



# **Changing the Choreography: When the Microdata Repository Gets a Macrodata Partner**

**San Cannon**

**Federal Reserve Board**

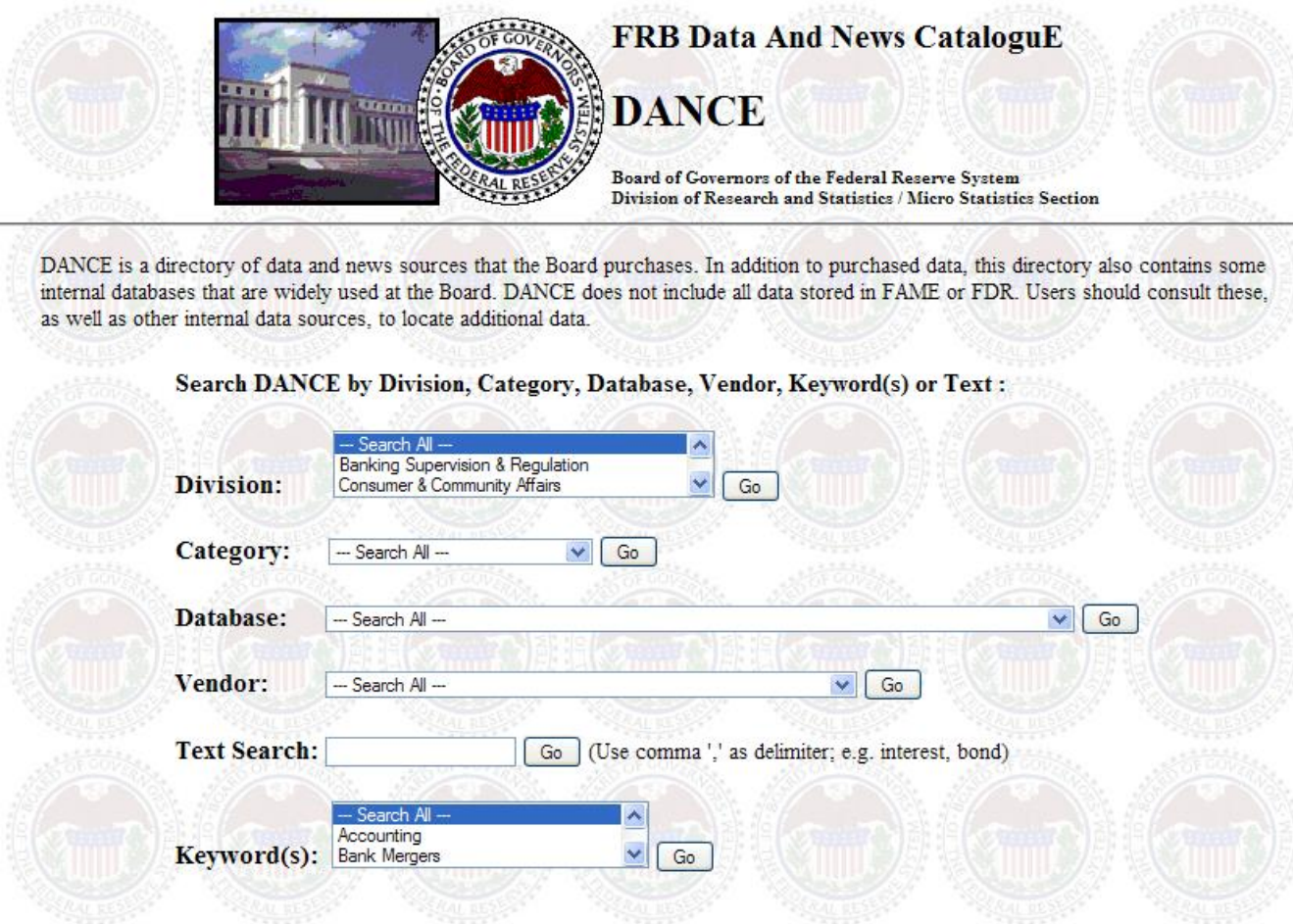
The views expressed are those of the author and do not indicate concurrence by the Board of Governors of the Federal Reserve System.



# The Story of DANCE

- **Federal Reserve Board uses LOTS of data, much of it purchased from private vendors.**
- **In 2005, created a basic database to hold the inventory of vendor data – the Data and News CataloguE (DANCE).**
- **Fairly basic metadata needs – started with homegrown tags but then adapted Dublin Core for most fields.**
- **Very successful! Full participation from across the Board.**

# “Browsing” the catalog



The image shows the 'FRB Data And News Catalogue' (DANCE) search interface. At the top, there is a header with the title 'FRB Data And News Catalogue' and 'DANCE' in large letters. Below the title, it says 'Board of Governors of the Federal Reserve System' and 'Division of Research and Statistics / Micro Statistics Section'. There are three circular logos: the Federal Reserve seal on the left, a photograph of the Federal Reserve building in the center, and the Federal Reserve seal on the right. Below the header is a paragraph of text describing DANCE. Underneath is a search section titled 'Search DANCE by Division, Category, Database, Vendor, Keyword(s) or Text :'. This section contains several search criteria, each with a dropdown menu and a 'Go' button. The 'Division' dropdown is open, showing 'Banking Supervision & Regulation' and 'Consumer & Community Affairs'. The 'Keyword(s)' dropdown is also open, showing 'Accounting' and 'Bank Mergers'. The 'Text Search' field is empty.

**FRB Data And News Catalogue**  
**DANCE**  
Board of Governors of the Federal Reserve System  
Division of Research and Statistics / Micro Statistics Section

DANCE is a directory of data and news sources that the Board purchases. In addition to purchased data, this directory also contains some internal databases that are widely used at the Board. DANCE does not include all data stored in FAME or FDR. Users should consult these, as well as other internal data sources, to locate additional data.

**Search DANCE by Division, Category, Database, Vendor, Keyword(s) or Text :**

**Division:**    
Banking Supervision & Regulation  
Consumer & Community Affairs

**Category:**

**Database:**

**Vendor:**

**Text Search:**   (Use comma ',' as delimiter; e.g. interest, bond)

**Keyword(s):**    
Accounting  
Bank Mergers



# Sample entry

**Division:** Research & Statistics

**Database:** [Dodge Construction Potentials Bulletin Service](#)

**Vendor:** [McGraw-Hill Construction Research & Analytics](#)

**Data Contact:** Vince Lombardi 1-800-GO-PACKERS

**License Contact:** Ray Nitschke 1-800-HIT-HARD

**License Information:** Site license for all Board staff

**Form of Access:** E-mail & CD-ROM

**Description:** Contract includes monthly Construction and Housing Review, total US 35-month construction start value data files, Monthly Dodge index 1980 to present, and total private construction data file.

**Status:** Active Contract

**Keyword(s):** Building Materials, Housing



# Success has its price

- **Metadata fields proved too limited – need more information on storage, access, usage rights.**
- **Coverage was too limited – a lot of the data we use is publicly available (not purchased) and wasn't covered in DANCE.**
- **Redesign begins: increase amount of information per entry. Subsequent decision to increase coverage.**
- **Some parts were easier than others!**



# Stage 1: Increasing the fields

- **Needed information on the following:**
  - **Storage:** Where do the data live? In what packages/formats are they stored?
  - **Access:** How to get to the data? Are there special security restrictions?
  - **Usage:** What can we do with the data? Can we store them? Merge them with other data? Do we need special permission for publication or redistribution?



# Stage 1: Continued

- **Production: When do we get the data?  
How do we load it?**
- **Administration: Who bought it? How much does it cost? What are the contract dates?**
- **Number of metadata fields increased from 16 to 85.**
- **No longer constrain fields to be unique: e.g. Location, Vendor Contacts and Internal Contacts allow multiple values.**



# **Are we done? Not really.**

- **Old catalog had a few purchased macro data sets but old fields presented no problem.**
- **New fields didn't really fit macro data well but as a small minority of datasets it was not really an issue.**
- **But... now we want to add large amount of macro level data and much of it is from public sources (e.g. CPI data from BLS)**
- **Microdata vendor model is strained to breaking point.**





## **Stage 2: Increasing the scope**

- **Only added one additional field**
- **Different data management styles and issues led to serious rethinking of field names and definitions.**
  - **Publisher/creator/vendor distinction less clear for more cases**
  - **Time series data identification and storage completely different than microdata**
  - **Updates and revisions treated differently**
  - **Usage rules have different dimensions**



## **Stage 2: Other issues**

- **Lots of clarification needed: relied heavily on Dublin Core where possible.**
- **Allow flexibility for contracting and citation information.**
- **Differences in security treatments – contracts often specify precise users which is different than public macrodata.**



# Stage 2: Example

Metadata field	BRM	CPI
Creator	Bankrate	Bureau of Labor Statistics
Vendor	Bankrate	None
Date range available	September 19, 1997 to Present	Mar 35 or Jan 47 or Jan 67 or Jan 78 or Dec 97 or.... To present
Update schedule	Weekly – append	Monthly – new observations with possible revision
Usage rules	Raw data cannot be shared. Aggregate data can be shared.	Time series data publicly available.



[About](#) [Control Panel](#) [Add to DANCE](#) [Request New Dataset](#)

Search for  in  and

## [BRM Data](#)

The Micro Statistics Section produces two weekly SAS datasets of Bank Rate Monitor (BRM) data. Both datasets are surveys of the current checking and CD rates around the country in specific markets. The majority of the data is collected through weekly surveys of financial institutions carried out by BRM's research department. However, some banks enter the information directly via a web portal. The data are then made available for download after a six-step quality control process. Once downloaded, the data then goes through additional quality control checks before being made available to the Board. [\[more\]](#)

Active Contract  
[Internal FR](#)  
[Wiki](#)

Keywords: Financial markets, Banking and financial institutions

Type: Micro

Geographic coverage: United States

Date range available internally: September 19, 1997 to Present

Related datasets:

[Loan Performance Benchmark \(Mortgage\)](#)

[Informa Bank Price Data](#)

**Documentation**

[Wiki](#)

[Where are these data stored?](#)

[How should these data be handled?](#)

[Where do these data come from?](#)

[Internal Contacts](#)





## Documentation

[Wiki](#)

## Where are these data stored?

Fedweb in <http://fweb.rsma.frb.gov/ms/c/BRM/index.htm>.

Access form(s): None  
Software name: SAS  
Input format: Unknown  
Updated: Monday and Thursday  
Update method: Vintaged  
Output: SAS Dataset

RSMA Network in [/rsma/microdata/brm/](#).

Access form(s): None  
Linux group: ms  
Software name: SAS  
Input format: Unknown  
Updated: Monday and Thursday  
Update method: Vintaged  
Output: SAS Dataset

## How should these data be handled?

[Internal FR](#)

Critical to function during contingency? No.

Can data be published?		
	Raw data	
	Chart	Table
Publication	No	No
Website	No	No

Can data be shared with?		
	Reserve Banks: Yes	Public: No
Reserve Banks: Yes		
Public: No		

Can data be published?		
	Aggregate data	
	Chart	Table
Publication	Yes	Yes
Website	Yes	Yes

Can data be shared with?		
	Reserve Banks: Yes	Public: Yes
Reserve Banks: Yes		
Public: Yes		

## License Summary:

The Federal Reserve may release to the public only aggregate series whether derived wholly or in part from these data. The Federal Reserve shall NOT publish institutional details from the data supplied on this website. The Federal Reserve shall clearly identify the Bank Rate Monitor as the source of data in any summary documents published as The Federal Reserve shall not resell, distribute, license or sub-license these data. Therefore, the Federal Reserve shall not share the microdata with associates employed outside of the Federal Reserve System including other government agencies and academics.

[License](#) [pdf]

## Bibliographic notation

The Federal Reserve shall clearly identify the Bank Rate Monitor as the source of data in any summary documents published as Source: Bank Rate Monitor®, N. Palm Beach, FL 33408, all rights reserved.





[About](#) [Control Panel](#) [Add to DANCE](#) [Request New Dataset](#)

Search for  in  and

### [Consumer Price Index](#)

The Consumer Price Indexes (CPI) program produces monthly data on changes in the prices paid by urban consumers for a representative basket of goods and services.

Active Contract  
[Public](#)  
[Wiki](#)

Keywords: Macroeconomics, Prices

Type: Macro

Geographic coverage: United States

Data Group: [cpi](#)

Date range available: Varied to Previous month

Date range available internally: Varied to Previous month

Related datasets:

[Blue Chip Economic Indicators](#)

[FX4Casts.com](#)

#### **Documentation**

[Wiki](#)

Internal documentation

[Data Sourcebook](#) (Internal Documentation)

[Where are these data stored?](#)

[How should these data be handled?](#)

[Where do these data come from?](#)

[Internal Contacts](#)





## Documentation

[Wiki](#)

Internal documentation

[Data Sourcebook](#) (Internal Documentation)

## Where are these data stored?

EIM Datawarehouse

Database: eim\_fame

Software name: Postges

Input format: FAME

Update method: Vintaged

Host: sqllx1

Output: FAME, CSV

US Database in /fame/data/database/us.db.

Database: US

Software name: FAME

Updated: Monthly

Update method: Partial replace and append

Host: RSMA Network

Output: FAME

US Database in distfame:/home/fame/dbs/.

Database: US

Software name: FAME

Host: distfame

Output: FAME

## How should these data be handled?

[Public](#)

Critical to function during contingency? No.

Can data be published?

	Aggregate data	
	Chart	Table
Publication	Yes	Yes
Website	Yes	Yes

Can data be shared with?

Reserve Banks: Yes

Public: Yes

No license information on file.





# Next steps

- **Dataset metadata fields still needs some fine tuning.**
- **Populating the database will take time as we fill in the new fields for existing datasets and add new datasets.**
- **Issues may recur with the ongoing development of the the VArIable Data Electronic Repository (VADER) and the Data Documentation Repository (DDR) which will contain all the variable level metadata.**





**Questions?**

**[Sandra.A.Cannon@frb.gov](mailto:Sandra.A.Cannon@frb.gov)**

**1 202 452 3710**