

Xubmit:

A system for managing text collections

*Dazhi Jiao, Tamara Lopez, and Jenn Riley
Indiana University Digital Library Program
DLF 2006 Spring Forum: Austin, Texas
April 10-12, 2006*

Overview

- IU Environment
- Case Studies
 - EAD *IU Finding Aids*
 - TEI *The Chymistry of Isaac Newton*
- Analysis
 - Process
 - Problems
- Xubmit
 - Technology overview
 - Demonstration
- Conclusions

IU Environment

- Digital Library Program provides a central home for library and faculty digital library activities
 - Consulting
 - Content creation
 - Metadata creation
 - Delivery
- Enormous diversity of projects
- Currently undergoing a complete redesign of our digital library software infrastructure

Case Study - EAD

IU Finding Aids

- Encoded by multiple repositories at IU
- Encoding done in <oXygen />
- XML template created and maintained by EAD Working Group
- Completed files delivered to DLP for online delivery
- Currently delivered via DLXS; soon to be delivered via XTF
- Tested RLG's EAD Report Card for improving encoding workflow

Case Study - TEI

The Chymistry of Isaac Newton

multi-stage, iterative

- transcription
- initial encoding
- XSLT transformation
- editing and proofing
- possible modification of encoding practice(s) and/or XSLT

repeat steps 3-5 until editors are satisfied.

Analysis - Process

Creation

- encoding
- transcription
- one person
- asynchronous

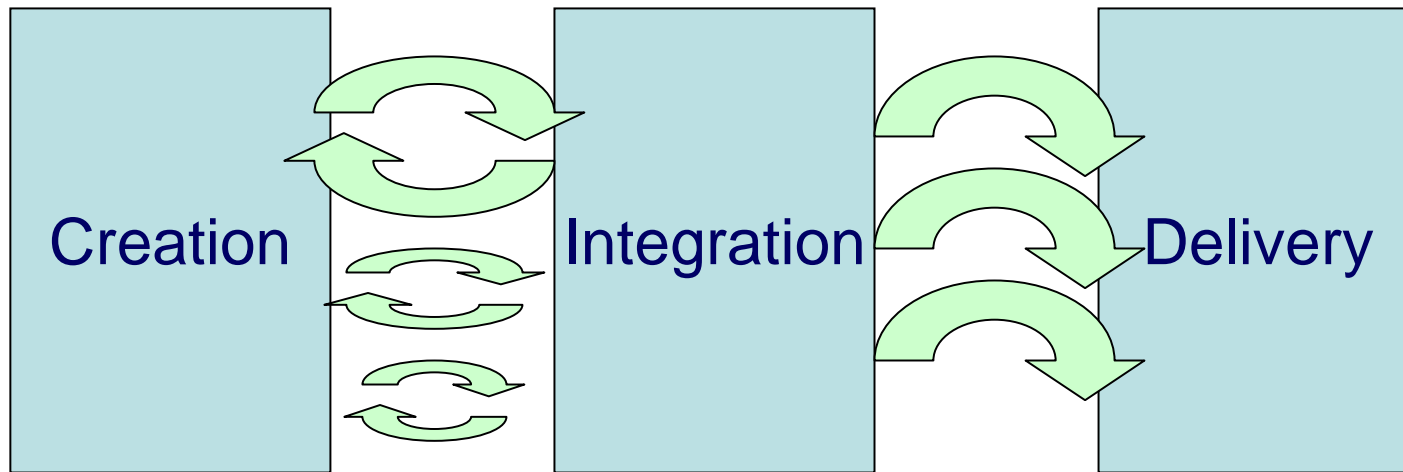
Integration

- validation
- preservation
- multiple people
- asynchronous

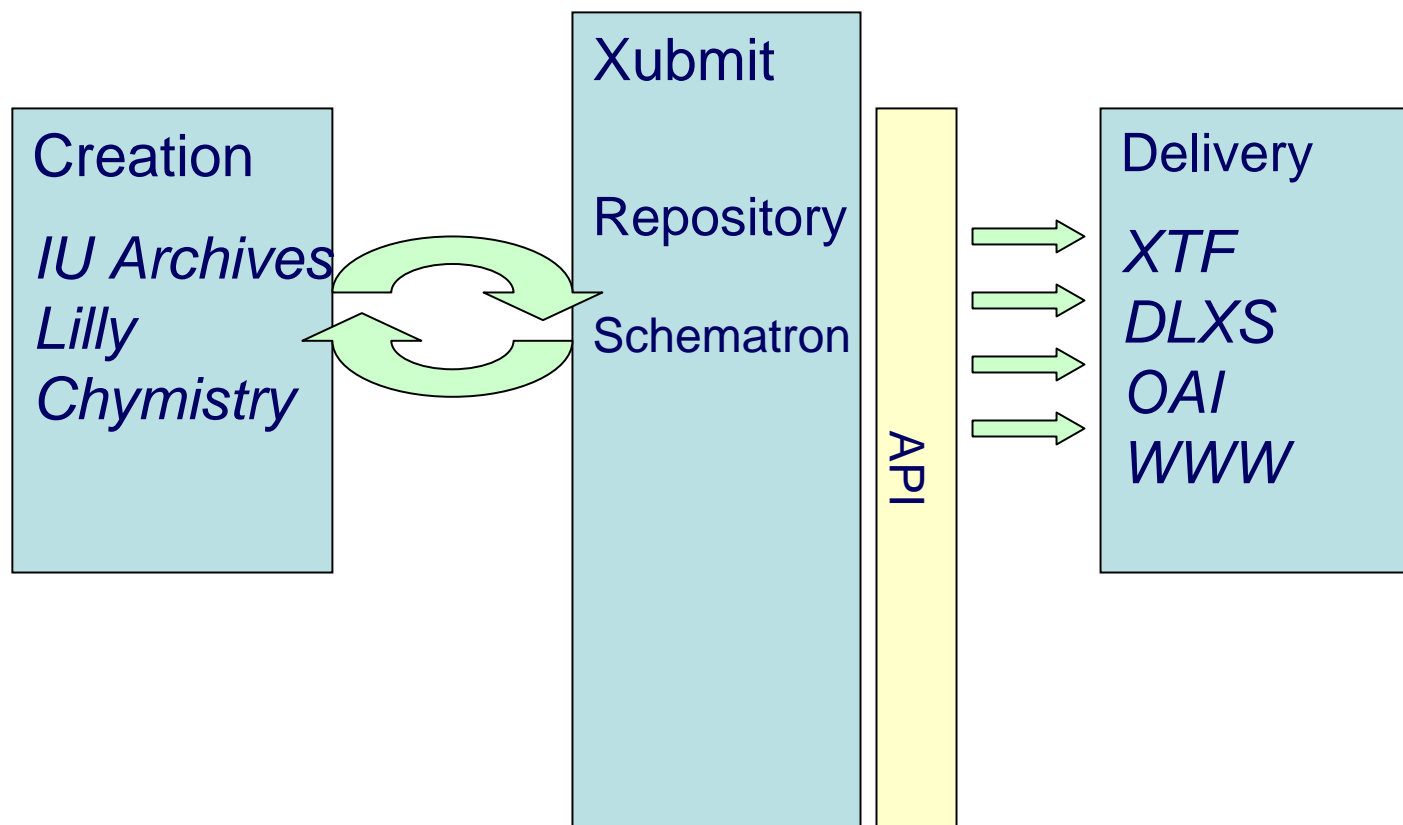
Delivery

- transformation
- access
- multiple systems
- asynchronous

Analysis - Process



Analysis - Architecture

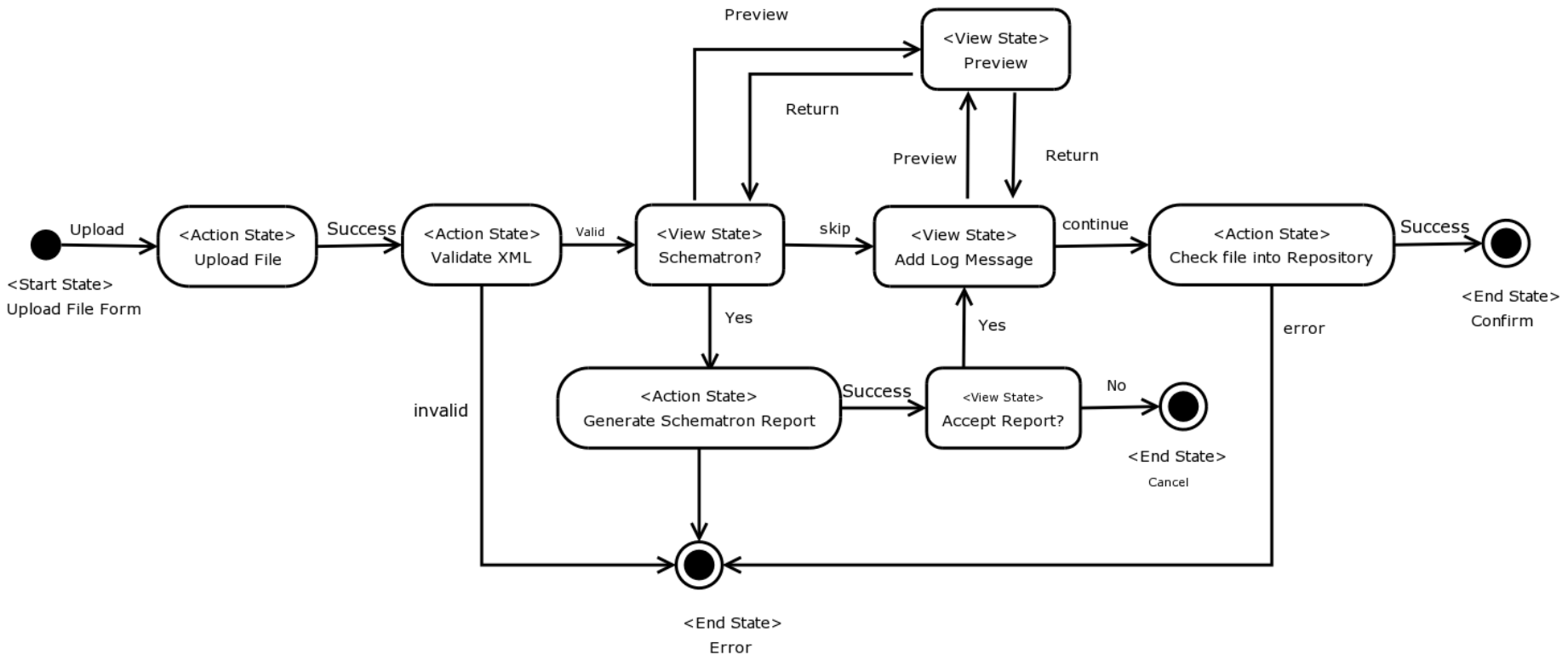


Xubmit :: Implementation

- Submission Flow
 - Spring Webflow
- Repository
 - JRCS, Platform Independent
- XML Validation / Transformation
 - Xerces, Saxon, XML Catalogs
- Web application
 - Struts, JSP, Javascript

Xubmit :: Submission Flow

- Flow UML Diagram



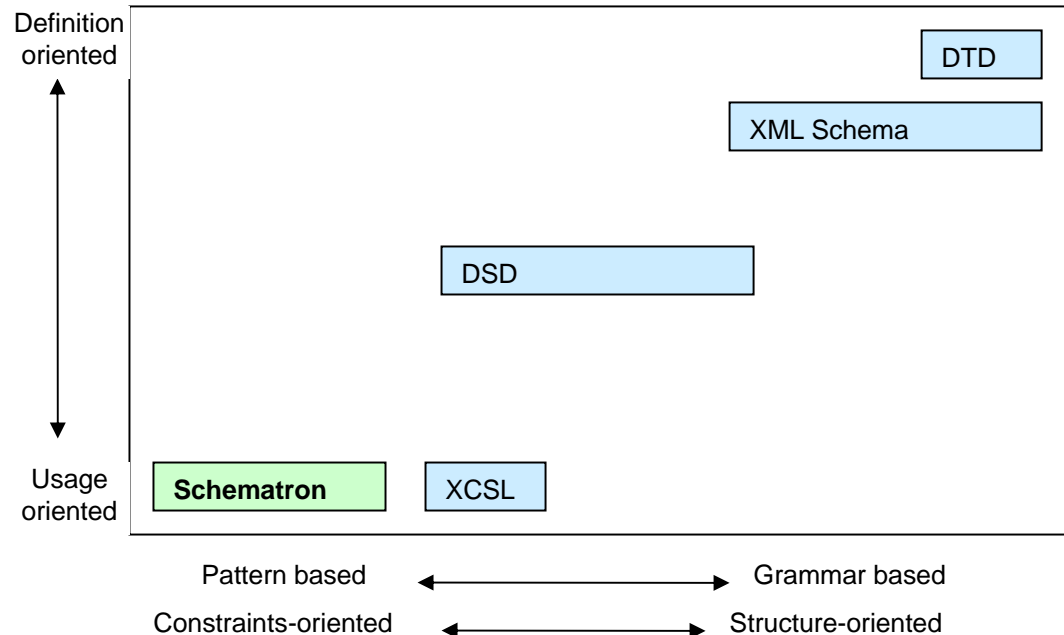
Xubmit :: Schematron

Pros

- *Simple vocabulary*
- *Allows for robust reporting*
- *Uses standards: XPATH, XSLT*
- *Becoming an ISO standard*

Cons

- *Verbose*
- *Maintenance*
- *No support for variables*
- *Not an ISO standard yet.*



Source: Adapted from Lee, D. & Chu, W. (2000),
Jacinto, M. et. al. (2002)

Xubmit :: Demonstration



Resources and References

Resources

- JRCS: <http://www.suigeneris.org/kb/display/jrcs/home>
- Saxon: <http://www.saxonica.com>
- Schematron: <http://www.schematron.com/>
- Spring Webflow: <http://opensource.atlassian.com/confluence/spring/display/WEBFLOW/Home>
- Xerxes: <http://xerces.apache.org/xerces-j/>
- XML Catalog: <http://www.oasis-open.org/committees/download.php/14809/xml-catalogs.html>

References

- Baecker, R. M, Nastos, D., Posner, I.R. & Mawby, K.L.(1993). The user-centered iterative design of collaborative writing software. Proceedings of the SIGCHI conference on Human factors in computing systems. Amsterdam, The Netherlands, 1993. pp. 399 - 405
- Dillon, A. (1993) How collaborative is collaborative writing? An analysis of the production of two technical reports. Published in M.Sharples (Ed) Computer Supported Collaborative Writing, London: Springer-Verlag, pp 69-86. Retrieved April, 2006 from: <http://www.ischool.utexas.edu/~adillon/BookChapters/collaborative%20writing.htm>
- Jacinto, M., Librelotto, G. R., Ramalho, J.C.L & Henriques, P.R. (2002). Constraint specification languages: comparing XCSL, Schematron and XML-Schemas. XML Europe, 2002. Retrieved April, 2006 from: http://www.idealliance.org/papers/xml02/dx_xml02/papers/03-03-02/03-03-02.pdf
- Jelliffe, R. (2000). Getting Information into markup: the data model behind the schematron assertion language. A technical whitepaper for GeoTempo, Inc. October 19, 2000. Retrieved April, 2006 from: <http://www.sinica.edu.tw/~ricko/schematron.PDF>
- Kim, H. & Eklundh, K. (2001). Reviewing Practices in Collaborative Writing. Computer Supported Cooperative Work. 10: 247-259, 2001.
- Lee, D. & Chu, W. (2000). Comparative Analysis of Six XML Schema Languages1. ACM SIGMOD Record 29(3), September 2000.Retrieved April, 2006 from: <http://cobase-www.cs.ucla.edu/tech-docs/dongwon/sigmod-record-00.html>