



FEDERAL RESERVE BANK *of* ST. LOUIS  
CENTRAL TO AMERICA'S ECONOMY®

# Rolling Back Link Rot

**Using Automation and Project Management for  
File URL Error Detection and Cleanup on an Economic Research Web Application**

Scott St. Louis

Midwest Business Librarian Summit

Purdue University

July 12, 2023

<https://fedinprint.org>

**The views expressed in this presentation do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.**

# Flying Under the Radar

- You know about **FRED**...
  - *Economic data aggregation service, compiling and presenting more than 800,000 data series from more than 100 public and private sector sources.*
- You know about **FRASER**...
  - *Digital library of United States economic, financial, and banking history, containing more than half a million pages of historical documents.*
- Do you know about **Fed in Print**?
  - *Web application indexing roughly 50,000 Federal Reserve System publications, papers, and speeches, presenting metadata to the larger discovery service known as Research Papers in Economics, or RePEc.*



# Driving Discoverability



50,000 items  
Specific to Federal Reserve

Research Documents  
Information Format  
(ReDIF) files



RePEc discovery service  
4,400,000 items  
Worldwide

# Persistent Challenges

- **Fed in Print indexes material from more than a dozen websites.**
  - Content migrations (website updates) cause links to break! Broken links propagate to RePEc.
  - Without read access and admin-level write access to item records programmatically, how do you assess the scale of the broken link problem and develop an efficient solution?
    - *Answer: You don't. Competing priorities in a busy environment pull your attention elsewhere!*
      - *Nobody has time to manually fix thousands of broken links with our admin interface alone.*
- **API to the rescue!**
  - Deployed in 2022, the Fed in Print Metadata Gateway API makes possible just enough automation to push this project into the realm of viability.
  - We have observed and documented some API bugs and implementation shortcomings that will be addressed ASAP. Please bear with us!



# API Enables Large-Scale URL Cleanup

- 1) **Pull** all item records in JSON using the API. Roughly 500 JSON files with 100 item records each.
- 2) **Test** approximately 50,000 file URLs for 404 (“resource not found”) errors and other discoverability problems. Around 7,500 problematic URLs.
- 3) **Organize** problematic URLs into 13 spreadsheets for each individual Bank and the Board of Governors.
- 4) **Fix** problematic URLs with Python-generated cURL requests imported to Postman.
- 5) **Snapshot** good URLs in the Wayback Machine to minimize the urgency of doing all this work again in a few years.



**PULL**



**TEST**



**ORGANIZE**



**FIX**



**SNAPSHOT**

*More than 1,200 links fixed in our production environment to date, and more than 13,000 links snapshotted.*