



Yale University

YARD: Yale Application for Research Data Curation Tool Integration

Bethany Seeger

Technical Lead for Research, Learning, and Scholarly Resources, Yale Library

Limor Peer

Associate Director for Research and Strategic Initiatives at the Institution for Social and Policy Studies & Lead, Open and Reproducible Research Program at the Data-Intensive Social Science Center, Yale University

Dataverse Community Meeting, May 12-15, 2026, Barcelona, Spain

ISPS

Institution for Social and Policy Studies

<https://isps.yale.edu/>



ISPS was founded in 1968 as an interdisciplinary center to support social science and public policy at Yale.

ISPS and Yale Library have collaborated on Yale's Dataverse since the start of it, with ISPS being a key stakeholder.

The ISPS Data Archive includes 147 Studies (replication packages) comprising 2,547 Files.

All materials are processed in YARD before being published.

ISPS requested to move the collection to Dataverse.

Presented about YARD at Dataverse Community Meeting in 2019.

What is YARD?

<https://yard.yale.edu/>

<http://doi.org/10.5334/dsj-2020-028>



Yale Application for Research Data (YARD) is a workflow tool created by ISPS and used for reviewing and enhancing research outputs before publication.

It generates DDI Lifecycle metadata (study-, file-, variable-level).

In addition, YARD prompts curators to take actions specific to file type, for example, to perform PII checks and to verify computational reproducibility, and logs these actions.

In this way, YARD complements the functionality of other potential curation tools currently integrated with Dataverse.

YARD has been actively used since 2011, publishing reviewed data and code to a Drupal website.

<https://github.com/Colectica/curation>

What is YARD?

<https://yard.yale.edu/>

<http://doi.org/10.5334/dsj-2020-028>



Web-based application to deliver highly curated research bundles to designated repositories.

Guides depositors and curators through a series of tasks for reviewing and enhancing research outputs, tracks these curation tasks, and generates rich metadata.

Designed to be fully customizable to accommodate various workflows (e.g., curation tasks, user roles, standardized metadata schema, study-level information).

Why YARD?

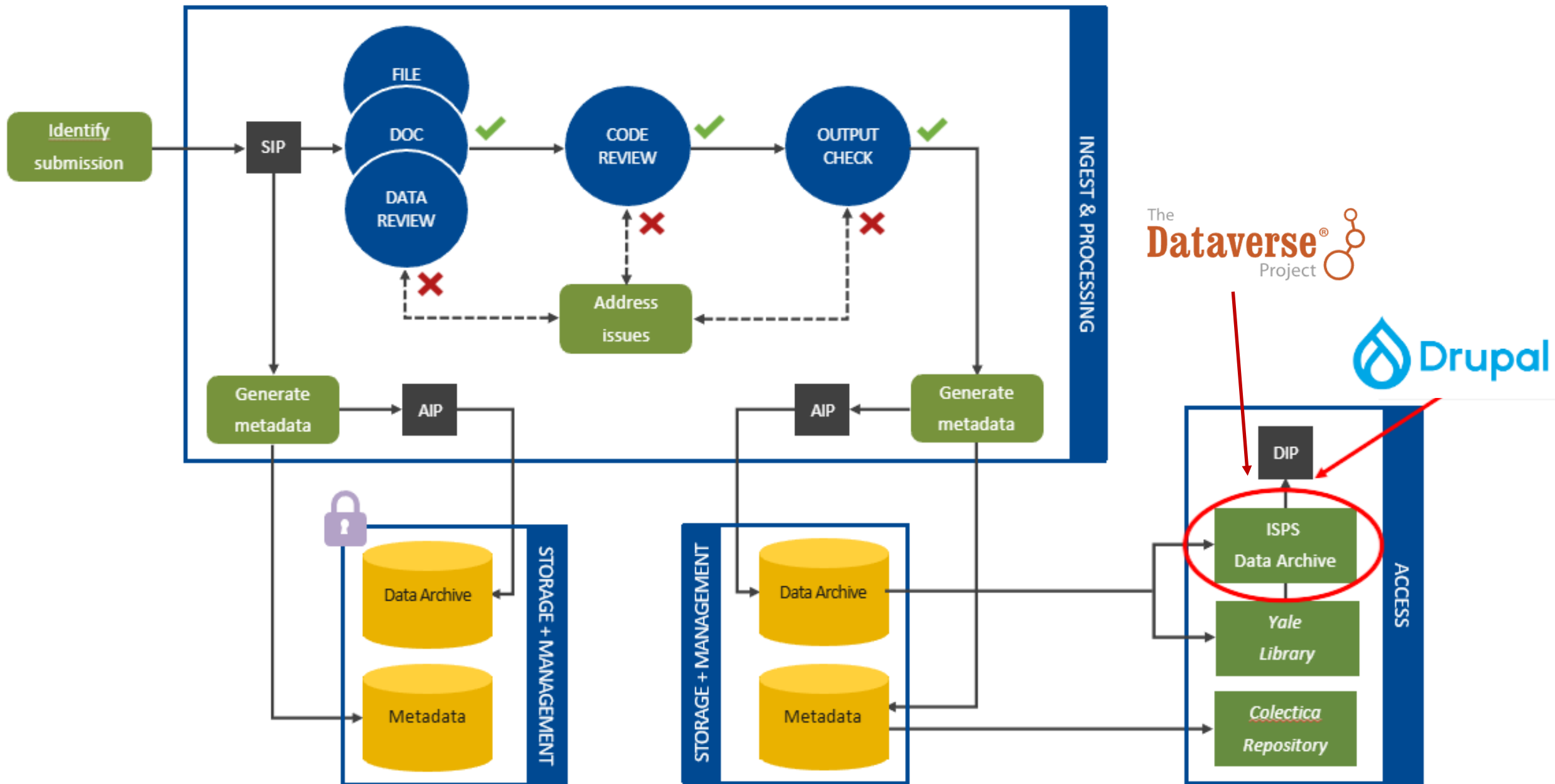
Provides steps for (some automated, but mostly manual) data quality review to help prevent:

- Insufficient documentation
- Missing variable labels
- Deviations in number of observations
- Unavailable software extensions
- Omitted code
- Sharing personally-identifiable information

Gold standard data curation + code review: YARD enables better reproducibility

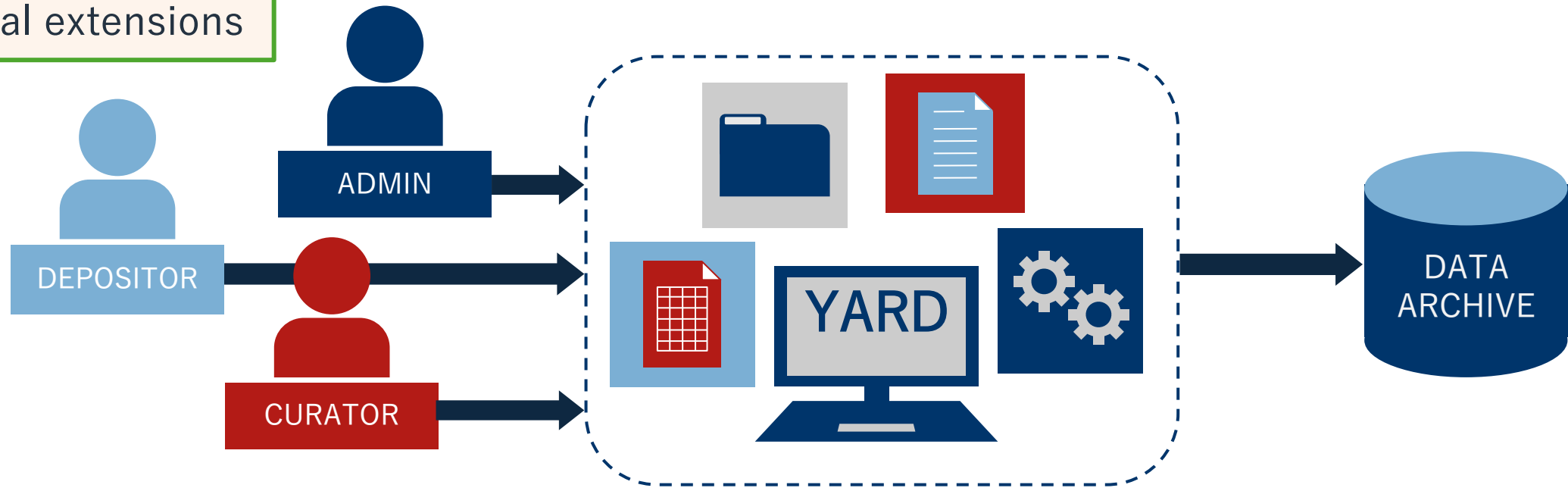
Main features:

- Curation workflow actions are logged
- Rich DDI Metadata is produced
- User roles: Depositor, Curator, Administrator



YARD deployment

Potential extensions



Refine^{OPEN}

GitHub

EaaS

OpenSci

OSF

The Dataverse[®]
Project

DRYAD

CODE OCEAN

DataCite

ReproZip

open
ICPSR

figshare
credit for all your research

Integration Process

Team:

- Limor Peer, Product Owner
- Colectica developer
- Technical Lead for Dataverse
- Metadata Specialist

Metadata was the most important part and took the most time to get right.

Methods:

- Regular weekly meetings
 - Aligned on vision and scope
- Crafted metadata mapping to create a block
- Shared meeting notes
- GitHub for Issue tracking

Work done:

On the Dataverse side

- Metadata needed to be augmented by adding a new block specific to ISPS data.
- Minor tweaks added, like file categories and facets changed
- Reprint Handles to new DOIs

On the YARD side:

- Internal changes to fields and how files were managed.
- Integration code needed to be written to push material to Dataverse when it was published in YARD.

Metadata

Mapping Metadata

- Metadata cleaning up and refining data
 - Cleanup / refinement of choices in YARD.
 - Normalization of some fields.
 - Required fields that really didn't need to be required ("none", "n/a", "none known")
 - Values that didn't fit Dataverse style
 - "other" with text box
- ISPS Metadata Block
- Blocks enabled to cover mapping:
 - ISPS – new block
 - CodeMeta
 - Social Science and Humanities
 - Geospatial

ISPS Metadata ^	
Archive Date ?	2026-04-23
Certified Date ?	2026-04-23
ISPS Version ?	1
Research Design ?	Survey Experiment
Mode of Data Collection ?	Survey - Web
Review Type ?	Partial - YARD data or code review
Unit of Observation ?	Individual
Outcome Measures ?	All subjects received a battery of outcome questions asking to what extent they governor (Study Four) to do actions that violate democratic principles; Candidate in Studies 2-4; Candidate electability in Studies 3-4
Randomization Procedure ?	subjects were randomly assigned to a control condition or a treatment condition. (made-up) politician bemoaning a unique issue: the Coronavirus (Study One), st or accurate vote counting (Study Four).
Unit of Randomization ?	Individual
Treatment ?	Each condition included a speech excerpt from a (made-up) politician bemoanin student debt (Study Two), the filibuster (Study Three), or accurate vote counting
Treatment Administration ?	Web

Curation

- Dashboard
- Catalog Records
- Depositor Guide
- Curator Guide
- Organization Admin
- Administrator Guide
- Site Admin
- Site Admin Guide

Candidate Choice Without Party Labels: New Insights from Conjoint Survey Experiments

Processing Curator Approver



General **Methods** Files Review **6** Request Publication

Methods

Research Design *

Other Research Design

Mode of Data Collection *

Field Dates * Year/Month 2015-03 range

Study Time Period * Year 2015 range

Location * United States

Location Details The whole nation of the US

Unit of Observation *

Other Unit of Observation

Sample Size * MTurk: 12,032 individual-election pairs (1,204 individuals) You

Inclusion/Exclusion Criteria * Observational study: People who voted in the 1010 US mayo

Randomization Procedure Subjects were assigned to conjoint profiles using simple rand

Unit of Randomization

Intervention Treatments are defined as the profiles of candidates subjects

Intervention Administration

Other Intervention Administration

Outcome Measures *

Separate multiple outcome measures with a comma.

Curation

- Dashboard
- Catalog Records
- Depositor Guide
- Curator Guide
- Organization Admin
- Administrator Guide
- Site Admin
- Site Admin Guide

Candidate Choice Without Party Labels: New Insights from Conjoint Survey Experiments

Processing Curator Approver

General Methods Files **Review 6** Request Publication

Status Notes

Status: Processing
Review Type: Full
Curators: Limor Peer

Task	Status	Completed
Collection		
Created	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 8/4/2018
Accepted	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 8/4/2018
Processing		
Review Observation Count	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 2/2/2020
Compare Questionnaire, Codebook, and Data in Data File	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 2/2/2020
Check for Personally-Identifiable Information (PII) in Data File	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 1/20/2020
Identify Potential Errors in Data File	<input checked="" type="checkbox"/>	<input type="checkbox"/> [redacted]@yale.edu on 2/2/2020
Confirm Code Executes		5 / 6
Confirm Code Replicates Reported Results		5 / 6
Publish		
Request Publication	<input type="checkbox"/>	
Publication Approved	<input type="checkbox"/>	
Published	<input type="checkbox"/>	
Archive		
Archived	<input type="checkbox"/>	

Future Work Highlights

<https://dataverse.yale.edu/dataverse/isps>

- Review required metadata fields in YARD
- Fix/cleanup some fields – Related Materials is a string but should be a link. Drupal will render this, but Dataverse cannot.
- Reorganize Dataset Terms in YARD to be at Dataset level.
- Files
 - Allow "Restricted" files in YARD, indicate Terms on file
 - Add directory structure for ingest in Dataverse
 - Enables having sub folders
 - File variables
- Dataverse placeholder records to retrieve DOI

Hopeful: work with the community on a reproducibility metadata block

Eventually expand usage at Yale, beyond the social science group.

Questions?

Bethany Seeger

[*bethany.seeger@yale.edu*](mailto:bethany.seeger@yale.edu)

Limor Peer

[*limor.peer@yale.edu*](mailto:limor.peer@yale.edu)

*Youn Noh, Research Information
Management Librarian*

[*Youn.noh@yale.edu*](mailto:Youn.noh@yale.edu)

Yale University



Photo by Paul Coco