Orientation to the *Data Warehouse*

IT Orientation – June 24, 2009
What is Data Administration?

- The Office of Data Administration manages data as a valuable University asset 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 continuously improve the availability, quality, security, and ease of access for authorized users.
The term Data Warehouse was coined by Bill Inmon in 1990, which he defined in the following way: "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process".
What is a Data Warehouse?

- **Subject Oriented:**
  - Data that gives information about a particular subject instead of about a company's ongoing operations. Specifically structured for query and analysis.

- **Integrated:**
  - Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole.

- **Time-variant:**
  - All data in the data warehouse is identified with a particular time period.

- **Non-volatile**
  - Data is stable in a data warehouse. More data is added but data is never removed. This enables management to gain a consistent picture of the business.
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 business systems, such as SRS (Students Record System) and BEN Financials. The data is filtered, made consistent, and aggregated in various ways. Use of the Warehouse is restricted to authorized individuals of the Penn community. Access to the Warehouse does not require formal programming skills; desktop tools such as Business Objects support analysis and reporting.

- Technical Information:
  - The University's Data Warehouse is an Oracle (10g) relational database management system (RDBMS). The database resides on a four processor IBM p630 Symmetric Multiprocessor (SMP) server with 8 GB of memory and over 500 GB disk space, running under AIX operating system.
## 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>FacilityCenter</td>
</tr>
<tr>
<td></td>
<td>coming soon: Work Requests</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>Express Mail</td>
<td>Express Mail application</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 coming soon: Penn Works</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>Course Eval</td>
<td>Penn Course Review</td>
</tr>
<tr>
<td></td>
<td>Student; Worksheet</td>
<td>SRS</td>
</tr>
<tr>
<td></td>
<td>Tuition Distribution; DTA</td>
<td>Warehouse Student, HR, GL &amp; BRS data</td>
</tr>
<tr>
<td></td>
<td>coming soon: Graduate Admissions</td>
<td>Apply Yourself</td>
</tr>
<tr>
<td>Training</td>
<td>Learning Management</td>
<td>Knowledge Link &amp; the Penn Profiler</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 (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.

© 2005
the PennERA Proposals data collection: requested and awarded funding for sponsored projects (e.g., grants and contracts); used in ad hoc and Web-based, parameter-driven reporting

the Sponsored Projects data collection: data per the legacy system (the Research Services System, or RSS, used through Sept., 2003); also available for ad hoc reporting

BRIM

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.
<table>
<thead>
<tr>
<th>Protocol Status Report</th>
<th>helps researchers &amp; their staff ensure that the protocols (research methods) used in their sponsored projects are board approved.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIS On-line</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>PD Pipeline Report</td>
<td>these two reports access live PennERA data on proposals routed for internal review, to help users identify the most urgent proposals in order to prioritize handling</td>
</tr>
<tr>
<td>PD Grants.gov Report</td>
<td>includes some Web-based, parameter-driven reports accessing Research data (in the Warehouse, unless otherwise specified)</td>
</tr>
</tbody>
</table>
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. *Used in ad hoc and Web-based, parameter-driven reporting.*

includes some Web-based, parameter-driven reports accessing Learning Management data in the Warehouse

**Trainees**
- Current Required Course Status
  - by Supervisor
  - by Trainee
  - by Course
- Current Status of Required and Optional Courses, for the Org./Business Unit

**Courses**
- Class Roster--Required and Optional Courses
- Count by Course--Required and Optional Courses
IT Orientation June 24, 2009

Atlas Data Warehouse Collection

- nightly refresh
  - Bio/Death/Prospect Data full copy
  - Gifts Data incremental load weekly, full load weekly,
  - 395 application tables
  - 175 reporting tables and views

Atlas Reports/Info\view only users.
- About 400 users
- Link directly from Atlas Portal

Business Office/Desks Users
- Advance Universe
- Gifts Universe in development
- About 20 users now*

DAR Report Developers
- 80 Crystal Reports
- Write reports for Atlas users.
- Posting reports in Info\view data
  - hoc emailed directly
  - About 12 Users

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

Student Collection
Atlas Collection
General Ledger

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?