Orientation to the *Data Warehouse*

IT Orientation – April 3, 2013
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.
What is a Data Warehouse?

- 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 indepth 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 members of the Penn community. Access to the Warehouse does not require formal programming skills; tools such as Business Objects support analysis and reporting.

- Technical Information:
  - The University's Data Warehouse is an Oracle 11.2.0.2 (11g) relational database management system (RDBMS). The database resides on a eight processor IBM P6 Series Symmetric Multiprocessor (SMP) server with 64 GB of memory and over 1.25 TB disk space, running under AIX 5.3.
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>Work Requests</td>
<td>AiM (formerly 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>International</td>
<td>Global Activities Registry (GAR)</td>
<td>Global Activities Registry application</td>
</tr>
<tr>
<td></td>
<td>Int’l Student and Scholar Services</td>
<td>iOffice</td>
</tr>
<tr>
<td>Personnel/Payroll</td>
<td>Salary Management</td>
<td>PennWorks, 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>SRS, Mainframe Billing &amp; Receivables System (BRS)</td>
</tr>
<tr>
<td></td>
<td>Course Eval</td>
<td>SRS, Penn Course Evaluation</td>
</tr>
<tr>
<td></td>
<td>Student; Worksheet</td>
<td>SRS; REQBUILD, AdvisorInTouch</td>
</tr>
<tr>
<td></td>
<td>Tuition Distribution; DTA</td>
<td>SRS, Warehouse Student, HR, GL &amp; BRS data</td>
</tr>
<tr>
<td></td>
<td>Graduate &amp; Undergraduate Admissions</td>
<td>SRS, Apply Yourself, OASIS</td>
</tr>
<tr>
<td></td>
<td>Student Financial Aid</td>
<td>Mainframe Student Aid Management (SAM)</td>
</tr>
<tr>
<td></td>
<td>Career Tracker</td>
<td>School of Arts &amp; Science Career Tracker system</td>
</tr>
<tr>
<td>Training</td>
<td>Learning Management</td>
<td>Knowledge Link</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.
Student Data Flows

Mainframe Applications
- **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)
- **REQBUILD**: Requirement Builder (degree requirement definitions and maintenance)

**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

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

**PennCommunity**
- assigns PennIDs
- one member record per PennID; multiple affiliations (STU, TEMP, etc.)

**PennKey** -> PennID authentication

**Batch feeds**: Housing & Dining, Admissions, Schools, Faculty/Staff payments, OIP, Student Health, Loan System...

**SRS** to/from PennCommunity

**email**

**Bulkloads**

**Direct queries**

**Web applications**

**Student Data from the warehouse goes to users via various systems and query tools**

**Student Data Flows**

**Mainframe and web systems users**

**InTouch** to/from **SRS**

**Directory** to/from **SRS**

**OCR**

**PennPortal**

**Penn InTouch, Advisor InTouch**

**Data Extracts**

**Nightly Extracts**
- from SRS/BRS
- based on data changes
- (PLOG) to DW.
- The student nightly job replaces entire record each time.

**Other extracts**

**Term Snapshots**
- appended to existing snapshot data from previous terms.

**Course Eval**

**ORACLE account holders using**
- Business Objects and InfoView
- Access, TOAD, other ODBC tools

**BlackBoard**

**Courses InTouch**

**IRQDB**

**WebCafe**

**PennCard**

**PennKey** -> PennID authentication
the PennERA Proposals data collection: current data on requested and awarded funding for sponsored projects (e.g., grants and contracts). Used in ad hoc and Web-based, parameter-driven reporting. This data collection also includes annual snapshots of the tables going back to Fiscal Year 2004. The snapshots are used in ad hoc reporting only.

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. Used in ad hoc reporting.
includes some Web-based, parameter-driven reports that present Research data (from the Warehouse, *unless otherwise specified*)

| **PD Pipeline Report** | 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 |
| **PD Grants.gov Report** |

| **Protocol Status Report** | helps researchers & their staff ensure that the protocols (research methods) used in their sponsored projects are board approved |
| **AIS On-line** | the Account Information Sheet: the GL account, the associated sponsored project, the responsible investigator, the amount available by budget period, etc. |
| **GRAM** | the Grant Reporting and Management System: enables researchers & their staff to track spending for their sponsored project award funds |
Data concerning training for members of the Penn Community. Covers trainees, courses, assignments, registrations, and completions.

the Knowledge Link LMS data collection: data from the current Knowledge Link platform (SuccessFactors, in use since October 9, 2012).

Used in ad hoc and Web-based parameter-driven reporting (via Business Objects – InfoView).

the legacy Learning Management data collection: data from the legacy Knowledge Link platform (Gen21, used from March, 2005 through October 4, 2012) Formerly used in ad hoc and Web-based, parameter-driven reporting (via BEN Reports and via Business Objects – InfoView). As part of the move to the current Knowledge Link platform, the legacy data collection is no longer being updated.
Atlas Data Warehouse Collection

Advancement

April 3, 2013

Nightly refresh
- 325 application tables
- 175 reporting tables and views

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

Business Object/WEBI Users:
- Advance Universe
- About 20 Users now*

DAR Report Developers:
- BO Crystal Reports
- Write reports for Atlas users.
- Posting reports in InfoView/Ad Hoc emailed directly
- About 12 Users

QlikView:
- 6 developers
- 100 users

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

Student Collection
Atlas Collection
- Alumni/Friends
- Prospect Management
- Giving History

General Ledger

Students/Alums
Gifts
What is Required for Access?

- Permission (formal authorization to use the data)
  - Electronic access request forms are located at: http://www.upenn.edu/computing/da/eforms/index.html

- A query tool (software)

- Training (in both the use of the tool and data)

- Password
  - Users set their own passwords using password changer: http://www.upenn.edu/computing/da/orapass.html
Tools of the (Warehouse) Trade

- **Business Objects**
  - *Web Intelligence (aka Webi, XI3)* browser-based tool for ad hoc & canned queries
  - *InfoView* for refreshing and manipulating reports
- **BEN Reports (available via U@Penn)**
  - Canned web query (Facilities, Financial, Payroll, Sponsored Projects, GRAM)
- **QlikView**
  - Data visualization and exploration
- **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**
Warehouse Information & Communications

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

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

Any questions?