colectica

MANAGING PRIVATE AND PUBLIC VIEWS OF DDI METADATA REPOSITORIES

EDDI2012

In This Talk

- Presenter: Dan Smith

- Challenges in confidentiality

  - How to expose metadata to the public

- Colectica Workflow Services

  - User Experience

  - Technical Implementation

- Future Directions

About the Speaker

- Former: IBM Advanced Internet Technology Group
  - Semantic Web (OWL-S and SWRL)
  - Persistent Identifiers (LSID)

- Present: Colectica
  - Many DDI 3 based software tools and customizations
  - Workshops and DDI training
- DDI Technical Implementation Committee

Challenges – Metadata Driven

- Metadata now drives production systems
  - Everything is recorded
- Some metadata used for processing may not be public
  - Intermediary versions of metadata
  - Internal locations of holdings
  - Organization personnel due to privacy concerns
  - Variable statistics may differ internally and externally

Challenges - Legal

- □ Legal issues in distributing metadata
- □ Privacy laws regarding particular studies
- □ Patient protections
- □ Licensed code lists that can not be republished

Challenges - Organizational

- Specific organization decisions about metadata release

- An organization may want to publish only a portion of their information
  - Only data file descriptions
 - Only questionnaires

Colectica Workflow Services

- Colectica Workflow Services manage a publication process
  - Approvals
  - Redacting, filtering, and pruning published metadata
  - Audit log of approval process
- Included with Colectica Repository
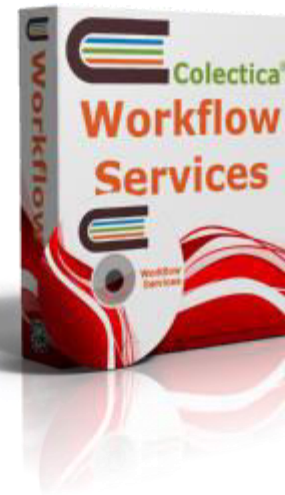
Publication Process - Selection

- ☐ User selects a metadata product to publish in Colectica Designer
  - ◻ Chooses where to publish the item
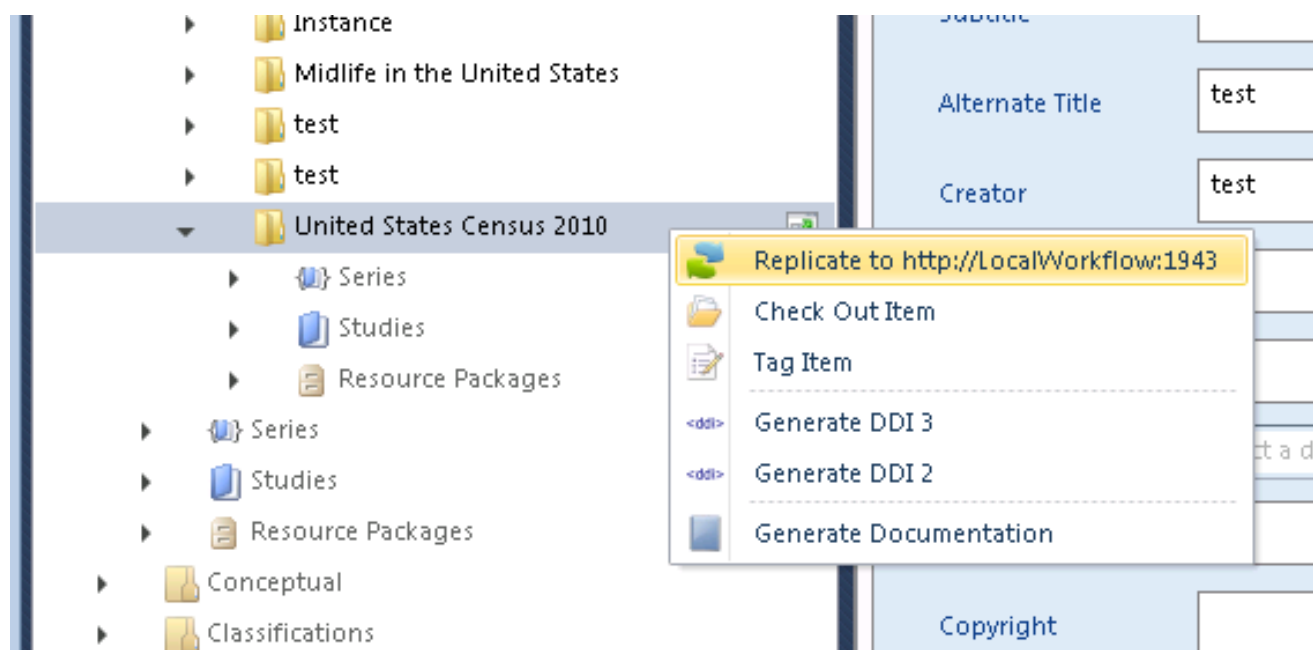
Publication Process - Request

□ A publication request is submitted to the Workflow Services

| Colectica Workflow Management | Requests ▾ | Processes ▾ | Published Items | Welcome, Denise\dan | Log Off |

Pending Requests

4 requests await review.

Review Requests »

Replication Status

🕐 Pending	0
⟳ Working	0
✔ Complete	6

Publication Process - Review

- Another user with appropriate permissions can view and approve the publication

Publication Process - Publish

□ Workflow Services redacts/filters and publishes the metadata

Publication Process - Provenance

☐ Audit trail of published metadata

Technical

- Rules are created in the Workflow Service that control what information becomes public
- Two types of rules
    - Per item rules control content within individual items
    - Per publication rules control whole items and their relationships
- Combining these two rule types gives an organization full control.

Technical – Included Samples

- Sample rules are included

- Filter published questionnaire flow

- Filter custom fields of a specific type
  - Example - Don't disclose the "Data Custodian"

- Filter by intended audience
  - Colectica allows multiple audiences to be defined for many fields, filter the "Internal Audience"

Technical – External view

- Public metadata has same identification as the internal metadata
  - Some fields may not be available
  - Some items may be entirely hidden
- Consistent identity information allows an organization to benefit from reuse
  - External reuse is still visible
  - Third party metadata can be linked to internal view

Future Directions

- ☐ More sample filters based on customer feedback

- ☐ Improved searching within publication audit trail

- ☐ Ability to publish to XML files in addition to another repository instance

- ☐ Public implementation at Statistics New Zealand in April 2013

colectica

Thank you

	Web	colectica.com
	Blog	blogs.colectica.com
	Twitter	@Colectica
	YouTube	youtube.com/colectica