

# YARD Work: Data Archiving and Publishing at Yale

**Limor Peer, Joshua Dull, Themba Flowers**

*Yale Digital Conference*

June 15, 2018

# Outline

Yale Institution for Social and Policy Studies

[isps.yale.edu](http://isps.yale.edu)

ISPS Data Archive

Curation Tool: Yale Application for Research Data (YARD)

YARD-Drupal integration

# Institution for Social and Policy Studies

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





# Institution for Social and Policy Studies

ADVANCING RESEARCH • SHAPING POLICY • DEVELOPING LEADERS

Research

Events

News

Team

About

Programs



## Study Shows Op-Eds Change Minds

Alex Coppock's new research in the Quarterly Journal of Political Science

[READ MORE ▶](#)

# ISPS Data Archive

An open access digital collection of social science experimental data, metadata, code, and associated files produced by ISPS researchers, for the purpose of replication of research findings, further analysis, and teaching.

Peer, L., & Green, A. (2012). Building an Open Data Repository for a Specialized Research Community: Process, Challenges, and Lessons. *International Journal of Digital Curation* 7(1), 151–162.  
<http://dx.doi.org/10.2218/ijdc.v7i1.222>

# IS PS Data Archive

Author: - Any -  
Discipline: - Any -  
Keywords: - Any -  
Area of Study: - Any -  
Location: - Any -  
Year: -Year -  
Research design: - Any -  
SEARCH

| <u>TITLE</u>                                                                                                            | <u>AUTHOR(S)</u>                                  | <u>YEAR ARCHIVED</u> ▾ |
|-------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|------------------------|
| <a href="#">Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment</a> | Alexander Coppock                                 | 2017                   |
| <a href="#">Trading Barriers: Firms, Immigration and the Remaking of Globalization</a>                                  | Margaret E. Peters                                | 2017                   |
| <a href="#">The Majority-Minority Divide in Attitudes Toward Internal Migration: Evidence from Mumbai</a>               | Nikhari Gaikwad and Gareth Nellis                 | 2017                   |
| <a href="#">Chocolate Scents and Product Sales: A Randomized Controlled Trial in a Canadian Bookstore and Café</a>      | Mary C. McGrath, Peter M. Aronow, Vivien Shotwell | 2017                   |

Since 2011  
Open access  
Creative Commons license  
Website integration

Specialized community  
Experimental design  
Quantitative data

90 studies  
1,400 files  
15 GB

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

# Replication Standard

The *replication standard* holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author.

King, G. (1995). Replication, replication. *PS: Political Science & Politics*, 28(3), 444–452. <http://doi.org/10.2307/420301>

# Common Replication Problems

- Insufficient documentation
- Missing variables
- Deviations in number of observations
- Unavailable software extensions
- Omitted code
- Incompatible datasets

# Curating for Reproducibility

Data Curation: Maintaining, preserving and adding value to digital research data throughout its lifecycle... [which] reduces threats to their long-term research value and mitigates the risk of digital obsolescence.

Digital Curation Center <http://www.dcc.ac.uk/digital-curation/what-digital-curation>

Reproducibility: Calculation of quantitative scientific results by independent scientists using the original datasets and methods.

Stodden, V. (Ed.), Leisch, F. (Ed.), Peng, R.D. (Ed.). (2014). *Implementing Reproducible Research*. New York: Chapman and Hall/CRC.

# Curating for Reproducibility

## Data Quality Review



## Data Quality Review



- ✓ Assign persistent identifier
- ✓ Create study citation and study-level metadata record
- ✓ Record file size details
- ✓ Check for presence of all files
- ✓ Verify content of files matches expected format
- ✓ Create non-proprietary versions of files
- ✓ Implement migration strategy for file formats

## Data Quality Review



- ✓ Confirm presence of comprehensive descriptive information necessary for informed reuse
  - Data definitions
  - Variable construction
  - Methodology
  - Sampling information
  - Original data source citation
  - Analysis software version
- ✓ Link to related research products

# Curating for Reproducibility

## Data Quality Review



- ✓ Check for undocumented variable and value information
- ✓ Examine data for inconsistencies and errors
  - Discrepancies in number of observations
  - Out-of-range or wild codes
  - Undefined null values
- ✓ Review data for confidentiality issues

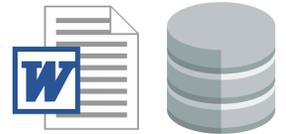
## Data Quality Review



- ✓ Convert absolute file paths to relative file paths
- ✓ Check code for presence of non-executable comments that document analysis processes
- ✓ Identify packages required to execute code
- ✓ Execute code to ensure code is error-free
- ✓ Compare code output to findings presented in article

# Typical ISPS 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



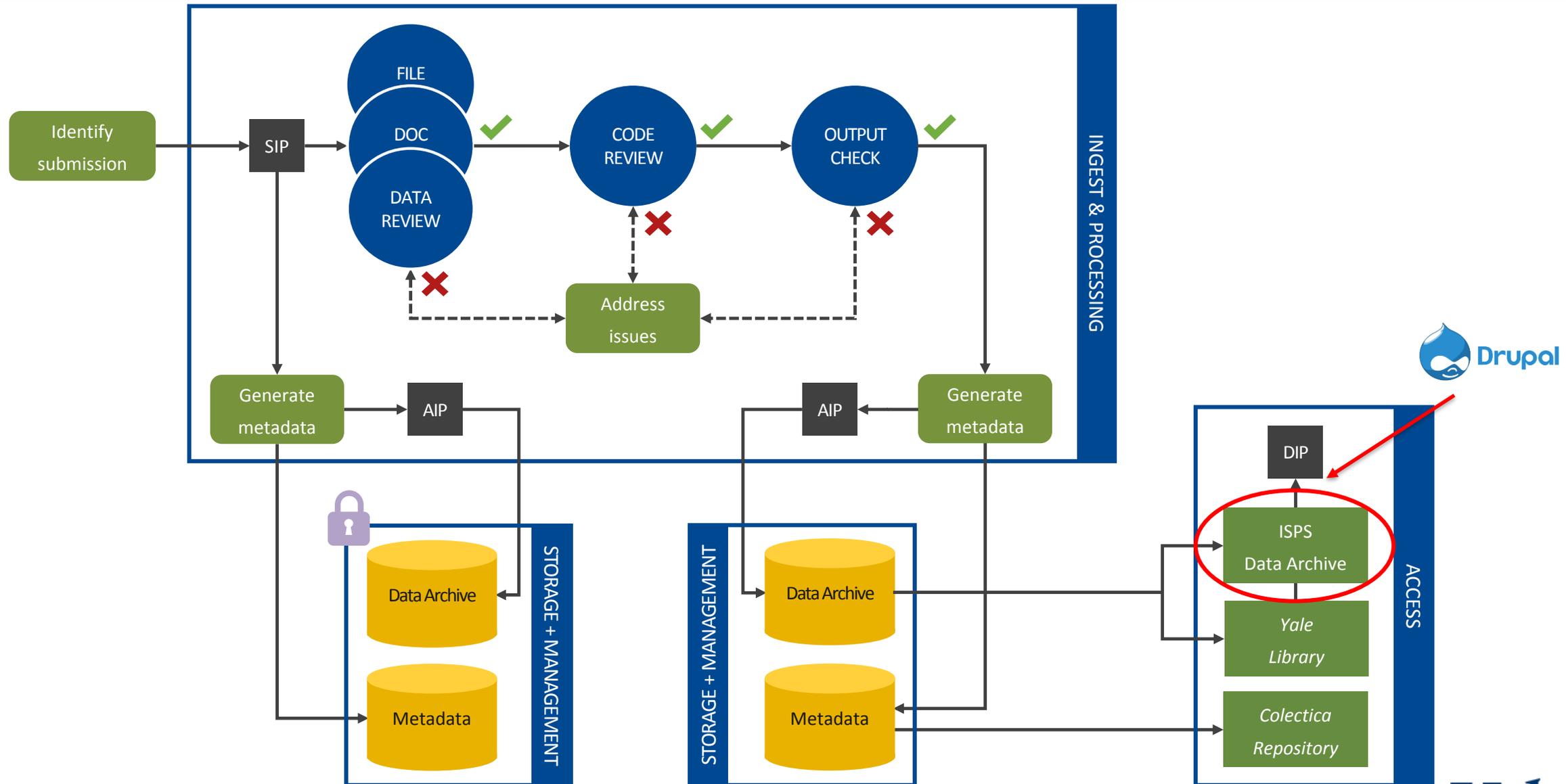
Storage@Yale



Yale

**... and now share the data!**

# Curating for Reproducibility at IS PS



# Curation Tool: YARD

## **YARD: Yale Application for Research Data**

A new workflow tool that allows Depositors, Curators, and Administrators to submit, review, process, and publish data within one system.

The software structures the curation and review workflow and all actions are recorded in the system.

The tool integrates and captures DDI metadata production with data and code review and cleaning.

Processed data packages are directed to pre-specified destinations.



# Curation Tool: YARD

## Yale Application for Research Data

Log in

---

**Yale**   
**ISPS**

Log in to the ISPS Data Curation Tool with your username and password.

Don't have a ISPS Data Curation Tool account?  
[Create an account.](#)

**Email**

**Password**

Remember me

[Log in](#)

[Forgot your password?](#)

Documentation: <https://docs.colectica.com/curation/>

# Curation Tool: YARD

Three roles:

- Depositor: Any User
- Curator: Trained Archive Staff
- Administrator/Approver: Archive Manager

# Curation Tool: YARD

## Data Deposit

### Create New Catalog Record

Draft

General Methods Files Submit for Curation

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

#### General

Title \*

Authors \* Add or remove authors

Description \*

Number

Keywords \*   
Separate multiple outcome measures with a comma.

#### Citation

Funding \*

#### Access

Access Statement \*

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Deposit

Research Study 2 Draft 

General **Methods** Files Notes Submit for Curation

### Methods

**Research Design \*** Natural experiment

**Mode of Data Collection \***  Interview: Face to Face  Interview: email

**Field Dates \*** Date   range

**Study Time Period \*** Date   range

**Location \*** Africa

**Location Details**

**Unit of Observation** Individual

**Sample Size** 5000

**Inclusion/Exclusion Criteria**

**Randomization Procedure** None

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Deposit

The screenshot shows the 'Research Study 2' interface in a 'Draft' state. A navigation bar at the top includes 'General', 'Methods', 'Files' (the active tab), 'Notes', and 'Submit for Curation'. Below this is a 'Files' section with a table listing four files. Each file row includes a file icon, the filename, version number, type, status, and a delete button. At the bottom of the table are two buttons: 'Add or Update Files' and 'Download All Files'.

| File                              | Version | Type                    | Status       |  |
|-----------------------------------|---------|-------------------------|--------------|--|
| 02_PerformAnalysis.do             | 1       | Program                 | Accepted<br> |  |
| POQ_PublicReplicationDatafile.dta | 1       | Data                    | Accepted<br> |  |
| testdattext.docx                  | 1       | Supplementary Materials | Accepted<br> |  |
| README.txt                        | 1       | Supplementary Materials | Accepted     |  |

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Deposit

Research Study 2 Draft 

General Methods Files Notes **Submit for Curation**

### Submit for Curation

Your record is ready for curation. Click the button below to submit it.

### Deposit Agreement

ISPS Data Archive: Data Deposit Agreement  
—Updated October 5, 2016—  
—Under General Counsel review—  
Terms of Data Deposit Agreement

Your decision to deposit data in the ISPS Data Archive requires that you accept the terms of this Data Deposit Agreement. This agreement grants permission to Yale University, on behalf of ISPS, to (a) archive files associated with scholarly studies, and (b) to distribute these files for the purpose of replication (hereafter, Replication Files).

Specifically, by using the ISPS Data Archive, you accept the terms of this agreement, and you (or your proxy, on your behalf) agree to the following (please check each item if you agree):

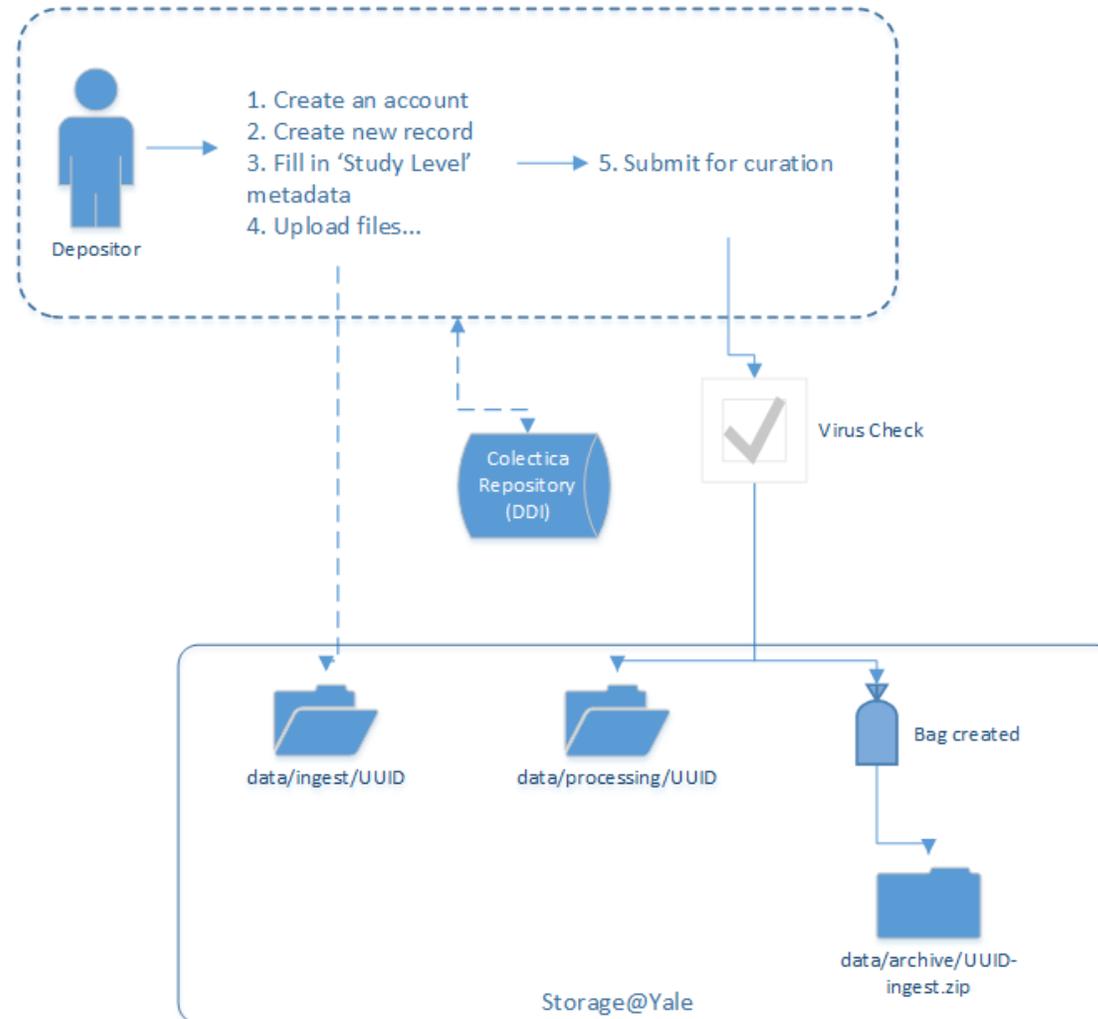
- (1) You accept and agree to abide by ISPS Terms of Use <link to <http://isps.yale.edu/research/data/terms-of-use> >.
- (2) You agree that you will comply with all applicable laws and licenses.
- (3) If you are not the data owner, you represent that you have obtained the unrestricted permission of the owner to grant Yale University, on behalf of ISPS, the rights required by this license, that you will inform the data owner that ISPS administrators have access to all content, for purposes of maintaining the system and assisting users, in accordance with the policies specified here (ISPS does not make data available by other means such as FTP, removable media, etc.), and that the owner of the material will be clearly identified and acknowledged within the text or content of the submission.

Important: The ISPS Data Archive expects all data contributors to comply with applicable copyright laws. Use of third-party copyrighted material without permission or that does not fall under one of the statutory exceptions, including but not limited to fair use, is illegal, unethical, and exposes Yale University to significant financial liability.

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Deposit



# Curation Tool: YARD

## Data Curation

The screenshot shows the 'Research Study' interface in the YARD tool. At the top, it indicates the study is in 'Processing' status and is being handled by a 'Curator'. A progress bar shows the current step is 'Review' with 12 items, following 'General', 'Methods', and 'Files', and preceding 'Request Publication'. Below the progress bar, there are tabs for 'Status' and 'Notes'. The 'Status' section shows: Status: Processing, Review Type: Full, and Curators: Curator Demo.

| Task                                                             | Status                              | Completed                             |
|------------------------------------------------------------------|-------------------------------------|---------------------------------------|
| <b>Collection</b>                                                |                                     |                                       |
| Created                                                          | <input checked="" type="checkbox"/> | deposit.demo@example.edu on 5/12/2017 |
| Accepted                                                         | <input checked="" type="checkbox"/> | deposit.demo@example.edu on 5/12/2017 |
| <b>Processing</b>                                                |                                     |                                       |
| Review Observation Count                                         | <input type="checkbox"/>            |                                       |
| Check Missing Labels                                             | <input type="checkbox"/>            |                                       |
| Compare Questionnaire, Codebook, and Data in Data File           | <input type="checkbox"/>            |                                       |
| Check for Personally-Identifiable Information (PII) in Data File | <input type="checkbox"/>            |                                       |
| Identify Potential Errors in Data File                           | <input type="checkbox"/>            |                                       |
| Confirm Code Executes                                            | <input type="checkbox"/>            |                                       |
| Confirm Code Executes, Reported Results                          | <input type="checkbox"/>            |                                       |

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Curation

The screenshot displays the YARD Data Curation interface for the file 'POQ\_PublicReplicationDatafile.dta'. The breadcrumb trail is 'Research Study > Processing > Curator'. The navigation tabs are 'General', 'Variables', 'Notes', and 'Review' (with a '5' indicator). A settings gear icon is also present.

The left sidebar lists several variables:

- weight (case weight)
- yal070 (interpersonal trust)
- yal071 (government trust)
- yal072 (political efficacy - external)
- yal073 (political efficacy - internal)
- yal130 (voting private matter)** - currently selected

The main content area shows the details for the selected variable 'yal130':

- Name:** yal130
- Label:** voting private matter
- Version:** 1
- Last Updated:** 5/12/2017
- Description:** Empty
- Response Unit:** Empty
- Analysis Unit:** Empty
- Variable Type:** Unspecified
- Representation:** Code

Below the details is a 'Frequencies' section with a table:

| Value | Label                                                                         | Frequency |
|-------|-------------------------------------------------------------------------------|-----------|
| 1     | democracy works best when people treat their vote choices as personal matters | 556       |

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Curation

The screenshot displays the 'Request Publication' step in the YARD Data Curation tool. The interface includes a top navigation bar with 'Research Study', 'Processing', and 'Curator' tabs, and a user profile icon. A progress bar below shows steps: General, Methods, Files, Review (with a '11' notification), and Request Publication (highlighted in blue). A settings gear icon is also present. The main content area is titled 'Request Publication' and contains instructions and a list of incomplete tasks.

**Research Study**  
Processing Curator

General Methods Files Review **11** Request Publication

### Request Publication

When curation is complete, you can request publication.

Please complete the following fields before requesting publication.

- [Incomplete task for 02\\_PerformAnalysis.do: Confirm Code Executes](#)
- [Incomplete task for testdattext.docx: Create Preservation Format](#)
- [Incomplete task for POQ\\_PublicReplicationDatafile.dta: Check Missing Labels](#)
- [Incomplete task for POQ\\_PublicReplicationDatafile.dta: Identify Potential Errors in Data File](#)
- [Incomplete task for POQ\\_PublicReplicationDatafile.dta: Check for Personally-Identifiable Information \(PII\) in Data File](#)
- [Incomplete task for 02\\_PerformAnalysis.do: Confirm Code Replicates Reported Results](#)
- [Incomplete task for POQ\\_PublicReplicationDatafile.dta: Compare Questionnaire, Codebook, and Data in Data File](#)
- [02\\_PerformAnalysis.do: Software](#)
- [testdattext.docx: Software](#)
- [POQ\\_PublicReplicationDatafile.dta: Data Type](#)
- [testdattext.pdf: Software](#)
- [README.txt: Software](#)

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Publication

The screenshot displays the YARD Curation Tool interface. On the left is a dark sidebar with navigation options: Curation, Dashboard, Catalog Records, Organization Admin, and Site Admin. The main content area shows the file name 'Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta' and a breadcrumb trail: <a href='\"#\">Test October 22</a>. Below this is a workflow progress bar with steps: General, Variables, Notes, and Review (5). A settings gear icon is also present. The 'File History' section contains three entries, each with a timestamp, user email, and a task description:

- Edit a File**  
Data Type changed from blank to Survey/interview (e.g., ANES)  
Source changed from blank to blank  
[- Butler\\_Dynes\\_AJPS\\_2015a\\_StateLegislators.dta](\"#\">Test October 22</a><ul><li>Butler_Dynes_AJPS_2015a_StateLegislators.dta</li></ul></li><li><b>Accept task: Identify Potential Errors in Data File</b><br/>No obvious data errors to flag<br/><a href=)
- Accept task: Check for Personally-Identifiable Information (PII) in Data File**  
No PII

Depositor  
Curator  
Approver

# Curation Tool: YARD

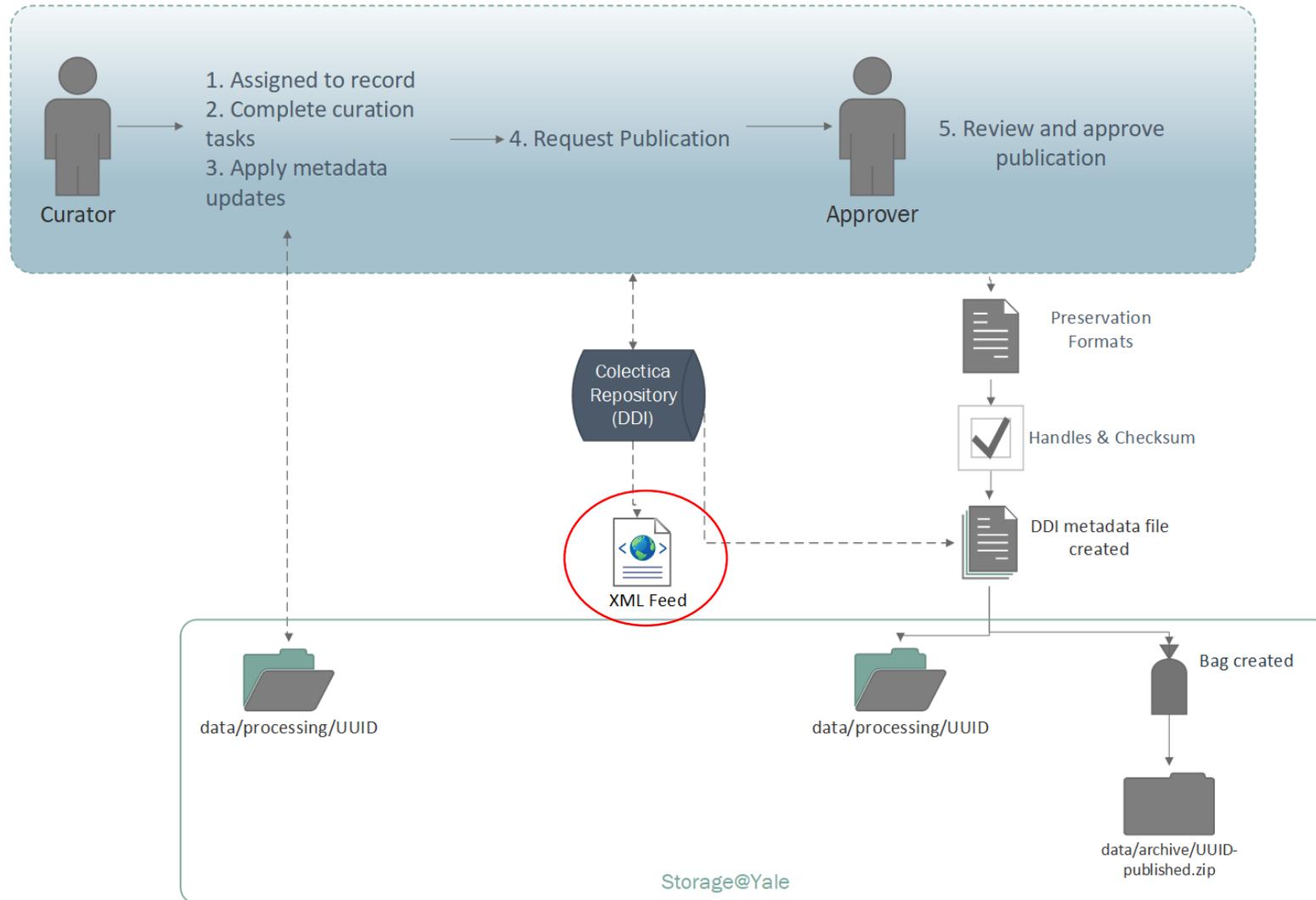
## Data Publication

The screenshot shows the 'Publish and Archive' step in the YARD curation tool. The interface includes a dark sidebar on the left with navigation icons. The main content area is titled 'Research Study 2' and features a breadcrumb trail: 'Publication Requested' (orange), 'Curator' (blue), and 'Approver' (blue). Below this is a progress bar with steps: 'General', 'Methods', 'Files', 'Review 1', and 'Publish and Archive' (highlighted in blue). A settings gear icon is visible to the right of the progress bar. The 'Publish and Archive' section contains two buttons: a green 'Approve Publication' button with a checkmark and a red 'Reject Publication' button with an 'X'.

Depositor  
Curator  
Approver

# Curation Tool: YARD

## Data Curation & Publication



# Drupal Integration

## XML Feed

```
▼<publishedCatalogRecords>
  ▼<Record>
    <Guid>a5e1a0ef-b84e-446c-bd5d-1a8541333123</Guid>
    ▼<Title>
      Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment
    </Title>
    <Author>Alexander Coppock</Author>
    <Owner/>
    ▼<Description>
      Explanations for the failure to predict Donald Trump's win in the 2016 Presidential election sometimes include the "Shy Trump Supporter" hypothesis, according to which some Trump supporters succumb to social desirability bias and hide their vote preference from pollsters. I evaluate this hypothesis by comparing direct question and list experimental estimates of Trump support in a nationally representative survey of 5,290 American adults fielded from September 2 to September 13, 2016. Of these, 32.5% report supporting Trump's candidacy. A list experiment conducted on the same respondents yields an estimate 29.6%, suggesting that Trump's poll numbers were not artificially deflated by social desirability bias as the list experiment estimate is actually lower than direct question estimate. I further investigate differences across measurement modes for relevant demographic and political subgroups and find no evidence in support of the "Shy Trump Supporter" hypothesis.
    </Description>
    <StudyID>D149</StudyID>
    <StudyIDLower>d149</StudyIDLower>
    ▼<RelatedPublication>
      Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment
    </RelatedPublication>
    <RelatedProject>none</RelatedProject>
    <RelatedDatabase>none</RelatedDatabase>
    <keywords>List Experiment,Election Polling</keywords>
    <CreateDate>2017-06-30T14:36:30.247</CreateDate>
    <ResearchDesign>Survey experiment</ResearchDesign>
    <DataType>Survey/interview (e.g., ANES)</DataType>
    <DataSource>Author</DataSource>
    <DataSourceInformation>isps(at)yale(dot)edu</DataSourceInformation>
    <CatalogRecordDataType>Survey/interview (e.g., ANES)</CatalogRecordDataType>
    <CatalogRecordDataSource>Author</CatalogRecordDataSource>
    <CatalogRecordDataSourceInformation/>
    <PersistentId/>
    <FieldDates>2016-09</FieldDates>
    <Location>United States</Location>
    <LocationDetails>United States</LocationDetails>
    <UnitOfObservation>Individual</UnitOfObservation>
    <SampleSize>5290</SampleSize>
    <InclusionExclusionCriteria>18+</InclusionExclusionCriteria>
    <RandomizedProcedure>Simple random assignment</RandomizedProcedure>
    <Treatment>treatment and control list</Treatment>
    <TreatmentAdministration>Web delivered</TreatmentAdministration>
    <OutcomeMeasures>Number of things a respondent "would do"</OutcomeMeasures>
    <ArchiveDate>2017-12-13T00:00:00</ArchiveDate>
```

# Drupal Integration

## Study Level

```
▼<publishedCatalogRecords>
  ▼<Record>
    <Guid>a5e1a0ef-b84e-446c-bd5d-1a8541333123</Guid>
    ▼<Title>
      Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment
    </Title>
    <Author>Alexander Coppock</Author>
    <Owner/>
    ▼<Description>
      Explanations for the failure to predict Donald Trump's win in the 2016 Presidential election sometimes include the "Shy Trump Supporter" hypothesis, according to which some Trump supporters succumb to social desirability bias and hide their vote preference from pollsters. I evaluate this hypothesis by comparing direct question and list experimental estimates of Trump support in a nationally representative survey of 5,290 American adults fielded from September 2 to September 13, 2016. Of these, 32.5% report supporting Trump's candidacy. A list experiment conducted on the same respondents yields an estimate 29.6%, suggesting that Trump's poll numbers were not artificially deflated by social desirability bias as the list experiment estimate is actually lower than direct question estimate. I further investigate differences across measurement modes for relevant demographic and political subgroups and find no evidence in support of the "Shy Trump Supporter" hypothesis.
    </Description>
    <StudyID>D149</StudyID>
    <StudyIDLower>d149</StudyIDLower>
    ▼<RelatedPublication>
      Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment
    </RelatedPublication>
    <RelatedProject>none</RelatedProject>
    <RelatedDatabase>none</RelatedDatabase>
    <keywords>List Experiment,Election Polling</keywords>
    <CreateDate>2017-06-30T14:36:30.247</CreateDate>
    <ResearchDesign>Survey experiment</ResearchDesign>
    <DataType>Survey/interview (e.g., ANES)</DataType>
    <DataSource>Author</DataSource>
    <DataSourceInformation>isps(at)yale(dot)edu </DataSourceInformation>
    <CatalogRecordDataType>Survey/interview (e.g., ANES)</CatalogRecordDataType>
    <CatalogRecordDataSource>Author</CatalogRecordDataSource>
    <CatalogRecordDataSourceInformation/>
    <PersistentId/>
    <FieldDates>2016-09</FieldDates>
    <Location>United States</Location>
    <LocationDetails>United States</LocationDetails>
```

HOME > RESEARCH > DATA

## Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment

ISPS Data Archive: Terms of Use

By using, contributing, and/or downloading files associated with scholarly studies available on the ISPS Data Archive, you agree to these terms and conditions.

Please read the ISPS Data Archive [Terms of Use](#).

**Author(s):** Alexander Coppock

**Suggested citation:**

Alexander Coppock (2017). replication materials for, 'Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment.'

<http://hdl.handle.net/10079/zw3r2f9>. ISPS Data Archive.

**ISPS ID:** D149

**Related publications:**

[Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment](#)

**Keyword(s):** [List experiment](#)

[Election Polling](#)

**Research design:** [Survey experiment](#)

**Data type:** Survey/interview (e.g., ANES)

# Drupal Integration

## File Level

```
▼<FileElement>
  ▼<File>
    <id>35c2543f-124b-4410-942a-5c2c48a02360</id>
    <FileSize>519</FileSize>
    <FileUrl>http://hdl.handle.net/10079/z34tn2k</FileUrl>
    <FileNumber>D149F01</FileNumber>
    <FileDescription>ReadMe file</FileDescription>
    <FileFormat>.txt</FileFormat>
    <PublicFile>1</PublicFile>
    <CatalogRecordId>D149</CatalogRecordId>
  </File>
  ▼<File>
    <id>3ac6bb4f-8c10-4d2b-aa20-993cb38396fc</id>
    <FileSize>26771</FileSize>
    <FileUrl>http://hdl.handle.net/10079/jsxkt0z</FileUrl>
    <FileNumber>D149F04</FileNumber>
    <FileDescription>Metadata (DDI 3.2)</FileDescription>
    <FileFormat>.xml</FileFormat>
    <PublicFile>1</PublicFile>
    <CatalogRecordId>D149</CatalogRecordId>
  </File>
  ▼<File>
    <id>02e67b9a-9383-4653-a6e9-c1d8fd6bef4a</id>
    <FileSize>481797</FileSize>
    <FileUrl>http://hdl.handle.net/10079/tb2rc20</FileUrl>
    <FileNumber>D149F02</FileNumber>
    <FileDescription>Data file</FileDescription>
    <FileFormat>.csv</FileFormat>
    <PublicFile>1</PublicFile>
    <CatalogRecordId>D149</CatalogRecordId>
  </File>
  ▼<File>
    <id>e4aca988-2b09-4288-8a5a-dfa6cddfdbcb</id>
    <FileSize>11384</FileSize>
    <FileUrl>http://hdl.handle.net/10079/pk0p31f</FileUrl>
    <FileNumber>D149F03</FileNumber>
    <FileDescription>Program file</FileDescription>
    <FileFormat>.r</FileFormat>
    <PublicFile>1</PublicFile>
    <CatalogRecordId>D149</CatalogRecordId>
  </File>
</FileElement>
```

| <a href="#">DATA FILE NUMBER</a> | <a href="#">DESCRIPTION</a> | <a href="#">FILE FORMAT</a> | <a href="#">SIZE</a> | <a href="#">FILE URL</a>      |
|----------------------------------|-----------------------------|-----------------------------|----------------------|-------------------------------|
| D149F01                          | ReadMe file                 | .txt                        | 519                  | <a href="#">Download file</a> |
| D149F02                          | Data file                   | .csv                        | 481797               | <a href="#">Download file</a> |
| D149F03                          | Program file                | .r                          | 11384                | <a href="#">Download file</a> |
| D149F04                          | Metadata (DDI 3.2)          | .xml                        | 26771                | <a href="#">Download file</a> |

# Drupal Integration

## Required Modules

|                                     |                                |                |                                                                                                                                                                                                                                                                                                                                                |
|-------------------------------------|--------------------------------|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> | <b>Field Collection Feeds</b>  | 7.x-1.0-alpha4 | Provides feeds integration for field collection.<br>Requires: Field collection ( <b>enabled</b> ), Entity API ( <b>enabled</b> ), Feeds ( <b>enabled</b> ), Chaos tools ( <b>enabled</b> ), Job Scheduler ( <b>enabled</b> )<br>Required by: Drupal (Feeds is currently using this module for one or more <b>importers</b> )                   |
| <input checked="" type="checkbox"/> | <b>Field Validation</b>        | 7.x-2.6        | Add validation rules to fields.<br>Requires: Field ( <b>enabled</b> ), Field SQL storage ( <b>enabled</b> ), Token ( <b>enabled</b> ), Chaos tools ( <b>enabled</b> )<br>Required by: Field validation extras ( <b>disabled</b> ), Field Validation UI ( <b>enabled</b> )                                                                      |
| <input type="checkbox"/>            | <b>Field validation extras</b> | 7.x-2.6        | Extra validators for Field validation 7.x-2.x.<br>Requires: Field Validation ( <b>enabled</b> ), Field ( <b>enabled</b> ), Field SQL storage ( <b>enabled</b> ), Token ( <b>enabled</b> ), Chaos tools ( <b>enabled</b> )                                                                                                                      |
| <input checked="" type="checkbox"/> | <b>Field Validation UI</b>     | 7.x-2.6        | UI for Field Validation.<br>Requires: Field Validation ( <b>enabled</b> ), Field ( <b>enabled</b> ), Field SQL storage ( <b>enabled</b> ), Token ( <b>enabled</b> ), Chaos tools ( <b>enabled</b> )                                                                                                                                            |
| <input checked="" type="checkbox"/> | <b>Pathauto</b>                | 7.x-1.3        | Provides a mechanism for modules to automatically generate aliases for the content they manage.<br>Requires: Path ( <b>enabled</b> ), Token ( <b>enabled</b> )<br>Required by: Cultura ( <b>disabled</b> ), Easy Breadcrumb ( <b>disabled</b> ), ELTI General ( <b>disabled</b> ), Yale Common ( <b>enabled</b> ), projects ( <b>enabled</b> ) |

# Drupal Integration

## Feeds Importers

[Home](#) » [Administration](#) » [Structure](#) » Feeds importers

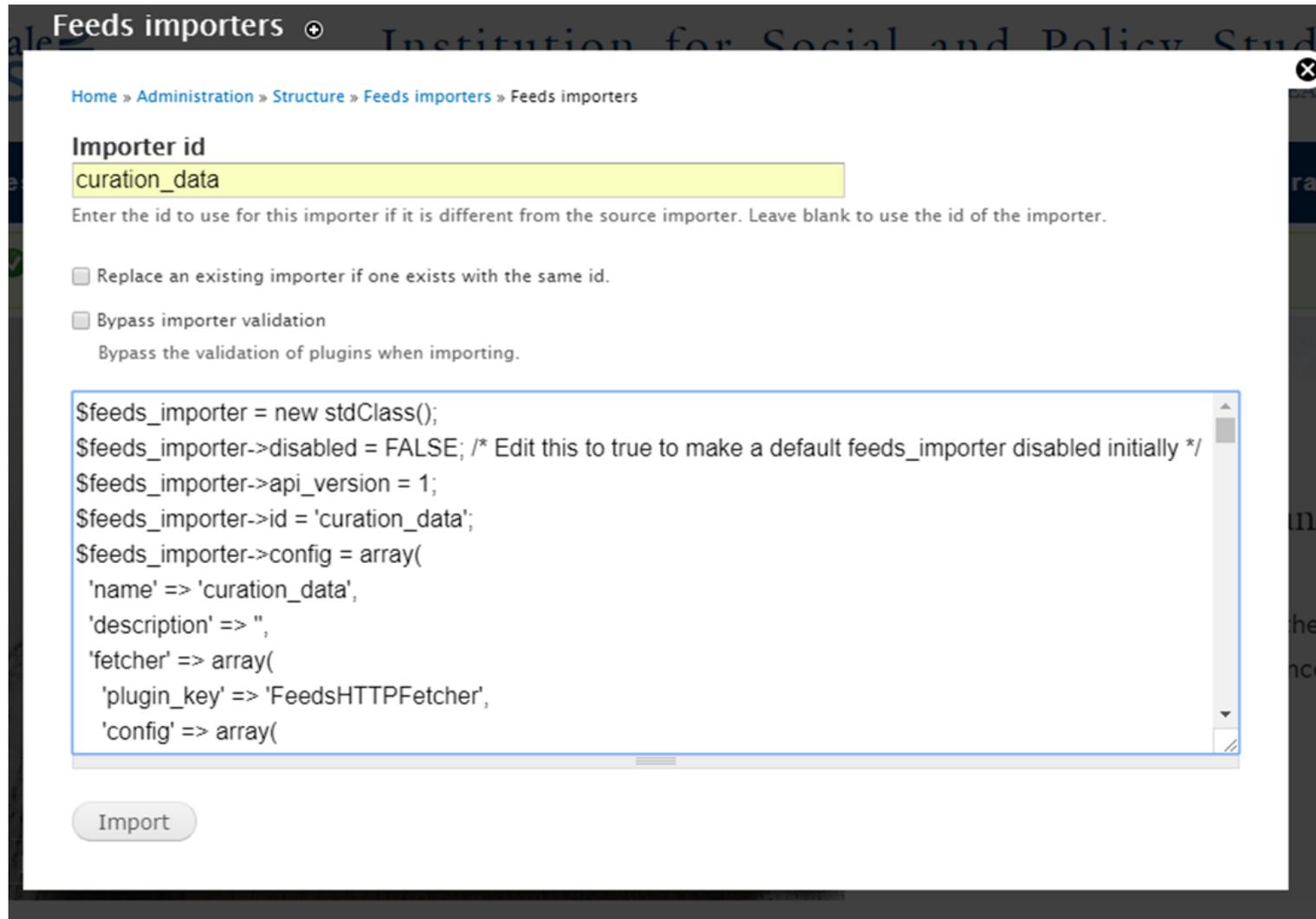
Create one or more Feed importers for pulling content into Drupal. You can use these importers from the [Import](#) page or – if you attach them to a content type – simply by creating a node from that content type.

[+ Add importer](#) [+ Import importer](#)

| NAME                     | DESCRIPTION | ATTACHED TO | STATUS | OPERATIONS                                                                                                              |
|--------------------------|-------------|-------------|--------|-------------------------------------------------------------------------------------------------------------------------|
| curation_data            |             | [none]      | Normal | <a href="#">Edit</a>   <a href="#">Export</a>   <a href="#">Clone</a>   <a href="#">Delete</a>   <a href="#">Tamper</a> |
| curation_data_files      |             | [none]      | Normal | <a href="#">Edit</a>   <a href="#">Export</a>   <a href="#">Clone</a>   <a href="#">Delete</a>   <a href="#">Tamper</a> |
| curation_keywords        |             | [none]      | Normal | <a href="#">Edit</a>   <a href="#">Export</a>   <a href="#">Clone</a>   <a href="#">Delete</a>   <a href="#">Tamper</a> |
| curation_location        |             | [none]      | Normal | <a href="#">Edit</a>   <a href="#">Export</a>   <a href="#">Clone</a>   <a href="#">Delete</a>   <a href="#">Tamper</a> |
| curation_research_design |             | [none]      | Normal | <a href="#">Edit</a>   <a href="#">Export</a>   <a href="#">Clone</a>   <a href="#">Delete</a>   <a href="#">Tamper</a> |

# Drupal Integration

## Feeds Importers



The screenshot shows the 'Feeds importers' configuration page in a Drupal administration interface. The breadcrumb trail is 'Home » Administration » Structure » Feeds importers » Feeds importers'. The 'Importer id' field is highlighted in yellow and contains the text 'curation\_data'. Below this field is a text instruction: 'Enter the id to use for this importer if it is different from the source importer. Leave blank to use the id of the importer.' There are two unchecked checkboxes: 'Replace an existing importer if one exists with the same id.' and 'Bypass importer validation'. Below the second checkbox is the text 'Bypass the validation of plugins when importing.' A large text area contains PHP code for configuring a feeds importer. At the bottom left, there is an 'Import' button.

Feeds importers

Home » Administration » Structure » Feeds importers » Feeds importers

**Importer id**  
curation\_data

Enter the id to use for this importer if it is different from the source importer. Leave blank to use the id of the importer.

Replace an existing importer if one exists with the same id.

Bypass importer validation  
Bypass the validation of plugins when importing.

```
$feeds_importer = new stdClass();  
$feeds_importer->disabled = FALSE; /* Edit this to true to make a default feeds_importer disabled initially */  
$feeds_importer->api_version = 1;  
$feeds_importer->id = 'curation_data';  
$feeds_importer->config = array(  
  'name' => 'curation_data',  
  'description' => "",  
  'fetcher' => array(  
    'plugin_key' => 'FeedsHTTPFetcher',  
    'config' => array(  

```

Import

# Drupal Integration

```
'processor' => array(
  'plugin_key' => 'FeedsNodeProcessor',
  'config' => array(
    'expire' => '-1',
    'author' => 0,
    'authorize' => 1,
    'mappings' => array(
      0 => array(
        'source' => 'xpathparser:0',
        'target' => 'title',
        'unique' => FALSE,
        'language' => 'und',
      ),
      1 => array(
        'source' => 'xpathparser:28',
        'target' => 'field_data_isps_id',
        'unique' => 1,
        'language' => 'und',
      ),
      2 => array(
        'source' => 'xpathparser:1',
        'target' => 'field_data_keywords',
        'unique' => FALSE,
        'language' => 'und',
      ),
    ),
  ),
),
```

## Mapping to Drupal

[Home](#) » [Administration](#) » [Structure](#) » [Feeds importers](#) » [curation\\_data](#)

### Basic settings

Attached to: [none] [Settings](#)  
Periodic import: every 1 day  
Import on submission

### Fetcher

[Change](#)

**HTTP Fetcher** [Settings](#)  
Download content from a URL.

### Parser

[Change](#)

**XPath XML parser** [Settings](#)  
Parse XML using XPath.

### Processor

[Change](#)

**Node processor** [Settings](#) [Mapping](#)  
Create and update nodes.

### Mapping for Node processor

Define which elements of a single item of a feed (= *Sources*) map to which content pieces in Drupal (= *Targets*). target can only occur once. E. g. only one item with the URL <http://example.com/content/1> can exist.

| SOURCE           | TARGET                                       |
|------------------|----------------------------------------------|
| + xpathparser:0  | Title (title)                                |
| + xpathparser:28 | ISPS ID (field_data_isps_id)                 |
| + xpathparser:1  | Keyword(s) (field_data_keywords)             |
| + xpathparser:29 | Author(s) (field_data_authors)               |
| + xpathparser:30 | Owner (field_data_owner)                     |
| + xpathparser:2  | Location (field_data_location)               |
| + xpathparser:5  | Research design (field_data_research_design) |

# Drupal Integration

## Mapping Issues

### RELATED DATABASE FIELD SETTINGS

These settings apply to the *Related database* field everywhere it is used. Because the field already has data, some settings can no longer be changed.

#### Number of values

Maximum number of values users can enter for this field. 'Unlimited' will provide an 'Add more' button so the users can add as many values as they like.

#### Maximum length \*

The maximum length of the field in characters.

Save settings

### FIELD DATE FIELD SETTINGS

These settings apply to the *Field date* field everywhere it is used. Because the field already has data, some settings can no longer be changed.

#### Number of values

Maximum number of values users can enter for this field. 'Unlimited' will provide an 'Add more' button so the users can add as many values as they like.

#### Date attributes to collect

Year  Month  Day  Hour  Minute  Second

Select the date attributes to collect and store. Changes to date attributes only effects new or updated content.

Collect an end date

End dates are used to collect duration. E.g., allow an event to start on September 15, and end on September 16.

Cache dates

Date objects can be created and cached as date fields are loaded, rather than when they are displayed, to improve performance.

Save settings

# Drupal Integration

## Schedule Importers

The screenshot shows the configuration page for a Drupal importer named 'curation\_data'. The page is divided into a left sidebar with category links and a main content area with configuration fields. At the top right, there are action buttons: EDIT, EXPORT, CLONE, DELETE, and TAMPER. The breadcrumb trail is 'Home » Administration » Structure » Feeds importers » curation\_data'. The left sidebar contains sections for 'Basic settings', 'Fetcher', 'Parser', and 'Processor', each with a 'Change' link and a 'Settings' link. The main content area has a 'Basic settings' section with a 'Name' field (value: 'curation\_data'), a 'Description' field, an 'Attach to content type' dropdown (value: 'Use standalone form'), and a 'Periodic import' dropdown (value: 'Every 1 day'). Below these are two checkboxes: 'Import on submission' (checked) and 'Process in background' (unchecked). A 'Save' button is at the bottom left of the main content area.

**curation\_data** [EDIT] [EXPORT] [CLONE] [DELETE] [TAMPER]

Home » Administration » Structure » Feeds importers » curation\_data

**Basic settings** [Settings]

Attached to: [none]  
Periodic import: every 1 day  
Import on submission

**Fetcher** [Change]

**HTTP Fetcher** [Settings]  
Download content from a URL.

**Parser** [Change]

**XPath XML parser** [Settings]  
Parse XML using XPath.

**Processor** [Change]

**Node processor** [Settings] [Mapping]  
Create and update nodes.

**Basic settings** [Help]

**Name \***  
curation\_data  
A human readable name of this importer.

**Description**  
A description of this importer.

**Attach to content type**  
Use standalone form ▾  
If "Use standalone form" is selected a source is imported by using a form under <http://tst.isps.yale.edu/import>. If a content type is selected a source is imported by creating a node of that content type.

**Periodic import**  
Every 1 day ▾  
Choose how often a source should be imported periodically. [Requires cron to be configured.](#)

Import on submission  
Check if import should be started at the moment a standalone form or node form is submitted.

Process in background  
For very large imports. If checked, import and delete tasks started from the web UI will be handled by a cron task in the background rather than by the browser. This does not affect periodic imports, they are handled by a cron task in any case. [Requires cron to be configured.](#)

[Save]

# YARD Work: Data Archiving and Publishing at Yale

**Limor Peer**

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

**Joshua Dull**

[joshua.dull@yale.edu](mailto:joshua.dull@yale.edu)

**Themba Flowers**

[themba.flowers@yale.edu](mailto:themba.flowers@yale.edu)

Acknowledgments: Innovations for Poverty Action, Colectica,  
Digital Lifecycle Research & Consulting, Yale University Library, YaleSites



# Technical Components & Support at Yale

## Yale ITS

- Hardware – Windows Server (VM), 32GB RAM minimum (8 Cores), 100GB local disk for OS, applications and swap files
- Software – Colectica repository, statistical software, integrated APIs
- Storage – storage@yale start at 500GB read, write, no-execute access to one or more directories
- Application hosting – WCF application and ASP.NET MVC web application on IIS, plus a SQL Server database (10GB), a Windows Service
- Security – Federated identification

## Yale Library

- Persistent links – handle service
- *Long-term preservation – Fedora Commons / Hydra\**
- *Discovery – Blacklight\**

*\*inactive*