



Ensuring Long Term Access to Quality Data:

A DDI-based Curation System

Two Partners, a Consultant, and a Software Developer



Digital Lifecycle Research & Consulting



Research

The ISPS KnowledgeBase is the gateway to all ISPS data, projects, and publications. It is an integrated database which provides a one-stop-shop for ISPS-related research products.

Search the KnowledgeBase or browse recent additions.

 [SEARCH ISPS](#)

RESEARCH FUNDING

ISPS invites proposals for important and well-crafted field experiments in the social sciences and related policy issues. Field experiments are fully-randomized research designs in which observations found in a naturalistic setting -- voters, patients, welfare recipients, community organizations, government entities, and the like -- are assigned to treatment and control conditions ([see more here](#)).

[> Apply for a field experiment grant](#)

[> Additional funding opportunities](#)

Yale ISPS KnowledgeBase

Data	Projects	Publications
Terms of use About the ISPS data archive		
AUTHOR	AREA OF STUDY	DISCIPLINE
<input type="text" value="- Any -"/>	<input type="text" value="- Any -"/>	<input type="text" value="- Any -"/>
YEAR	LOCATION	KEYWORDS
<input type="text" value="- Year -"/>	<input type="text" value="- Any -"/>	<input type="text" value="- Any -"/>
RESEARCH DESIGN		
<input type="text" value="- Any -"/>		
See all data		Search

FEATURED BOOKS BY ISPS FACULTY



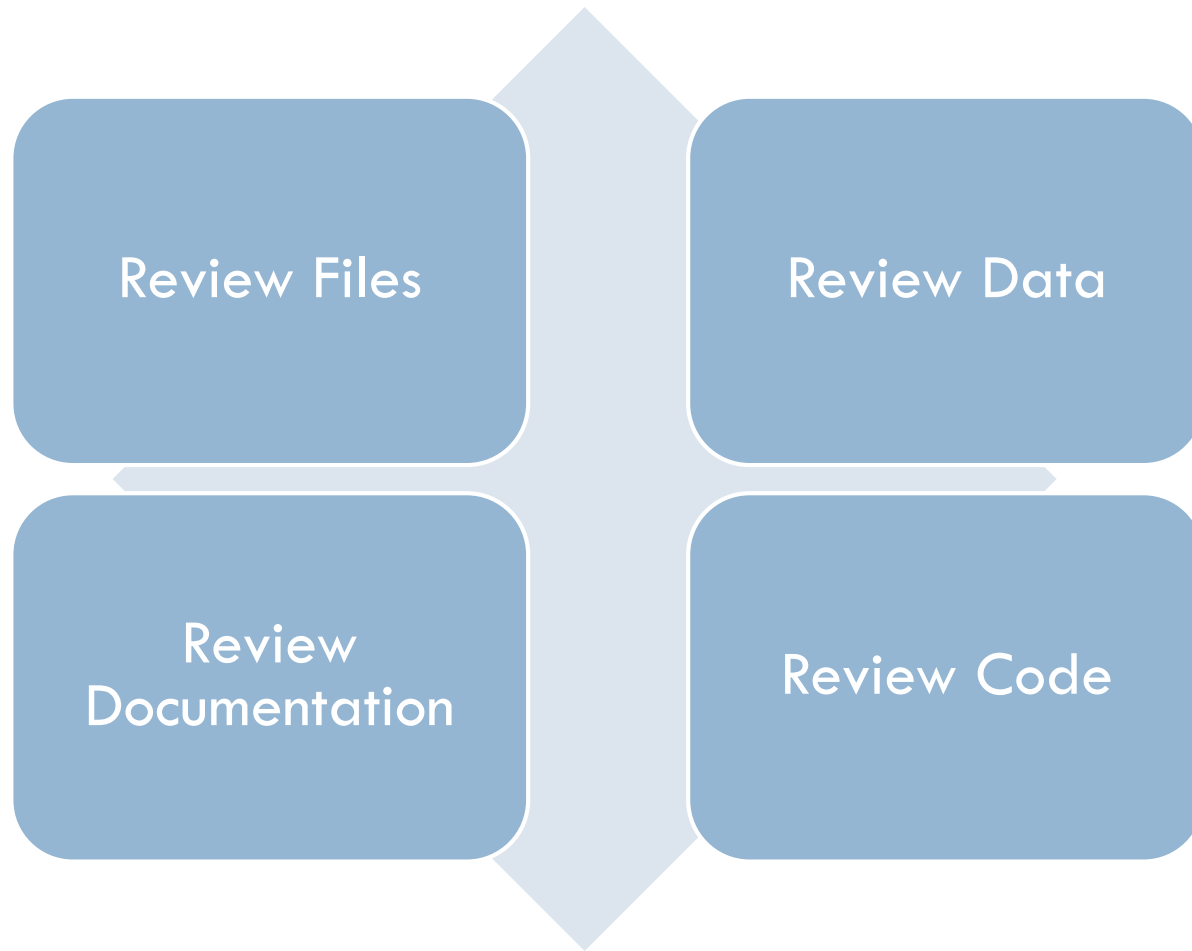


STRATEGIC PLAN >> 2013–2018

MORE EVIDENCE, LESS POVERTY

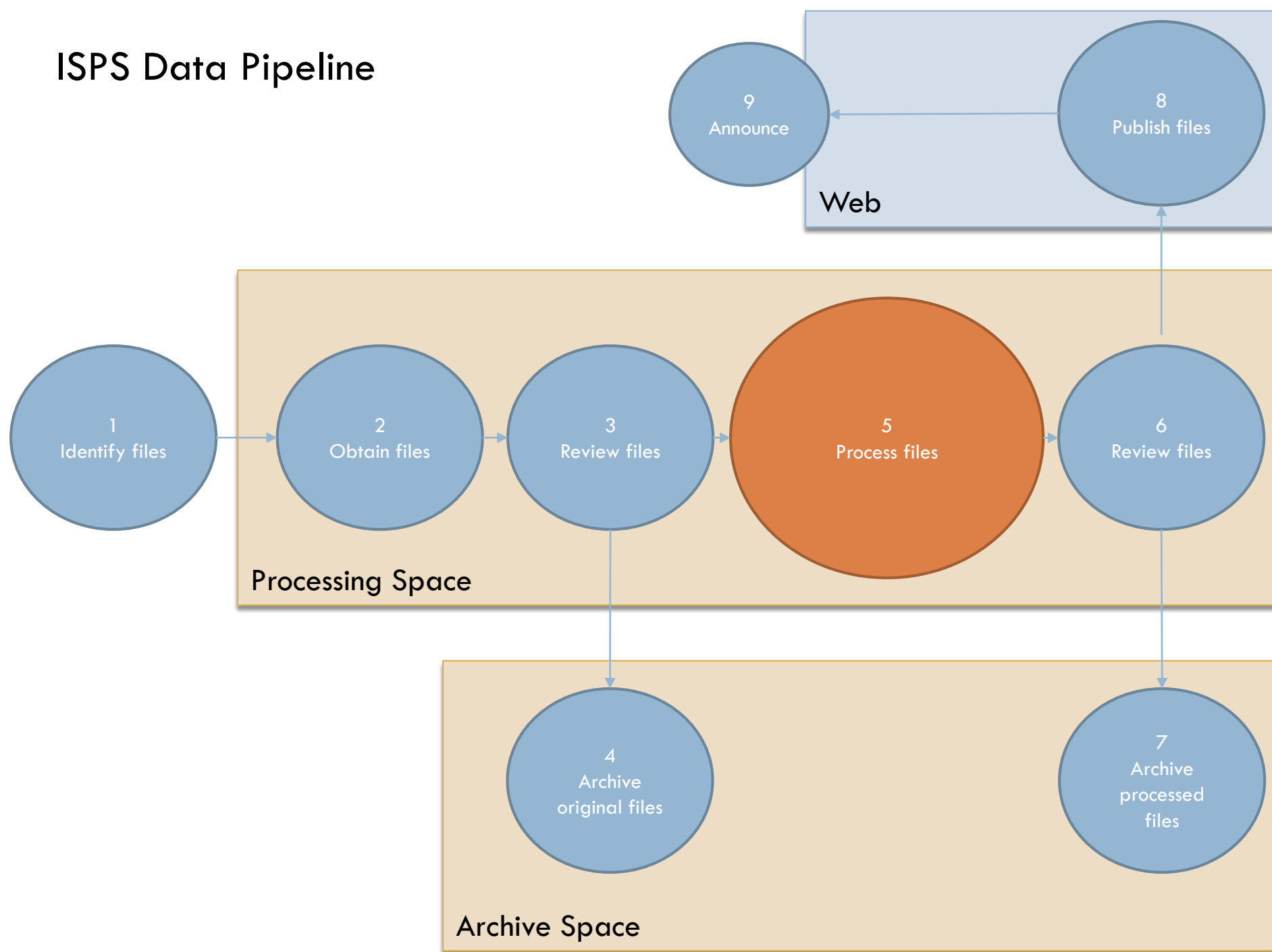


Data Quality Review



Source: Peer, Green, and Stephenson. 2014. Committing to Data Quality Review. International Journal of Digital Curation, 19(1): 263-291. doi:10.2218/ijdc.v9i1.317

ISPS Data Pipeline



Processing Steps

1. Assign staff to study and files
2. Move original files to Archive space
3. Make copies of processed files and move to collaborative space
4. Identify related publications and projects
5. Rename all copied files for public dissemination according to ISPS Data Archive naming conventions
6. Check and complete variable-level metadata for each data file
7. Compare variable information, check for additional variables and recoded variables, check variable/value labels
8. Check all files for confidential and other sensitive information
9. Run the statistical code and check against published results
10. Re-write statistical code in R and check replication
11. Communicate with PI as needed
12. Create new DDI-XML file with variable-level information
13. Create additional files by converting to readable formats (e.g., ASCII, PDF)
14. Update study- and file-level metadata record
15. Update tracking documents: process record / general study database / status document

Starting Point



Tools for curation



Tools for publication

Starting Point

- Curation guides
- Excel audit logs
- Shared file systems

Processes and ad hoc tools

The STATA logo, featuring the word "STATA" in a blue, sans-serif font with a registered trademark symbol.The StatTransfer logo, which includes the word "StatTransfer" in white on a black rectangular background, flanked by blue arrows pointing outwards.

Tools for curation

The logo for The Dataverse Network Project, which includes the text "The Dataverse Network Project" in orange and black, accompanied by a circular icon with a network of nodes.

Tools for publication

Goal: Introduce Standards



- Curation guides
- Excel audit logs
- Shared file systems

Processes and ad hoc tools



Tools for curation



Tools for publication

Goal: Introduce Standards



- Curation guides
- ~~Excel audit logs~~
- Shared file systems

Processes and ad hoc tools



Tools for curation



Tools for publication

Goal: Introduce Standards



- Curation guides
- ~~Excel audit logs~~
- ~~Shared file systems~~

Processes and ad hoc tools

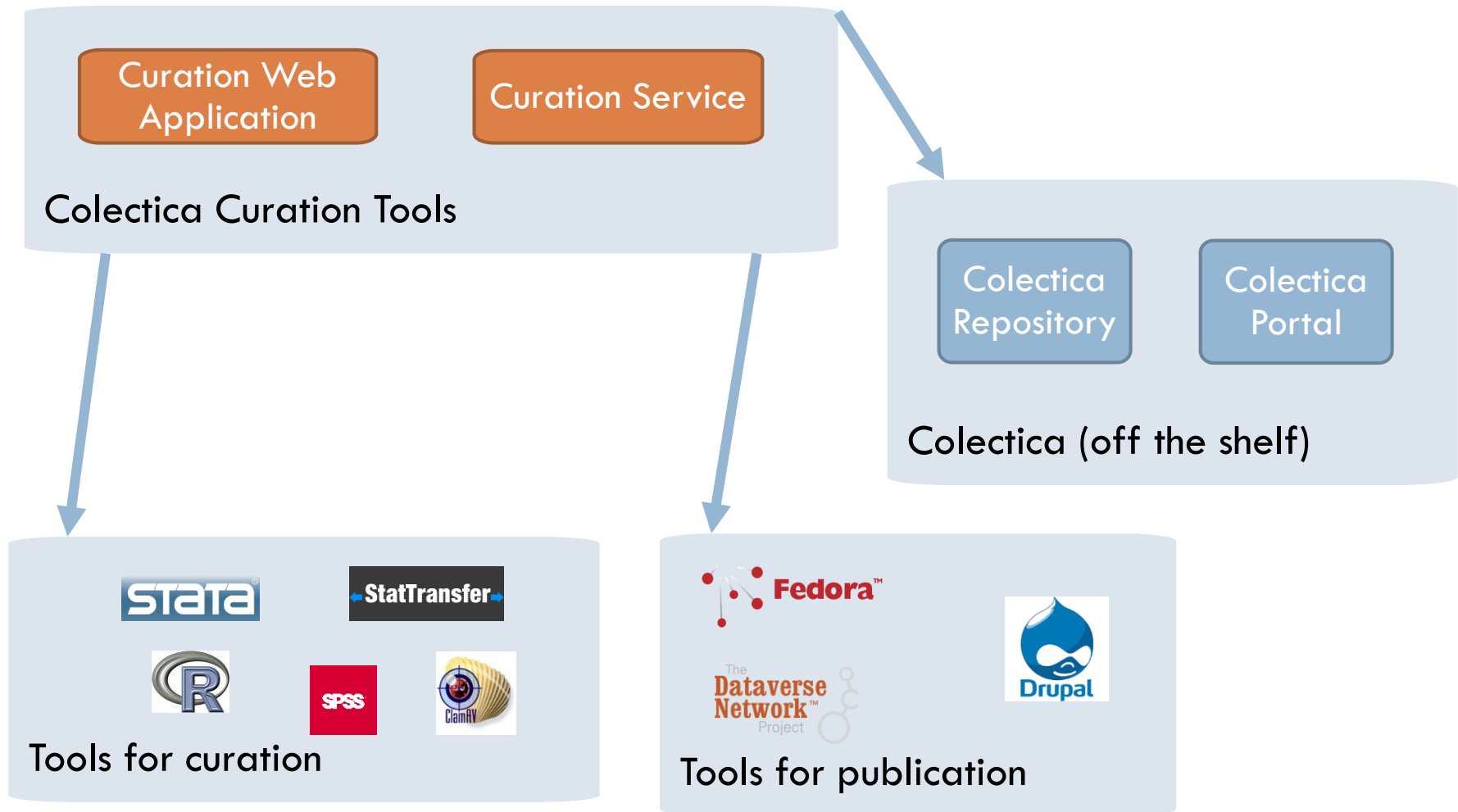


Tools for curation



Tools for publication

Result: Colectica Curation Tools



Mix: Commercial and Open Source

Curation Web
Application

Curation Service

Colectica Curation Tools

Colectica
Repository

Colectica
Portal

Colectica (off the shelf)

STATA

StatTransfer



Tools for curation



The
Dataverse
Network™
Project



Tools for publication

Neat Features

- ❑ Built on DDI 3.2
- ❑ Web-based
- ❑ Open Source



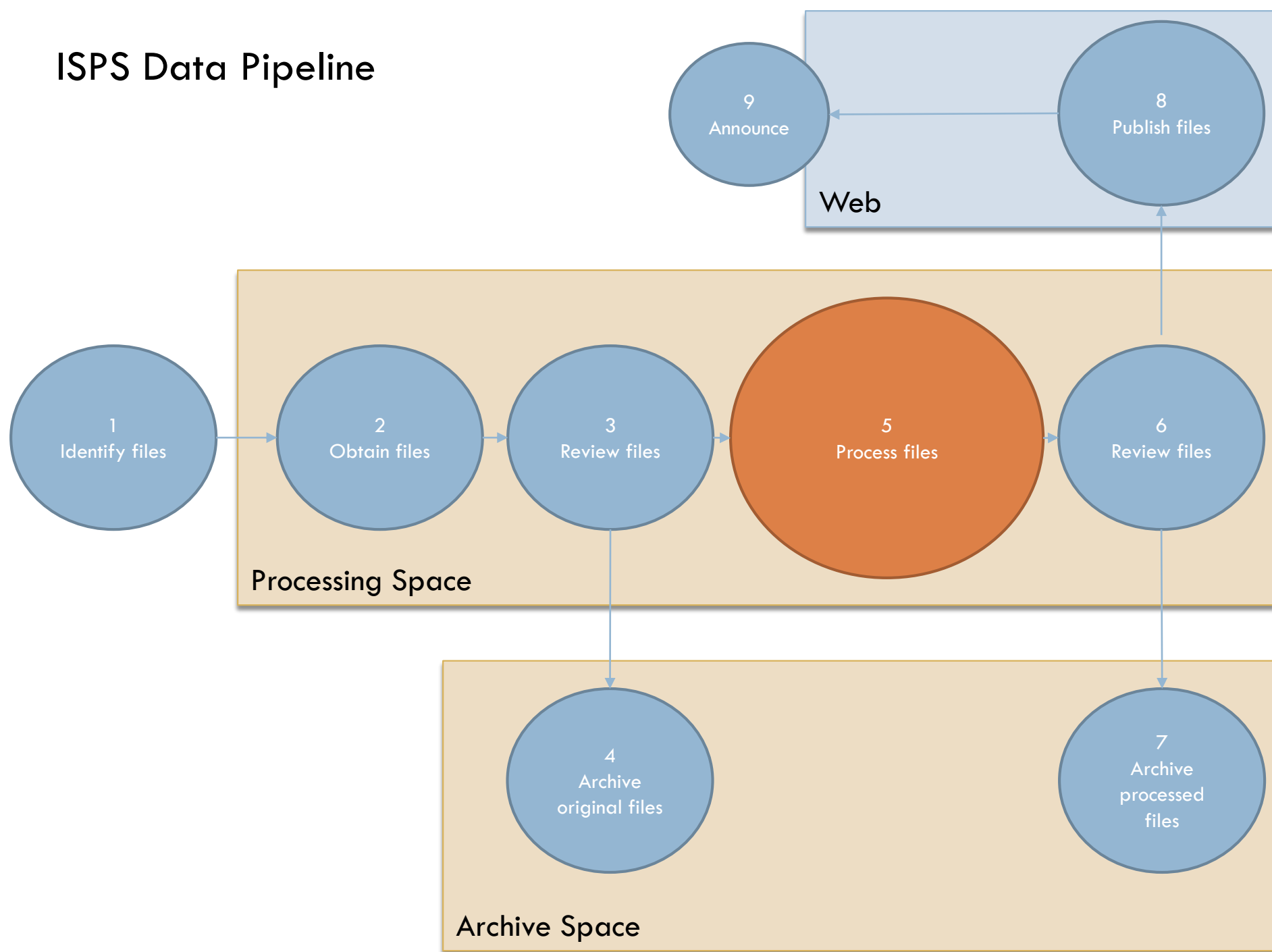
Target Users

- Depositor
- Curator
- Administrator
- Machines
- Researchers

ISPS and IPA Requirements

- ❑ Curation workflow management (dashboard)
- ❑ Track changes to files (provenance)
- ❑ Integrate metadata production with data and code review and cleaning
- ❑ Preservation metadata and formats
- ❑ Secure storage and access
- ❑ Smooth transition to public dissemination of content
- ❑ Preference for open source solutions

ISPS Data Pipeline



Mix: Commercial and Open Source

Curation Web
Application

Curation Service

Colectica Curation Tools

Colectica
Repository

Colectica
Portal

Colectica (off the shelf)

STATA

StatTransfer



Tools for curation



The
Dataverse
Network™
Project



Tools for publication

Extensibility

- Other organizations might find this useful, too
- Requirements may differ

Extensibility: Actions

- New Catalog Record Created
 - ▣ Current: Create git storage, create DDI StudyUnit

Extensibility: Actions

- New File Added

- Current

- Virus scan
 - Extract variable level metadata
 - Store metadata in standardized repository

Extensibility: Actions

- Submit for Curation
 - Current: Send email notifications

Publication

1. Create preservation formats for proprietary data files.
2. Request persistent identifiers from the configured Handle service and store them with the metadata.
3. Create checksums for each file and store them with the metadata.
4. Create DDI metadata for all files and the catalog record itself.
5. Create an archive package and place it in the configured archive ingest location.
6. Mark the catalog record as published.

Publication Extensions

- Persistent Identifiers

- Current: Create Handles for all files and the catalog record
- Possible: Create DOIs or other identifiers

Publication Extensions



- Preservation Formats
 - ▣ Current: Use Stat/Transfer to create CSV data files
 - ▣ Possible: Convert other formats, too

Publication Extensions



- Finalize Metadata
 - ▣ Current: Create DDI 3.2 XML metadata
 - ▣ Possible: Create other formats, too

Publication Extensions

- Disseminate
 - ▣ Current: Fedora Commons, Hydra, Colectica Portal
 - ▣ Soon: Drupal, Dataverse
 - ▣ Possible: Other dissemination platforms

Demonstration

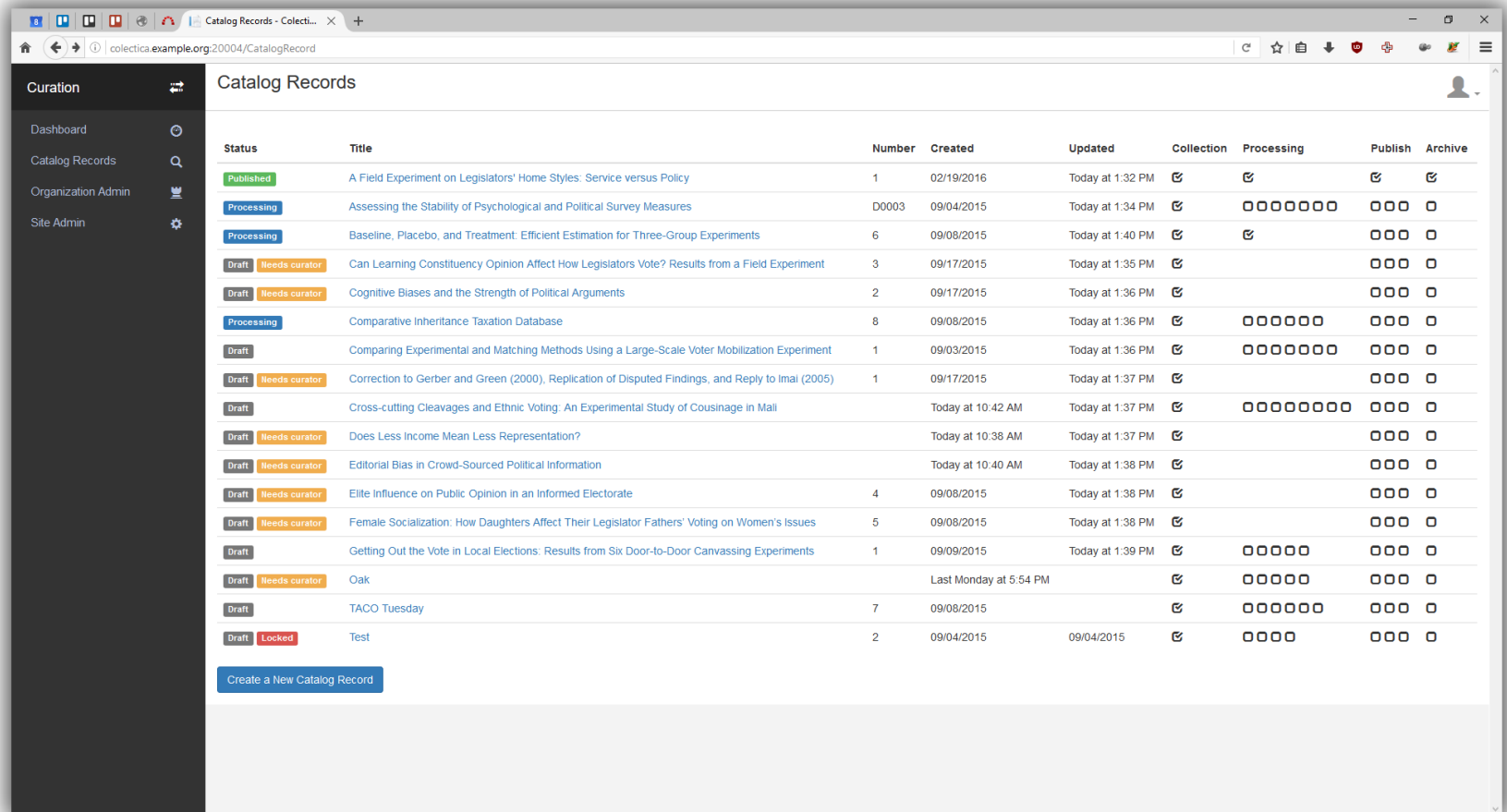


Learn More: Documentation



docs.colectica.com/curation

Catalog Records



Status	Title	Number	Created	Updated	Collection	Processing	Publish	Archive
Published	A Field Experiment on Legislators' Home Styles: Service versus Policy	1	02/19/2016	Today at 1:32 PM				
Processing	Assessing the Stability of Psychological and Political Survey Measures	D0003	09/04/2015	Today at 1:34 PM		0000000	000	0
Processing	Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments	6	09/08/2015	Today at 1:40 PM			000	0
Draft	Needs curator	Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment	3	09/17/2015	Today at 1:35 PM		000	0
Draft	Needs curator	Cognitive Biases and the Strength of Political Arguments	2	09/17/2015	Today at 1:36 PM		000	0
Processing	Comparative Inheritance Taxation Database	8	09/08/2015	Today at 1:36 PM		000000	000	0
Draft	Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment	1	09/03/2015	Today at 1:36 PM		0000000	000	0
Draft	Needs curator	Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005)	1	09/17/2015	Today at 1:37 PM		000	0
Draft	Cross-cutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali		Today at 10:42 AM	Today at 1:37 PM		00000000	000	0
Draft	Needs curator	Does Less Income Mean Less Representation?		Today at 10:38 AM	Today at 1:37 PM		000	0
Draft	Needs curator	Editorial Bias in Crowd-Sourced Political Information		Today at 10:40 AM	Today at 1:38 PM		000	0
Draft	Needs curator	Elite Influence on Public Opinion in an Informed Electorate	4	09/08/2015	Today at 1:38 PM		000	0
Draft	Needs curator	Female Socialization: How Daughters Affect Their Legislators' Voting on Women's Issues	5	09/08/2015	Today at 1:38 PM		000	0
Draft	Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments	1	09/09/2015	Today at 1:39 PM		00000	000	0
Draft	Needs curator	Oak		Last Monday at 5:54 PM		00000	000	0
Draft	TACO Tuesday	7	09/08/2015			000000	000	0
Draft	Locked	Test	2	09/04/2015	09/04/2015		0000	000

Create a New Catalog Record

New Record

The screenshot shows a web browser window with the address bar displaying `colectica.example.org:20004/CatalogRecord/Create`. The page title is "Create New Catalog Record". On the left is a dark sidebar with navigation links: "Curation" (with a double-headed arrow icon), "Dashboard" (with a clock icon), "Catalog Records" (with a magnifying glass icon), "Organization Admin" (with a group of people icon), and "Site Admin" (with a gear icon). The main content area has a header "Create New Catalog Record" and a progress bar with four steps: "General" (active, blue), "Methods", "Files", and "Submit for Curation". Below the progress bar, a note states: "Fields marked with * are required before submitting a catalog record for curation." The "General" section contains several form fields: "Title *" with the value "NADDI 2013 Evaluation"; "Authors *" with a dropdown menu showing "Larry Hoyle" and a message "No results found for Larry Hoyle" with a "Create Larry Hoyle" button; "Description *" (empty); and "Number" (empty). The browser's tab bar shows the current tab as "Create New Catalog Recor..." and a plus sign to add more tabs.

Create New Catalog Record

General Methods Files Submit for Curation

Fields marked with * are required before submitting a catalog record for curation.

General

Title * NADDI 2013 Evaluation

Authors * Add or remove authors

Larry Hoyle

No results found for Larry Hoyle

Create Larry Hoyle

Description *

Number

Add Files

The screenshot shows a web application interface for uploading files. The browser address bar displays a URL from 'colectica.example.org'. The left sidebar, titled 'Curation', contains links to 'Dashboard', 'Catalog Records', 'Organization Admin', and 'Site Admin'. The main content area is titled 'Upload Files' and features a green '+ Add files...' button. Below this, the 'File Details' section lists two files with their respective metadata and form fields.

File Name	Created	Size	Type	Public File Name	Source	Source Information	Rights	Public Access
Future topics suggested.docx	4/8/2016, 9:01:34 AM	14 KB	Supplementary Materials	Future topics suggested.docx				Yes
NADDI2013Evaluation.sav	4/28/2015, 2:55:04 PM	530 KB	Data	NADDI2013Evaluation.sav				

Assign Curators

The screenshot displays a web browser window with the address bar showing the URL: `colectica.example.org:20004/CatalogRecord/Permissions/186ef42d-b3b7-4899-81f0-882a08adc671`. The page title is "NADDI 2013 Evaluation".

Left Sidebar (Navigation Menu):

- Curation (active, with a double arrow icon)
- Dashboard (with a clock icon)
- Catalog Records (with a magnifying glass icon)
- Organization Admin (with a crown icon)
- Site Admin (with a gear icon)

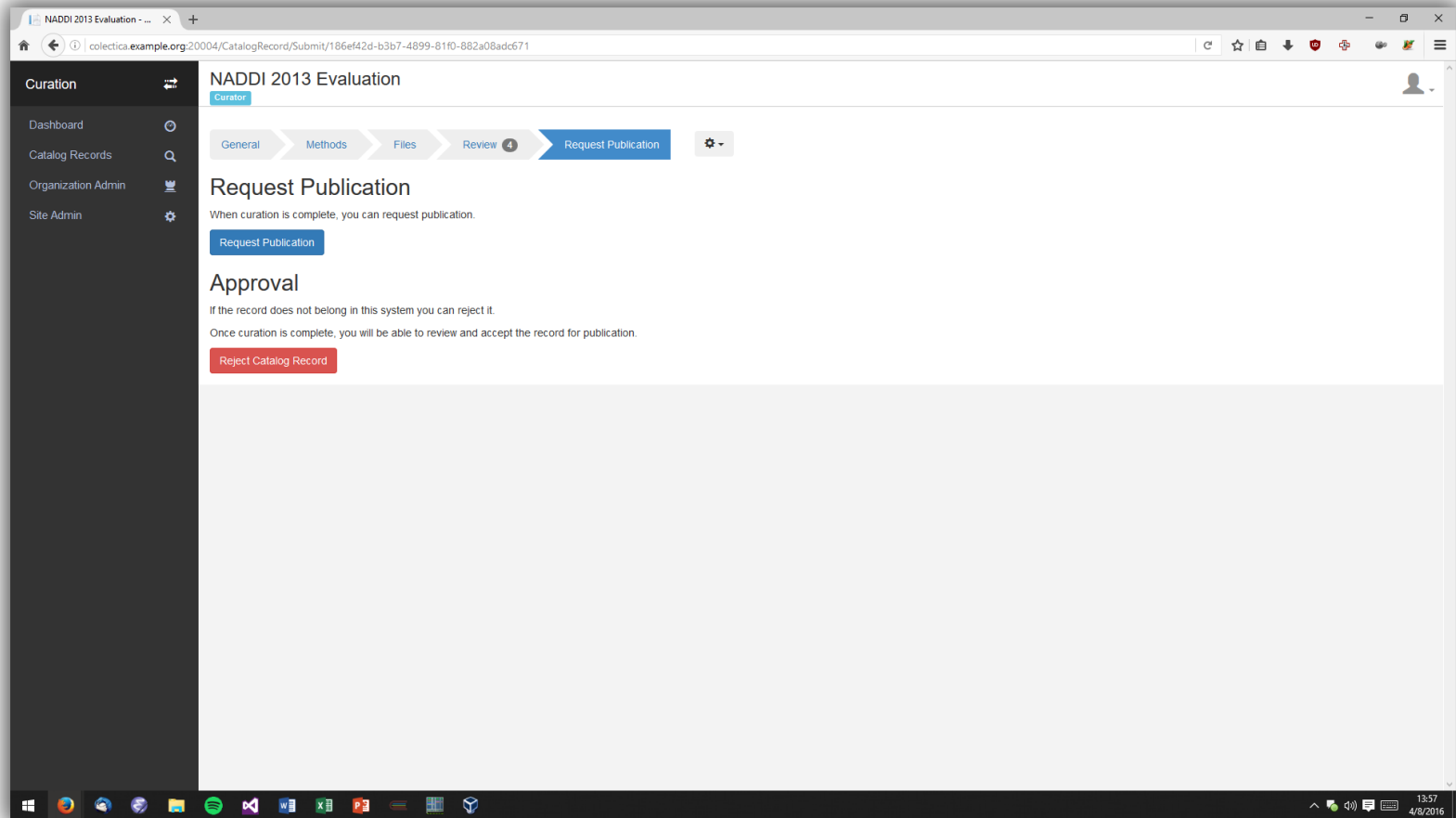
Main Content Area:

The main content area is titled "NADDI 2013 Evaluation". It features a horizontal navigation bar with the following steps: General, Methods, Files, Review (12), and Submit for Curation. A settings gear icon is located to the right of the navigation bar.

Below the navigation bar, there are three panels:

- All Users:** A list of users with their names and email addresses:
 - Jeremy Iverson (jeremy@colectica.com)
 - Test102 Tester102 (test102@jeremyIverson.com)
- Curators:** A panel for assigning curators, currently empty.
- Approvers:** A panel for assigning approvers, currently empty.

Request Publication



Review Status

The screenshot shows a web application interface for managing catalog records. The browser address bar displays the URL: `colectica.example.org:20004/CatalogRecord/Status/186ef42d-b3b7-4899-81f0-882a08adc671`. The page title is "NADDI 2013 Evaluation".

Navigation Sidebar:

- Curation
- Dashboard
- Catalog Records
- Organization Admin
- Site Admin

Breadcrumb Trail: General > Methods > Files > Review 12 > Submit for Curation

Buttons: Status, Notes, Settings (gear icon)

Status: Draft

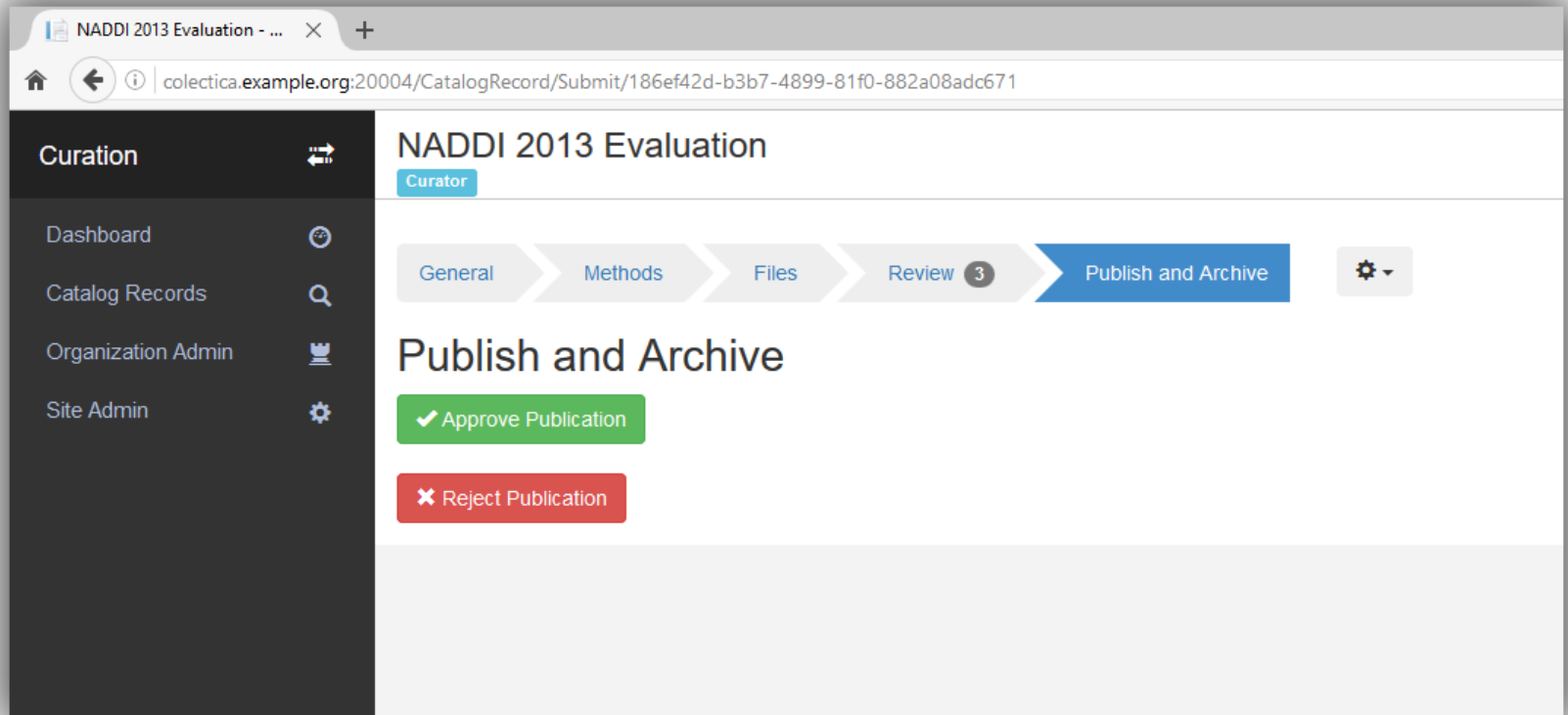
Review Type:

Curators: No curators are assigned.

Task	Status	Completed
Collection		
Created	✓	jeremy@colectica.com on 4/8/2016
Accepted	✓	jeremy@colectica.com on 4/8/2016
Processing		
Review Observation Count	○	
Check Missing Labels	○	
Compare Questionnaire, Codebook, and Data in Data File	○	
Check for Personally-Identifiable Information (PII) in Data File	○	
Identify Potential Errors in Data File	○	
Confirm Code Executes	○	
Confirm Code Replicates Reported Results	○	
Create Preservation Format	○	
Publish		
Request Publication	○	
Publication Approved	○	
Published	○	
Archive		
Archived	○	

The Windows taskbar at the bottom shows the time as 13:48 on 4/8/2016.

Approve Publication



Published

The screenshot displays a web application interface for a catalog record. The browser address bar shows the URL: `colectica.example.org:20004/CatalogRecord/Status/186ef42d-b3b7-4899-81f0-882a08adc671`. The interface features a sidebar on the left with navigation links: Curation, Dashboard, Catalog Records, Organization Admin, and Site Admin. The top navigation bar includes tabs for General, Methods, Files, Review (selected), and Publish and Archive. The main content area shows the record's status as 'Published' and includes a table of tasks and their completion status.

General **Methods** **Files** **Review** **Publish and Archive**

Status: **Published**
Review Type: Full
Date Certified as Curated: 4/8/2016
Curators: Jeremy Iverson

Task	Status	Completed
Collection		
Created	✓	jeremy@colectica.com on 4/8/2016
Accepted	✓	jeremy@colectica.com on 4/8/2016
Processing		
Review Observation Count	✓	jeremy@colectica.com on 4/8/2016
Check Missing Labels	✓	jeremy@colectica.com on 4/8/2016
Compare Questionnaire, Codebook, and Data in Data File	✓	jeremy@colectica.com on 4/8/2016
Check for Personally-Identifiable Information (PII) in Data File	✓	jeremy@colectica.com on 4/8/2016
Identify Potential Errors in Data File	✓	jeremy@colectica.com on 4/8/2016
Confirm Code Executes	✓	jeremy@colectica.com on 4/8/2016
Confirm Code Replicates Reported Results	✓	jeremy@colectica.com on 4/8/2016
Create Preservation Format	✓	jeremy@colectica.com on 4/8/2016
Publish		
Request Publication	✓	jeremy@colectica.com on 4/8/2016
Publication Approved	✓	jeremy@colectica.com on 4/8/2016
Published	✓	jeremy@colectica.com on 4/8/2016
Archive		
Archived	✓	jeremy@colectica.com on 4/8/2016

Thank you

Contributor	Organization	Email
Ann Green	Independent Consultant	green.ann@gmail.com
Jeremy Iverson	Colectica	jeremy@colectica.com
Niall Keleher	Innovations for Poverty Action	nkeleher@poverty-action.org
Limor Peer	Yale University	limor.peer@yale.edu
Dan Smith	Colectica	dan@colectica.com
Stephanie Wykstra	Innovations for Poverty Action	swykstra@poverty-action.org