Orientation to the *Data Warehouse*

IT Orientation – February 27, 2008
What is Data Administration?

- The Office of Data Administration manages data as a University resource so members of the Penn community have the information they need, when they need it, and in the form in which they need it to do their jobs. The objectives of Data Administration are to improve the quality, security, ease of access, and reduce the redundancy of the data. Specific functions of the Office of Data Administration focus on data architecture as a framework for decision making, the creation of a common language of standards, the arbitration and negotiation of agreements, and a commitment to education and advocacy.
ISC-AIT Data Administration
2/22/2008

Jim Johnson
Director, Data Administration

Amy Miller
Team Lead – Financial, HR, Research & Facilities Information and Data Warehouse Reporting Infrastructure

Ed Stemmler
Team Lead – Student Information

Lisa Campeau
Team Lead – Identity Management Information and System Operations

Bill Branen
Project Lead – Identity Management Initiatives

Lori Rotsjczak
Ann White
Susan Collins
Mike Budischak

Vicki Fullam
Deb Doll

Deb Sklawer
David Gold

***Rashma reports to Lisa for Penn Community conflict resolution activities and to Jim for security forms processing activities.***
What is a Data Warehouse?

“A data warehouse is a copy of transaction data specifically structured for query and analysis.”

Ralph Kimball
The Data Warehouse Toolkit

“The form of the stored data has nothing to do with whether something is a data warehouse…

Data warehousing is not necessarily for the needs of ‘decision makers’ or used in the process of decision making.”

Larry Greenfield
Dwinfocenter.com
Penn's Data Warehouse

- Implemented in 1995 after an in depth pilot study, the University's Data Warehouse makes Penn's institutional data available to decision makers for query, analysis, and reporting. The Warehouse stores selected data from Penn's existing and new business systems, such as SRS (Students Record System) and BEN Financials. The data is filtered, made consistent, and aggregated in various ways. Restricted to authorized individuals of the Penn community, Warehouse access does not require formal programming skills, and desktop tools such as Business Objects support analysis and reporting.

- Technical Information:
  - The University's Data Warehouse is an Oracle 10.2 (10g) relational database management system (RDBMS). The database resides on an four processor IBM p630 Symmetric Multiprocessor (SMP) server with 8 GB of memory and over 500 GB disk space, running under AIX 5.2.
# The Data in the Warehouse

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Collection</th>
<th>Source System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alumni &amp; Giving</td>
<td>Advancement/Gifts</td>
<td>ATLAS (PeopleSoft CR)</td>
</tr>
<tr>
<td>Facilities</td>
<td>Space@Penn</td>
<td>FacilityFocus</td>
</tr>
<tr>
<td>Faculty</td>
<td>Faculty</td>
<td>Faculty Information System (FIS)</td>
</tr>
<tr>
<td>Financial</td>
<td>Assets</td>
<td>BEN Assets</td>
</tr>
<tr>
<td></td>
<td>General Ledger</td>
<td>BEN Financials</td>
</tr>
<tr>
<td>Personnel/Payroll</td>
<td>Salary Management</td>
<td>Payroll &amp; Salary Management</td>
</tr>
<tr>
<td>Research</td>
<td>PennERA Proposals</td>
<td>PennERA</td>
</tr>
<tr>
<td>Student/Teaching</td>
<td>BRS</td>
<td>Mainframe Billing &amp; Receivables System</td>
</tr>
<tr>
<td></td>
<td>DTA (Departmental Teaching Analysis)</td>
<td>Warehouse Student, GL &amp; Salary Management</td>
</tr>
<tr>
<td></td>
<td>Student; Worksheet</td>
<td>SRS</td>
</tr>
<tr>
<td></td>
<td>Tuition Distribution</td>
<td>Warehouse Student, GL &amp; BRS data</td>
</tr>
</tbody>
</table>
How it’s used – the Warehouse out in the field

Users run hundreds of queries a day, with a wide range of goals:

- Ad hoc query and analysis – longitudinal studies (employees, courses, finances)
- Day-to-day management by school/center administrative personnel
- Principal Investigators monitoring their grants and protocols
- Source for other applications (PennPortal/U@Penn, PennCommunity, CoursesInTouch)

….. and many more!
Generally speaking, the owner of the source system of a collection is considered its **Data Steward**, and provides guidance on access controls, permissible usage of data, and training requirements for the collection.

**Security** is administered by DA and SEO; access is granted by database roles, and can be applied down to the data element level (e.g., ORG code).

**Training** varies by collection, and may be driven by source system access and training, school/center business process, and end-user need.
Salary Management Loads Schedule:

- Payments and Encumbrances for Weekly runs are loaded to Salary Management and the Data Warehouse Monday nights; loaded from Salary Management to BEN Financials on the following Thursday night.
- Payments and Encumbrances for Monthly runs are loaded Monday night of the third week of the month to Salary Management and the Data Warehouse; loaded from Salary Management to BEN Financials on the second-to-last working day of the month.
Student Data Flows

- SRS: Student Records System (student bio/demo, matriculation, degree, registration info, course inventory and term data maintenance)
- BRS: Billing and Receivables System (tuition, fees & charges, payments & credits)
- REGBUILD: Requirement Builder (degree requirement definitions and maintenance)

Mainframe Applications
- Penn InTouch, Advisor InTouch
- Penn Portal

Nightly Extracts from SRS/BRS
- Term Snapshots used on data changes from SRS to DW
- nightly job replaces online record each time

Online Directory
- multiple records per person (ex: "preferred email")
- student info is restricted by default; student can "opt in"

PennId and email added to extract

Data Warehouse Student Collection
- student data refreshed nightly
- student term snapshots
- billing and receivables
- academic progress worksheet
- course evaluation data
- student/course/faculty-related collections

Direct queries
- Web applications: batch, and online query

Students receive data from the warehouse
- Access, TOAD, etc.
- BlackBoard
- WebCafe
- Courses InTouch
- IROQI6
- WLRG

PennKey/PennId authentication

PennCommunity
- assigns PennIds
- one member record per PennId; multiple affiliation records (STU, TEMP, etc.)
<table>
<thead>
<tr>
<th><strong>PennERA Proposals</strong></th>
<th>the PennERA Proposals data collection: requested and awarded funding for sponsored projects (e.g., grants and contracts); <em>used in ad hoc and Web-based, parameter-driven reporting</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protocol Status Report</strong></td>
<td>helps researchers &amp; their staff ensure that the protocols (research methods) used in their sponsored projects are board approved</td>
</tr>
<tr>
<td><strong>GRAM</strong></td>
<td>the Grant Reporting and Management System: enables researchers &amp; their staff to track spending for their sponsored project award funds</td>
</tr>
<tr>
<td><strong>AIS On-line</strong></td>
<td>the Account Information Sheet: the GL account, the associated sponsored project, the responsible investigator, the amount available by budget period, etc.</td>
</tr>
<tr>
<td><strong>PennERA</strong></td>
<td></td>
</tr>
</tbody>
</table>
| **ben reports** | includes some Web-based, parameter-driven reports accessing Research data in the Warehouse
the BRIM data collection: data from the Billing and Receivables Information Management system, which tracks invoices, automatic payments, and cash receipts for funds for sponsored projects whose sponsors do not send award payments via a Letter of Credit. Available for ad hoc reporting by Office of Research Services staff only.

Coming soon—the Learning Management data collection: data concerning training for members of the Penn community. Covers courses that are required (based on one’s academic activities and job responsibilities), and courses that are optional.
Atlas Data Warehouse Collection

Nightly refresh
- Bio/Dena/Prospect Data full copy
- Gifts Data incremental load weekdays, full load weekly.
- 325 application tables
- 175 reporting tables and views

Atlas Reports/infowiew only users:
- About 400 users
- Link directly from Atlas Portal

Business Obj/Dev/Dev Users:
- Advance Universe
- Gifts Universe in development
- About 20 Users now.

DAR Report Developers:
- BO Crystal Reports
- Write reports for Atlas users
- Posting reports in infowel/itd
  hrc emailed directly
- About 12 Users

External Applications:
- Custom views
- Regular loads to external apps.

Student Collection
Atlas Collection
General Ledger

Student/Alums
Gifts

Alumni/Friends
Prospect Management
Giving History
What is Required for Access?

- Permission (formal authorization to use the data)
- A query tool (software)
- Training (in both the use of the tool and data)
- Password
  - Use the password changer to reset passwords:
    http://www.upenn.edu/computing/da/orapass.html

Access request forms are located at:
http://www.upenn.edu/computing/da/dw/forms.html
Tools of the (Warehouse) Trade

- **Business Objects**
  - *Desktop Intelligence (aka Deski, XIR2)* client/server tool for ad hoc & canned queries
  - *InfoView* for refreshing canned reports over the web

- **BEN Reports**
  - Canned web query (Financial, Payroll, Sponsored Projects, GRAM, Learning Management)

- **SQL*Plus** (primarily developers)

- **TOAD** (primarily ISC and some school/ctr developers)

- **MS Office via ODBC links** (school/ctr end-users and developers)

- **Local school/center applications**
Support email addresses:
- da-staff@isc.upenn.edu: Data Warehouse questions; Business Objects issues
- oraclepass@pobox.upenn.edu: Oracle password change issues
- Data collection listservs (FINQUERY-WH, STUDENT-WH, etc.)

On the web:
- Warehouse web site: http://www.upenn.edu/computing/da/dw
- Load status page: https://medley.isc-seo.upenn.edu/warehouse/jsp/fast.do
- Password changer application: http://www.upenn.edu/computing/da/orapass.html
- U@Penn Business tab (Business Administration)

In person:
3401 Walnut Street, Suite 265C
Philadelphia, PA 19104-6228
Thanks for your attention!

Any questions?