Orientation to....

the *Data Warehouse*

IT Orientation – February 8, 2017
ISC Enterprise Information & Analytics

- **What we do:**
  - Query best practices and tool support, enabling and facilitating access to Penn’s administrative information assets (residing in the Data Warehouse and other sources).
  - Foster analytical capabilities to support data discovery culture, and strategic use of information to further School, Center, and institutional goals.

- **Data Warehouse:**
  - Penn’s core administrative and institutional data available to decision makers for query, analysis, and reporting
  - Data collections from numerous business systems, including BEN Financials, PennERA, Pennant/SRS, PennWorks/Payroll, Space@Penn, and many others

- **Centrally supported tools:**
  - SAP BusinessObjects (currently includes Web Intelligence and Crystal Reports)
  - QlikView (numerous applications, including Advancement, P2P, Federal 990, Global Reporting)
  - Splunk (supports Security Logging Service and other log analysis)
Organization Chart - AIS
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".

February 8, 2016
IT Orientation  February 8, 2016
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

- 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 BEN Financials, PennWorks/Payroll, and Pennant/SRS.
  - 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 12.1.0.2.5 (12c) relational database management system (RDBMS). The database is over 2 TB in size and resides on a Dell PowerEdge R530 server with 64 GB of memory and 32 CPUs, running under Red Hat Linux 6.7 (64-bit).
The Data in the Warehouse

Alumni / Giving
- Advancement

Faculty
- Appointments
- Recruitment
- Grad Groups

Facilities
- FRES Work Requests
- Space@Penn

Financial
- General Ledger
- Assets
- BBR
- Concur Expense
- Express Mail
- ISC Billing
- Planning

International
- Global Activities
- Global Travel
- Immigration/Visa

Personnel / Payroll
- Salary Mgt
- Faculty Comp
- Position Inventory
- PAR
- Change Log

Research
- Proposals
- Human Subjects
- Protocols

Student / Teaching
- Student bio, enrollment
- Financial Aid
- Admissions
- Billing
- Financial Aid
- Grad Funding
- Worksheet
- Tuition Distribution
- Course Evaluation
- Career Tracker

Source Systems:
- Alumni / Giving: ATLAS
- Faculty: AIM (FRES work orders)
  - Space@Penn
- Facilities: FIS (Faculty Information System)
- Faculty Search (People Admin)
- International: GAR (Global Activities Registry)
- Global Travel (travel reservation)
  - iOffice (student/scholar immigration)
- Financial: BEN (Oracle Fixed Assets & GL)
- BBR (grants billing / receivables)
- Concur Expense
- UPS Express Shipping
- ISC Billing (mainframe, network, telecom)
- Planning (Hyperion RevExp & Salary)
- Identity Management: Penn Community (Penn ID)
- Online Directory (email)
- Penn People (race/ethnicity)
- Personnel / Payroll: PennWorks (bio, roles, faculty comp)
- Payroll (mainframe)
- Position Inventory
- Research: PennERA (proposals, protocols)
- Training: Knowledge Link (learning mgt)
- Student / Teaching: Pennant (student billing; grad funding)
- SRS (mainframe registration)
- SAM (mainframe student aid mgt)
- Slate & OASIS (undergrad admissions)
- Apply Yourself (grad admissions)
- Penn Course Evaluation
- AdvisorInTouch
- REQBUILD (worksheet requirements)
- SAS Career Tracker

Extract, Transformation & Load processing – daily, periodically, or other process-based
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!
Stewardship, Security & Training

- 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 ISC; 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.
the PennERA Proposals data collection: current data on requested and awarded funding for sponsored projects (e.g., grants and contracts). Annual snapshot data goes back to Fiscal Year 2004.

the BEN Billing and Receivables (BBR) data collection: current data for awarded sponsored projects and their 5- funds, from MAY-14 onward. Includes the budgeted and billable amounts, revenue, invoices, and applied receipts, as well as details on dunning and information on financial reports for sponsors. For transactions, monthly snapshot data is also available. Successor to the BRIM data collection.

the Human Subjects Protocols data collection: information on research protocols monitored by an Institutional Review Board (IRB) because they involve human beings, their tissue, or their private data. Includes data about Clinical and Translational Research Center (CTRC) services. Access request eForm available upon request.
Online Reports (formerly BEN Reports)

Research includes parameter-driven reports that present Research data (from the Warehouse, *unless otherwise specified*):

**PD Pipeline Report**

accesses *live PennERA* data on proposals routed for internal review, to help users identify the most urgent proposals in order to prioritize handling.

**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).

*Successor to the legacy Learning Management data collection.*

**the legacy Learning Management data collection:** data from the legacy Knowledge Link platform (Gen21, used from March, 2005 through October 4, 2012).

*As part of the move to the current Knowledge Link platform, the legacy data collection is no longer being updated.*
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](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](http://www.upenn.edu/computing/da/orapass.html)
Tools of the (Warehouse) Trade

- **BusinessObjects**
  - *Web Intelligence (aka Webi, BI4)* browser-based tool for ad hoc & canned queries
  - *BI Launch Pad* for refreshing and manipulating reports

- **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:** [http://www.upenn.edu/computing/da/dw](http://www.upenn.edu/computing/da/dw)
  - FAQs & HowTos: [http://www.upenn.edu/computing/da/bo/webi/faq_tips.html](http://www.upenn.edu/computing/da/bo/webi/faq_tips.html)
  - 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?