



MANAGING PRIVATE AND PUBLIC VIEWS OF DDI METADATA REPOSITORIES

EDDI2012

In This Talk



2

- 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

3

- 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

4

- 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

5

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



Challenges - Organizational

6

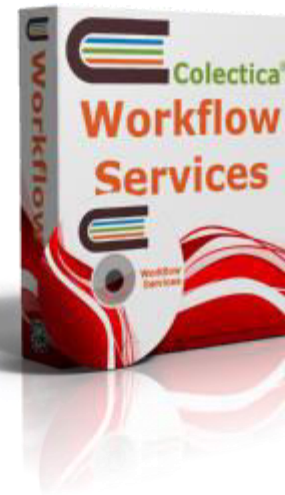
- Specific organization decisions about metadata release
- An organization may want to publish only a portion of their information
 - ▣ Only data file descriptions
 - ▣ Only questionnaires



Colectica Workflow Services

7

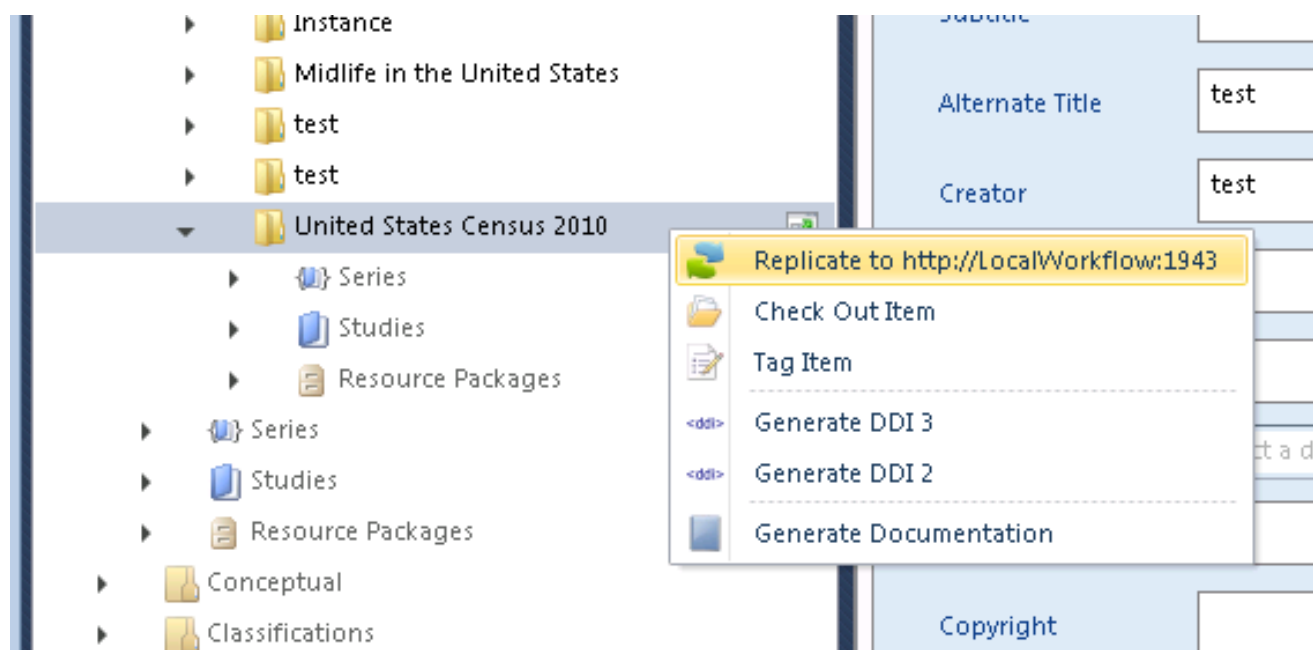
- Colectica Workflow Services manage a publication process
 - ▣ Approvals
 - ▣ Redacting, filtering, and pruning published metadata
 - ▣ Audit log of approval process
- Included with Colectica Repository



Publication Process - Selection

8

- User selects a metadata product to publish in Colectica Designer
 - ▣ Chooses where to publish the item



Publication Process - Request

9

- A publication request is submitted to the Workflow Services

The screenshot displays the 'Colectica Workflow Management' dashboard. The top navigation bar includes links for 'Requests', 'Processes', and 'Published Items', along with a user greeting 'Welcome, Denise\dan' and a 'Log Off' option. The main content area is divided into two sections. The first section, 'Pending Requests', indicates that '4 requests await review' and features a blue button labeled 'Review Requests »'. The second section, 'Replication Status', contains a table with the following data:

Replication Status	Count
⌚ Pending	0
🔄 Working	0
✓ Complete	6



Publication Process - Review

10

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

Colectica Workflow Management

Requests ▾













Processes ▾

Published Items

Welcome, Denise\dan

Log Off

Requests

		Owner	Created	Updated	Name		
 Publication	 Pending	 Denise\dan	9/26/2012 5:37:45 PM	9/26/2012 5:37:45 PM	Instance v.2	Approve	Deny
 Publication	 Pending	 Denise\dan	9/26/2012 5:36:06 PM	9/26/2012 5:36:06 PM	Example Study v.1	Approve	Deny
 Publication	 Pending	 Denise\dan	9/18/2012 7:37:22 PM	9/18/2012 7:37:22 PM	Example Study v.1	Approve	Deny
 Publication	 Pending	 Denise\dan	9/18/2012 6:56:13 PM	9/18/2012 6:56:13 PM	Example Study v.1	Approve	Deny



Publication Process - Publish

11

- Workflow Services redacts/filters and publishes the metadata

Colectica Workflow Management

Requests ▾




Processes ▾

Published Items

Welcome, Denise\dan

Log Off

Processes

	Owner	Label	Created	Updated	Message
 Replication	 Pending	 Denise\dan	Example Study v.1	12/4/2012 1:51:36 AM	12/4/2012 1:51:36 AM



Publication Process - Provenance

12

□ Audit trail of published metadata

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

Published Items

ColecticaRepositoryPublic

Example Study (v.1) Denise\dan 9/6/2012 6:40:54 PM [Remove](#)



Technical

13

- 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

14

- ❑ 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

15

- 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

16

- 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





17

Thank you

Web colectica.com



Blog blogs.colectica.com



Twitter @Colectica



YouTube youtube.com/colectica