

# Building a Strong Foundation with Dublin Core

Linda Powell  
DCMI 2009

The views expressed in this paper are those of the author and do not necessarily represent those of the Board of Governors of the Federal Reserve System.



# Background

- Board of Governors of the Federal Reserve System
  - U.S. central bank – engages in monetary policy and banking supervision
- Improvements in technology led to improved data availability (more and better)
- Changes in the financial markets resulted in the need for more data



# Historically

- Purchased Data were used for specific research projects
  - Therefore, economists that purchased data maintained the data themselves
- Data collected by the Federal Reserve System were used for policy work
  - Therefore, data maintenance was centralized and well documented



# Business Problem

Currently no matter how data enter the Federal Reserve System (FRS) they may be used for policy work as well as research projects

Data collected by the FRS was well documented and centralized

Data purchased was decentralized and not centrally documented



As of October 2009

The Board has three metadata repositories

1. Collection level metadata (DANCE)
2. FRS collected data (MDRM)
3. Vendor metadata repository

All three serve a purpose but they didn't work together



Users want more and they want it in one retrieval

As soon as DANCE was implemented in 2004 users wanted more:

- Lots more metadata
- A way to make centralized notes about datasets
- More datasets

It was time to look at international standards

So we decided to look at enhancing all three metadata repositories



# What Standards did we Review?

Although additional standards were reviewed, five candidates for the Board's metadata warranted in-depth evaluation;

- SDMX and XBRL (which are familiar to Board Staff)
- MARC (which is familiar to library staff)
- Dublin Core and DDI (which were not familiar to Board staff)



## We Couldn't Find a Perfect Fit

SDMX - geared toward aggregate time series data

DDI - good collection and variable level metadata but is geared toward social science rather than financial data and wasn't (at the time of review) persistent over time.

MARC - more bibliographic than we needed.

None was perfect but Dublin Core has a robust set of collection level data and XBRL provided guidance for the firm level financial data.

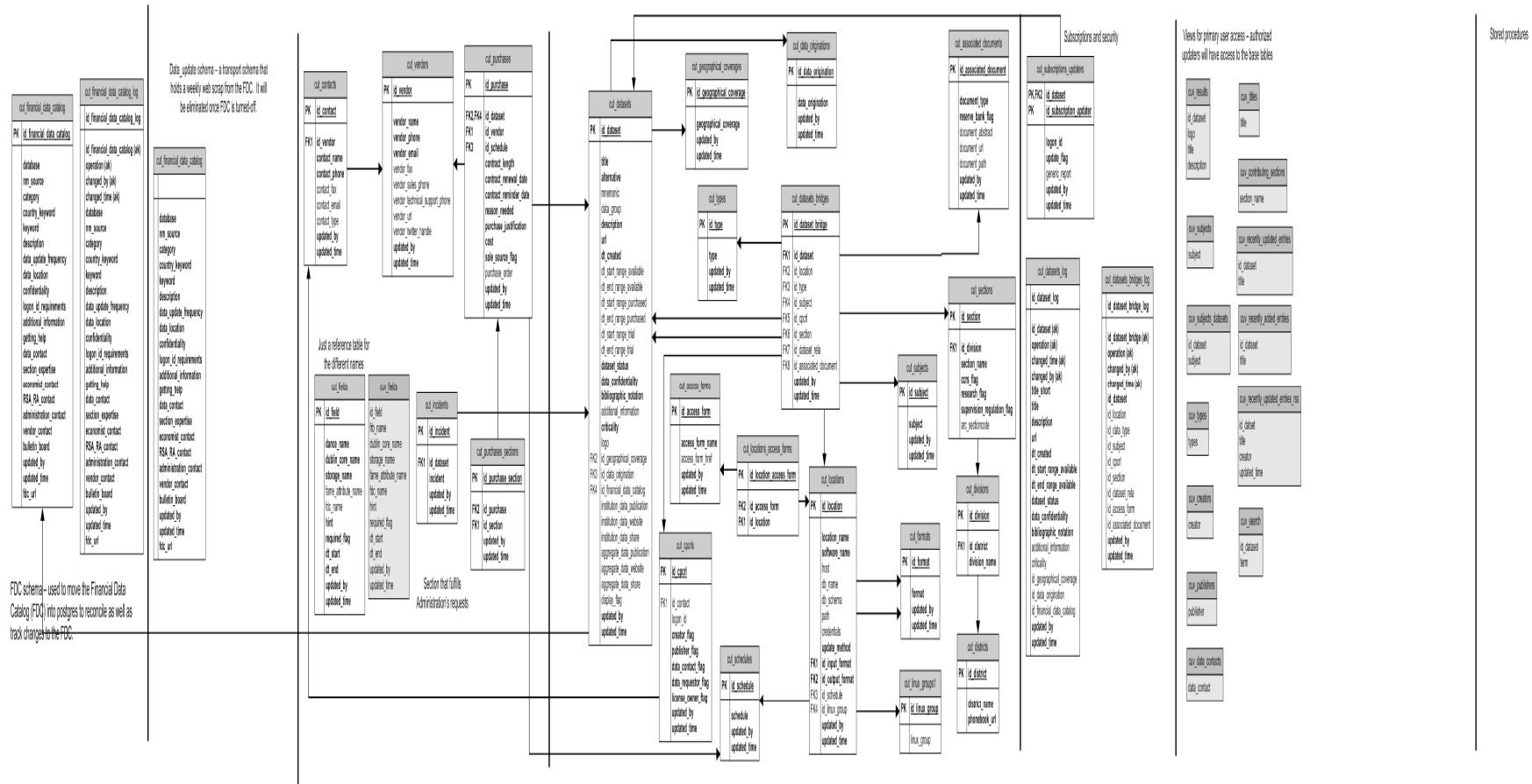
# Dance Variables

## Original vs. Dublin Core

Original Variables	FRB Name of Retained/ New Variables	Dublin Core Element
category		
database	title	title
	alternative	alternative
database URL	product url	
description	description	description
keyword(s)	keyword	subject
vendor	publisher	publisher
vendor URL	vendor url	
	creator	creator
	record id	identifier
	date created	date
	date range available	available
	geographical coverage	coverage
status	dataset status	accrual policy
	data origination	source
	related resources	relation

Original Variables	FRB Name of Retained/ New Variables	Dublin Core Element
form of access	data location	
form of access	format	format
	type	type
	physical medium	medium
	update method	accrual method
	software name	software
	update schedule	accrual periodicity
Data Contact	data contact	
Division	purchasing division	
	data requestor	
	contributing section	contributor
	purchasing section	
	vendor contacts	
license contact	license owner	rights holder
license information	data confidentiality	rights
	bibliographic notation	bibliographic citation
	license agreement	license
	help files	instructional method

# DANCE Schema



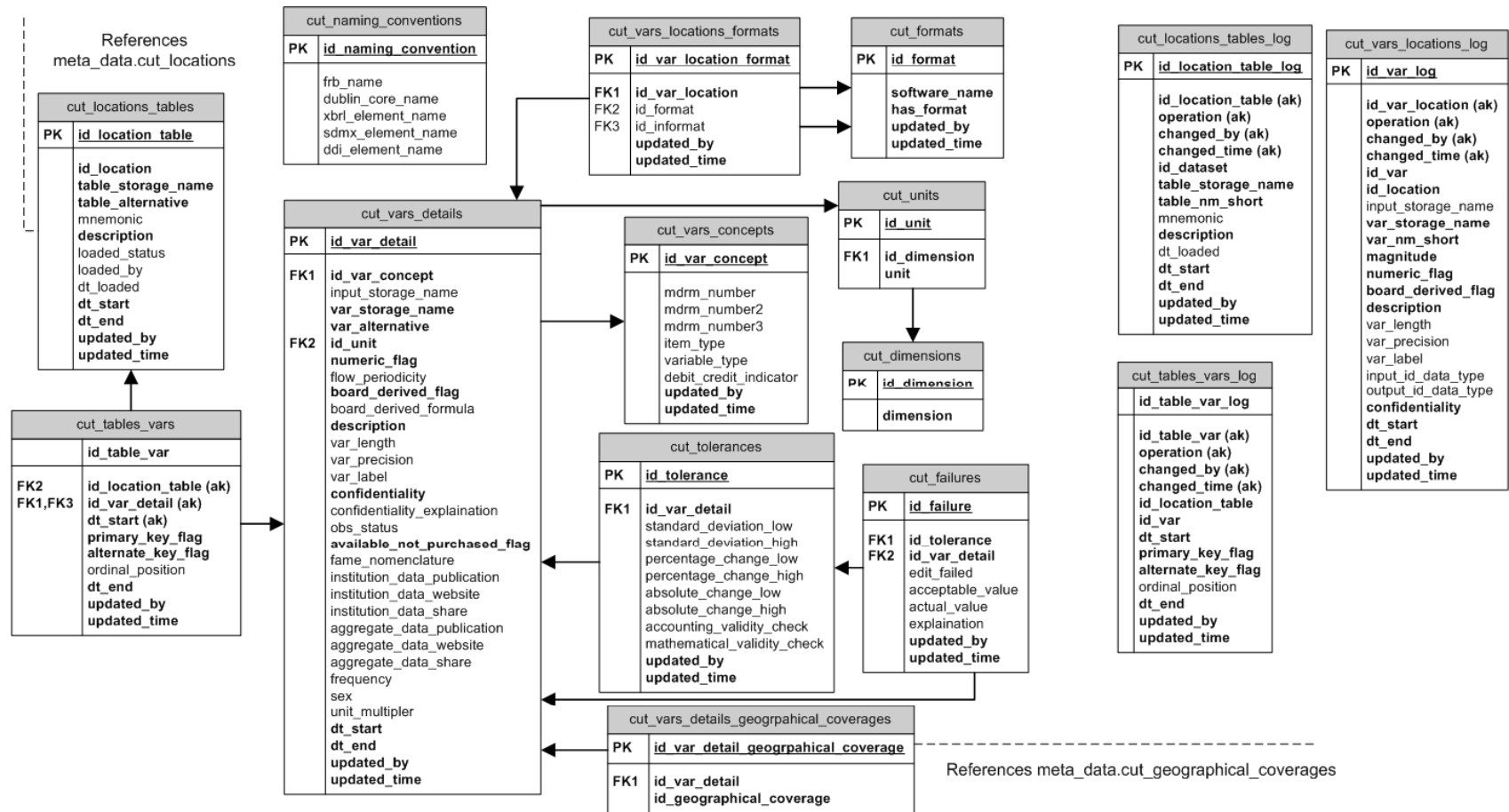
# Vendor Metadata Repository

## DC and XBRL Variables

FRB Name of Retained/ New Variables	Dublin Core Element	XBRL Element
variable storage name	title	
variable name short	alternative	
mnemonic		
description	description	
mdrm number		
id variable	identifier	
date start		
date end		
ordinal position		
flow-stock		flow-stock
flow periodicity		flow periodicity
debit / credit indicator		debit / credit indicator
data type		data type

FRB Name of Retained/New Variables	Dublin Core Element	XBRL Element
multiplier		
numeric flag		
derived flag		
Formula		formula
variable length		
variable precision		
confidentiality		
format	hasformat	
record last updated		
record updated by		

# Vendor Metadata Schema





# Open Source Software

Explored using an open source repository

- EPrints,
- FEDORA, and
- DSpace – most promising and used a PostgreSQL database

To meet the requirements for the new DANCE application the open source system would need significant customization.

- Integrate with data security information
- Integrate with Vendor Metadata repository
- Allow end users to drill down through the application for related information
- Provide a wiki bulletin board

Decided to build our repository in-house



# Conclusion

Advantage of using the Dublin Core standard

- Benefit of collective experience
- Knowledge of hundreds of metadata professionals
- Interoperability between systems that support Dublin Core
- Better final product and fewer changes after implementation