Creating a Research Data Inventory: How Much You Can Get from Researchers

Federal Reserve Bank of Chicago, Deng Pan
deng.pan@chi.frb.org

One of the action items developed out of the approved Chicago Fed's Research Data Life Cycle Strategy (RDLCS) is to develop a Research data inventory survey to identify the data we have owned/used and what has been “hidden”

1. Develop Action Item

One of the action items developed out of the approved Chicago Fed's Research Data Life Cycle Strategy (RDLCS) is to develop a Research data inventory survey to identify the data we have owned/used and what has been “hidden”

2. Conduct Survey

- **Part I: Identify current data use:**
  Ask users to select from a list of data products acquired by Research Department and/or made accessible to the researchers. Most of these data have been cataloged in the FRS catalog (DataFinder)

- **Part II: Identify “hidden” data:**
  Ask researchers to report preliminary information on any data that are they have used in the past five years but are not included in Part I.

  - Survey were sent to 81 researchers
  - Used FedSurvey (an internal tool) to administer the survey

3. Survey Results

- **Part I: Identify current data use:**
  - 75 data products were selected.
  - 7 data products were reported no use in the past 5 years.

- **Part II: Identify “hidden” data:**
  - 30 “hidden” datasets were reported (Chart 1 and Chart 2)

4. Develop Metadata List

- Several resources were consulted:
  - DataFinder – Fed System catalog for contracted data
  - ICPSR
  - Colleagues in Fed System

- Bring the metadata list to Research Data and Computing Committee (CDC) for discussion and approval

- List has 42 metadata elements
  - Most important to users are:
    - dataset name; author/publisher; data use permission; dataset owner; dataset steward; expert user; time coverage; brief description

5. Follow up with Researchers

- **Test case:** 4 datasets (one from government/public; two from private company; one from university)

- 15 researchers were given the metadata list to report details on the “hidden data”

- 3+2+1 rule (3 weeks + 2 weeks +1 week)

- Found more “hidden data” when interviewing researchers

6. Catalog the “Hidden Data”

- **Use SharePoint to host inventory:**
  - Pro: easy set up, low cost
  - Cons: search capabilities are limited

- All the reported “hidden data” have been cataloged in the inventory

7. What We Have Learned

- Be flexible

- Connect to researcher’s interest

- Follow up, follow up and follow up!

- Incorporate into a normal process
  - “Self-report” survey mechanism is now part of the economist onboarding process