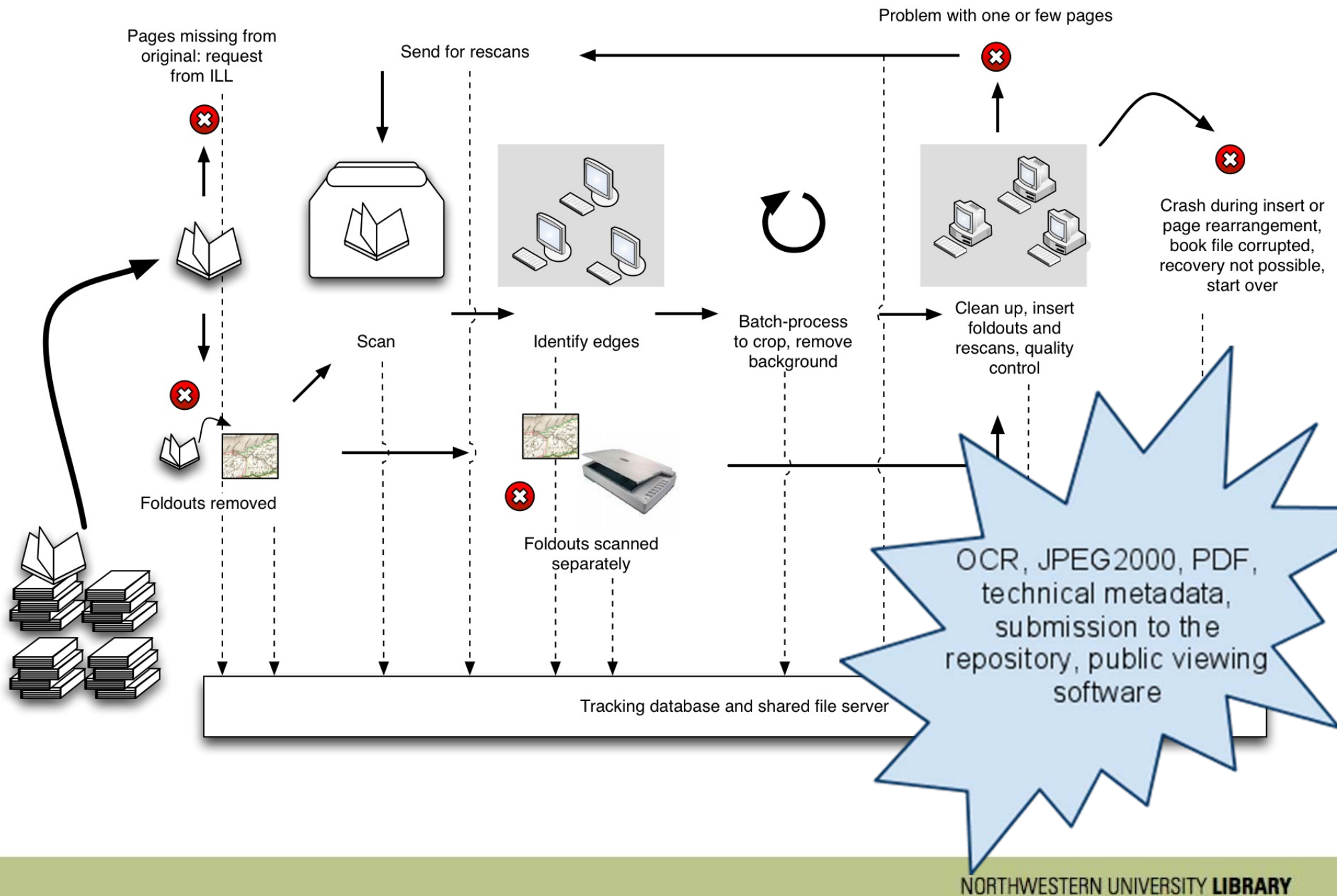
Workflow reality



Workflow reality over time



Workflow reality over time



Scope of the Mounting Books project

- Streamline and extend the book scanning workflow
- Publish digitized books online, link from the library catalog
- Share as many of our software components as possible

Demonstration

Finding books in Voyager

List of in-process book jobs

Book Jobs Admin Log Out									
Book Job Instances for "mountBook" v11									
Action	Bib # ▾	Title ▾	Vol # ▾	Priority ▾	Assigned ▾	Status ▾	Schedule ▾	# Imgs ▾	Project(s) ▾
Filter: On / Off	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
 Open Job	0005328948	Rural science for standards III & IV : new syllabus / by G.P. Chidia.	n/a	normal	n/a	Scan	3	0	3
 Open Job	0005328580	The fugitive and other literary works in prose and poetry / of G.M. Woodward.	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005328526	Conforto de gli afflitti condannati a? morte / del dottor collegiato Gio. Battista Gargiaria. Con l'indice delle cose piu? singolari e notabili.	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005328447	La crise des transports : illusions & re?alite?s	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005313912	Proceedings ... annual convention. [electronic resource]	9	normal	n/a	Scan [Error]	3	0	2
 Open Job	0005313912	Proceedings ... annual convention	8	normal	n/a	Touchup	3	0	3
 Open Job	0005313763	Lis amouro de ribas / culido pe?r la Felibresso do?u Cauloun.	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005313570	Informacio?n de estado	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005313543	Chicago traction : a history, legislative and political / by Samuel Wilber Norton.	n/a	normal	n/a	Book Builder	2	258	3
 Open Job	0005313540	The new Slovakia / by R. W. Seton-Watson.	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005313347	ha-Eshkol. [electronic resource]	1	normal	n/a	Touchup	3	0	3
 Open Job	0005313347	ha-Eshkol	3	normal	n/a	Touchup	3	0	3
 Open Job	0005313247	Panama : a personal record of forty-six years, 1861-1907 / by Tracy Robinson.	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0005313053	Knaves of diamonds : being tales of mine and veld / Illustrated by E. F. Sherie and T. A. Ayton Symington.	n/a	normal	n/a	Book Builder	3	290	3
 Open Job	0005313034	The social evil in Chicago : a study of existing conditions with recommendations	n/a	normal	n/a	Touchup	3	0	3
 Open Job	0001132936	Knaves of diamonds : being tales of mine and veld / Illustrated by E. F. Sherie and T. A. Ayton Symington.	n/a	normal	n/a	Batch [Error]	3	n/a	3
Start a new book job									

Book job details

Book Jobs » 0005313543 (Job #30) Admin | Log Out

StatusBook DetailsBook Job DiagramProcess Vars

Bib #: 0005313543 *

Barcode:

OCLC: (OCoLC)ocm00271661

Volume:

Title: Chicago traction : a history, legislati *

Author: Author: Norton, Samuel Wilber. *

Date: 1907. *

Location: TRANSPORTATION Library *

Call Number: HE 4491.C4 N888 *

Schedule: 2009 Winter *

Projects: Collection *

Patron

Brittle

Priority: Normal *

Job Type: Kirtas ☒ Vendor ☐

VisibilityPublic ☐ Private ☒

Has Foldouts: ☐

Requester's contact information:

Name: Brittle *

Email: preservation@northwestern.edu *

Close job

The following Notification and Confirmation fields are for future use.

Requester notified date:

Notification method:

Notified by:

Book routed date:

Confirmed receipt by:

Confirmation date:

Copyright

The following Access and Copyright fields are for future use.

Access:

World: ☐

U.S. only: ☐

Northwestern only: ☐

Custom (select all that apply): ☐

Time-limited link: ☐

Northwestern group(s): ☐

Northwestern user(s): ☐

Copyright status:

Copyrighted: ☐

Public domain: ☐

Unknown: ☐

Notes and description of copyright analysis:

Book Structure

Root

Save Book Details

Suspend this book job

End this book job

Delete this book job

NORTHWESTERN UNIVERSITY LIBRARY

Status message: book files moving

[Book Jobs](#) » [0000490771 \(Job #29\)](#)

Status Book Details Book Job Diagram Process Vars

This book job is currently at the following state:
MoveProcessingFiles and the state **is running**.
Please refresh to see the most current status.

Suspend this book job

Add a note

Add note

The Queue Server

The Queue Server is notified of jobs that need to be processed, and queues them up to avoid overloading the server.

Each set of images has the following processes:

- Conversion to JPEG2000 using Aware software.
- MIX metadata generated with JHOVE
- Exif extraction from original camera images
- Three OCR derivatives are extracted (image plus text PDF, plain text, and XML) using ABBYY OCR software.
- The Queue Server also triggers FOXML generation and ingestion into Fedora.

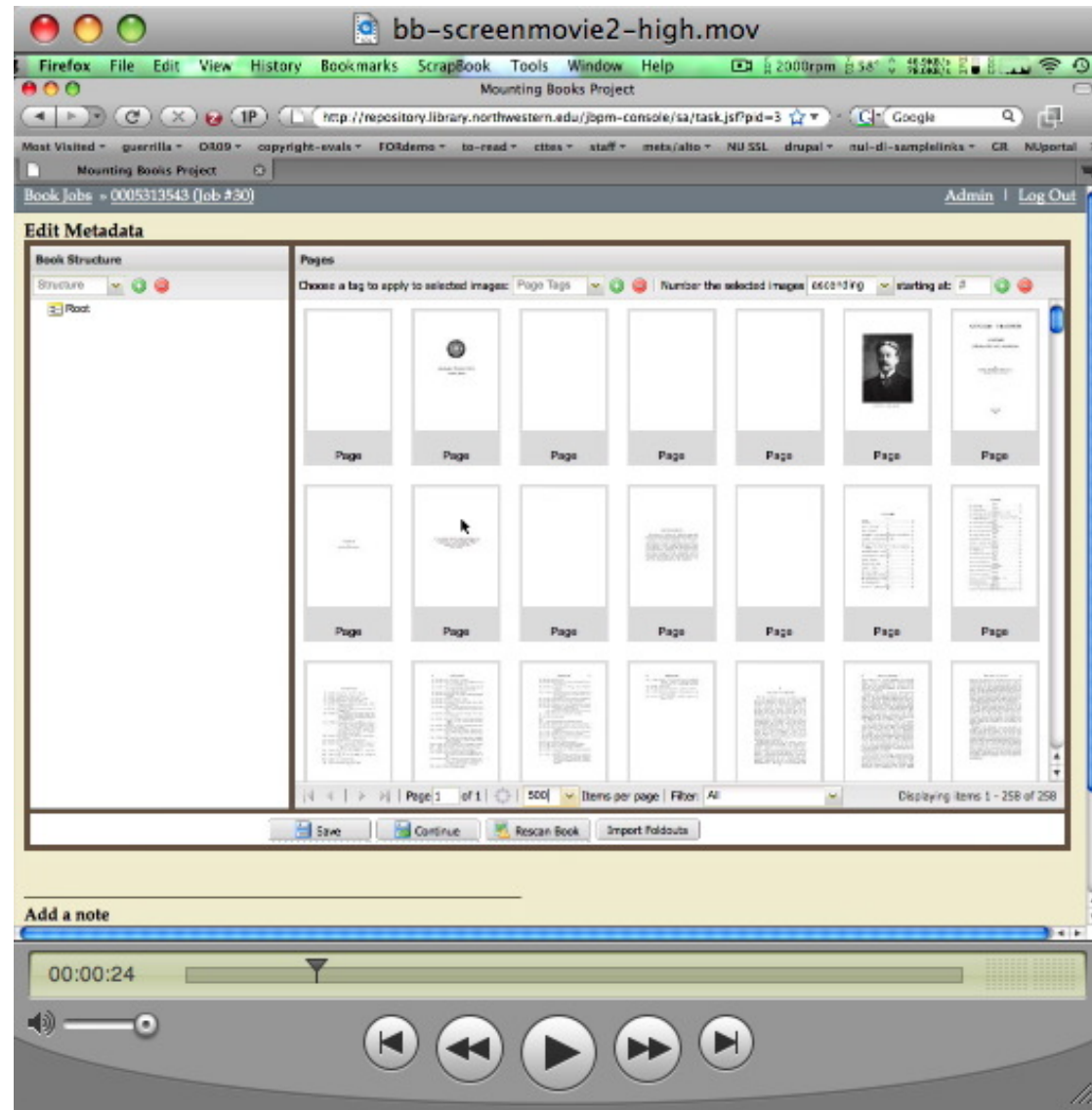
The Book Builder

As part of the workflow, operators create and enhance structure for digitized books.

Features of the Book Builder:

- Operators can create chapters and other subdivisions.
- Each part of the structure (such as a chapter) can be given a number and a label.
- Automatic page numbering in descending or ascending order. Optionally, operators can apply their own custom page numbers.
- Application of informational tags indicating the presence of other elements such as figures and blank pages. These tags will be useful in the future for scholarly analysis and collection maintenance.
- Information about the structure is stored as METS.

Book Builder demonstration



Book approval

- Generate a Handle that links to book in public interface
- Update Voyager (Library catalog) with 856 URL to Handle
- Update the Search index
- Update PREMIS (future feature)

Fedora Datastreams and Disseminators

Datastreams are either pieces of content, or pointers to external content that are stored within one Fedora object.

For example, a datastream may be a JPEG image, a Dublin Core XML file, or a redirect to a streaming movie on a server.

Disseminators are services that can provide a different view of the stored data.

For instance, a dissemination may include the ability to retrieve a different size of an image that is stored. The service takes the stored image, resizes it, and sends it to the user. The original image is not modified.

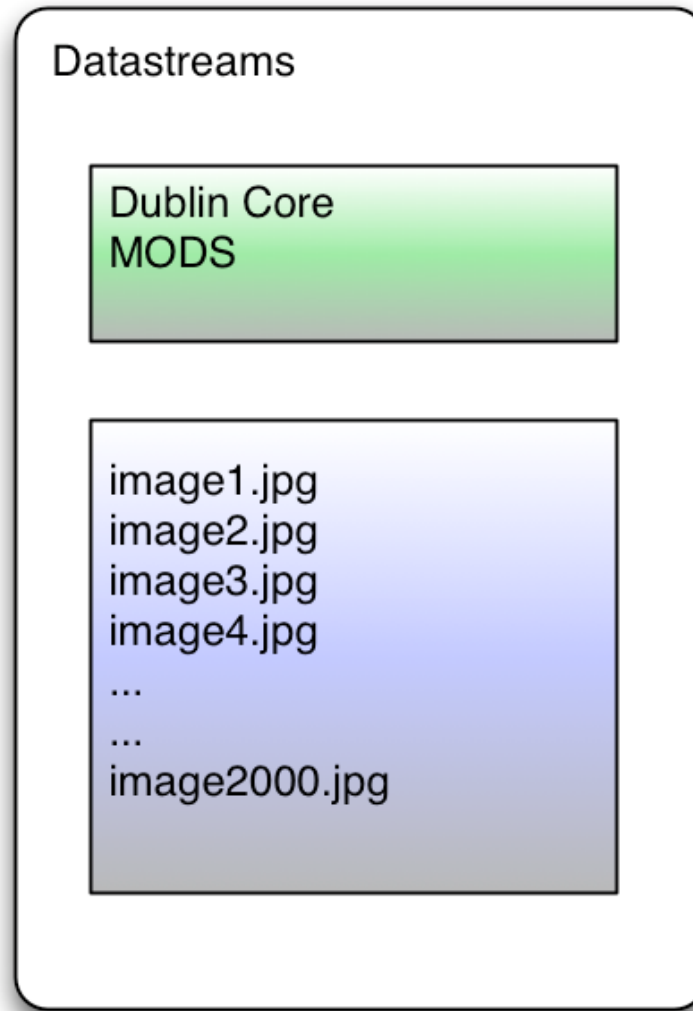
Atomistic vs. Compound Approach

Compound objects:

- Include many datastreams within one object
- Have information about the whole and its parts
- While the information is stored in one big object, it's harder to take advantage of Fedora's object model
- Objects can be as easily reused

Atomistic vs. Compound Approach

An example Compound object (a book object):



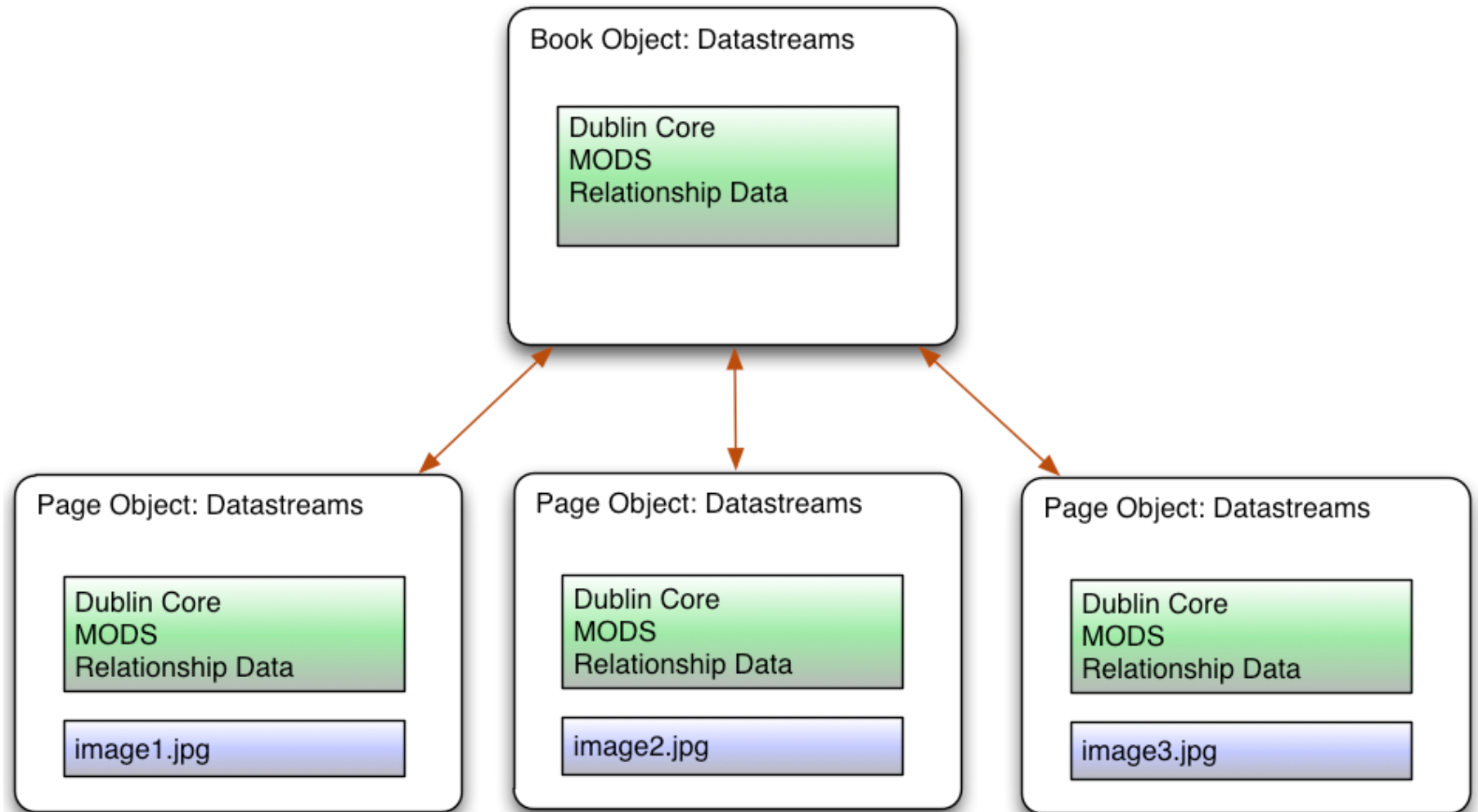
Atomistic vs. Compound Approach

Atomistic objects:

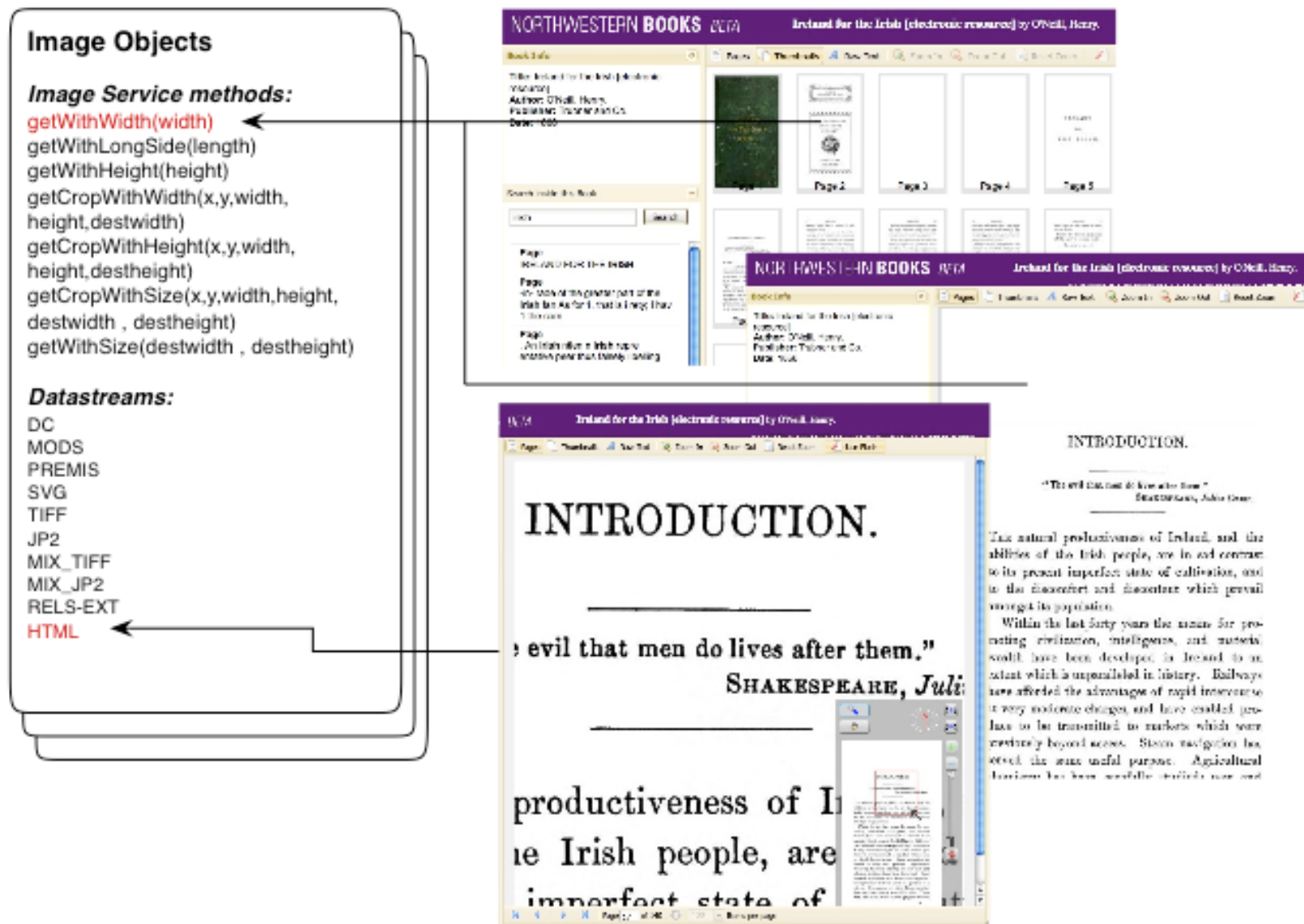
- Have a small number of datastreams per object
- Relationships are stored so the objects can be recognized as one entity.
- Many objects may be needed, but this takes best advantage of Fedora's object model.
- Best for object reuse; for example, all image objects in the repository can share the same dissemination services.

Atomistic vs. Compound Approach

Example Atomistic model (a book object and related page objects):



Each page object shares the same basic access model
(In fact, all of our Fedora image objects share the same dissemination access model)



Page Imaged Book Model

Datastream LABEL	ID	MIMETYPE	Schema
Dublin Core	DC	text/xml	OAI_DC
MODS	MODS	text/xml	MODS
MARC	MARC	text/xml	MARC XML
XACML Policy file	POLICY	text/xml	XACML
Kirtas job tar file	KIRTAS_(1,2,3,...)	application/x-tar	KIRTAS JOB FILES
Tagged book Structure	BOOKSTRUCT	text/xml	METS
Relationships	RELS-EXT	text/xml	RELS-EXT
HTML viewer	HTML	text/html	XHTML

Page Image Model

Datastream LABEL	ID	MIMETYPE	Schema
Dublin Core	DC	text/xml	OAI_DC
MODS record	MODS	text/xml	MODS
Archival image format	ARCHV-IMG	image/jpeg	
Archival MIX	ARCHV-TECHMD	text/xml	MIX
Archival Exif Metadata	ARCHV-EXIF	text/xml	Exif as XML
Corrected Image format	PROC-IMG	image/jpeg	
Corrected MIX	PROC-TECHMD	text/xml	MIX
Delivery image format	DELIV-IMG	image/jp2	
Delivery MIX	DELIV-TECHMD	text/xml	MIX
Delivery metadata	DELIV-OPS	text/xml	SVG
Digital Provenance	PROVENANCE	text/xml	PREMIS
XACML Policy file	POLICY	text/xml	XACML
Relationships	RELS-EXT	text/xml	RELS-EXT
OCR in XML	OCR_XML	text/xml	Abbey OCR
OCR in PDF	OCR_PDF	application/pdf	
OCR in plain text	OCR_TEXT	text/plain	
HTML viewer page	HTML	text/html	XHTML

Web-enabled services

Web-enabled services allow us to more easily use software components for other projects. The web-enabled services used for this project include:

- Voyager (Integrated Library System) updater
- Handle persistent URL creator
- Queue server
- Yaz Proxy Z39.50 server
- SOLR search service
- Fedora digital repository

Other Technologies used in the project

- **METS** - used to store the book structure
- **Ext-JS** - JavaScript library, includes interface and AJAX
- **JSON** - Used to communicate between JavaScript and server-side Java
- **ABBYY** - OCR technology
- **Aware** - Generates and serves up derivatives from JPEG2000 files
- **JHove** - For generating MIX data from images

Book Info

Title: History of the Roman empire.
[electronic resource]
Author: Pococke, Edward
Publisher: J.J. Griffin and Co.; [etc., etc.]
Date: 1853

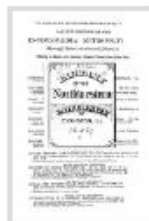
Search inside this Book

Explore this Book Link to this Book Download/Print this Book

Pages Thumbnails Raw Text Zoom In Zoom Out Reset Zoom Use Flash



Page 1



Page 2



Page 3



Page 4



Page 5



Page 6



Page 7



Page 8



Page 9



Page 10



Page 11



Page 12



Page 13



Page 14



Page 15



Page 16



Page 17



Page 18



Page 19



Page 20



Page 21



Page 22



Page 23



Page 24



Page 25



Page 26



Page 27



Page 28



Page 1 of 7 100 Items per page

"we" =

Northwestern University Library

Administration, Bibliographic Services, Digital Collections, Library Technology, Preservation, Business & Finance, Special Collections and University Archives, Africana, Science and Engineering Library, Transportation, GovInfo, Music Library, Circulation ... and many more

Northwestern University Academic & Research Technologies
Flash viewer development

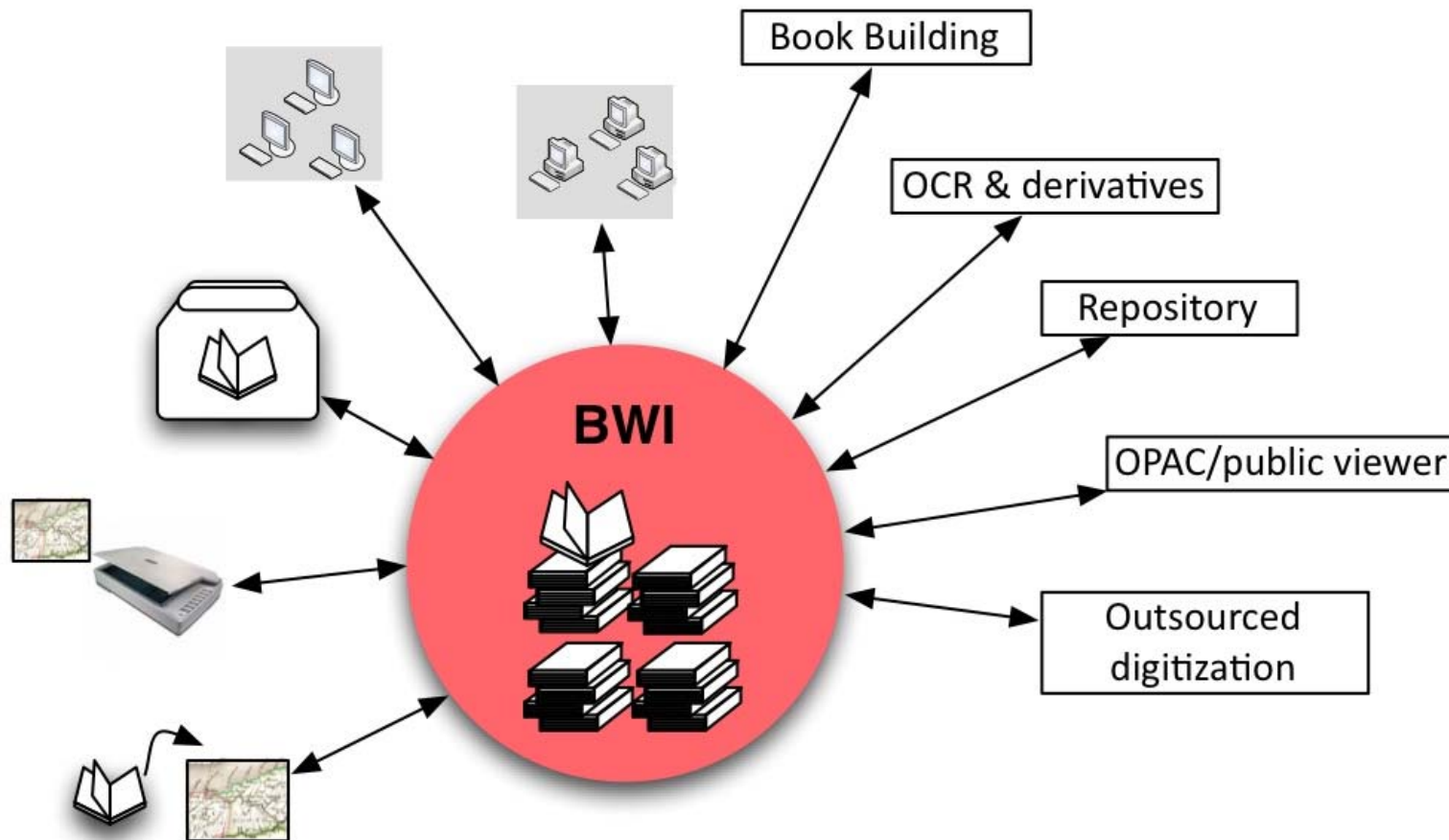
OpenSky Solutions

James Chartrand

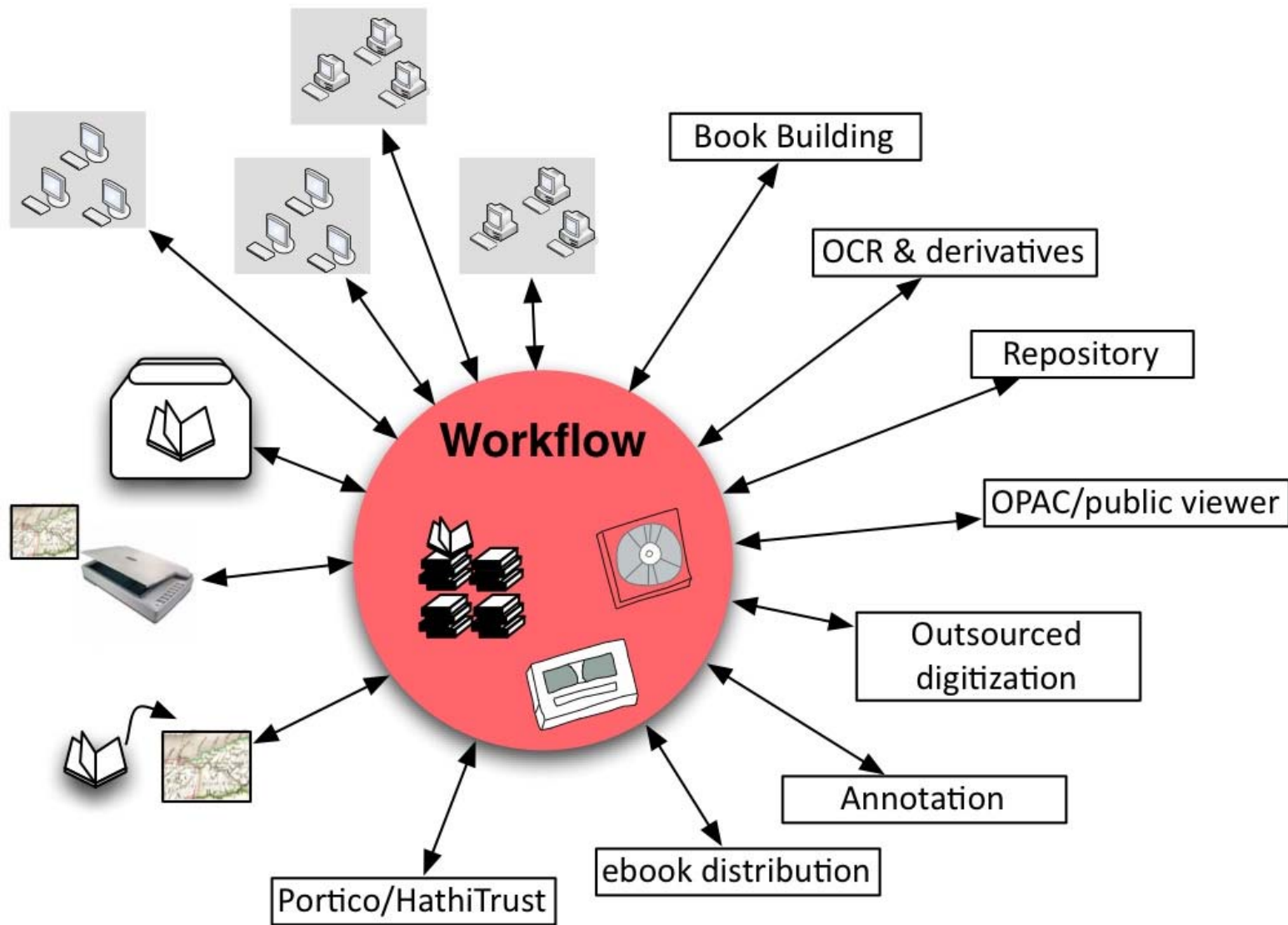
Project funded by the Andrew W. Mellon Foundation

Thank you

where we are now...



future growth...



What next?

- PDF for download and printing, OCR tweaks
- Minor fixes to metadata display, OPAC linking
- Fine-tuning Book Builder navigation
- Northwestern Books beta launch early April
- Mounting Books software available June 30

then...

ebook support, print-on-demand, expanding access to Book Builder, METS/ALTO & TEI, integration with Portico, Google content/HathiTrust?