

From Trickle to Flood: An Extensible Approach to Large-Scale Digitization of Manuscript Collections at UNC Chapel Hill Libraries

Laura Clark Brown, Coordinator of the Digital Southern Historical Collection
 Timothy Shearer, Head Application Development Team, Library Systems

Southern Historical Collection in The Wilson Special Collections Library:
 16 million items in 4,600 archival collections
 Papers, letters, photographs, etc.

Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections
 Mellon grant (2007-2009)

- Year 1: Determine interest in digitization of manuscript materials
- Year 2: Explore sustainable models for prioritization and workflow

Grant Outcomes

- Researchers see immense value in manuscript digitization (surprise surprise)
- Digitize everything, not selected items
 - At least at the container level
 - “show me everything”
- Discovery via the finding aid
- Digitization priorities (matrix):
 - Material condition
 - Research value
 - Convenience
 - Etc.
- Also influencing priorities:
 - Digitization on demand program
 - Special digitization projects (grants)

Goals

- Create sustainable workflows to digitize manuscript materials
- Provide a robust and extensible system that supports discovery of and access to digitized items

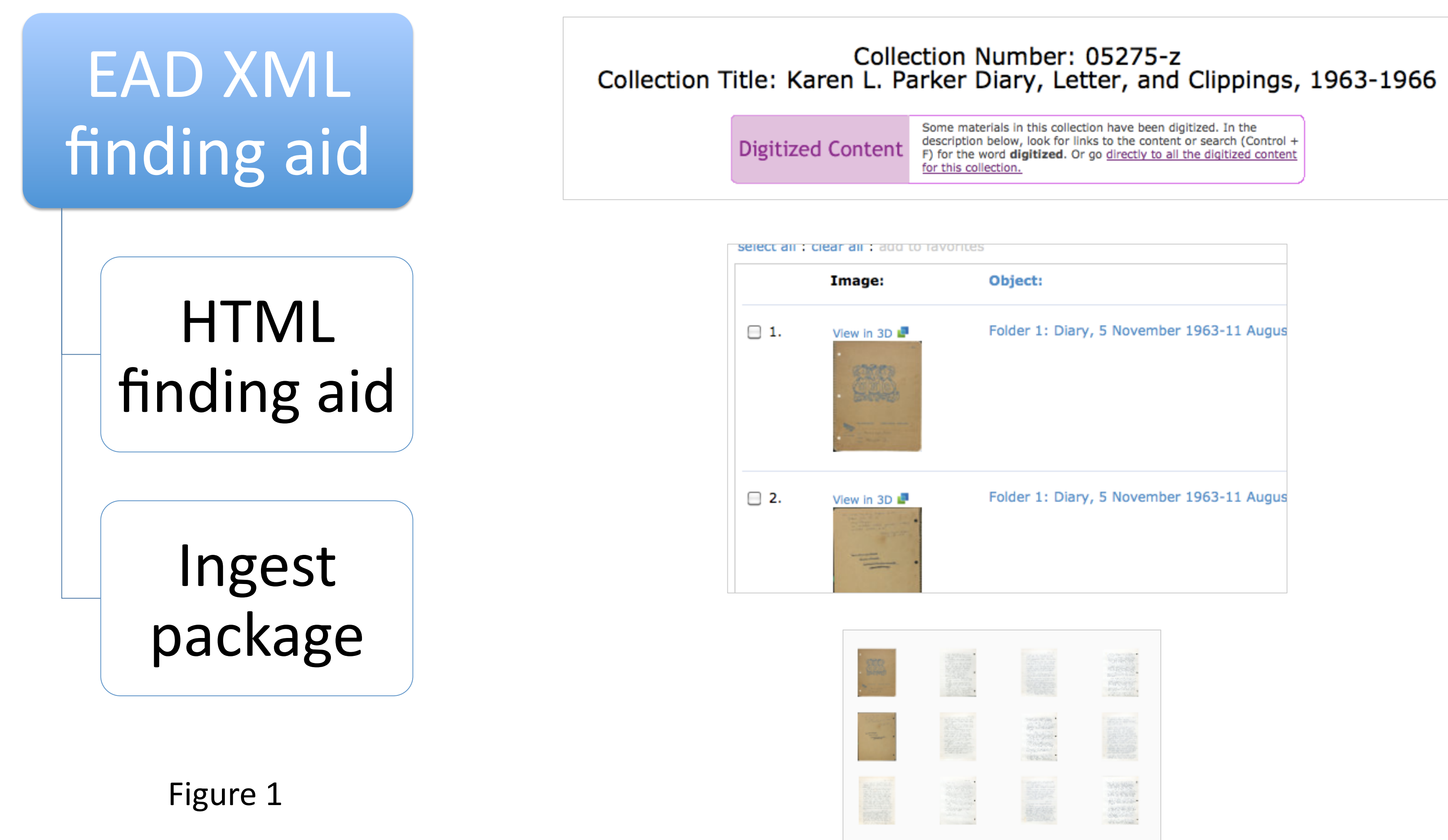


Figure 1

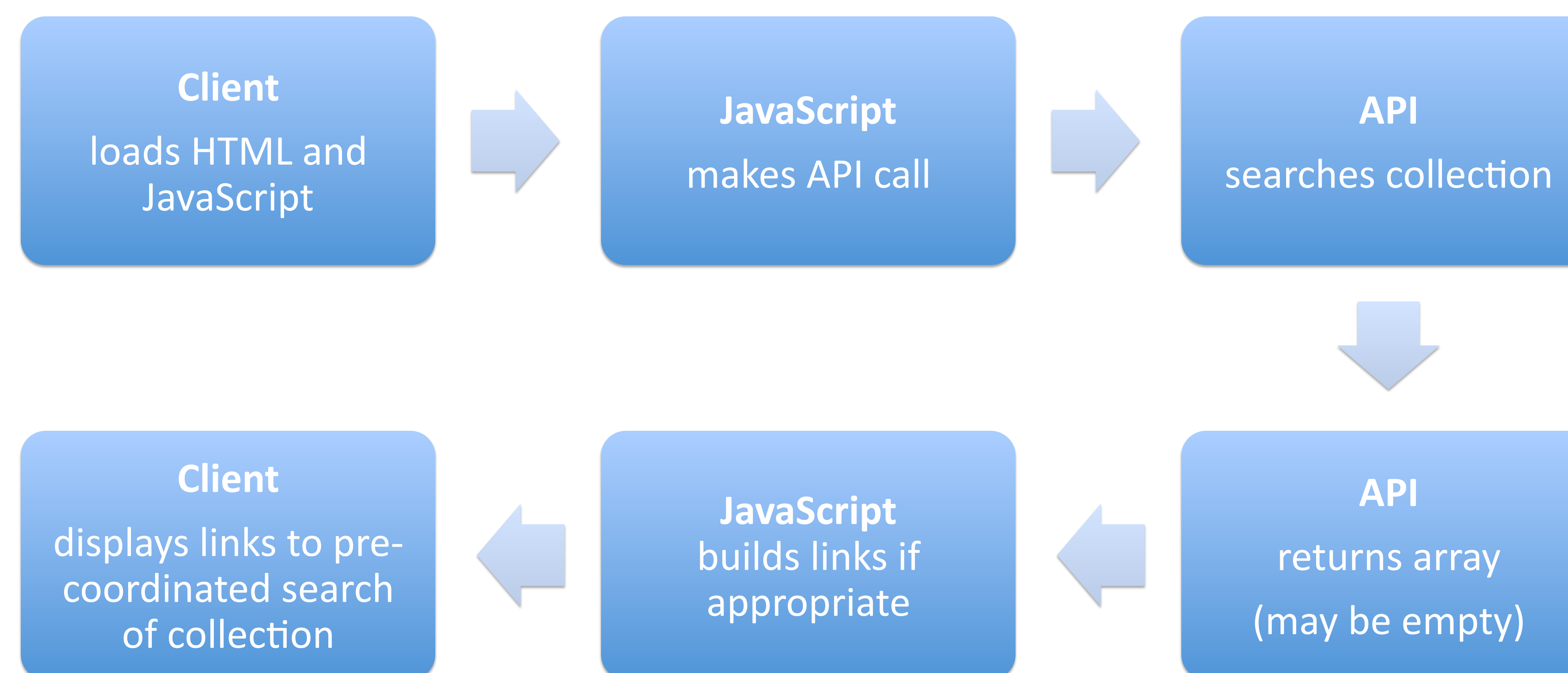


Figure 2

Constraints

- Discovery via HTML versions of EAD finding aids
- Minimal metadata for scanned items
- CONTENTdm as digital object repository
- Must support variable rates of ingest based on available resources
- Digitization and ingest must be accomplished by graduate students who can work independently with minimal training
- Should work transparently with little or no management or maintenance

Possibilities

- Very consistent metadata
- Tight workflow and policies from technical services staff
- CONTENTdm is easy to use for ingest
- Unique IDs for collections and containers within collections

Solution

- HTML finding aids and ingest packages built from XSL transforms of base file (figure 1)
- Both contain unique identifiers
- API created to query CONTENTdm collections and return results
- JavaScript added to every HTML finding aid.
- AJAX query for content and create links if appropriate (figure 2)

Growth

- Began ingest in December 2009
- Went live January 14, 2010
- Contains ~110,000 objects as of October 2010
- 200 collections currently represented
- ~8,500 scans/month