

# *The ISPS Data Archive: Mission, Work, and Some Reflections*



New Article: "Getting Out the Vote Is Tougher Than You Think"

[More ▶](#)



#### UPCOMING EVENTS

**Comparative Approaches to Immigration, Ethnicity, and Integration**

June 14, 2016

[Read more ▶](#)

**CSAP Annual American Politics Conference**

June 16, 2016 to June 17, 2016

#### LUX ET DATA: ISPS BLOG

**Relationships Matter for Poverty, But Not in Isolation**

Thomas Lyttelton, March 30, 2016

Marriage, births, geography, and the "diverging destinies" for children in America.

[Read more ▶](#)

**Hello. World: The FBI. Apple. and the**

 [SEARCH ISPS](#)

#### FROM THE DIRECTOR



It's fellowship time at ISPS. Vesla Weaver, head of ISPS's Center for the Study of Inequality, has received a prestigious Carnegie Award for her proposal "The Faces of American Democracy." ISPS's economics phenom

Dean Karlan has received a Guggenheim Fellowship. Meanwhile, two of our law school stars, Eric Chung and Aisha Saad, received the George Soros New American Fellowship. We're still waiting for our first Oscar. In the meantime, my new book with Paul Pierson has attracted some (mostly friendly) attention, Greg Huber and Alan Gerber take another look at all the get-out-the-vote research that ISPS has pioneered, and Andy Papachristos and Michael Sierra-Arevalo report that an alternative approach to law enforcement—Project Longevity—has significantly reduced violent crime in New Haven.

WHAT IS ISPS?

# Archive Embedded in ISPS Website

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 ►

# ISPS Data Archive: Basic Facts





- A digital repository for research produced by scholars affiliated with ISPS, with special focus on experimental design and methods. Associated data are primarily quantitative and gathered from a combination of administrative records, surveys, and observations.
- The Archive accepts research outputs underlying scientific publications including data, code, output, documentation, and other research artefacts.
- The Archive provides free and public access to research materials and accepts content for distribution only under a Creative Commons license.
- The Archive was launched in September 2010 as a pilot at Yale. (We learned a lot, and still are.)
- The Archive is managed by a full time professional with knowledge of the specific research domain and graduate student research assistants who handle most of the data curation.

# ISPS Data Archive: Mission

The ISPS Data Archive was built to create 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.**

# Curation for Quality & Reproducibility

## Data Quality Review Framework

<p><b>REVIEW FILES</b></p>  <p>Assign persistent IDs * Create a citation to the study and a study level metadata record * Record file details (size, format, checksums) * Check that all files are present * Verify that content of files matches expected format * Create non-proprietary versions of the files * Implement migration strategy for file formats * Monitor bits</p>	<p><b>REVIEW DATA</b></p>  <p>Check for undocumented variable and value information or out of range codes * Review data for confidentiality issues</p>
<p><b>REVIEW DOCUMENTATION</b></p>  <p>Confirm comprehensive descriptive information for informed reuse including methodology and sampling information * Link to other research products</p>	<p><b>REVIEW CODE</b></p>  <p>Check and verify code for data analysis and replication</p>

# Curation Tasks for Quality & Reproducibility

Performed by the ISPS Data Archive pre-publication:

1. Check for missing variable labels and value codes.
2. Review observation count
3. Identify potential data errors
4. Compare questionnaire, codebook, & data
5. Ensure there is no personally-identifiable information (PII)
6. Confirm code executes
7. Confirm code produces reported results
8. Create open formats

# New Curation Tool: Curator @Yale



- ✓ Structures and tracks the curation workflow
- ✓ Helps automate parts of the review pipeline



- ✓ Captures all metadata throughout the process
- ✓ Pushes out relevant information to pre-determined destinations
  - i.e., a user, the archive administrators, a Web based dissemination system, or preservation systems
- ✓ Can fit into repository and research workflows



# New Curation Tool: Curator @Yale



- **Leverages DDI Lifecycle**
  - Machine executable, open structured format supports research transparency
  - Plays a part in review tasks - study, file & variable levels
- **Modular, open-source**
  - Could be adapted to changing needs, research methods, dissemination platforms, and preservation solutions
  - Could be used by repositories, researchers, and research staff
  - Flexible deployment and configuration options

# New Curation Tool: Curator @Yale



test-isps.yard.yale.edu/CatalogRecord

Most Visited Getting Started

**Curation**

- Dashboard
- Catalog Records
- Organization Admin
- Site Admin

## Catalog Records

Status	Title	Number	Created	Updated	Collection	Publish	Archive
Processing	Test Dec 22	1	12/22/2015	12/22/2015	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Rejected Needs curator	ERJ Test Record	7	11/02/2015	11/02/2015	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Processing	TEST	3	05/18/2015	10/14/2015	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Processing	Test	1	03/26/2015	05/19/2015	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Processing	Test.2	2	05/07/2015	05/17/2016	<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>
Published	Test October 22	6	10/22/2015	05/13/2016	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/>
Draft Needs curator Locked	Chocolate scents and product sales: a randomized controlled trial in a Canadian bookstore and café		05/24/2016		<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>

# New Curation Tool: Curator @Yale



test-isps.yale.edu/CatalogRecord/Files/fff9c520-8ede-4a2f-b471-9932de5e41b3

Most Visited Getting Started

**Curation**

- Dashboard
- Catalog Records
- Organization Admin
- Site Admin

## Test October 22

Curator

General > Methods > **Files** > Review 0 > Publish and Archive

### Files

File	Version	Type	Status	
Butler_Dynes_AJPS_2015a_StateLegislators.dta	1	Data	Accepted 	
replication-data_experiment2_public.dta	1	Data	Accepted 	
Butler_Dynes_AJPS_2015a_labels.do	1	Code	Accepted 	
Butler_Dynes_AJPS_2015a_readme.pdf	1	Other	Accepted	
fff9c520-8ede-4a2f-b471-9932de5e41b3.ddi32.xml	0	Codebook	Accepted	

Add or Update Files Download All Files

# New Curation Tool: Curator @Yale



test-isps.yard.yale.edu/File/Status/e9724de4-5f04-4604-b25b-7cbc01bc8ef9

Search

Most Visited Getting Started

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

Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta

« Test October 22 Curator

General Variables Notes Review 5

### Curation Status

Download Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta Download All Files

Status	Task	Completed
✓	Review Observation Count	limor.peer@yale.edu on 10/22/2015 <input type="button" value="Undo"/>
✓	Check Missing Labels	limor.peer@yale.edu on 5/13/2016 <input type="button" value="Undo"/>
✓	Compare Questionnaire, Codebook, and Data in Data File	limor.peer@yale.edu on 5/13/2016 <input type="button" value="Undo"/>
✓	Check for Personally-Identifiable Information (PII) in Data File	limor.peer@yale.edu on 5/13/2016 <input type="button" value="Undo"/>
✓	Identify Potential Errors in Data File	limor.peer@yale.edu on 5/13/2016 <input type="button" value="Undo"/>

# New Curation Tool: Curator @Yale



Most Visited Getting Started

Curation



Dashboard



Catalog Records



Organization Admin



Site Admin



Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta

« Test October 22



General

Variables

Notes

Review

5



## File History



Friday, May 13, 2016 10:37 AM by limor.peer@yale.edu

### Edit a File

Data Type changed from blank to Survey/interview (e.g., ANES)

Source changed from blank to blank

Test October 22

- Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta



Friday, May 13, 2016 10:36 AM by limor.peer@yale.edu

### Accept task: Identify Potential Errors in Data File

No obvious data errors to flag

Test October 22

- Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta



Friday, May 13, 2016 10:36 AM by limor.peer@yale.edu

### Accept task: Check for Personally-Identifiable Information (PII) in Data File

No PII

# New Curation Tool: Curator @Yale



The screenshot shows a web browser window with the URL `test-isps.yale.edu/File/History/e9724de4-5f04-4604-b25b-7cbc01bc8ef9`. The interface features a dark sidebar on the left with navigation options: Curation, Dashboard, Catalog Records, Organization Admin, and Site Admin. The main content area displays a list of tasks:

- Finished applying metadata updates** (Friday, May 13, 2016 10:15 AM by limor.peer@yale.edu)
  - Test October 22
    - Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta
- Applying metadata updates** (Friday, May 13, 2016 10:15 AM by limor.peer@yale.edu)
  - Test October 22
    - Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta
- Accept task: Review Observation Count** (Thursday, October 22, 2015 11:22 AM by limor.peer@yale.edu)
  - Test October 22
    - Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta
- Virus Scan** (Thursday, October 22, 2015 11:16 AM by limor.peer@yale.edu)
  - Clean virus scan on Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta
  - Test October 22
    - Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta
- Upload: Initial Upload** (Thursday, October 22, 2015 11:16 AM by limor.peer@yale.edu)
  - Test October 22
    - Butler\_Dynes\_AJPS\_2015a\_StateLegislators.dta

# Reflections: What's Next?

- Max Weber
- German sociologist, philosopher, political economist
- Inevitable and linear trend toward increasing rationalization, systemization, and routinization



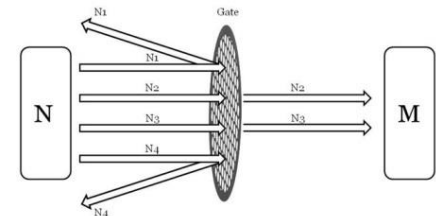
**Will curation for quality & reproducibility become routinized?**

# Curation for Quality & Reproducibility Will Become Routinized When...

## 1. Research data policies & culture mature

- ✓ Wide recognition that curating for quality and reproducibility is a valued goal
- ✓ Researchers are rewarded for quality of materials shared (e.g., badges)
  - OSF, ACM-DL
- ✓ Gatekeeping is acknowledged as essential function

VIEW THE **BADGES:**



N – Source of news item     $N_{1,2,3,4}$  – News items    M – Audience  
 $N_{1,4}$  – Discard item     $N_{2,3}$  – Selected item

Gate Keeping Theory



# Curation for Quality & Reproducibility Will Become Routinized When...

## 2. It is easier and cheaper to do

- ✓ It can be better automated
  - Curator tool, APIs
- ✓ Computational methods evolve to readily capture the entire workflow and preserve it
  - R knitr, R Sweave, Docker, ReproZip

Curator  
@Yale

 R Studio®

 docker

# Curation for Quality & Reproducibility Will Become Routinized When...

3. It is a part of active research process



<http://www.thesleuthjournal.com/military-co-option-media-entertainment/>

# Curation for Quality & Reproducibility Will Become Routinized When...

1. Research data policies & culture mature
2. It is easier and cheaper to do
2. It is a part of active research process

**Can Curation Prevent the Next Data Sharing Disaster?**

<http://isps.yale.edu/news/blog/2016/06/can-curation-prevent-the-next-data-sharing-disaster>

# Thank you!

Limor Peer

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

@l\_peer