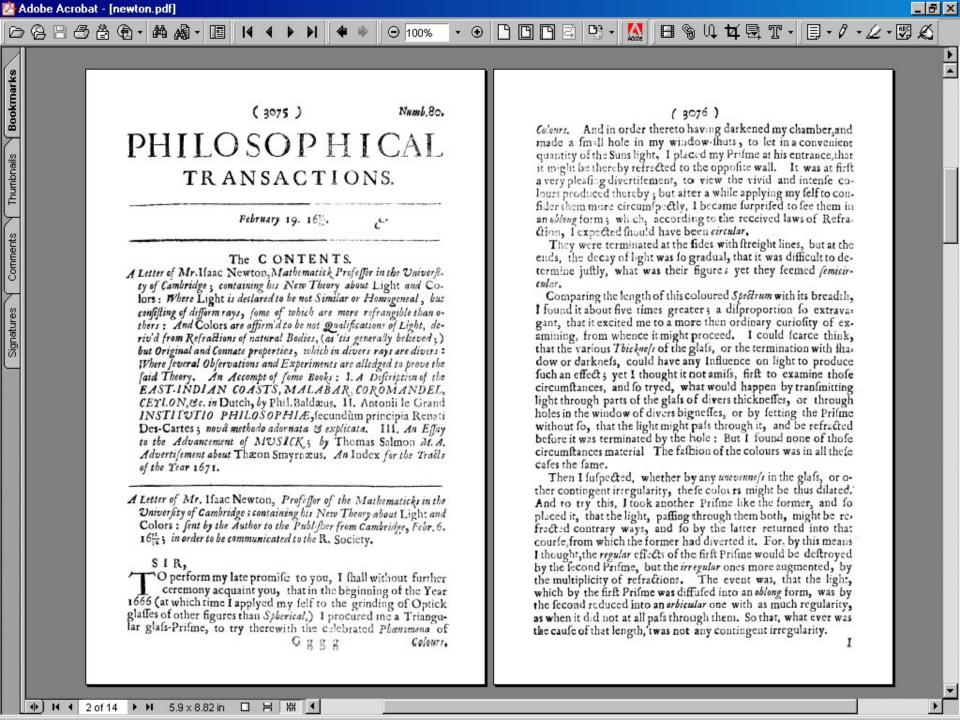


## Michael Krot, Data Manager and David Yakimischak, CTO

krot@jstor.org, davidyak@jstor.org http://www.jstor.org







## JSTOR Mission

- JSTOR is a not-for-profit organization with a mission to help the scholarly community take advantage of the advances in information technology. This includes: (1) building a reliable and comprehensive **archive** of core scholarly journals, and (2) dramatically improve **access** to this scholarly material
- In pursuing its mission, JSTOR takes a systemwide perspective, seeking benefits for libraries, publishers and scholars



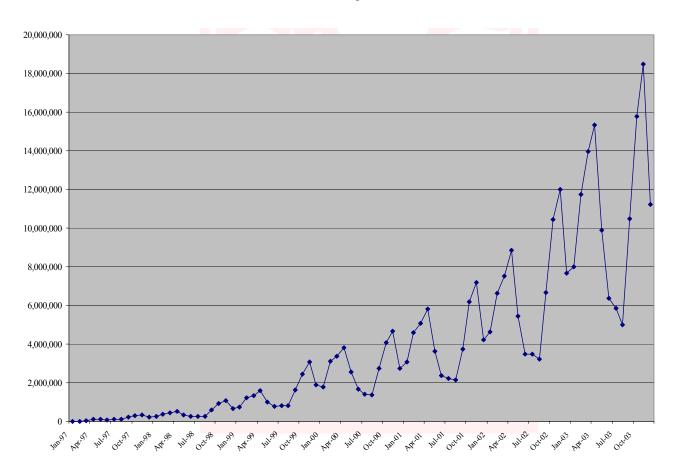


- ➤ Over 1,000 U.S. Participating Sites
- > Over 700 International Participating Sites
- > Over 200 Participating Publishers
- ➤ Over 300 Publications
- > Broad coverage of disciplines
- > 14 million pages scanned (and counting)
  - > (average 10,000 pages per day)



## Monthly Usage

#### Meaningful Accesses





# OAI-PMH Project Background

- JSTOR has shared metadata for some applications
- However we use proprietary data formats and transmission methods
- OAI-PMH had the right characteristics
- But, we are re-writing our system
- Gave us a chance to learn new techniques
- Forced the separation of server from data



# Purpose of this Presentation

- Overview of JSTOR OAI-PMH System
- Constraints
- Process
- Design
- Sharing our observations



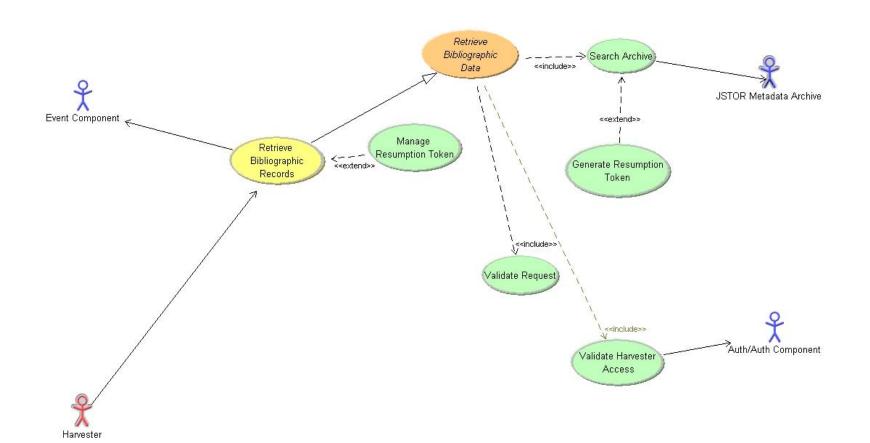
- Large amount of data (2.5 million articles)
- Content restricted by subscription
- Authorization System in transition
- Metadata store in transition
- Code must be sharable with others, in Java
- Lots of uncertainty!



- Initial Requirements Gathering
  - No existing software for our needs
  - Current JSTOR System inadequate
- Unified Process/UML
- Outside advisors (Object Insight)
- Create pluggable parts to handle uncertainty



# Use Case Diagram

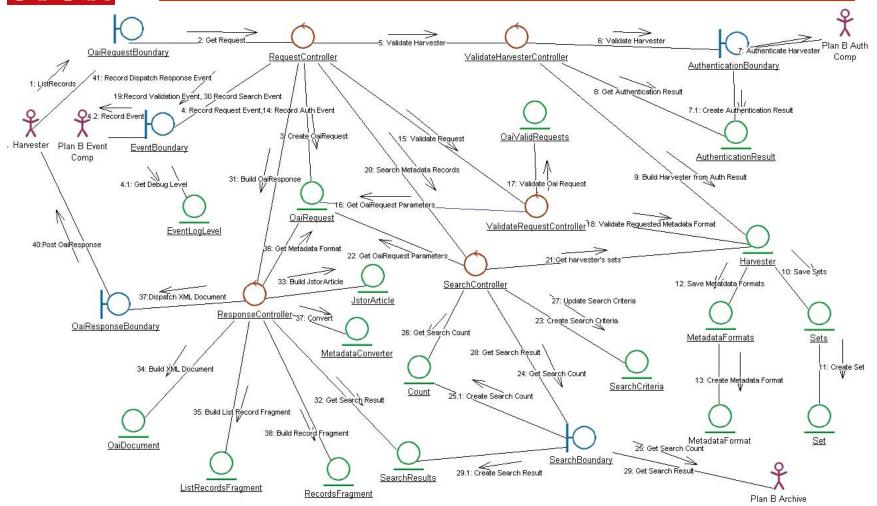




- 'Retrieve Bibliographic Records' Use Case
- Use cases gave insight into Search/Auth requirements
  - Repository would have to handle increments, counts
  - Auth would have to know about harvester sets
- Use Case Analysis using Collaboration Diagram (MVC)

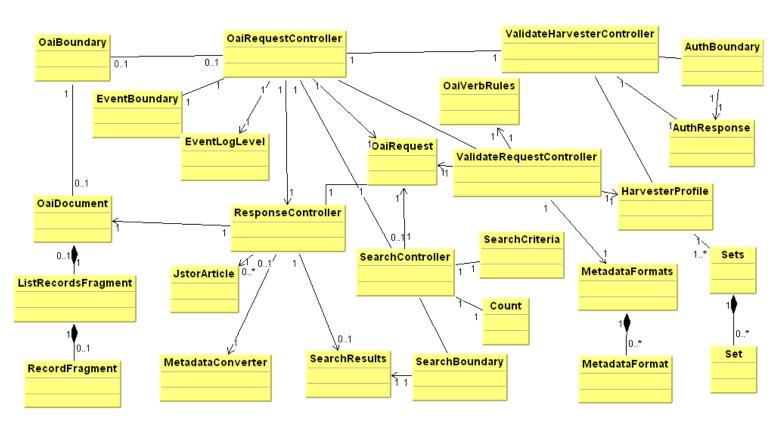


### Retrieve Bibliographic Records Collaboration Diagram





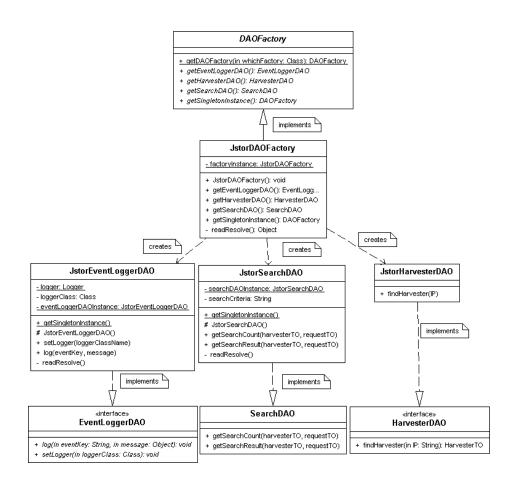
# View of Participating Classes



Retrieve Bibiliographic Data - VOPC Class Diagram



## J2EE Design Patterns





# Current Issues, Questions

- What's "new to repository" vs. "new to subscription"
- Resumption Tokens
- Compression
- Associating metadata formats with types of objects returned from search
- Development nears completion

## Conclusion

- Constraints, Process, Sharing our Findings
- Load Testing has been helpful
- Internal Use First
- How and when to introduce externally
- Possibility of sharing code, UML
- Need for a harvester, internally and externally
- Paper Available
- krot@jstor.org, davidyak@jstor.org
- Questions?