



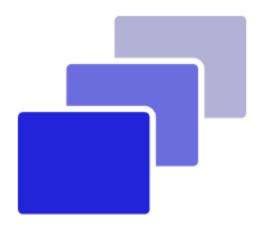
# Finding licensed data resources: the user experience

Adrienne Brennecke May 26, 2017



## **Outline**

- Background
- What I learned from user interviews
- Catalog features users liked
- Most popular interfaces for locating data
- Lessons learned and action items





# Question

What is the best way to let users know about the data resources available to them?





# Previous data catalog

Browse Datasets by Keyword Agriculture Banks and depository institutions Consumption, saving, and investment Corporate: finance and ownership Corporate: risk and risk management Demographic economics Derivatives, futures, and options Development economics Equity Fixed income Government expenditures: national Government expenditures: state and local Health, education, and welfare Household behavior and family economics Insurance and insurance companies

mational finance

#### Results for: "crsp"

#### Center for Research and Securities Prices - San Francisco

CRSP U.S. stocks 1925- database provides unique historical indices for NYSE, Amex, NASDAQ, and Arca markets. The Indices product also includes S&P 500, NASDAQ composite, but we do not have US Trea [more]

#### Compustat Bank and Segment Back Data - Board

The current version of this record is maintained in DataFinder-Board. Compustat is a database of information about corporations. The data include: 1. Values for balance sheet and income statement it [more]

#### Compustat Bank and Segment Back Data - Boston

Compustat is a database of information about corporations. The data include: 1. Values for balance sheet and income statement items (accounting data), at both annual and quarterly frequencies, as well [more]

#### Compustat Bank and Segment Back Data - Chicago

Compustat is a database of information about corporations. The data include: 1. Values for balance sheet and income statement items (accounting data), at both annual and quarterly frequencies, as well [more]

#### Compustat Bank and Segment Back Data - Richmond

Compustat is a database of information about corporations. The data include: 1. Values for balance sheet and income statement items (accounting data), at both annual and quarterly frequencies, as w [more]



#### Compustat Bank and Segment Back Data - Richmond

Compustat is a database of information about corporations. The data include:

- 1. Values for balance sheet and income statement items (accounting data), at both annual and quarterly frequencies, as well as business segment data.
- 2. For firms included in the database that have publicly traded equity, month-end equity prices, as well as dividend payment information and earnings information. Price-to-earnings ratios can be computed from data in this "price dividend earnings" or PDE file.

The bank data package is a specialty database containing additional segment data history beyond the current Compustat North American database from 1978 up to current 7 year history.

Also known as: CRSPBANK, CRSPBDS

Keywords: Corporate: finance and ownership

Type: Micro

Geographic coverage: United States

Related datasets:

Compustat Execucomp - Richmond

#### Documentation

Knowledge Exchange

S&P Support Center (Vendor Documentation)

For S&P product support questions, product downloads, and supporting documentation on all enterprise products, including Xpressfeed and Research Insight, visit the Support Center website. Users can submit their questions online directly to the Client Support Department; inquiries will be answered within 24 hours during regular business hours. If you require more immediate assistance, contact S&P at: 800.523.4534 (6:00 a.m. 6:00 p.m., US and Canada, MST) e-mail: clientsupport@standardandpoors.com website: https://support.standardandpoors.com

Where are these data stored?

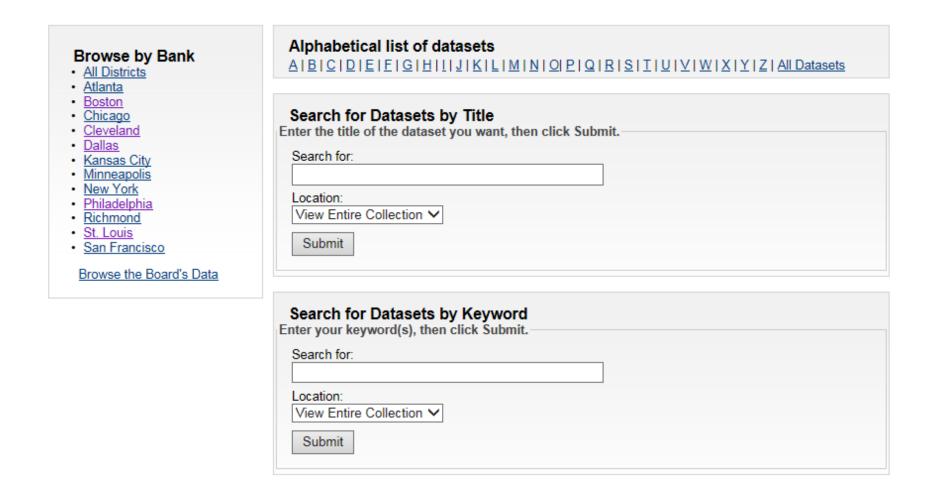
How should these data be handled?

Vendor/Provider & Purchasing Information

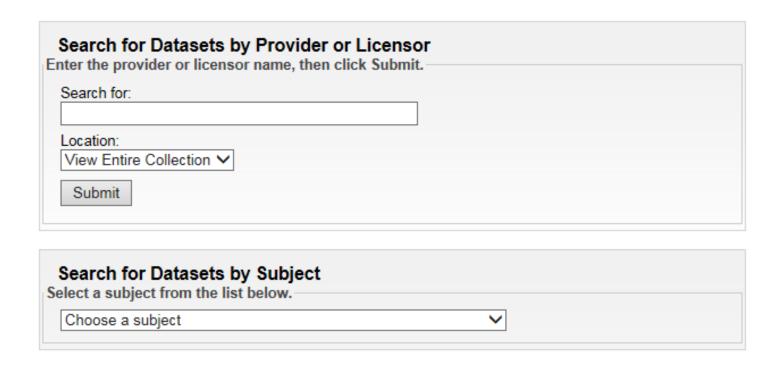
**Internal Contacts** 



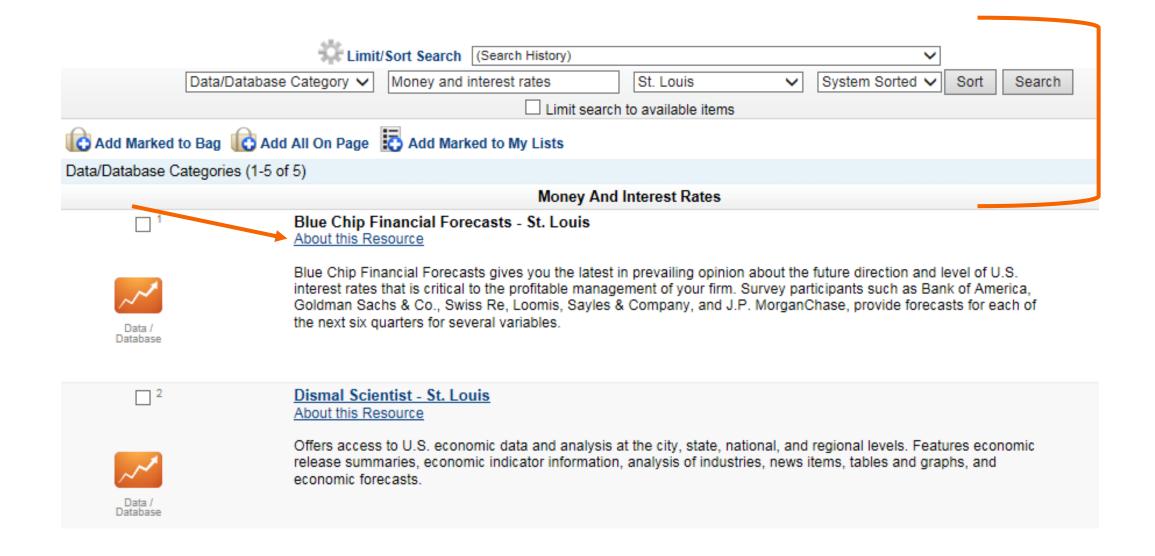
## **DataFinder**















Resource Link

Resource Name Blue Chip Financial Forecasts - St. Louis

Data Provider Aspen Publishers Author Aspen Publishers

Access Files are received monthly by email and stored on Research's shared drive. The Library has

Information some print.

Description Blue Chip Financial Forecasts gives you the latest in prevailing opinion about the future

direction and level of U.S. interest rates that is critical to the profitable management of your firm. Survey participants such as Bank of America, Goldman Sachs & Co., Swiss Re, Loomis, Sayles & Company, and J.P. MorganChase, provide forecasts for each of the next six

quarters for several variables.

Date Coverage Library has Hard copy: Nov 1982-2007. PDF versions, 2004-: G:\Economic outlook\Blue chip

financial

January 2010-present.

Geographic Coverage

United States

Resource Type Data

Resource Excel & PDF Format

Category Data Resource

Data B

Mathematical and quantitative methods

Money and interest rates

Data collection, modeling, and quantitative methods

#### License

Authorized Users	FRS + Board		
Attribution requirements	Copyright © YEAR by Aspen Publishers, Inc. All rights reserved. 1. (f) "The Licensee agrees to provide reasonable and appropriate attribution to Licensor in Work Product that contains Licensed Content. For example, the Licensee will acknowledge any applicable Licensor copyright for raw Licensed Content that is reproduced in Work Product.		



### **User interviews**

- Script available on Google Drive
- Most fruitful questions:
  - When you look for data, is it typically something you know exists or does it more often require background research?
  - When you find information on a useful data set, what do you do next?
  - What are some of your most common challenges when working with data?





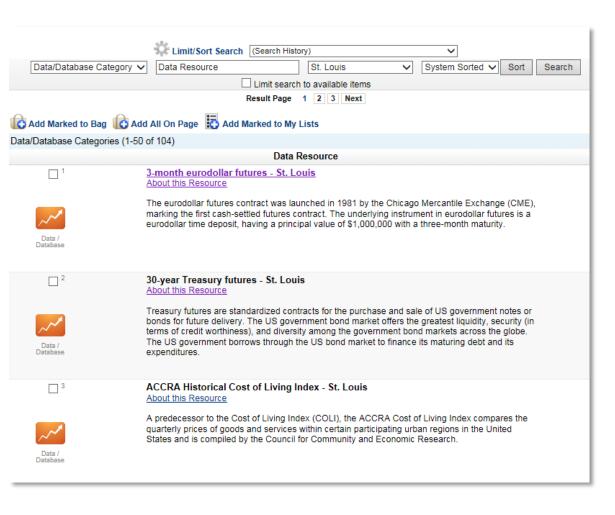
# Insights

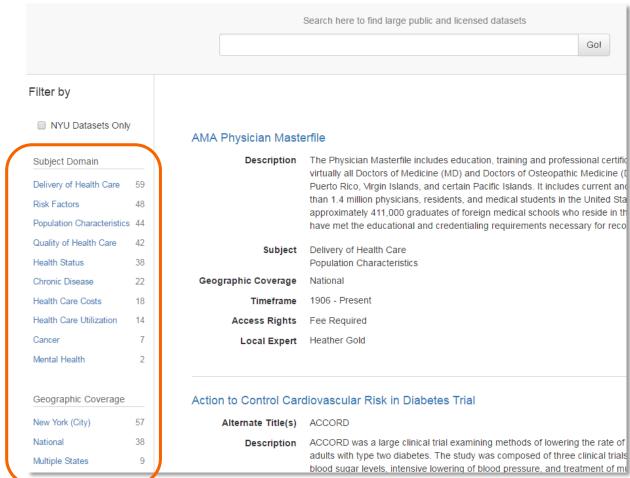


- Data users look for articles first
  - Articles based on the research question to see which data sources were used by others
  - Articles which describe the particular data they plan to use
- Historical data are not well-documented or easily accessible
- Better information management might save repetitive work
- Keep documentation as close to data as possible



# DataFinder vs NYU Data Catalog

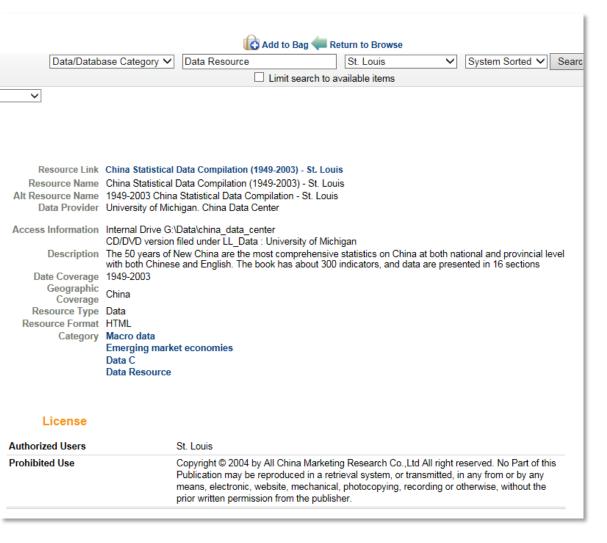




https://datacatalog.med.nyu.edu/



### **Comments**



#### AIVIA FITYSICIATI IVIASIETIIIE

UID: 10032

#### Description

The Physician Masterfile includes education, training and professional certification information on virtually all Doctors of Medicine (MD) and Doctors of Osteopathic Medicine (DO) in the United States, Puerto Rico, Virgin Islands, and certain Pacific Islands. It includes current and historical data for more than 1.4 million physicians, residents, and medical students in the United States. This figure includes approximately 411,000 graduates of foreign medical schools who reside in the United States and who have met the educational and credentialing requirements necessary for recognition.

#### Publisher

American Medical Association (AMA)

#### Timeframe

1906 - Present

#### Geographic Coverage

United States

#### Local Expert

Heather Gold

#### **Subject Domain**

Delivery of Health Care Population Characteristics

#### Keywords

Medical students
Physicians

#### Access via AMA Physician Masterfile

#### Access Restrictions

Fee Required

#### Access Instructions

Requires specific Unique Physician Identifica Numbers (UPIN) or National Provider Identificaccess Physician Masterfile data. To reques Masterfile data, one can fill out the form usin Lists website.

#### Data Type

Administrative

#### Dataset Format(s)

SAS, PDF, Microsoft Excel

#### PubMed Search

View articles which use this dataset

https://datacatalog.med.nyu.edu/



# Catalog features users like

- Filter by subject, multiple filters
- Subject, geographic coverage, and timeframe are all great filters
- Information on how to access and terms of use are appreciated
- Just one search bar
- Related keywords linked to related datasets





# Final survey

Imagine you want to know what data sets or databases are available to you within the Fed.

Which of these options would you most likely use to look for the information?

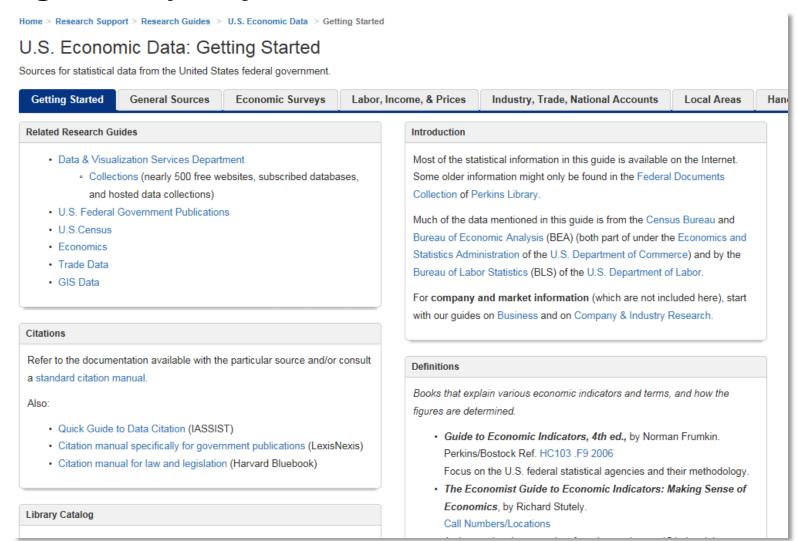


### 1. Alphabetical list of sources linking to a webpage with more information

#### E Elkon (Thomson Reuters) eMBS (Mortgage-Backed Securities) Online (via RADAR) eMentor (Risk Management Association) Equifax Credit Risk Insight Servicing McDash (CRISM) (via RADAR) Experian AutoCount FDIC (Federal Deposit Insurance Corp.) FR Y-14M Data - Board of Governors FRBNY/Equifax Consumer Credit Panel (via RADAR) FRED (Federal Reserve Economic Data) G GAAIT (Global Accounting and Auditing Information Tool) Global Banking Resource (TGBR) Global Financial Data (GFD) Global Financial Development Database (GFDD) Haver Analytics Historical Statistics of the US Home Mortgage Disclosure Act - Confidential Data (via RADAR) Home Mortgage Disclosure Act - Public Data (via RADAR) ICPSR (Inter-University Consortium for Political and Social Research) IEA Oil Market Report IMF (International Monetary Fund) IPUMS (Integrated Public Use Microdata Series)

In-Text Citation Format	Follow Chicago's Author-Date system for in-text references wherever practical. For example, if a dataset has a particular author include the last name in the in-text citation, e.g. (Butler).  If there's no author, the parenthetical reference uses the first few words of the citation on the References page, e.g. (U.S. Census Bureau).		
	Table/Graph Citation Format	For graph and table citations, eschew parentheses in favor of a list. Third party data should include the publisher or distributor in parentheses following the third party data provider. Example:	
Sources: Federal Reserve Bank of New York Consumer Credit Panel and Equifax data (Federal Reserve Bank of Philadelphia RADAR), International Monetary Fund (Haver Analytics), and U.S. Census Bureau.			
La TeX References Citation Format	See LaTeX Guidance for more information on LaTeX use and formatting.		
	Place the following citation between the tags \section*{Data References} and \end{document} in your LaTeX file.		
	Federal Reserve Bank of New York Consumer Credit Panel and Equifax. Year(s) of data publication or release. "Supplement or Subsection, when applicable". \textit(Dataset – as specific as possible, e.g. Household Debt: Total Debt Balance}, Date range, i.e. day(s), month(s) or quarter(s), if relevant, and year(s). Accessed through Publisher or Distributor, when applicable, i.e. Federal Reserve Bank of Philadelphia RADAR. Available at URL or DOI, if available.		
	Note: Do not include the internal RADAR link in citations.		
Word References Citation Format	Federal Reserve Bank of New York Consumer Credit Panel and Equifax. Year(s) of data publication or release. "Supplement or Subsection, when applicable". Dataset – as specific as possible, e.g. Household Debt: Total Debt Balance, Date range, i.e. day(s), month(s) or quarter(s), if relevant, and year(s). Accessed through Publisher or Distributor, when applicable, i.e. Federal Reserve Bank of Philadelphia RADAR. Available at URL or DOI, if available.		
	Note: Do not include the internal RADAR link in citations.		

### 2. Website organized by subject areas

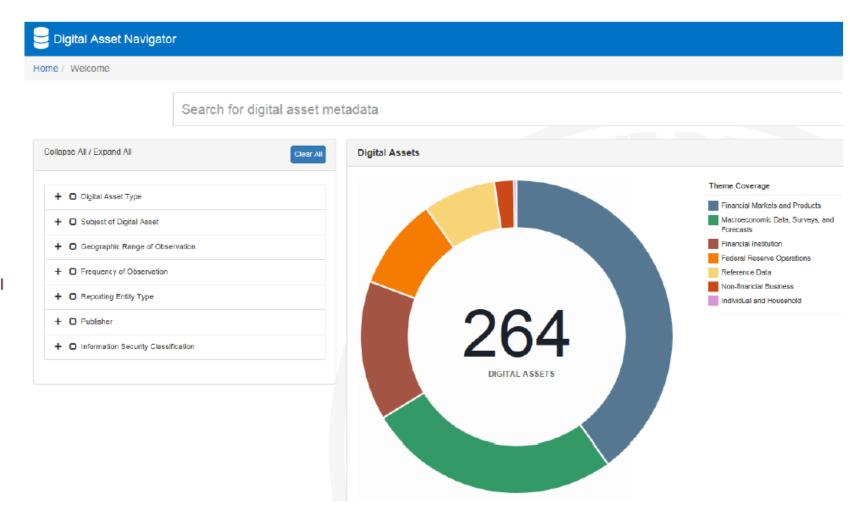


http://guides.library.duke.edu/economic\_data



# 3. Catalog organized by 7 main themes, searchable by keyword and filtered by subject, geography, etc.

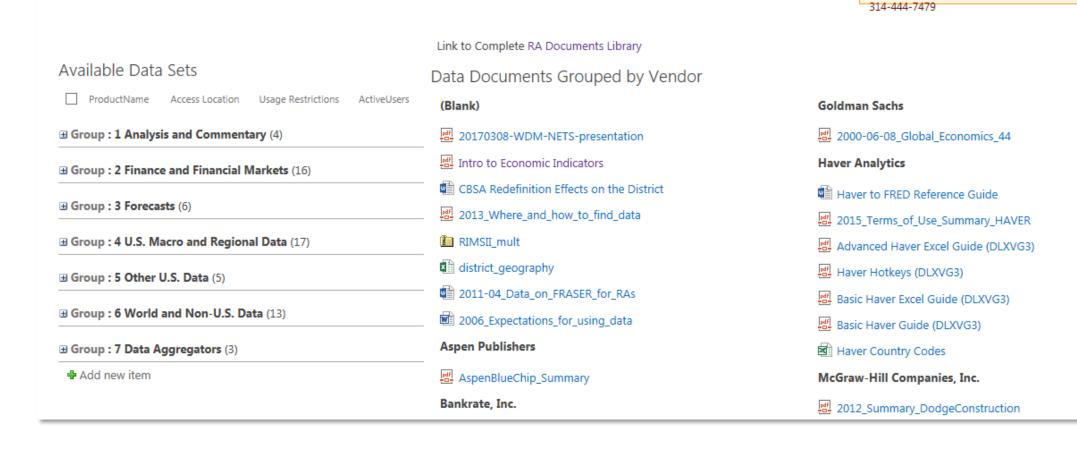
- Macroeconomic Data, Surveys and Forecasts This theme covers digital assets that are about macroeconomic subjects, such as data on macroeconomic indicators, forecasts, and analyses of conditions or trends.
- Financial Markets and Product This theme covers digital assets that are about financial instruments or products, financial markets characteristics or transactions, holdings, or issuances. Such digital assets include both assets and instruments such as analyses.
- Individual and Household This theme covers digital assets that are about people and households.
- Non-financial Business This theme covers digital assets that are about business in industries or sectors outside of the Financial Sector, such as manufacturing, tourism, and agriculture.
- Financial Institution This theme covers digital assets that are about a particular Financial Institution, or about Financial Institutions of some type.





### 4. SharePoint site

Data Sets & Related Documentation



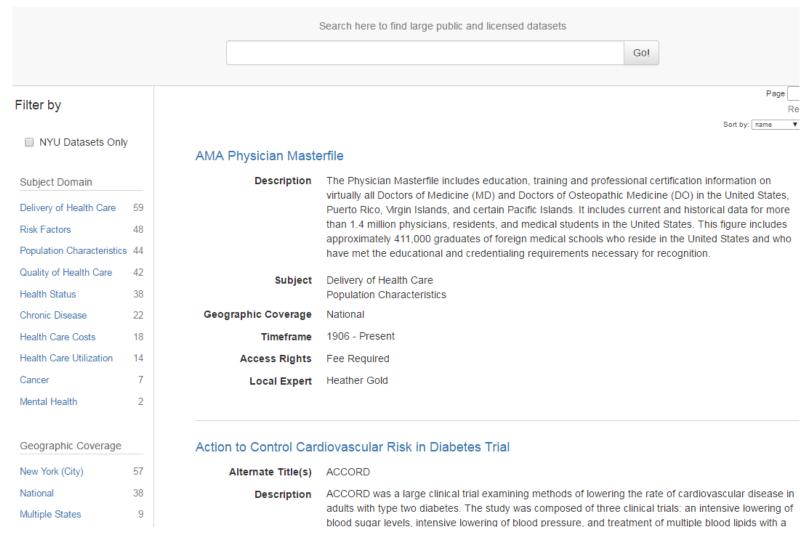
Questions? Contact:

Adrienne Brennecke Data Librarian

adrienne.j.brennecke@stls.frb.org



### 5. Catalog searchable by keyword and filtered by subject, geography, etc.



https://datacatalog.med.nyu.edu/

### 6. Catalog searchable by keywords

#### DataFinder - Banks

Search for datasets available at the 12 Federal Reserve Banks. Records tell you what the data are, which users have access, how to get to the data, and authorized uses for the data. To view what data are available at the Board, search their <a href="DataFinder-Board">DataFinder - Board</a> catalog. The Board's DataFinder catalog must be accessed from an authorized Bank network.

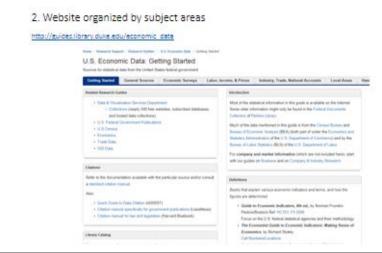
This catalog is currently in development. System librarians are working to enter records for all available datasets. Please contact your library if you cannot find the data you need.

Browse by Bank	Alphabetical list of datasets  A B C D E F G H   J K L M N O P Q R S T U V W X Y Z All Datasets
All Districts	VIDICIDIE I I DI II I I I I I I I I I I I I I
Atlanta     Boston     Chicago     Cleveland     Dallas     Kansas City	Search for Datasets by Title Enter the title of the dataset you want, then click Submit. Search for:
Minneapolis     New York     Philadelphia     Richmond     St. Louis     San Francisco	Location:  View Entire Collection ✓  Submit
Browse the Board's Data	
	Search for Datasets by Keyword Enter your keyword(s), then click Submit.  Search for:  Location:  View Entire Collection  Submit
	Search for Datasets by Provider or Licensor Enter the provider or licensor name, then click Submit.  Search for:

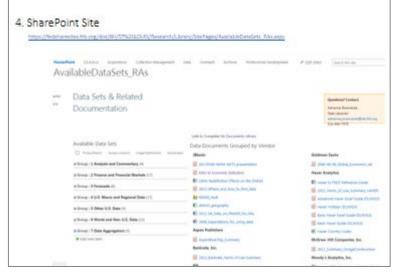


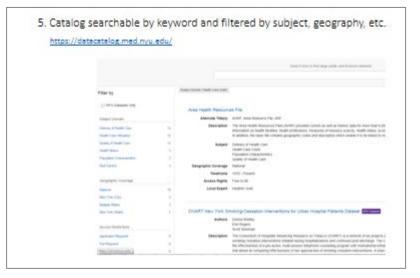
# Which did they prefer?

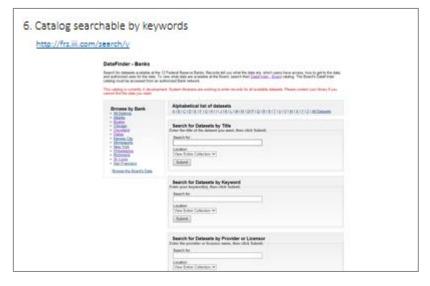




subject, geography,	etc.		
<ul> <li>Microeconomic Data, Surveys and Forecasts - This there covers digital assets that are about microeconomic subjects, such so data an macroeconomic indicators, forecasts, and analysis of subditions of termin.</li> </ul>	Cognitive Adment to Auropation		
<ul> <li>Financial Markets and Product: This thome covers digital exsets that are about financial instruments or</li> </ul>	Search for ingle	annet metalica	
products, financial markets characteristics or transactions, holdings, or insuences. Such digital assets include both assets and instruments such as	Inter to Lune to	Cryster Assets	
analyses.  • Individual and Household. This theme covers digital.	+ C ((a) (a) (a)		E book book or hour
scorts that are about people and households.	* Committee has		Management by heap on
<ul> <li>Non-financial Business - This thems covers sligital searts that are alread inscinencial industries or</li> </ul>	W Inspect Property Committee		
sections outside of the Financial Section, such as	# (B Treated of Treatment		- Name of the State of the Stat
manufacturing, tourism, and agriculture;  # Researce Institution - This thems covers digital	# E ferring this face		E modern to bear to
accets that are about a particular Financial	+ B tomo	264	
institution, or allow! Financial testitutions of some hore.	Consisting behind	204	
• Reference - This themse covers digital assets that provide standard representations of commonly used coveragin or ordinic. Exemples related onderests for group-pile regimes or lessions clearful-attines, or structures representations of legal duttine, and their primary characteristics, believence that a must are aften used to enote from our standard or from grained deteases, or in the encoding and decoding of information.		aura, aurain	
<ul> <li>Redard Reserve Operations - This theme covers digital exsets that are about Federal Reserve operations. It also includes digital assets that have output for such operations among their primary upon at the Bank.</li> </ul>			









# **Survey result**

- Both approaches are needed
  - One organized by subject AND another to go quickly to a known item





# Additional insights, future projects?

- Identified several areas where the library could add value
  - Curate and connect relevant papers to data catalog records
  - Digitize historical data, document methodology changes over time
  - Preserve internal information that will be needed again
  - Promote data citation and data management best practices



### **Action items**

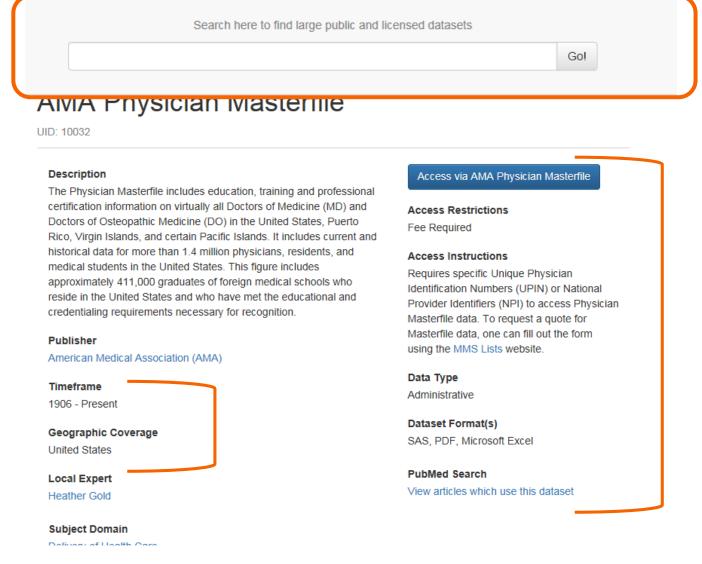
- Develop a LibGuides site
- NY Fed Digital Asset Navigator will it be an option for us?
- NYU Data Catalog is open source, can we repurpose?
- <u>RePEc Biblio</u> start curating data papers?
- Promote data citation best practices





# Future data catalog







### **Adrienne Brennecke**

Research Information Specialist adrienne.j.brennecke@stls.frb.org @abrennec