Everything but the Kitchen Sink:
Building a metadata repository for time series data at the Federal Reserve Board

San Cannon and Meredith Krug
Federal Reserve Board
IASSIST May 24, 2006
Survey vs. time series metadata

• Survey data observations are answers to questions:
  – Do you live alone? “Yes”
  – How much credit card debt? “$12,487.10”

• Aggregate time series data are different. Strictly numeric observations with a time component:
  – GDP for US in 2005 measured in thousands of US dollars was: $12,487.10
Different types of time series metadata

• **Structural**: the basic what, when, who, and how that define a series.

• **Reference**: the details behind what a series measures and how it was created.

• **Operational**: the administrative information necessary for storing a series.
Example: Producer Price Index for Electricity, April 2006 = 159.2

• **Structural:** Published by the Bureau of Labor Statistics (BLS); Index: 1982 = 100; Not seasonally adjusted; Not at an annual rate...

• **Reference:** Survey conducted on the Tuesday of the week that contains the 13th of the month; “Producers” are the businesses from which a given product is first purchased...

• **Operational:** April data released on May 17, 2006; Data pulled from BLS website using retrieval code WPU054; revisions occur annually in February with the release of the January data,...
Different metadata types have different audiences

• **Structural**: anyone who wants to know exactly what 159.2 measures.

• **Reference**: analysts who want more detailed information about how the BLS calculated 159.2.

• **Operational**: data managers who need to know how to store 159.2 in the database.
The old way

• **Structural**: detailed nomenclature system and 15 customized attributes stored in FAME database with data.

• **Reference**: Word processing documents converted to static HTML and updated when someone got around to it.

• **Operational**: Word processing documents converted to static HTML with cryptic instructions like “Hit enter twice”.
Examples:

- **Structural metadata**
- **Reference metadata**
- **Operational metadata**
- **Embarrassing metadata**
Management problems

• **Lack of consistency**: Information was “maintained” by different people in different formats, often on different platforms.

• **Poor maintenance**: Static files weren’t updated in a timely or organized fashion.

• **Missing clarity**: Often multiple copies existed with no clear indication of which was the “source”.
Housekeeping starts at home

• The wider the audience, the larger the task of creating a new system.
• Easiest to start with operational metadata since data managers are the only audience.
• Two main components of operational metadata:
  – Data Release Calendar
  – Data Documentation System
Time is of the essence for operational metadata

- **Our first tool**: [Data Release Calendar](#) – our first database driven product.
- The most important part of the operational metadata is knowing when the statistical release is being published.
- Moved static html pages to a relational database (SQL) and created web interface.
- Calendar contains all releases and their scheduled publication dates.
Ditching the ugly blue pages

• **Our second tool:** [Data Documentation system](https://example.com).

• Main page is an alphabetical listing with about 100 statistical releases and growing.

• Secondary pages provide details for each statistical product and what is done with the data once they are received.

• We identified the common elements of metadata needed by the data managers and stored them in a relational database with a web interface.

• “Restricted wiki” approach: all may read but only data managers may update.
Huge payoffs

• **Consistent**: Common categories are defined.

• **Well maintained**: Easy to use interface ensures that changes are made when necessary and not “when we have time”.

• **Clear**: All procedures are listed together and data managers have an organized way to communicate with each other.
Same framework for the larger reference metadata audience

- The **Data Sourcebook** is in the early stages of being transferred from static HTML to a table in the relational database.
- It is similar in structure to the Data Documentation system’s “restricted wiki” approach.
- We are still finalizing the secondary pages.
- Analysts as well as data managers can update these metadata.
Unfortunately for structural metadata, one size doesn’t fit all!

- The structural metadata are closely tied to the data.
- The primary source should be the FAME database with the data.
- We made a copy of the FAME attributes and stored them in a relational database and created a web interface.
Tying it all together

• Once we finish the Data Source Book all our metadata will be in a single physical repository… so now what?

• Challenges ahead:
  – **Time sensitive:** Adding time component to metadata to complete our vintage data collection.
  – **Searchable:** Working on building a search interface that will allow users to retrieve information regardless of the type of metadata.
  – **Meaningful display:** Determine how to clearly deliver different types of metadata in a useful interface.
The last slide

We’re not done yet, but we’re getting there!

Comments and suggestions are always welcome!

Thanks for your patience!

San Cannon (scannon@frb.gov)
Meredith Krug (mkrug@frb.gov)