Data Warehousing and Business Intelligence
Tools for IR: How do we get there?

Marilyn H. Blaustein
Alan H. McArdle
Banu Solak
Office of Institutional Research

Northeast Association for Institutional Research
November 10, 2009
Session Overview

- UMass at a Glance
- Current Systems
- Data Warehouse: How did we get here?
- Amherst Campus Project
- Delivered System Architecture
- Evaluating the Delivered Model
- Understanding the Delivered Model
- Adapting the system to our needs
- From Delivered Model to IR Models
Session Overview

- What we found – The Good and the Bad
- Data Quality
- Staff Issues
- Performance
- Security
- Delivered Tool Maintenance: Bundles
- OIR Development: Census Files
- Lessons Learned
- Where do we go from here?
UMass at a Glance

- Flagship of 5 campus system of 66,000 students
- In Fall 2009
  - 27,016 students
  - 5,380 employees
  - 1,388 instructional faculty
Current Systems

- Oracle/Peoplesoft for HR, Finance and Student
- Human Resources & Finance managed centrally
- Student
  - Amherst managed locally
  - Three campuses managed centrally
Data warehouse: How did we get here?

- Pre 2006...
- 2006: Amherst campus supports investment in reporting solution (data warehouse or data mart)
- 2007: UMass five campus collaboration to select tools
  - Online reporting tool (OBIEE) for system (data marts in production)
  - Separate campus solutions warehouse delivered data model for Amherst (student)
Amherst Campus Project

- Aspirations
  - Turnkey system (plug and play)
  - Use delivered model for IR and general reporting
  - Integration with HR and Finance
  - IR plays active role

- What did we get?
  - Data quality issues
  - Product problems
    - Conceptual design
    - ETL
    - Data model
  - Does not meet IR needs
Delivered System Architecture

OBIEE Architecture

PeopleSoft Financials
PeopleSoft HCM
PeopleSoft Campus Solutions

PeopleSoft EPM
Operational Warehouse Staging (OWS)
Operational Warehouse Enriched (OWE)

Multi-Dimensional Warehouse (MDW)

OBIEE Administrator
Presentation Layer
Business Mapping & Modeling
Physical Layer

OBIEE Presentation Services
Oracle BI Interactive Dashboards
Oracle BI Answers
Oracle BI Delivers

End User Interfaces
Interactive dashboards
Ad hoc analytics
Embedded analytics
Evaluating the Delivered Model

- Requires considerable skills
  - Database
  - Data warehouse
  - Business Intelligence
  - System Administration

- Data and models not ready once the tool is installed

- Almost as complicated as starting from scratch
  - Must learn the tools involved
  - Data models need to be adapted to our usage
  - Parts of delivered system are not yet mature
  - Documentation is inadequate
Understanding the Delivered Model

- Studied ETL jobs
- Checked dimensions for correctness
- Checked fact tables for correctness
- Checked for missing data
- Checked error tables
- Learned repository and how to use the Admin Tool
- Identified errors in model in physical, business and presentation layers
- Identified custom development needs
Adapting the system to our needs

- **IT Perspective**
  - Small projects with immediate payoff managed by BI staff
  - Eventual buy in from functional areas

- **OIR Perspective**
  - Converting existing data into warehouse form
  - Building a census system in the warehouse
  - Adapt existing reports to BI tool
  - Distribute tool to users
  - Integrate HR and Finance
Delivered model was not designed for OIR use.
  • No census capabilities
  • Data Marts were narrowly defined
  • Data marts were independent of each other

How we are creating OIR model
  • Make sure the delivered model works
  • Examine delivered model for recycling opportunities
  • Creating the model
    • Converting old OIR data
    • Creating new OIR data model
What we found – the Good and the Bad

**Good**
- Report delivery tool is very good
  - Easy to learn
  - Very powerful
- Dashboard capacity
- Flexible for custom development
- Metadata repository
- Vendor is paying attention to our successes and failures

**Bad**
- Data quality checking was inadequate
- Data mart subject areas not integrated
- Design flaws in modeling and implementation
Data Quality

- PeopleSoft is big and complicated
- Our configuration differed from vendor assumptions
- Data integrity problems existed in transactional system which caused missing data in warehouse
- Hard to detect without specialized data quality checks
Staff Issues

- Limited staff resources
  - Only 3 new FTE for the whole project
  - DBAs needed to learn how configure DB for DW
  - Training was limited
  - Consulting was inconsistent

- Persuading users to try it out
  - Other demands on functional offices
  - No central mandate to use DW or BI tools
  - Existing tools are familiar even though inadequate
Performance

• Performance was a problem from beginning
  • Slow delivery of results from data warehouse
  • ETL took too long

• DW Performance
  • Delivered system had no optimizations for speed
  • Database configuration needed work

• ETL Performance
  • Couldn’t do overnight loads (17+ hours)
  • Server configurations
  • SQL optimization
Security

- Vendor had a delivered solution
- We created our own security strategy
- Used LDAP for authentication
- Used PeopleSoft tables for row level security
- Still questions for IR use
Delivered Tool Maintenance: Bundles

• Keeping up with bundles is hard
• New bundle every 3 months
• Hard to combine bundles with custom changes
• Bundles always require testing
OIR Development: Census Files

- After almost 1 year
  - We know the tool
  - Corrected data, ETL, modeling problems
- Ready to create census files
### Census: Student Enrollment

#### Dimension Census

<table>
<thead>
<tr>
<th>ID</th>
<th>Term</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1097</td>
<td>12-Sept-09</td>
</tr>
<tr>
<td>2</td>
<td>1097</td>
<td>15-Sept-09</td>
</tr>
</tbody>
</table>

- **Student Enrollment**
- **Class**
- **Class Meeting Pattern**
- **Term Enrollment**
Lessons Learned

- Using delivered model is as hard as building your own
- IR has unique needs
- Adjusting expectations to reality
- Getting people’s attention is hard in a decentralized environment
- Everything takes longer than you hope
Where do we go from here?

- Continued collaborations with system and campus
- Establish advisory board for future development
- Wider access to web based tools
- More flexible reporting
  - Dynamic
  - Dashboards with drill-downs, graphics...
  - Applying what if scenarios to reports
- Greater integration with HR and Finance
- Importing external data
Questions?
Contact Information

- Marilyn Blaustein blaustein@oirp.umass.edu
- Alan McArdle amcardle@oirp.umass.edu
- Banu Solak bsolak@oirp.umass.edu