



Metadata and its Relationship to Journal Workflow

SSP Seminar in Philadelphia, PA

David Yakimischak, Chief Technology Officer

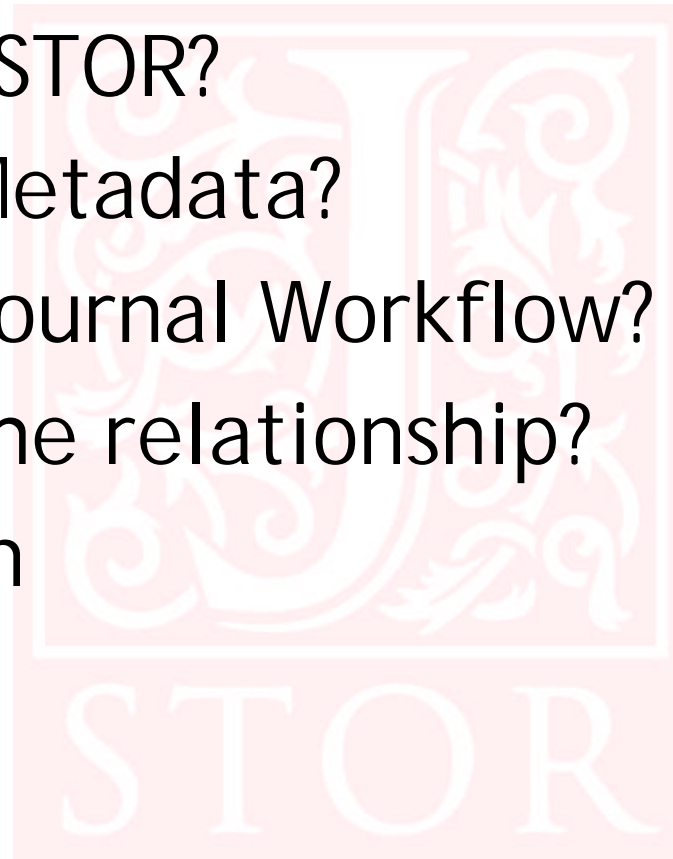
November 18, 2004

www.jstor.org



Agenda

- What is JSTOR?
- What is Metadata?
- What is Journal Workflow?
- What is the relationship?
- Discussion





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 system-wide perspective, seeking benefits for libraries, publishers and scholars



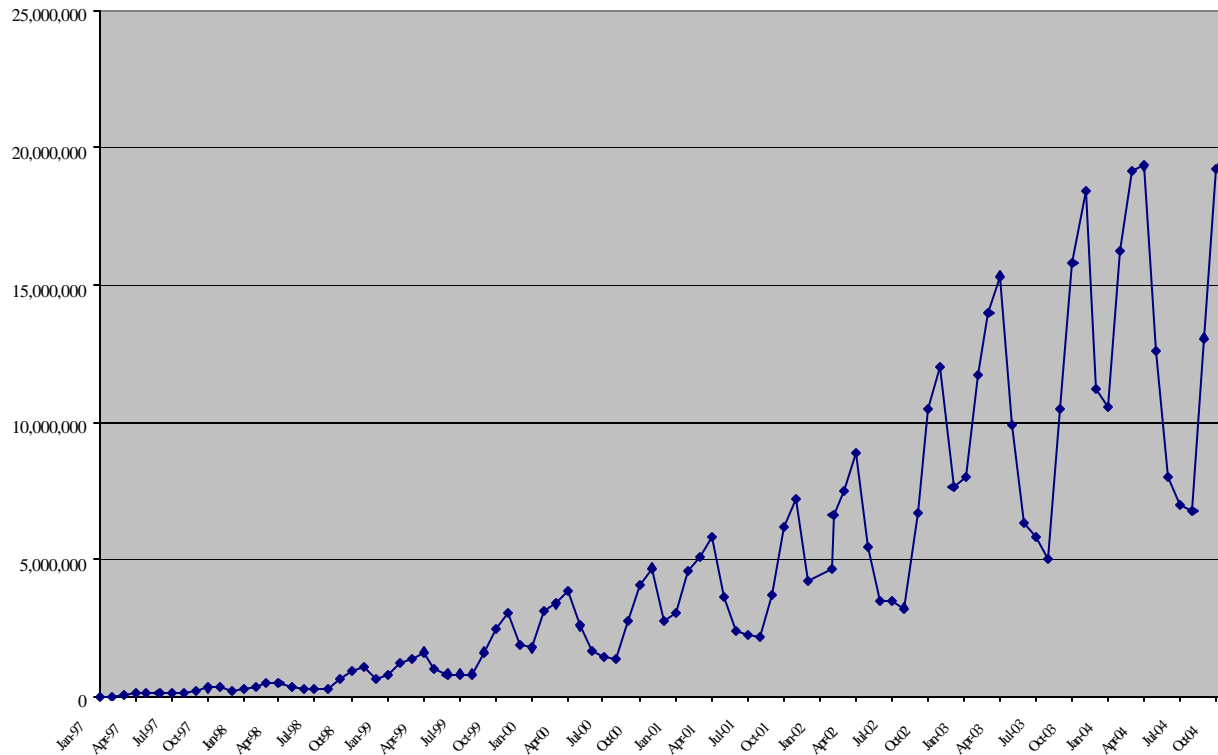
JSTOR Today

- 2,160 participating libraries
- 269 participating publishers
- 449 journals online
- 16,379,559 pages scanned (and counting!)
- Formed an Electronic Archiving initiative in 2003



JSTOR Monthly Usage

Meaningful Accesses per Month





What is Metadata?

- Data about data
- Internally we outlaw the unqualified use of the word metadata
- Three broad categories
 - Descriptive Metadata
 - Technical Metadata
 - Administrative Metadata



Descriptive Metadata

- Describes the content or underlying asset
- Bibliographic information
- Title, author, journal, date, page, etc.
- Used to create citations
- Use to categorize or organize information
- Creates the browse interface to content



Technical Metadata

- Describes the structure or format of an item
- Allows for verification or validation
- Typically used for internal operations
- A registry can be used to catalog these
- Commonly needs to be migrated, although this applies to all metadata
- Most suitable for automated processing



Administrative Metadata

- Generally deals with how an item has been handled
- Most relevant to the archival records
- Tracks history of change or access
- Closely related to workflow
- Can be gathered automatically and unobtrusively



What is Journal Workflow?

- The process of creating the finished product
- Can be formal or ad-hoc
- Automated or manual
- Stress-relieving or stress-inducing



What is the relationship?

- Metadata can be managed during the journal workflow process
 - Added, ingested, modified, calculated, stored, versioned
- Metadata is used to drive the workflow process itself
- The intended use must be considered carefully
 - Why are we gathering this?
 - Who is the audience? When? Why?
 - What about statistics, for example?
 - Connections into other business process systems



JSTOR Observations

- Metadata is created for every word, illustration, page, article, issue, title
- Authorizable units are described with metadata
- Users are described with metadata
- We have established internal guidelines and standards
 - Relationship to, but not the same as external exposure
- Quality assurance and quality control are involved in oversight
- Formalize workflows early when it is easier
- Get it right up front, it is expensive to go back



Discussion

