Integrating Business and Technical Metadata to Support Better Data Quality

ABSTRACT----------

How can integrating Business and Technical Metadata help in your efforts to gain better quality of your Enterprise wide information? How can it help to consolidate the multiple silos of information that your Company has and continues to produce? How can you link all of that information to a Data Steward or Business Owner who has the answers? At what Level (Enterprise or LOB) do those Business metadata elements get associated to the Technical Metadata? And how about Business Terms or Business Data Assets; does one description fit your enterprise? How about information that is outside of the ETL world; how will you integrate that metadata. Now, once these issues have been resolved, who is going to load/update these elements on an ongoing basis? And don't forget the ever growing demands for compliance and business decision support system information.

See how Citigroup's Center of Technical Excellence (CTE) group within the Enterprise-wide Citi Architecture and Technology Engineering (CATE) organization, used an extended ETL Metadata Repository to create an Enterprise-wide Metadata Repository. Complete with Business terms, Grouping, Data Models and Databases, Data Governance (Business Ownership and Data Stewardship) workflow process and approvals, as well as Source to Target traceability of Metadata Elements, and auditing aspects for compliance issues.

Although this presentation is made for the CIO, Enterprise Architect level, Bob's technical presentation style also appeals to the Data professional level and everyone will come out of the presentation with something useful.

BIOGRAPHY----------

Robert Schork
Citi Architecture and Technology Engineering (CATE) Center
for Technical Excellence (CTE) group
Citi

Bob is a team leader with the Citi Architecture and Technology Engineering (CATE) Center for Technical Excellence (CTE) group at Citi. Bob has over 25 years of IT experience and over 15 yrs of metadata management experience. He has installed both Platinum (CA) and Rochade (ASG) metadata repositories and has modeled Enterprise Data Warehouses and other enterprise architectures. He has also installed several packages which include tool evaluations, data classification, data mapping and API accesses from upstream and to downstream systems. Bob has also created several metamodels as well as developed the scanners for loading and extraction processes for reporting.

Bob has performed metadata analysis of several Fortune 500 corporations. He has a background as a BA, Developer, DBA, Data Analyst as well as a Data Administrator. He has implemented several Metadata solutions, including the creation of Standards and Procedures to help the enterprise processes function properly. He currently works in the group that is responsible for

setting enterprise-wide architectural, data, and metadata standards for Citi (formally called Citigroup). Bob is also a board member of the Metadata Professional Organization (MPO) and DAMA-NJ and has presented at several Data Management conferences and FIMA events.

Integrating Business Metadata into Technical Metadata

How Citi Extended an ETL Repository Enterprise-wide to support Information Stewardship and Data Quality

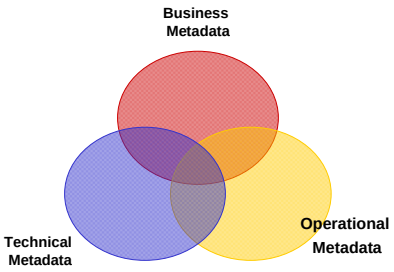
Who am I?

- Over 25 yrs of IT experience (including consulting)
- Over 15 yrs of Metadata experience including Metadata Analysis and Design for Fortune 100 corporations
- Implemented ASG Rochade and CA Platinum Repositories including the maintenance and reporting processes
- Created several Homegrown Metadata Repositories, Scanners and Governance processes.
- Current Board member of the DAMA NJ Chapter, and the Metadata Professional Organization (MPO)
- Former member of the IBM Data Governance Council and MeetTheBoss
- Presenter at several Metadata and FIMA conferences
- Former BA, Developer, DBA, Data Modeler, Data Architect

Current Assignment

TEKsystems® is the leading technical staffing and services company in North America. With 26 years of experience, we know the industry and the importance of deeply understanding each customer's business situation. We also know the right team is essential for a project to succeed. Our proven staffing and project delivery processes enable us to support your critical engagements with the best experts in the market – those who are thoroughly qualified to achieve the results you need.

What is Metadata?



Definition:

- Data about Data (Useless)
- Metadata is the *definition* and *usage* of the information about the data of your organization.®
- 3 Types
  - Business
  - Technical
  - Operational
- Few really understand it, but it is not Pie in the Sky
- Can not do Data Integration or true Data Quality without it.

AGENDA

Citi's Dilemma and Possible Information Stewardship Solutions

- Information Stewardship
- Getting Business Buy-In
- Extensions and Integration points
- Metadata Governance
- Future Directions

Information Stewardship

Definition: The willingness to be accountable for a set of business information (Data Assets) for the well-being of the larger organization by operating in service, rather than in control of those around us. Larry English

Why establish this first:

- **Does control of other groups generally work?**
- **Can you define who is accountable for your corporations Data?**
- **Does the information exist? If so, where is it?**
- **If you find the information, what does it mean?**
- **How does what you are doing, benefit their organization?**
- **Can the information be integrated?**
- **Can this support Data Profiling?**

Dilemma at Citi

Citi had some decisions to make

- Goal was to capture Business Metadata and associate that with the existing Technical and Operational Metadata.
- Need to incorporate that throughout the Enterprise that has several distinct Business Sectors or LOB.
- Should we build, buy, or use an existing system?
- A Metadata Repository was tried before and failed. Why?
- Could the Ab Initio ETL metadata repository be extended?
- How can we capture and incorporate Data Governance (Business Ownership and Data Stewardship) into the overall strategy.
- How can we sell this to the Business Customers
- Can the existing resources (people and tools) be utilized?

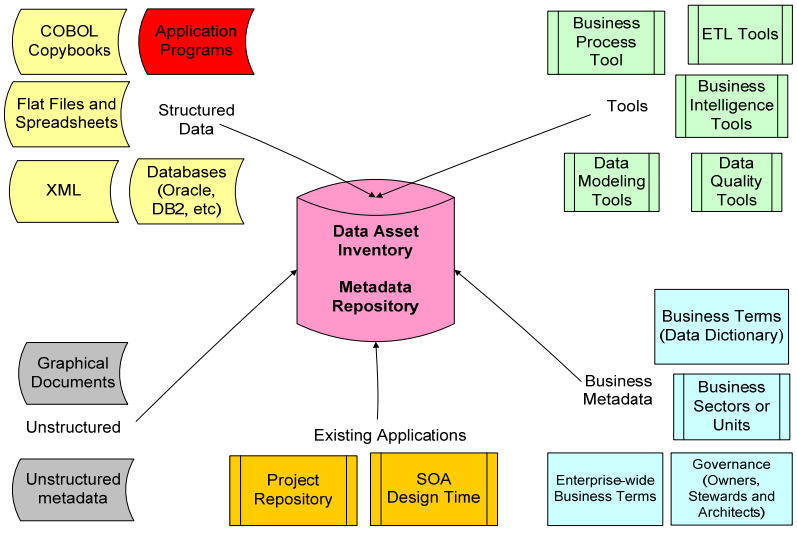
First Step Decisions

- Find a vehicle to capture Business Metadata in various forms.
- Define a process to associate that Business Metadata with the existing ETL Metadata elements.
- Architect distinct Enterprise and Business Sectors separations of Business Term Metadata.
- Implement a POC to determine if the ETL Repository was extensible, easily loadable, and reportable.
- Determine if the existing Business Process Manager workflow engine could communicate and support Business Ownership and Data Stewardship approvals. Unexpected Twist.
- Develop loading methods for each different metadata type.
- Sell it to a skeptical Sector base who has been burned before.

Metadata Analysis

- Train those who do not know metadata about metadata
- Define the scope of the Metadata elements
- What type of metadata to capture
- What vehicle to capture that metadata
- Test the scope
- The Business Users DO NOT define the repository scope
- Group the like metadata elements
- Test it for validity
- Create the API accesses (Loading and Reporting)
- Create the metadata Governance for each element
- Update the SDLC
- Auditing functions need to be established for compliance

Citi Data Stores



The Problem with ETL Repositories

- They are not a full Enterprise Repository, but they have advantages.
- Their enterprise view is limited to the metadata elements within their knowledge. Good meta modeling skills are needed for extension and integration.
- Metadata outside the ETL tool is not generally accounted for. You must create the scanning API accesses to load and integrate the outside metadata elements into the matching ETL Repository elements.
- Most are very weak at Business Metadata capturing and usually have an enterprise view of those terms.
- Problem with understanding and implementing the various complexities of Business Metadata to Technical and Operational metadata.
- In order to report on your extended metadata elements you may need to extend the reporting GUI that already exists.

AGENDA

- Information Stewardship
- Citi's Dilemma and Possible Solutions
- **Getting Business Buy-In**
- Extensions and Integration points
- Metadata Governance
- Future Directions

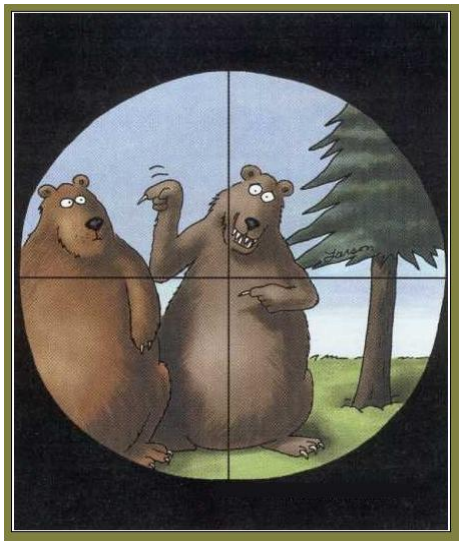
Management Reactions

- Some Management Reaction
- Buy Another tool?
- Who will lead it?
- What does the Business Customer want?

Citi Sector Management Reaction

Management View on Tool Purchases

Who Would Lead This Endeavor?



What does the Business want?

MIT IQIS Conference 2009 17

What does the Business really want?

TIPS

- Find out where their pain points are.
- Is there a manual process that a Repository can help with?
- What is their biggest time consumers?
- FORGET what the Vendors/Consultants have told them.
- Where do they see their groups 5 years from now?

TIPS FOR YOU

- Partner with them; don't do all the work
- Be Realistic: "I don't know yet" is a valid answer.
- Look for opportunities they may not know about.
- Promote and Sell to other groups.

MIT IQIS Conference 2009 18

The Team

- The Development Engineering group of our Enterprise Architecture group stepped up and said it can be done.
- The Center of Technical Excellence (CTE) group took the lead.
- CTE Team consisted of the following:
  1. Project Manager
  2. Metadata Architect
  3. Sr. ETL Architect
  4. Jr. ETL Architect
- Extend the ETL Metamodel.
- Develop a plan to load and report on Metadata
- Get the Vendor involved.

Citi Business Metadata Needs

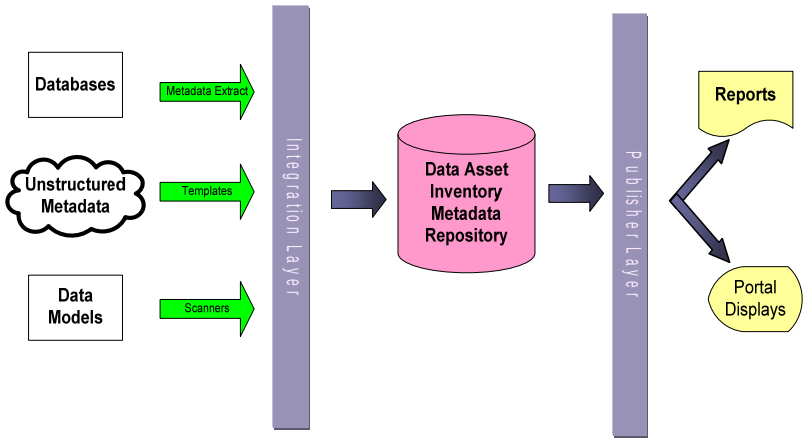
ACTION OPTIONS

- Integration between the Business and Technical metadata.
- Traceability of Data Elements throughout their applications.
- Wanted to know who owns what. (Information Stewardship)
- Other considerations that surfaced:
  - Data Redaction/Obfuscation
  - Information Security Classifications
  - Data Governance email metadata change notifications
  - Impact analysis for major initiatives

Key Repository Highlights

- **Data Asset Inventory (DAI) stores consistent consolidated views of Enterprise and Business Sector Data Assets linked to technical metadata**
  - Includes Data Models, Databases and ETL information
- **DAI incorporates Standard Templates and Scanners to load all metadata types.**
- **DAI stores, displays and provides at a glance, the contact information for the Data Governance workflow showing the Business Owners, Data Stewards and integrates that information with each metadata element (entity/table level).**
- **Allows easy reporting of metadata information via a Web Portal interface.**

Metadata Flow



AGENDA

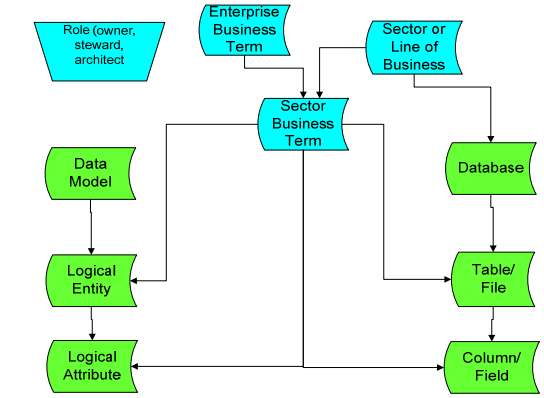
- Information Stewardship
- Citi's Dilemma and Possible Solutions
- Getting Business Buy-In
- **Extensions and Integration points**
- Metadata Governance
- Future Directions

Levels and Type of Metadata

The Metadata Repository currently supports the following major asset types:

- Enterprise Level Data Assets (Business Terms)
- Business Sector Level Data Assets
- Logical and Physical Technical Metadata
- Entity/Table/File Level information:
  - Business Ownership and Data Stewardship Metadata
  - Data Profiling Results and Valid Values
- Attribute/Column/Field level information:
  - Data Obfuscation/Redaction Information
  - Information Security Classification information
- Physical Operational Metadata with some parameters

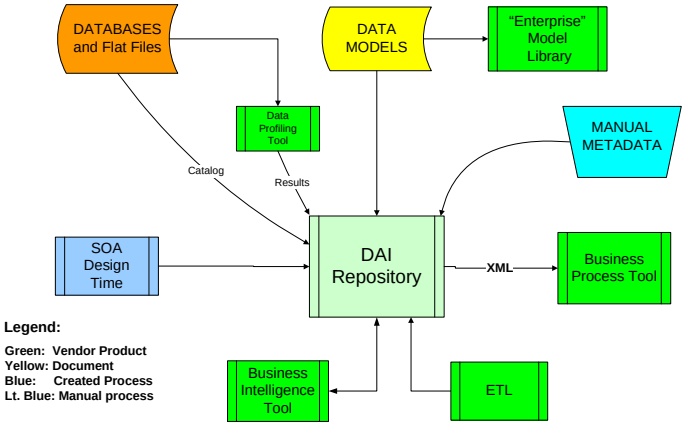
Extended Metamodel Components



Legend:

Green: Vendor product structure
Lt. Blue: Extended Metamodel Structure

Overall Architecture Components



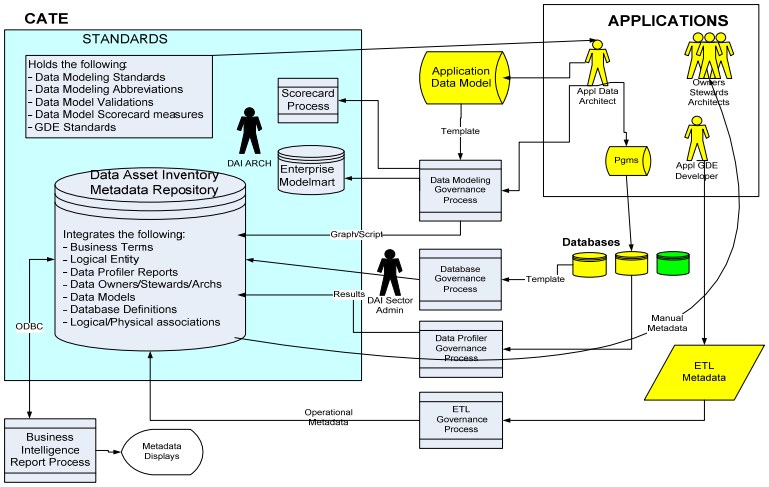
Legend:

Green: Vendor Product
Yellow: Document
Blue: Created Process
Lt. Blue: Manual process

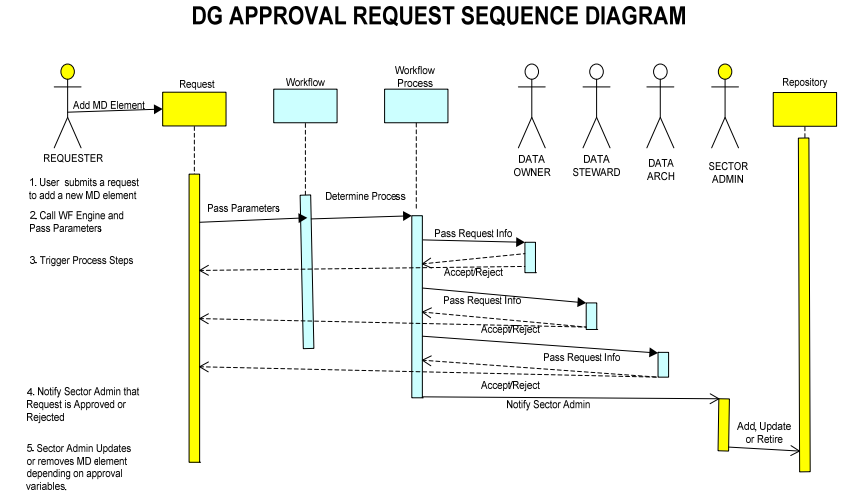
AGENDA

- Information Stewardship
- Citi's Dilemma and Possible Solutions
- Getting Business Buy-In
- Extensions and Integration points
- **Metadata Governance**
- Future Directions

Repository Loading Process



Data Governance Process



Features and Benefits

Features

- Metadata element search capabilities
- Contains Data Model information including graphical document of the Data Model
- Lists all the Data Assets that stakeholder is entitled
- Contains links to Business Terms, Logical and Physical elements
- Contains all audit trails and versions of the data element
- Initiates the approval process notification to the Owners and Stewards.

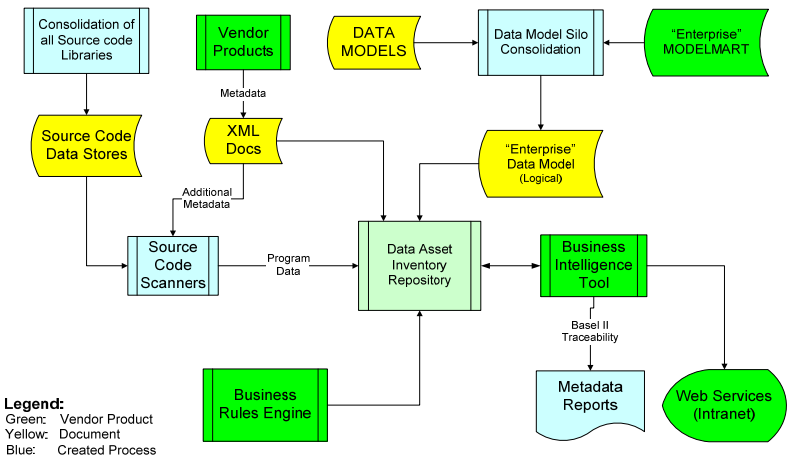
Benefits

- Lists all Enterprise Data Assets with its associations
- Enables control over Sector Business Data Assets
- Contains traceability information and its associations
- Aids in compliance tracking
- Can hold XML document information
- Can store and integrate Data Model elements outside of the Data Models
- Can hold and attach documents to any Metadata elements

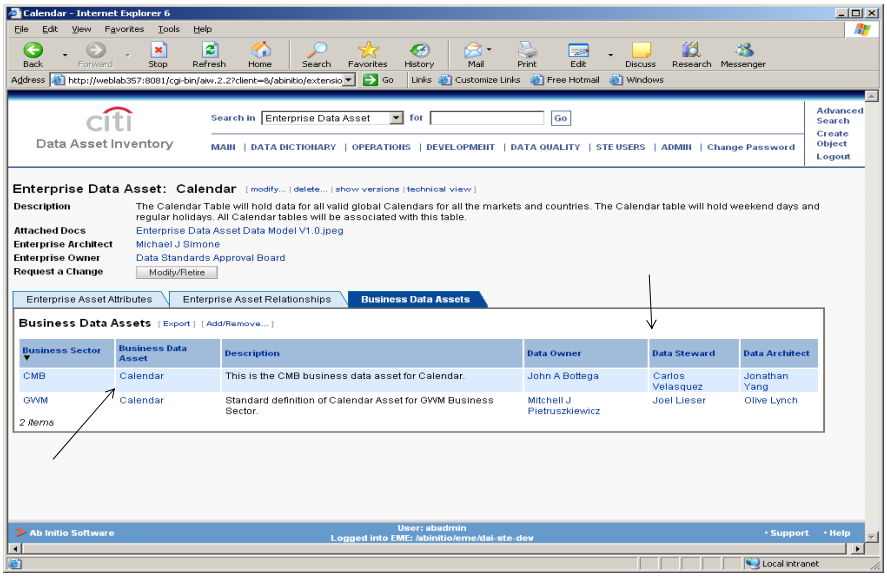
AGENDA

- Information Stewardship
- Citi's Dilemma and Possible Solutions
- Getting Business Buy-In
- Extensions and Integration points
- Metadata Governance
- **Future Directions**

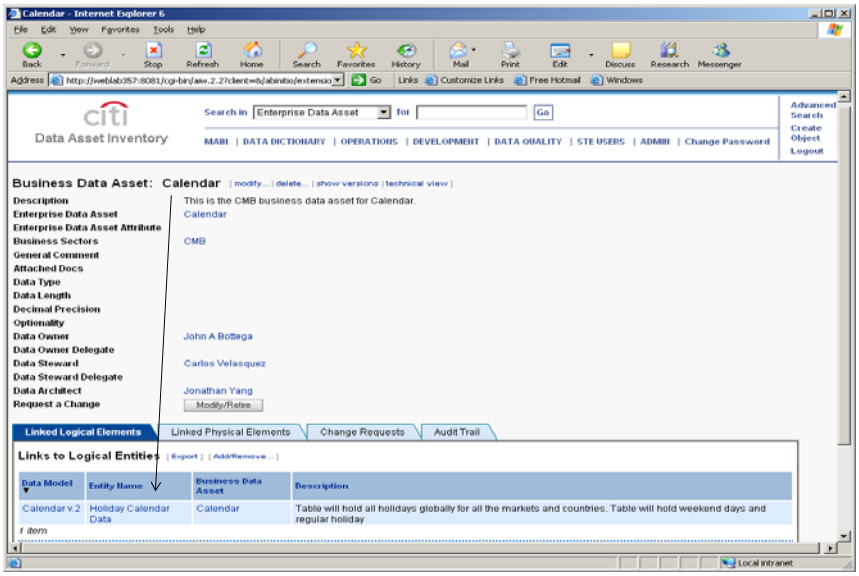
Future Architecture



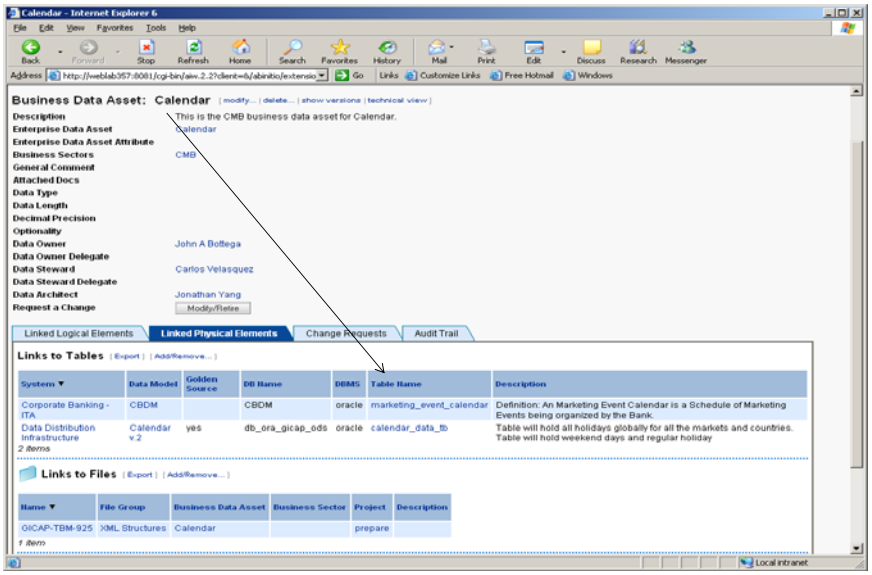
Enterprise Data Asset



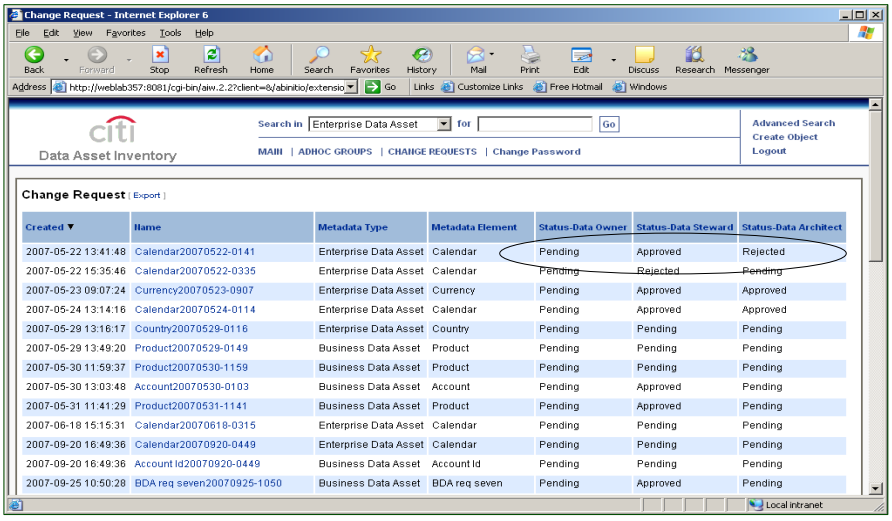
Business Sector Data Assets



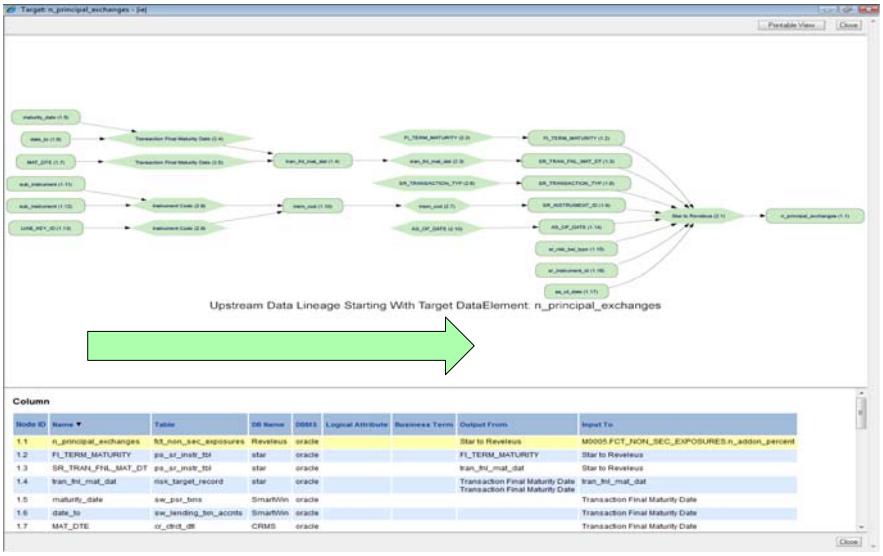
Business Data Asset Links



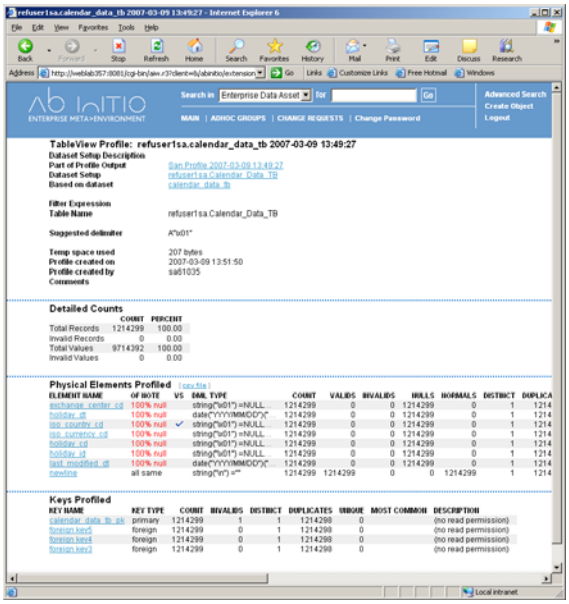
Change Requests



Traceability



Data Profiling Results



Other Considerations

- Metadata must be reported on the Production Level
- Modify SDLC procedures to enhance metadata capturing
- Must integrate all local ETL instances into the Global DAI.
- Consolidate Reference Data to create Enterprise Entities
- Provide Educational Services of Data Governance Procedures
- Implement Data Profiling and Data Quality processes for on boarding applications
- DO NOT let other areas dictate to you what should be in the Repository. They will want to use it as an application.
- SELL, SELL, SELL
- BUT, Keep in mind…

MIT IQIS Conference 2009 39

Pitfalls

NEVER TRY TO TEACH A PIG TO TALK

IT WILL WASTE YOUR TIME AND ANNOYS THE PIG!

MIT IQIS Conference 2009 40

Questions

Contact Information

Bob Schork

Formerly of Citigroup, Warren, NJ

Currently Consultant for TEKsystems, Charlotte, NC

Assigned to Bank of America as a Metadata Architect

robert.schork@bankofamerica.com

bobschork@hotmail.com