Data Governance Initiative

Tyler Crawford
Information Technology, Data Modeler
Why Data Governance?

We are here:
- Um ...
- 30,440
- Fall only?
- 28,891
- Enrolled or registered?
- 30K-ish
- Define "student" ...

We want to be here:
- How many students?

Data Governance is a process that requires participation and agreement from units across campus. We need your help.

- Interactive visuals
- Clean data
- Standardized definitions
PRINCIPLES OF DATA GOVERNANCE

- Auburn is the owner of all institutional data and will manage this data as a strategic asset.
- There is a common vocabulary and definition to describe each data element and data relationship. All data describing data (metadata) is stored and maintained in a shared repository.
- There is a single system of record for each University data element and unnecessary duplication of data across multiple information systems will be avoided.
- Institutional data must be stored in a secure manner. Access is provided to internal authorized users who have a legitimate need for the data based on their role in the University.
- Data is used only for the purposes for which use is authorized and authorization for access to data is not transferable.
- Decisions about data stewardship, access, quality, official system of record, and conflict resolution follow consistent processes.
- Management of data complies with legal requirements (FERPA, HIPAA, etc.), Auburn University internal policies, and privacy compliance.
STATUS OF DATA GOVERNANCE

Shared Responsibilities for the Data Life Cycle

**In Progress**
- **Acquire**
  - Create, capture, gather from activity:
    - Applicants
    - Students
    - Graduates
    - Alumni
    - Faculty
    - Staff
    - Billing
    - Vendors
    - Financial Aid
  - Responsible Parties:
    - Data Governance Council
    - Data Stewards
    - Data Custodians

**Next Steps**
- **Clean**
  - Organize
  - De-Duplicate
  - Annotate
  - Clean
  - Combine
  - Standardize
  - Correct
  - Responsible Parties:
    - Data Stewards
    - Data Custodians
- **Use and Reuse**
  - Analyze
  - Mine
  - Model
  - Derive
  - Visualize
  - Decide
  - Act
  - Responsible Parties:
    - Data Users
- **Publish**
  - Share within classification requirements:
    - Portals
    - Aggregations
    - Websites
    - Executive Reports
    - Visualizations
  - Responsible Parties:
    - Data Users
- **Preserve or Destroy**
  - Choose whether and how to preserve:
    - System of Record
    - Data Warehouse
    - Destroy
  - Responsible Parties:
    - Data Governance Council
    - Data Stewards
    - Data Custodians
University Data Environments
**Data Governance Structure**

**Data Governance Roles and Process**

- **Auburn University Data Governance Steering Committee**
  Provides oversight for the Data Governance Process.

- **Data Council Co-Chairs**
  Lead the Data Governance Council.

- **Data Stewards**
  Ensure that institutional data within their domains is identified, documented, and made available.

- **Data Modeler**
  In consultation with Data Stewards, leads the identification, documentation, and availability processes.

- **Data Custodians**
  Manage the technical implementation of data as defined by Data Stewards, and ensure data security and privacy.

- **Data Users**
  Follow the guidelines created by the Data Stewards using the systems provided by the Data Custodians.

Other university constituents, including faculty, staff, and students, generate data and use it incidentally or in a limited manner pertaining only to their duties, actions, and history with the university.
Central User Access Point
Proof of Concept

1. Data sources
   - Relational databases
   - File exports
   - Big data sources

2. Azure Data Factory
   - PaaS Hybrid ETL/ELT service
   - Can host SSIS packages

3. Azure Data Lake Store
   - Hyper-scale repository for big data analytic workloads
   - Ingest data of any size, type, velocity

4. Azure SQL Data Warehouse
   - Massively Parallel Processing (MPP) SQL PaaS
   - Partitions & dimensions
   - Petabyte scale
   - Pause compute resources

5. Azure Analysis Services
   - PaaS semantic model
   - Centralized calculations, hierarchies, KPIs, etc.
   - In-memory, compressed

6. Power BI

Auburn University POC Reference Architecture
METADATA

- Element Name
- Definition
- Sensitivity Classification
- Access Designation
- Stewardship Office
- Source
- Data Origin
- Associated Terms
- Availability
- Retention Requirements
- General Notes
Research Grant Proposal Pipeline

Pre-Award

Award

Post-Award
- Number of proposals in each stage by:
  - Faculty
  - Department
  - Research topic area
  - Research funding Source
- Track the success rates by various attributes
- Support reporting to funding agencies
  - Financial reporting
  - Programmatic reporting
  - Audit requests
QUESTIONS?