Digital Lifecycle Research & Consulting

Ensuring Long Term Access to Quality Data:

A DDI-based Curation System

NADDI 2016 – Edmonton

Two Partners, a Consultant, and a Software Developer

Digital Lifecycle Research & Consulting

Yale ISPS

Institution for Social and Policy Studies

ADVANCING RESEARCH • SHAPING POLICY • DEVELOPING LEADERS

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.

Yale ISPS KnowledgeBase

Data | Projects | Publications

Terms of use | About the ISPS data archive

AUTHOR: – Any –

AREA OF STUDY: – Any –

DISCIPLINE: – Any –

YEAR: –Year

LOCATION: – Any –

KEYWORDS: – Any –

RESEARCH DESIGN: – Any –

Search

See all data ►

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

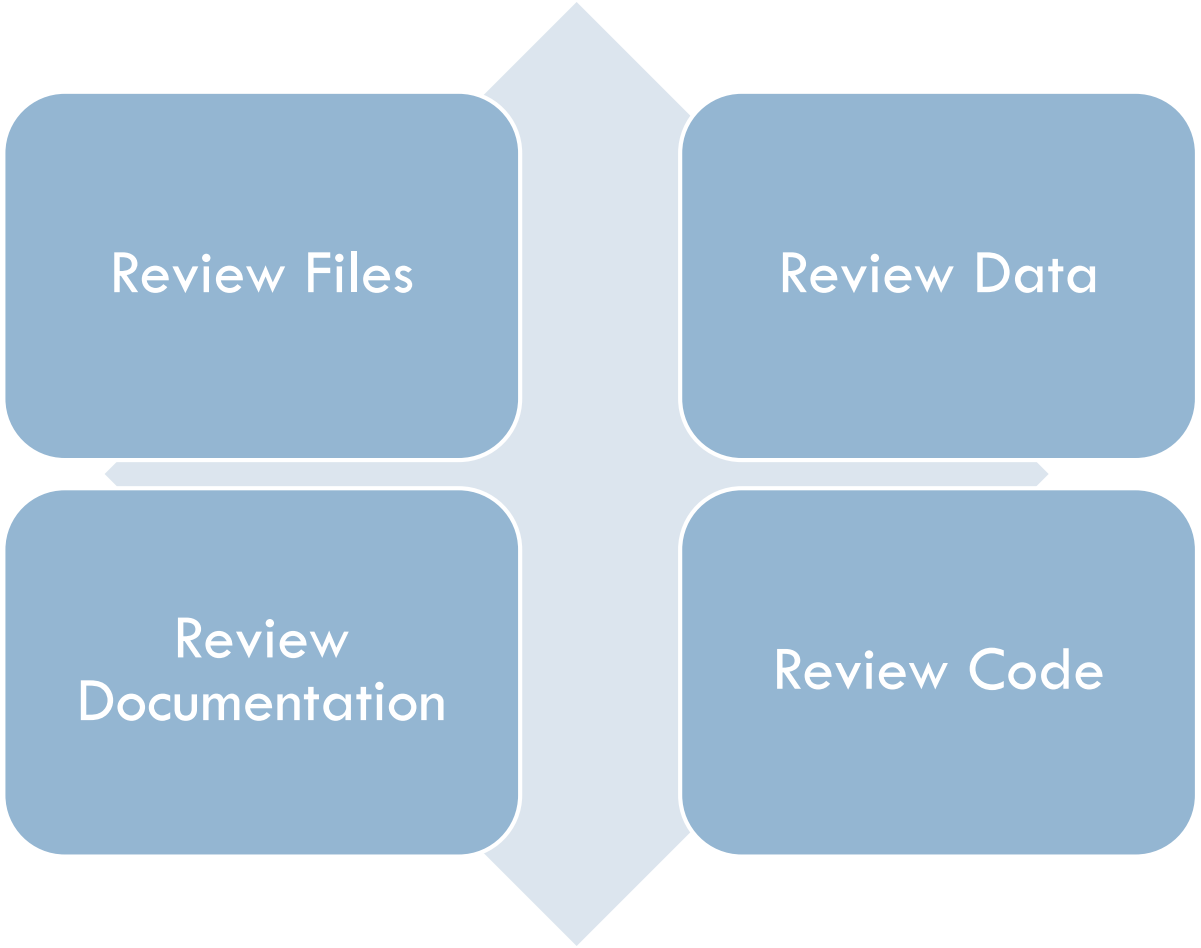
Featured Books By ISPS Faculty

STRATEGIC PLAN >> 2013–2018

MORE EVIDENCE, LESS POVERTY

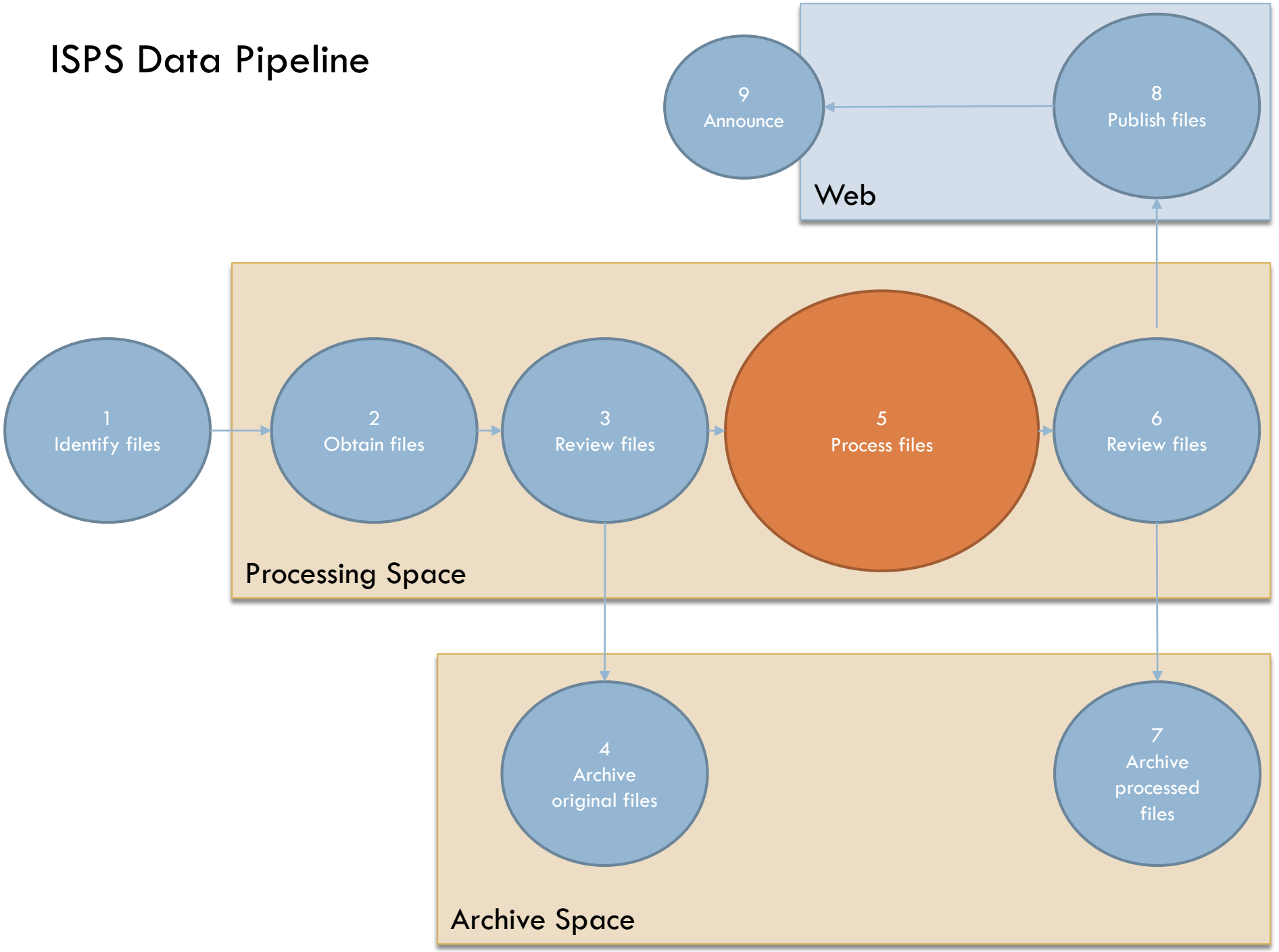
ipa
INNOVATIONS FOR POVERTY ACTION

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

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

Goal: Introduce Standards

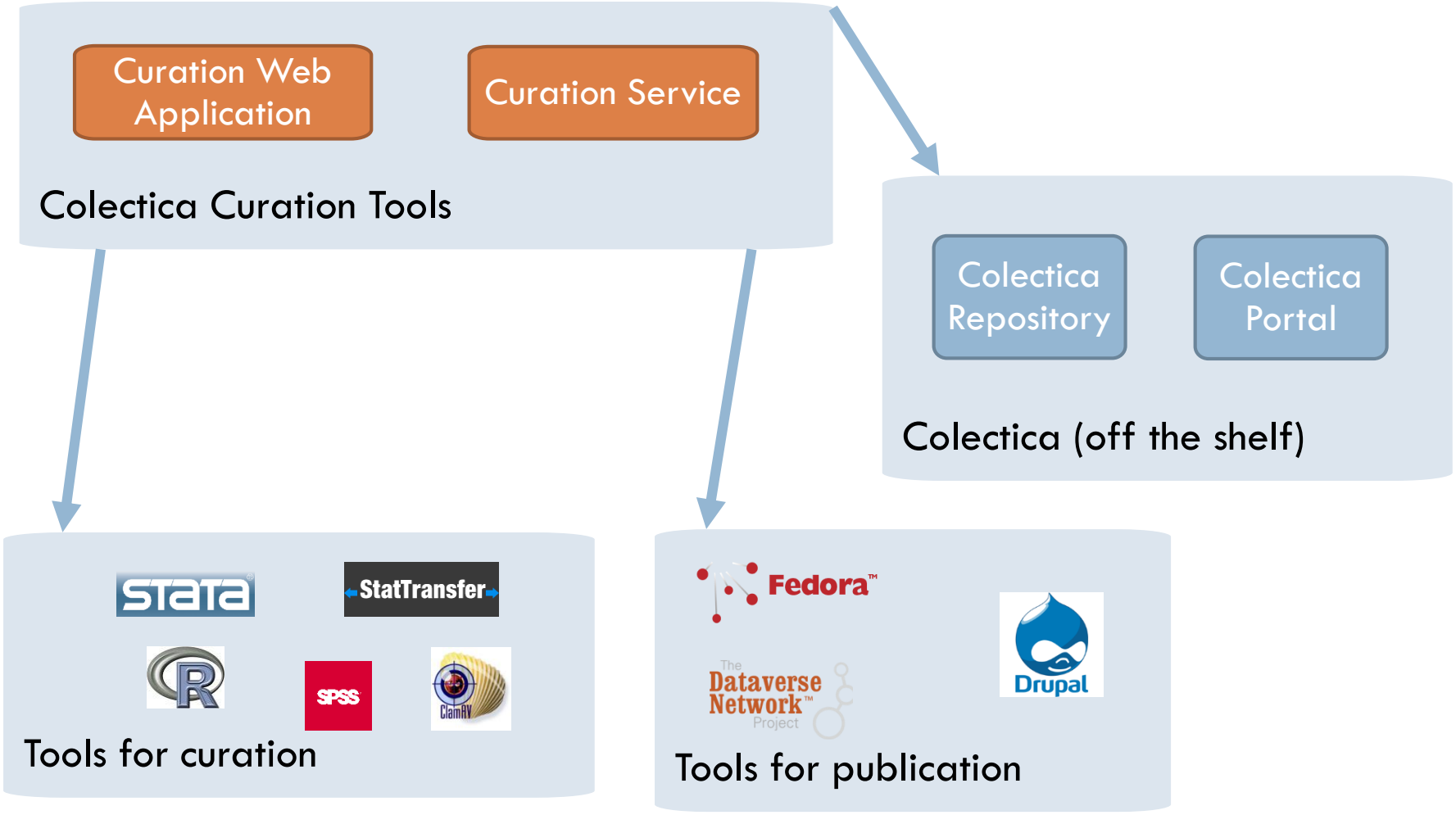
- Curation guides
- ~~Excel audit logs~~
- ~~Shared file systems~~

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)

Tools for curation

Tools for publication

Neat Features

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

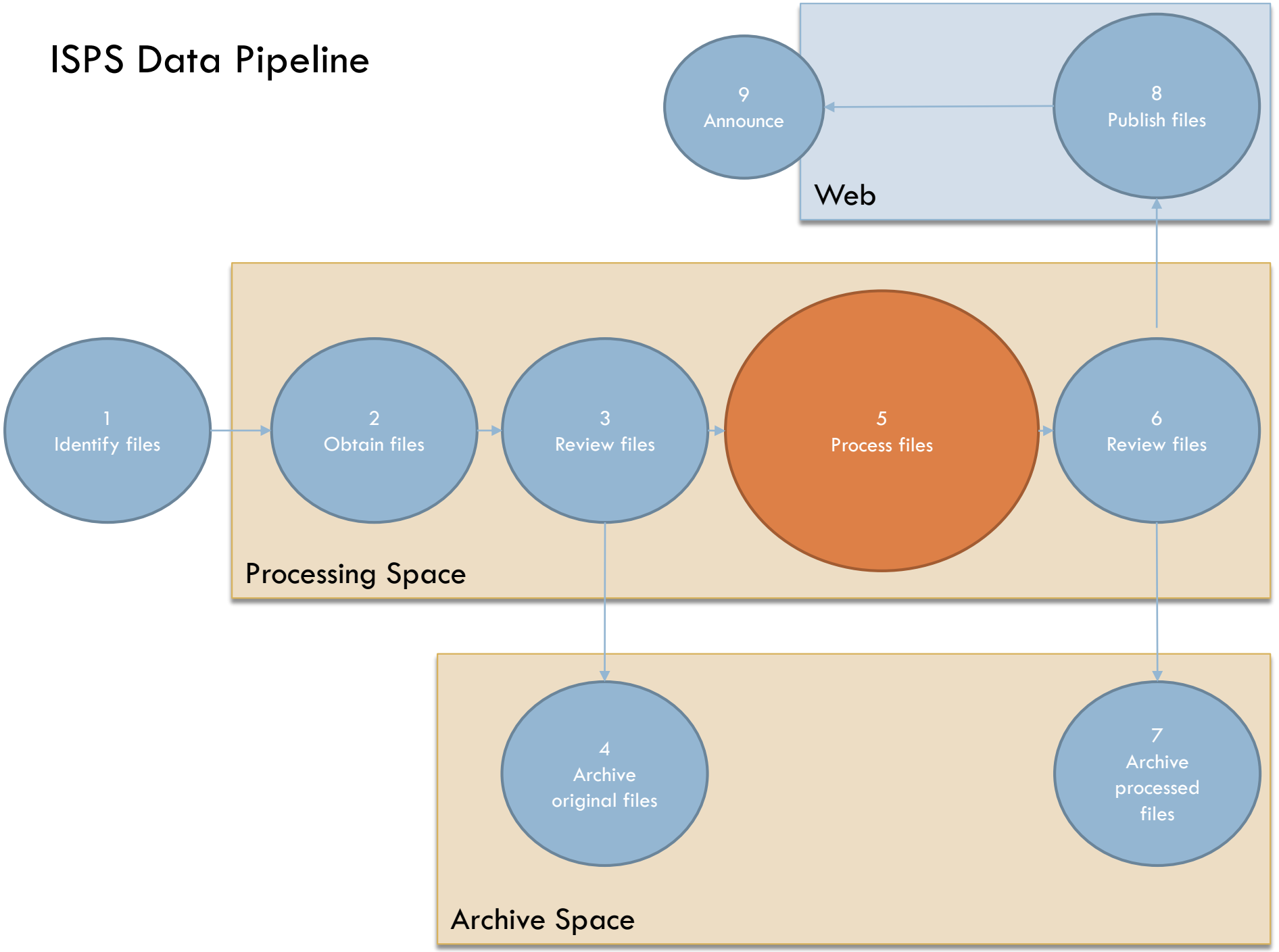
Target Users

- 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)

Tools for curation

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

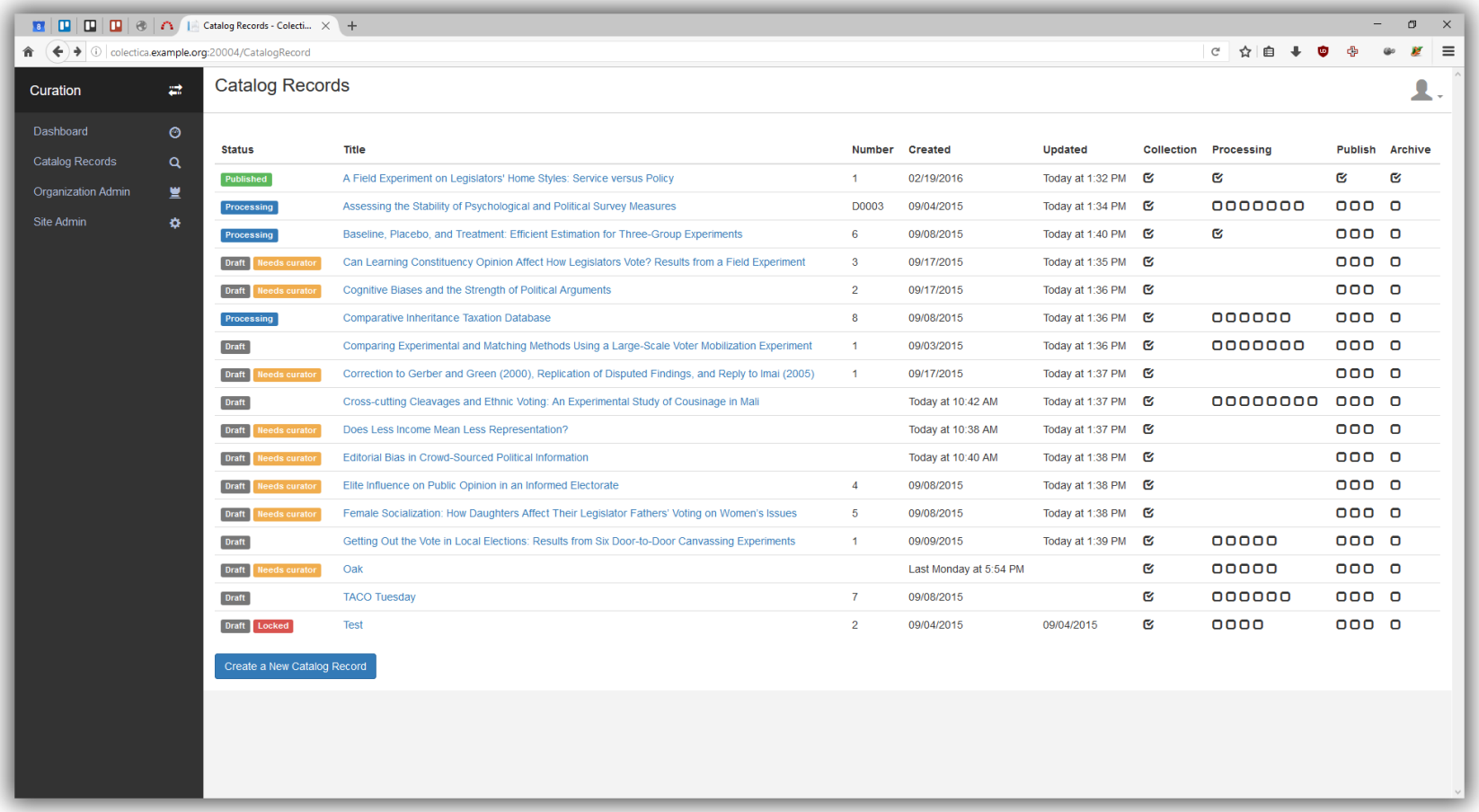
- Disseminate
  - Current: Fedora Commons, Hydra, Colectica Portal
  - Soon: Drupal, Dataverse
  - Possible: Other dissemination platforms

Demonstration

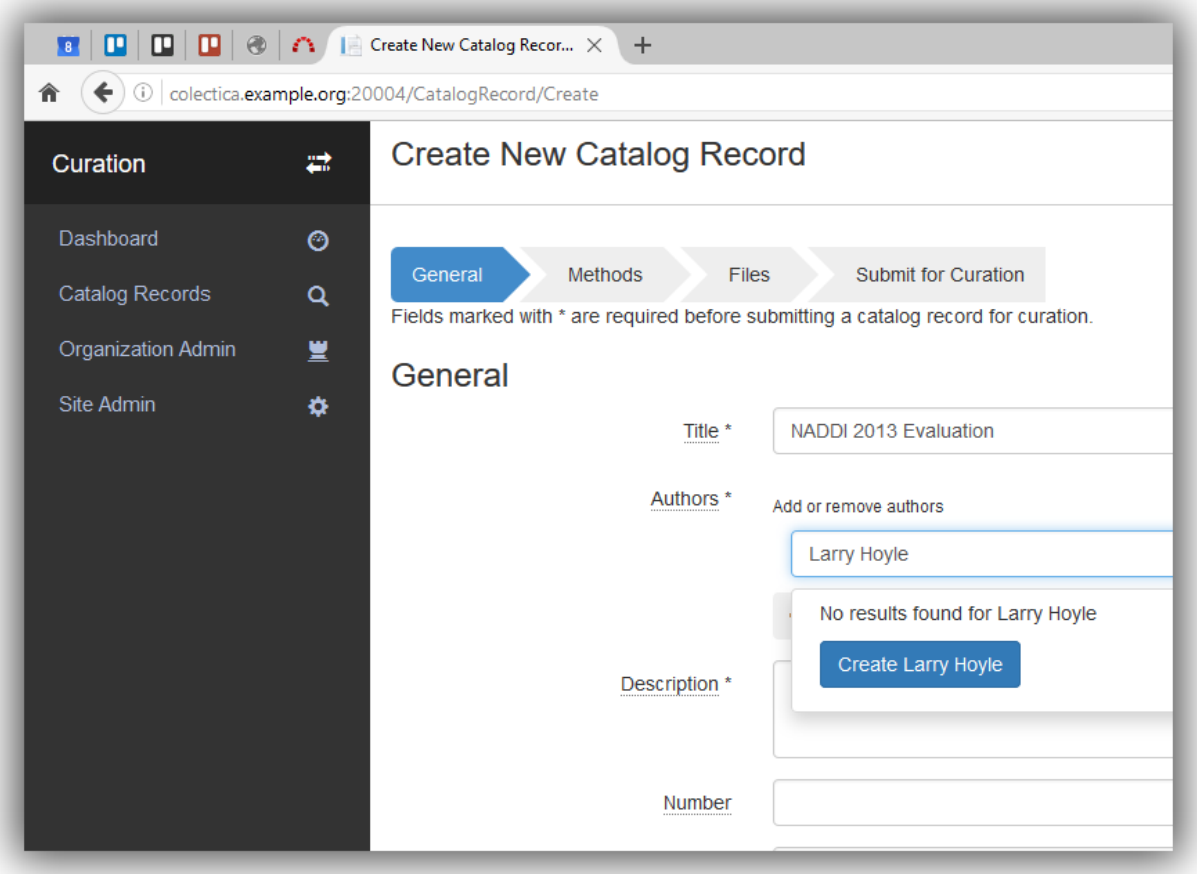
Learn More: Documentation

docs.colectica.com/curation

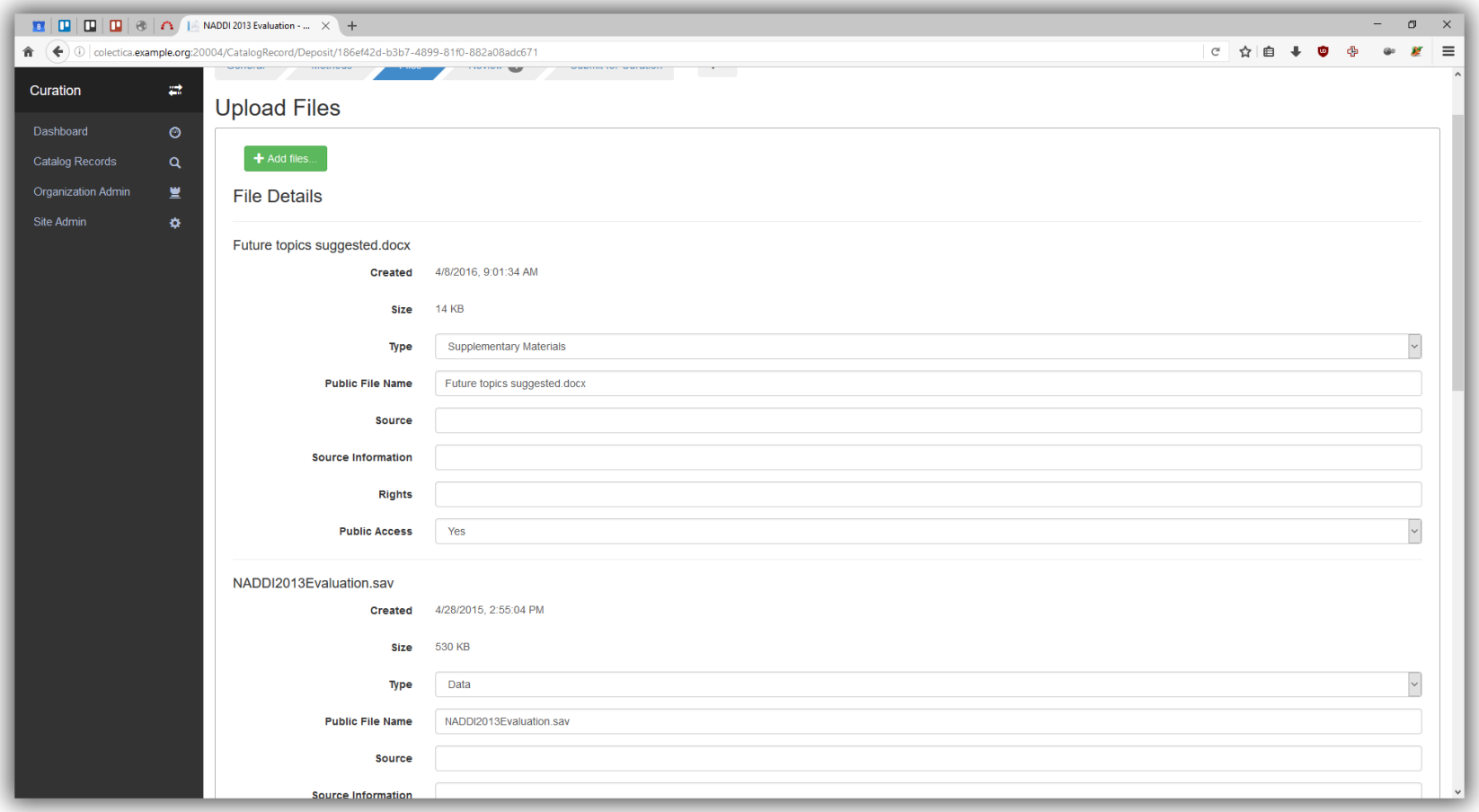
Catalog Records



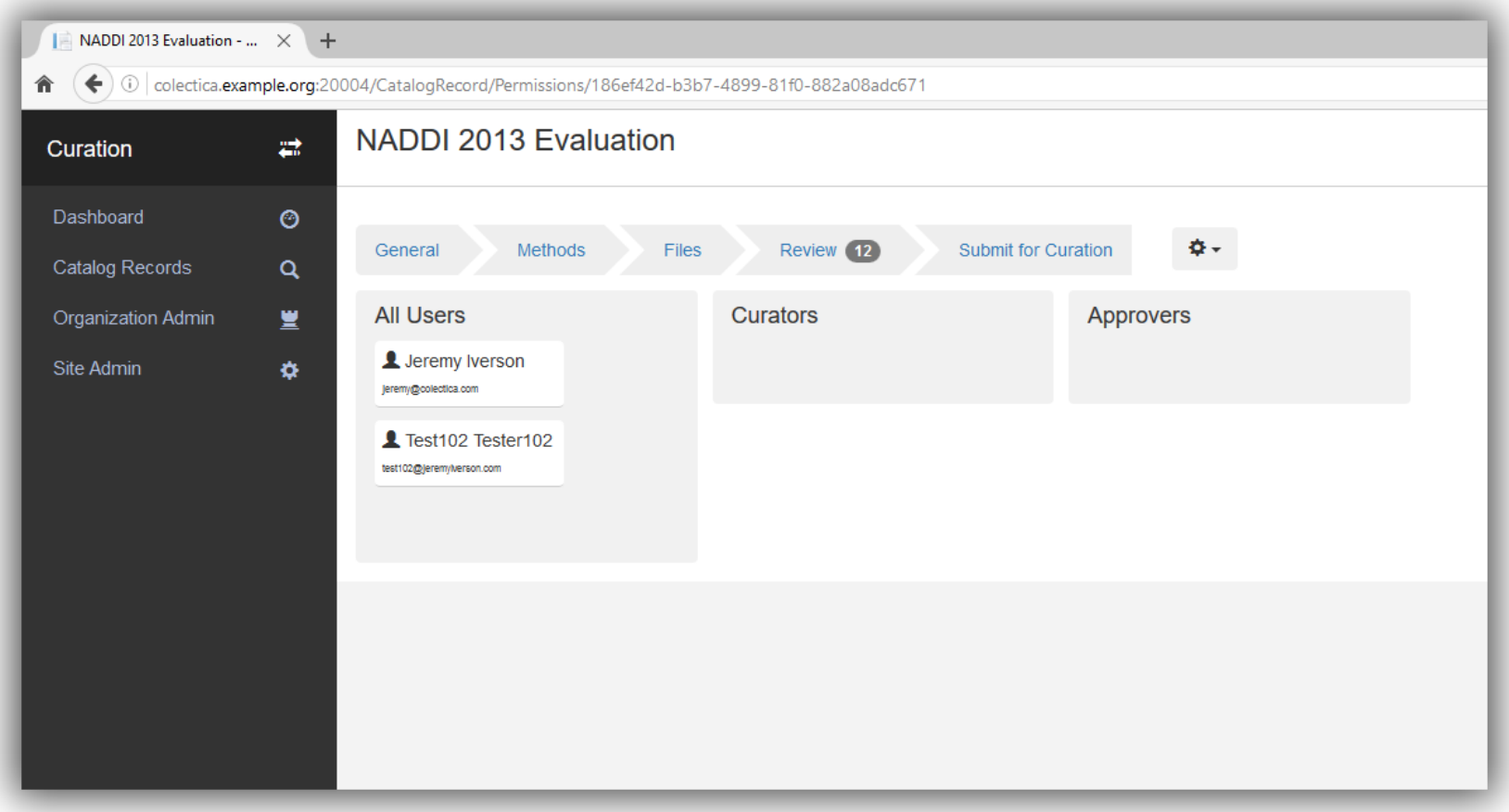
New Record


Create New Catalog Recor...
colectica.example.org:20004/CatalogRecord/Create
Curation
Dashboard
Catalog Records
Organization Admin
Site Admin
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

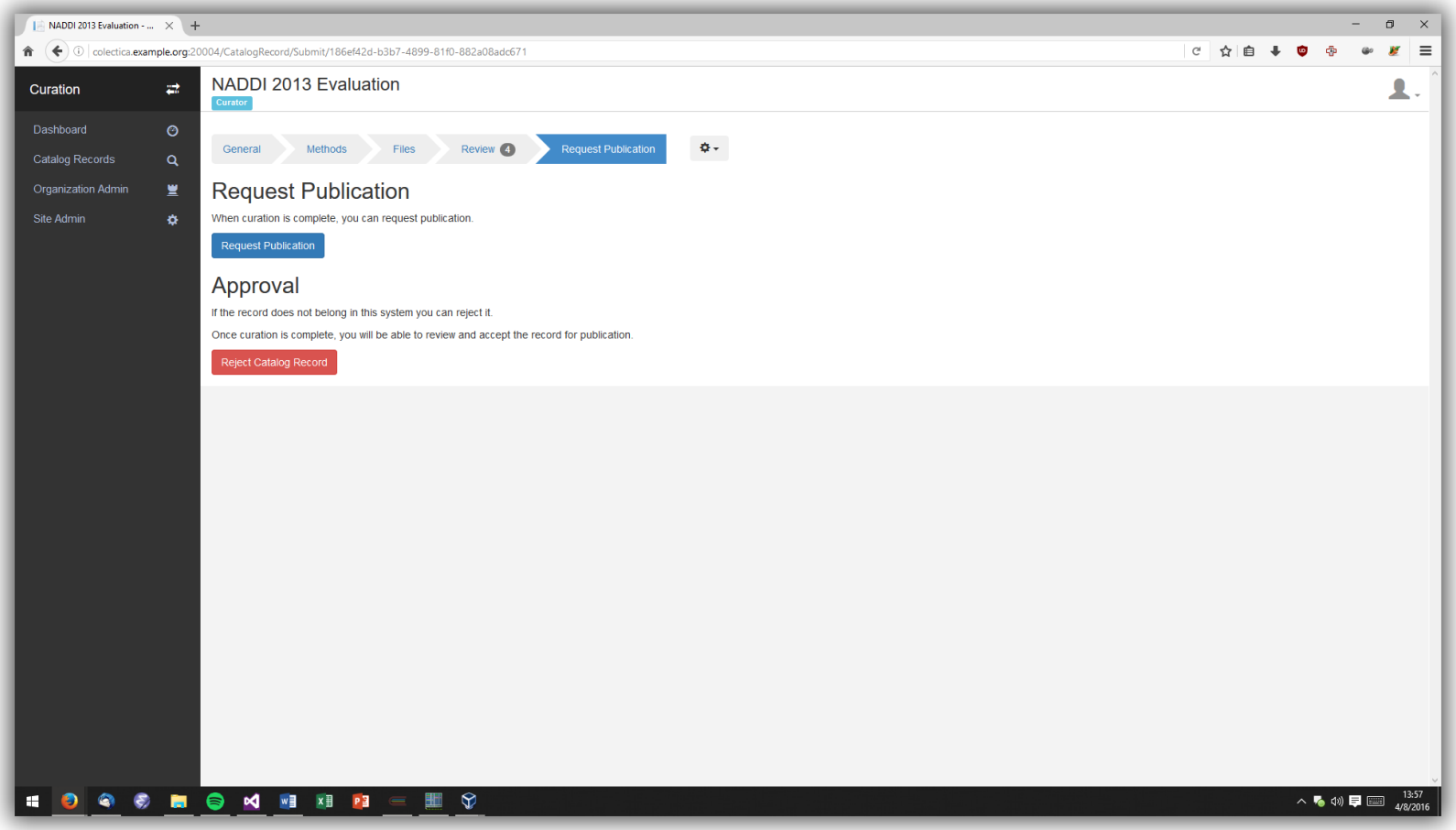
Add Files


NADDI 2013 Evaluation - ...
colectica.example.org:20004/CatalogRecord/Deposit/186ef42d-b3b7-4899-81f0-882a08adc671
Curation
Dashboard
Catalog Records
Organization Admin
Site Admin
Upload Files
+ Add files...
File Details
Future topics suggested.docx
Created
4/8/2016, 9:01:34 AM
Size
14 KB
Type
Supplementary Materials
Public File Name
Future topics suggested.docx
Source
Source Information
Rights
Public Access
Yes
NADDI2013Evaluation.sav
Created
4/28/2015, 2:55:04 PM
Size
530 KB
Type
Data
Public File Name
NADDI2013Evaluation.sav
Source
Source Information

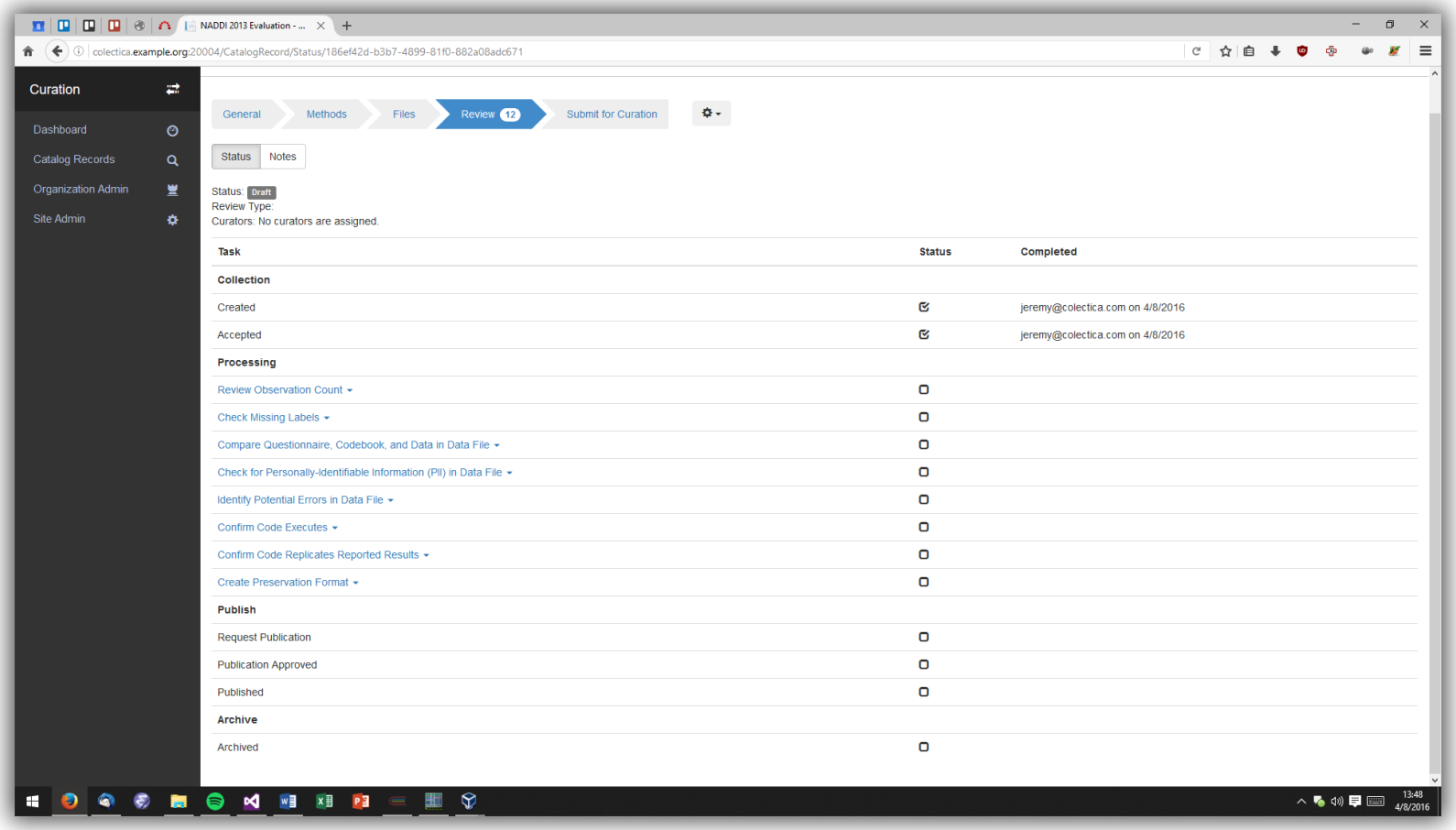
Assign Curators



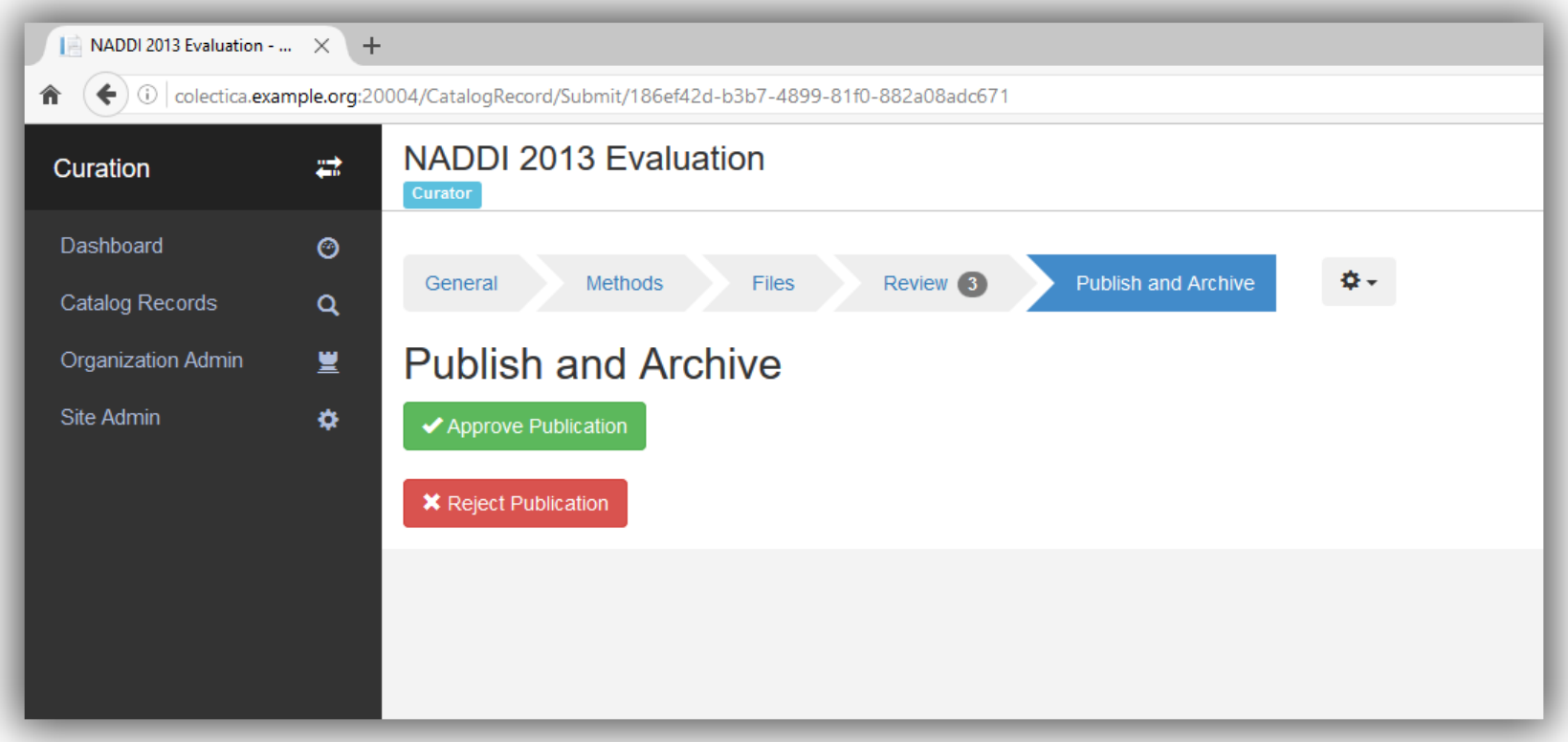
Request Publication



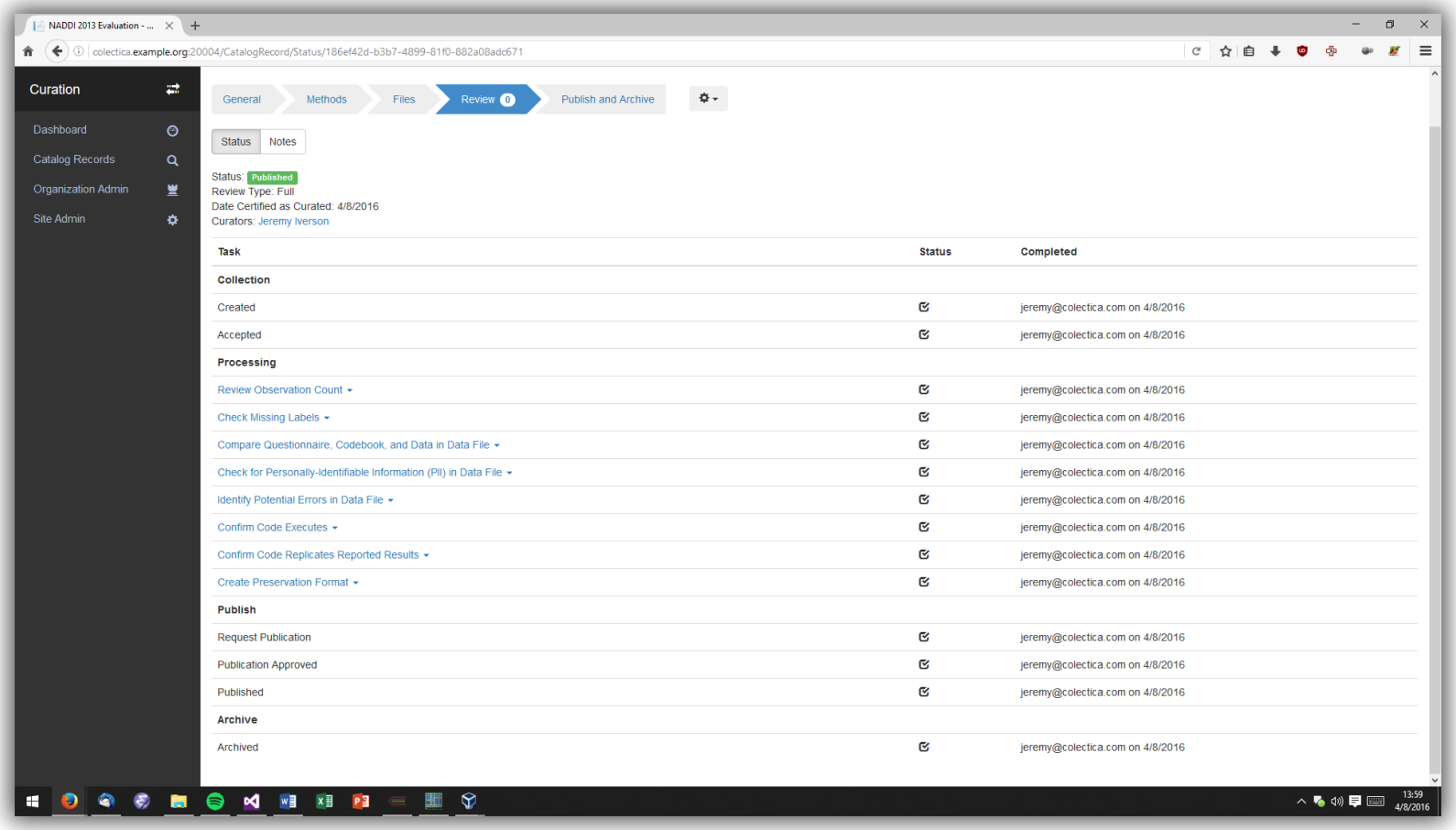
Review Status



Approve Publication



Published


NADDI 2013 Evaluation - ...
colectica.example.org:20004/CatalogRecord/Status/186ef42d-b3b7-4899-81f0-882a08adc671
Curation
Dashboard
Catalog Records
Organization Admin
Site Admin
General
Methods
Files
Review 0
Publish and Archive
Status
Notes
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
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
13:59
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