Metadata Management Best Practices and Lessons Learned

Presentation at

2006 DAMA / Wilshire Metadata Conference

Denver, CO

John R. Friedrich, II, PhD

Friedrich@metaintegration.net

Outline

- Recent developments in metadata management
- New opportunities
- New challenges and Lessons Learned
- Conclusion

Format of This Presentation

- Outline to “stay on the path”
- Background to “level the playing field”
- Example for clarity of understanding
- Real-time example for credibility

Recent Developments in Metadata Management

What is “new” out there?

Recent Developments: Metadata Exchange Supported by Vendors

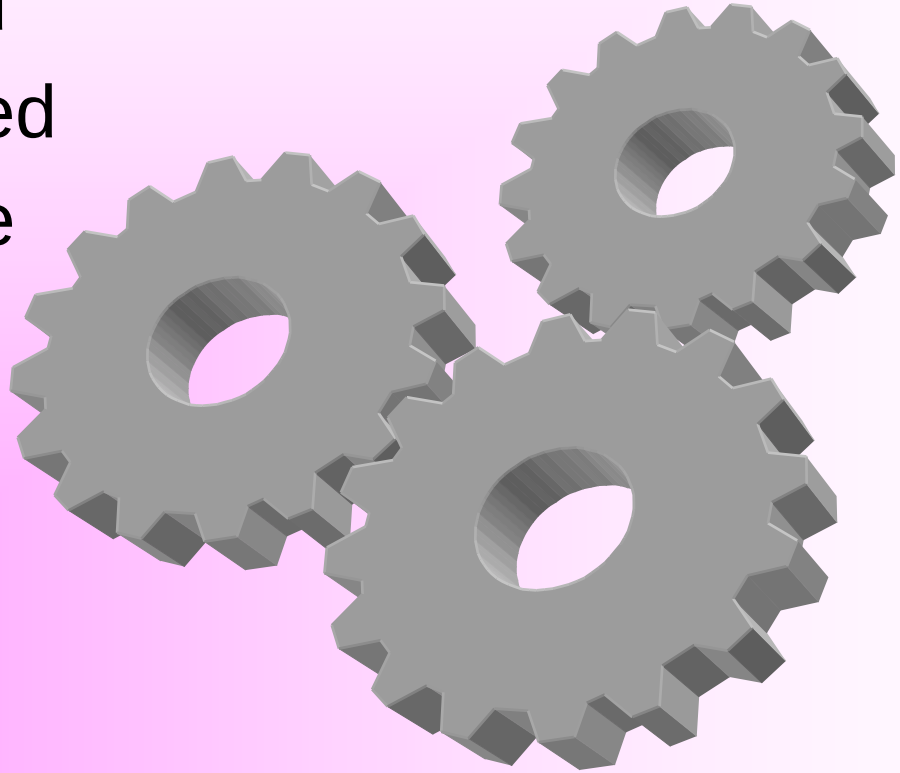
- Nearly all recognize the need for metadata exchange
  - Especially across different “types” of tools
    - Warehouse design to ETL or BI
    - ETL to lineage analysis tool
    - BI to Enterprise Reference Model
- E.g., Multi-Vendor panel with 14 panelist
  - Each one has metadata exchange capabilities
  - Most built in to the tools

Recent Developments: Multi-vendor Metadata Accessibility

- Metadata hubs with multi-vendor capabilities in one product
  - Over 90 products integrated into a tool
  - "Metadata services"
    - Not just "one stop shopping" for metadata, but for metadata accessibility services

Recent Developments: Automated and Efficient Metadata Access

- Not just services, but automation services
  - Server based
  - Process based
  - Customizable

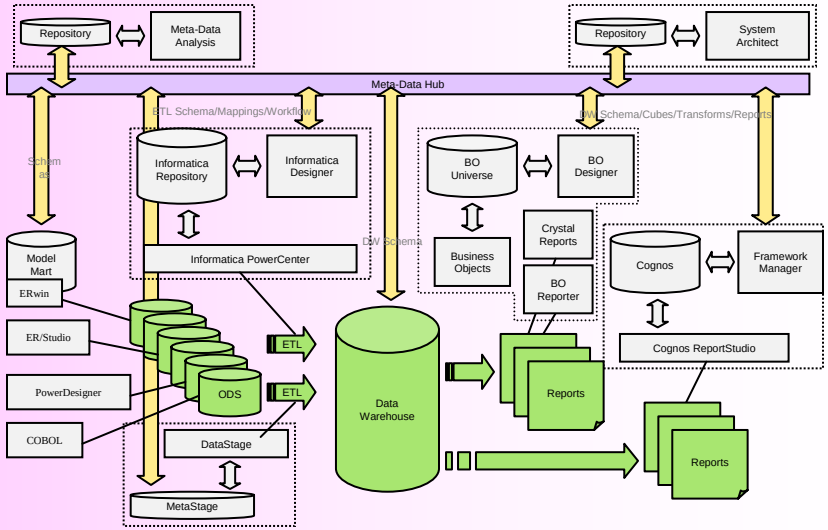


New Opportunities

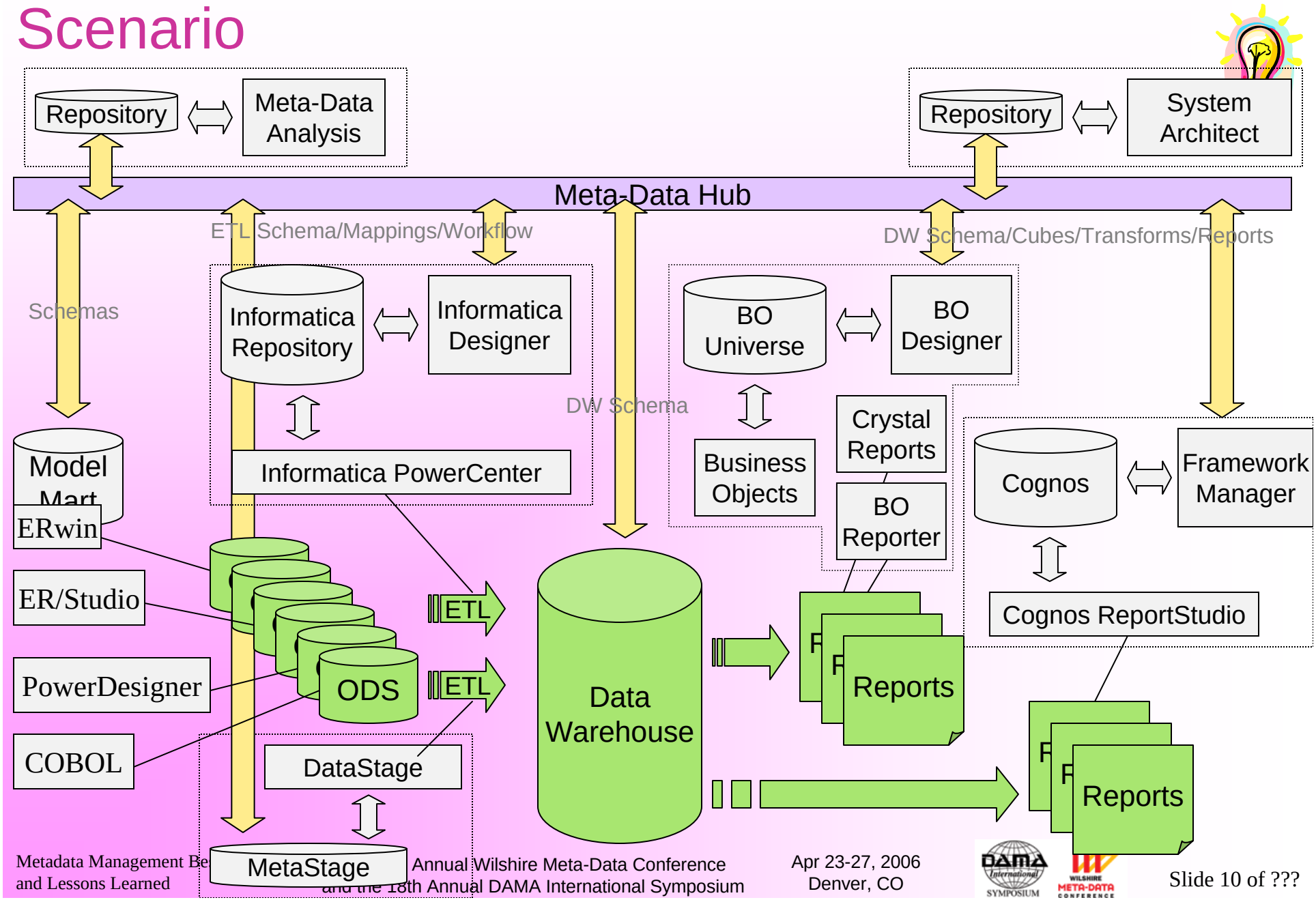
Out of these developments come opportunities.

New Opportunities: Multi-vendor Metadata Analysis

- Accessibility + Metadata Storage ➔
- Throughout the entire data lifecycle
  - Operational Data Stores
  - ERP
  - ETL
  - EAI
  - EII
  - DW
  - BI



New Opportunities: Multi-Vendor Metadata Scenario

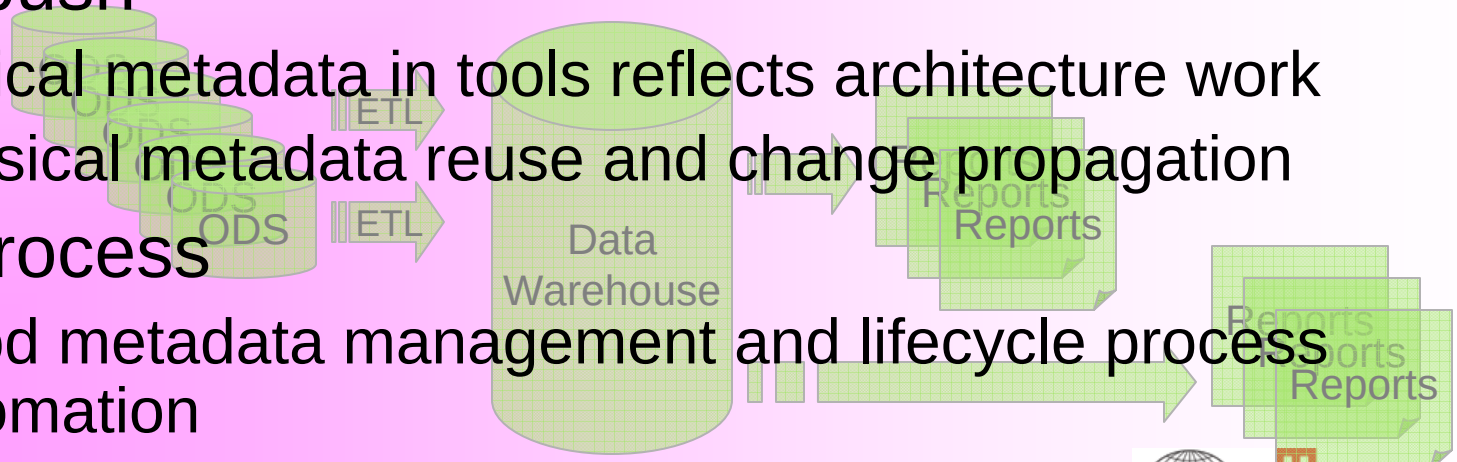


Show and Tell

Let us stop and build something here.

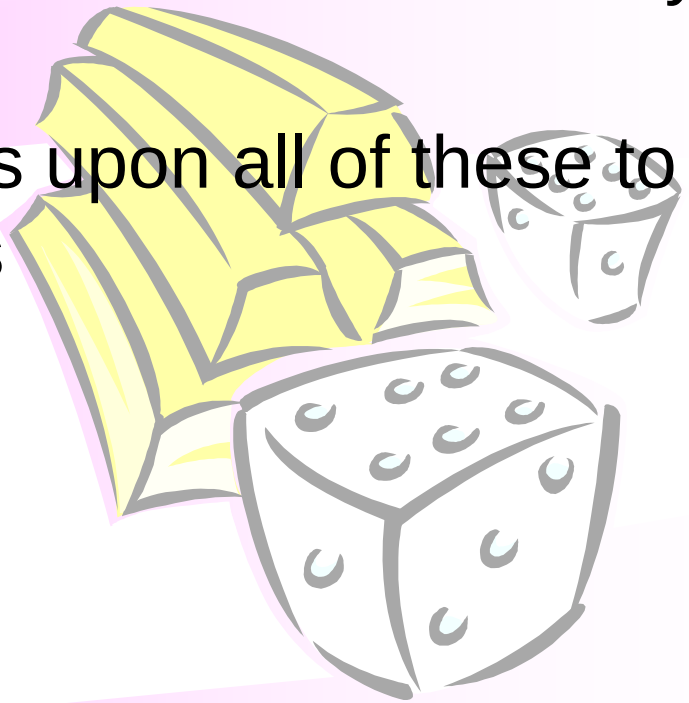
New Opportunities: Up-To-Date Physical (and Logical) Metadata

- Accessibility + Automation ➔
- The “pull”
  - “As close to the grove as you can get” physical metadata
  - Physical (real-world or data tool) driven data life-cycle
    - ETL transforms really can define the data flow in the repository
  - Logical lineage derived from physical “reality”
- The “push”
  - Logical metadata in tools reflects architecture work
  - Physical metadata reuse and change propagation
- The process
  - Good metadata management and lifecycle process automation



New Opportunities: What-If Impact Analysis

- Accessibility + Automation + Process ➔
  - Not *just* “one version of the truth”
  - Multiple future “configurations” of metadata may be captured
  - Analysis of change impacts upon all of these to be or proposed configurations
  - Deployment planning
  - Impact risk assessments



New Opportunities: Historical Business-Oriented Lineage Analysis

- Accessibility + Automation + Time ➔
  - Reverse lineage ("where *did* it come from") is often an historical question
  - Sarbanes-Oaxley is for a year, at least
  - BASEL II is up to five years of history
  - Last quarter's sales is last quarter
  - Today's "version of the truth" is not yesterday's, just as it is not tomorrow's (what if impacts)

New Challenges

If it can be done, it has been, in one form or another.

Only the unlikely or impossible are worth striving for.*

New Challenges: Multiple Repositories

Development Metadata Repositories

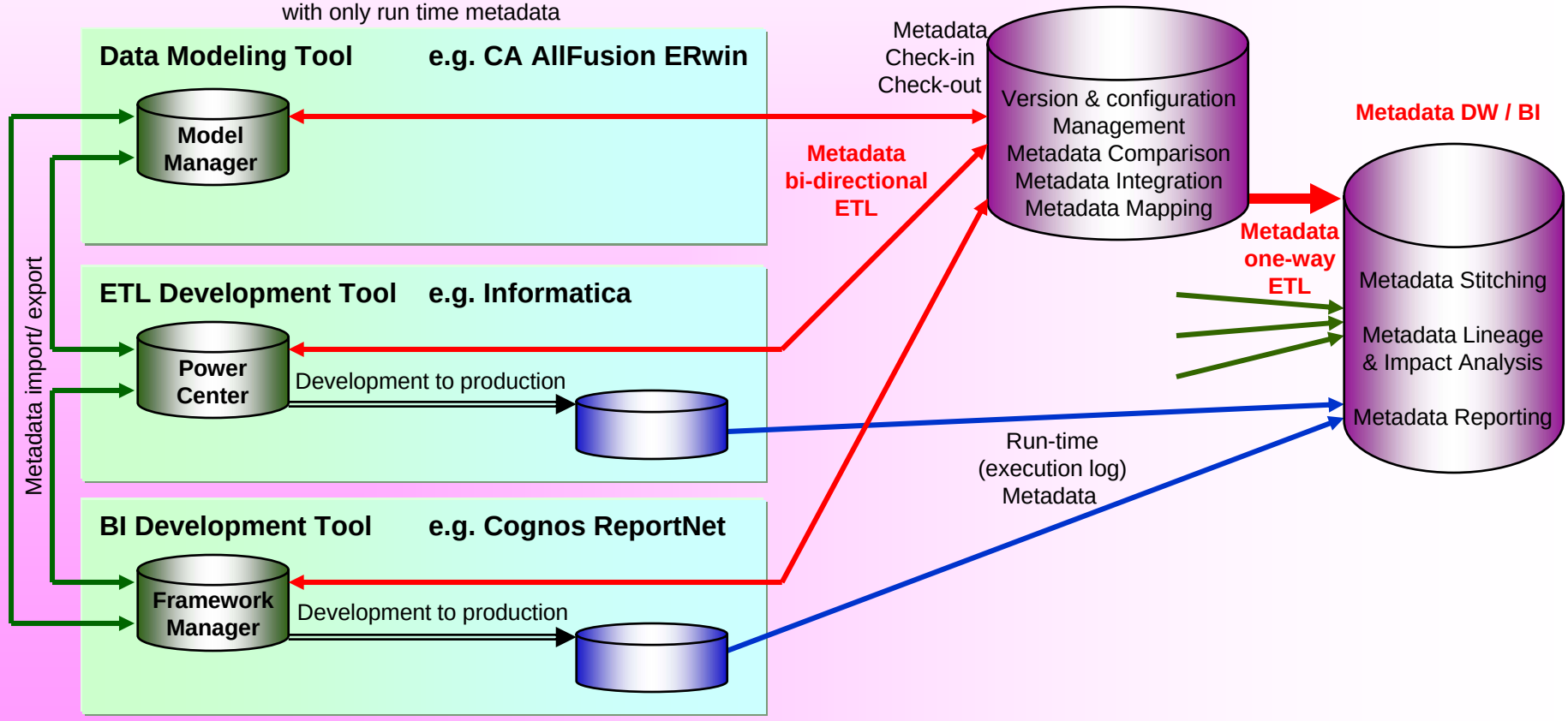
Operational Metadata Repositories

Life Cycle Metadata Repository

Analysis Metadata Repository

The development and operational metadata repositories can be the same product (development vs. production instance) or the operational repository can be a specific product with only run time metadata

The life cycle and analysis metadata repositories can be the same product.



Lessons Learned: Multiple Repositories

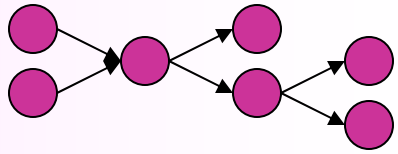
- Learn from the data lessons
  - A single grand repository, like a single grand database, is not going to happen
- "Embrace diversity":
    - Use the ETL tool to describe data movement transformations and workflows, the BI tool for Cubes and reports, the CASE tool for design, etc.
    - Pitfalls of the "round-trip"
    - Capture tool-specific metadata, share normalized metadata.
- Remember the word "standards" **always** has an "s" on the end of it!

New Challenges: Version Management

- Many repositories and tools x many models x time and change ➔
  - A version for each!
  - Several new dimensions to the repository
  - Answer the difficult questions, not the “single version of the truth” assumption-based ones

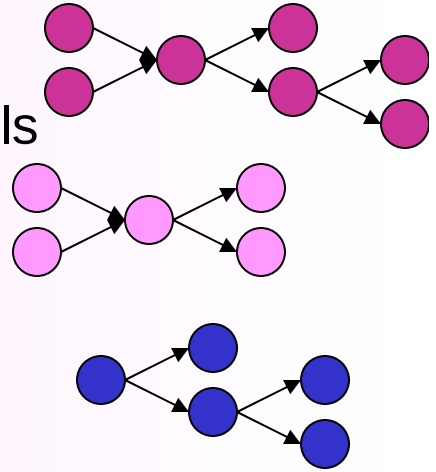
Lessons Learned: Version Management

- Need true version management
  - Maintain multiple versions, not just deltas
  - Historical path (version traceability)
  - Process (milestone) driven
  - Fully automated (don't muck around in the repository)
- Bonus: Process based metadata quality



New Challenges: Configuration Management

- Versions x deployments x what-ifs x organizational structure x . . . ➔
  - True configuration management with many configurations of many versions
  - Many dimensions of CM problem:
    - Multiple deployed versions of each of the source systems,
    - Multiple design, developmental, beta, etc.
    - Multiple version of standards and/or reference models
    - Multiple versions of data migration transformations
    - Multiple business organizational “cuts”
    - Multiple IT organizational “cuts”
    - And many, many more



Lessons Learned: Configuration Management

- There are many ways to slice it
- Must plan ahead
- Tie configuration organization to:
  - Data Flow!
  - IT deployment an responsibilities
  - Milestones
  - Business organization
- Manage fundamental (separately versioning) components *separately* in the data flow
- Most of your time will be spent telling the metadata what the separate tools did not understand about each other ➔ STITCHING

New Challenges: Automation, Processes and Metadata Quality

- Complexity of access processes, versions, and configurations ➔
  - Must automate
  - Must automate metadata management (which are data management driven) processes
  - Automation means making mistakes very quickly, so must ensure quality of metadata, version and configurations
  - Don't want to go to jail due to a bad SOX answer!

Lessons Learned: Automation, Processes and Metadata Quality

- This is meta-automation (I guess)
- Repository (metadata) administration is NOT very often administration of the repository (metadata)
- Repository is most often administration of the processes
- These processes must be derived from the data processes
- As with SOX, quality comes implicitly from, and is monitored by way of the process

Conclusion

Conclusion

- Recent Developments in Metadata Management
  - Multi-vendor Metadata Accessibility
  - Metadata Exchange
  - Automated and Efficient Metadata Access
- New Opportunities
  - Multi-vendor Metadata Analysis
  - Up-To-Date Physical Metadata
  - What-If Impact Analysis
  - Historical Lineage Analysis
- New Challenges and Lessons Learned
  - Multiple Repositories
  - Version Management
  - Configuration Management
  - Automation, Processes and Metadata Quality