Data Architecture

Administrative and Enterprise Applications
Office of Data Administration
Data Architecture

Introduction......................................................................................................................... 3
Data Standards and Guidelines ........................................................................................... 4
  Data Standards ................................................................................................................ 4
    Data Element Naming Standards ................................................................................. 4
    Standard Abbreviations ............................................................................................ 4
    SCAT (Standard Coding Access Tables).................................................................... 4
  Database Services for Core Applications .................................................................. 5
  Organizational Hierarchy ......................................................................................... 5
  Metadata ....................................................................................................................... 5
  Acronyms ..................................................................................................................... 5
  Data Security ............................................................................................................... 6
  Backup and Recovery/Disaster Recovery .................................................................. 6
Data Guidelines............................................................................................................... 6
  Institutional Data Management Policy ....................................................................... 7
  Data Management Structure and Procedures ............................................................ 7
  Data Committees ....................................................................................................... 7
Data Integration .................................................................................................................. 8
  Person Identifiers ............................................................................................................ 8
  University Person System (UPS) .................................................................................. 9
  Directory ....................................................................................................................... 9
  Data Modeling .............................................................................................................. 10
  Data Warehouse .......................................................................................................... 10
Data Access....................................................................................................................... 11
  Source Application System Access ............................................................................. 11
  Data Warehouse Access ............................................................................................. 11
  Exported Data from Source Applications ................................................................ 11
Data Quality...................................................................................................................... 12
  Requesting, Changing or Making Units Obsolete ...................................................... 13
  Database Service ......................................................................................................... 13
  Business Process Analysis ......................................................................................... 14
Introduction

The Office of Data Administration (ODA) was established in 1996 as a result of a continuous quality improvement committee recommendation that recognized the need to provide a data management framework for the institution. ODA manages the institutional data of the University of Maryland to provide reliable, accurate, secure and accessible data to meet the strategic and management needs at all levels of the University. "Institutional data" includes data that is relevant to planning, managing, operating, and auditing administrative functions of an administrative or academic unit of the University.

The Office of Data Administration plays a pivotal role in the planning and development of data platforms on the UM campus. It is ODA's responsibility to develop an effective data architecture to provide:

- A framework that supports efficient, secure business processes and is an integral component of the overall IT infrastructure.
- Policies for the management of institutional data and the responsibilities for the protection of those data.
- Data stewardship to clearly identify who has responsibility for institutional data subsets and the management of them.
- Data models that provide a consistent enterprise-wide data view and decomposition of the institution’s business processes.
- Standards for building and defining data elements to provide consistency across data subsets and avoid redundancy.
- Integration across data subsets to provide a foundation for information sharing among business processes.
- Accessibility of data by the appropriate method, when needed, in an easily understood format.
- Security to allow access to appropriate personnel and guarantee protection of data.
- Data quality programs to ensure trust in the accuracy of the data.

ODA's mission is to help data planners integrate business systems when possible, utilize existing structures when possible, and enhance their data infrastructures so they are cohesive and part of the larger University data systems environment. Of key importance is the development of a business driven data infrastructure that supports decision-making and data analysis.

The data architecture consists of several layers:

- Data Standards and Guidelines – this is the foundation layer that provides the structure and rules for storing data.
- Data Integration - this layer manages integration of data across databases.
- Data Access – this layer manages the methodologies for delivering data to users.
- User – this layer represents the users of data.
Surrounding the architecture and part of every layer is data quality, which addresses accuracy, consistency, and completeness. Additionally there are processes that facilitate use of the data and user support.

**Data Standards and Guidelines**

**Data Standards**

An integral part of the data architecture is the development of an infrastructure that enables integration of data across systems through the establishment and use of standards:

**Data Element Naming Standards**

http://www.oit.umd.edu/dataadmin/Standards/standards.html

Data object names comprise one or more prime words, optional qualifier words and one class word and fully describe the object. These words are used when constructing a data model name, database name, filename, table name or data element name. A **prime** word describes the topic or subject area of the data. A **class** word describes the type of data, i.e. the major classification of data. A **qualifier** further defines a prime or class word. (http://www.oit.umd.edu/units/dataadmin/Standards/standards.html).

**Standard Abbreviations**

http://www.oit.umd.edu/cgi-bin/acrolist.cgi

The University of Maryland Data Object Abbreviation Standards list is the official list of abbreviations to be used for data object (data model, database, file, table, data element, etc.) naming by application developers and information deliverers. Standard abbreviations are requested by users, researched via reference sources, reviewed by the Data Technical Advisory Committee, then formally adopted and published on the ODA web site.

**SCAT (Standard Coding Access Tables)**

Standard coding access tables (SCAT) hold standard code sets and their descriptions to describe a data object. For example, the racial identity data object is defined for consistent reporting purposes and must contain the same values and descriptions across business applications (PHR, SIS, etc.). ODA works with Service Offices to ensure that values for commonly shared data object are identical. ODA is becoming increasingly involved in the maintenance of these tables, particularly in the case when multiple unique tables exist for internal and external reporting and must be mapped to meet these requirements.
**Database Services for Core Applications**

Database services are provided for all systems developed by the OIT-Administrative & Enterprise Applications (AEA) unit. These services include designing/developing databases and providing technical assistance to application designers in the use of database management systems (standards, performance, tuning, tools, security, backup, recovery, etc). Database administrators in AEA-Administrative Application Support Services provide limited consulting to decentralized campus IT organizations on the use of Oracle databases, server connectivity, and installation of SQL*Net client software.

**Organizational Hierarchy**

The University of Maryland maintains a hierarchical organizational structure – Office of the President, Divisions, Colleges & Schools, Departments and Sub-Departments. There are approximately 400 active organizational units that comprise this university. They are a combination of academic departments, programs, centers & institutes, administrative offices and service offices. For institutional reporting and budgeting, many of these entities have codes that represent them in official campus databases. The Financial Records System (FRS), Payroll/Human Resources System (PHR), and Academic Resource System (ARS) all use the same underlying code structure provided by the FRS system to maintain a common hierarchy. Numerical logic had been built into the coding scheme to allow for consistency of codes, maintenance of the levels of the hierarchy, and to provide for institutional reporting rollups.

**Metadata**

Metadata are “data about the data”. To facilitate turning data into information, it is crucial to understand the meaning of a data object, how it is used, and the nuances surrounding its use. The Office of Data Administration has created an application that serves as the repository or encyclopedia for the metadata. Metadata exists for each institutional data element that currently resides in the University Data Warehouse as well as other critical data elements. For each data element, the following information is tracked via the metadata application – data element name, data definition, supplementary data definition, source system data, data warehouse tables, and keyword classification. Data definitions are available via search on the ODA website, [http://www.oit.umd.edu/dataadmin/DataDef](http://www.oit.umd.edu/dataadmin/DataDef), and to data warehouse users via the supported query tool.

**Acronyms**

Many entities and frequently used terms have acronyms to represent them. The Office of Data Administration maintains an acronym list of frequently used campus acronyms on its web site.

**Data Security**

As described in the Institutional Data Management Policy, the Institution’s data assets will be safeguarded/protected. As an institutional asset, data will be protected from deliberate, unintentional or unauthorized alteration, destruction and/or inappropriate disclosure or use in accordance with established institutional policies and practices and federal and state laws.

Authorized University of Maryland, College Park, business managers, administrators, service providers, institutional researchers, faculty and students can access data as appropriate for the tasks they perform. Access and authorization is based on a person’s need for the data and granted through approval by the employee's unit head as well as the appropriate data steward/manager.

Granting of access imparts responsibility to the user to use data only for the purpose of authorized University business. Unauthorized activity, including use for personal monetary gain, jeopardizing legitimate use, providing resources to unauthorized persons, or conducting illegal activities, may result in removal of access, disciplinary action, dismissal, and/or prosecution under the scope of applicable laws. Additionally, all users of data have an obligation to protect the data they extract and/or use.

OIT’s security architecture provides additional levels of protection for institutional data. The architecture includes access control devices and intrusion detection sensors as well as vulnerability assessment and security awareness education.

**Backup and Recovery/Disaster Recovery**

Regular backups are performed on a routine basis for systems managed by the Office of Information Technology. Databases and source system code are backed up according to a schedule. Backups are performed so that in the event that a disaster occurs (hardware failure, natural disaster, etc.) databases and systems can be restored.

**Data Guidelines**

The University of Maryland recognizes that data is a University asset that must be managed. The University formally adopted Policies and Guidelines on February 7, 2003, that provide the underlying philosophy and framework regarding data management:
Institutional Data Management Policy

http://www.inform.umd.edu/PRES/policies/vi2200a.html

The Institutional Data Management Policy states that university data will be protected, accessible, managed, identified and defined. A data architecture will be developed and data quality will be managed. Through a strong data stewardship program, the entire campus community plays a role in compliance.

Data Management Structure and Procedures

http://www.inform.umd.edu/PRES/policies/vi2300a.html

The Data Management Structure and Procedures define the levels of responsibility for managing institutional data. The University data management structure is defined by five levels within the institution:

1. Data Owner - The University is the data owner for all institutional data.
2. Data Trustee - Vice President of a major functional division of the institution, having responsibility for the collection of data subsets within that division.
3. Data Steward – Senior University officials having overall responsibility for a specific subset of data; e.g. Director of Personnel Services.
4. Data Manager - University officials and members of their staff who have operational level responsibility for a specific subset of data.
5. Data User – University faculty, staff, or student who has access to a specific subset of data; the data user may use specific subsets of data in the performance of his/her job (e.g. Department Chair) or may have access to information about himself/herself.

Key responsibility lies with the Data Steward who has policy level responsibility for the data subsets he/she manages, ensures the data quality of those subsets, and identifies access and authorization. ODA maintains the data management structure for the campus.

Data Committees

Data Policy Advisory Committee

http://www.oit.umd.edu/dataadmin/DataCommittees/policy.html

The Data Policy Advisory Committee (DPAC) handles matters related to institutional data, information policy, and information security. More specifically DPAC:

- Advise, make recommendations, and develop policy on matters related to data use, data policy, data privacy and data security.
• Make recommendations regarding the development of the data warehouse, the University Directory, and other efforts to provide use and access to University databases.

• Consult broadly with the University community regarding the utility and effectiveness of University databases to meet their needs.

Data Technical Advisory Committee

http://www.oit.umd.edu/dataadmin/DataCommittees/techadv.html

The Data Technical Advisory Committee (DATAC) is responsible for identification, discussion, and recommendations regarding data issues, and for integration and implementation of data policy. Data issues encompass data architecture, data standards, integration, security, and access.

Data Integration

Person Identifiers

A critical aspect of the data architecture is how we represent a person’s identity in our information systems. Traditionally a person’s Social Security number (SSN) has been used as the primary identifier in University databases and for login authentication. Currently, records in source application systems for students, faculty, and staff, can be identified by SSN (or alternate ID), University Identification Number (U ID), or both, depending on the system. When SSN is used as an identifier it is referred to by different element names in our databases:

• Student ID (SID) in SIS (students); SAM (Student Aid Management in OFSA)  
  Note: this is an ID number only in SAM but is the same number as the SIS SID; real SSN is stored in a different field.

• Faculty ID (FID) in SIS (SIS scheduling)

• Person ID in SAR (Student Accounts Receivable) Note: this is an ID number only in SAR and is not the SIS SID; there is an SSN field in SAR and the SIS SID is stored there.

• SSN_Id in UPS (stored in UPS_PERS table and used by PHR and many other applications); for students the SIS SID is stored here.

• UM_Id in the Data Warehouse.

• umId in the University Directory

Due to the need to protect the privacy of personally identifiable information and to prevent identity theft, the University is developing a plan to limit use of SSN and promote the use of alternate identifiers. A unique nine-digit identification number, University Identification Number (U ID), is currently assigned to each individual
(faculty, staff, student or affiliate) associated with the University. This number is for identification purposes only and in the future may be used as the primary key for records associated with an individual. Some University systems have already converted to use of U ID as the primary key for individuals (e.g. PHR), while others will be phased in over time. Social Security numbers or other tax identifying numbers are not intended for identification purposes.

A separate Directory ID is also being utilized for the purpose of authentication and authorization in the University Directory and University applications. (see section on Directory).

**University Person System (UPS)**

The University Person System (UPS) is a data/technical architecture to store and manage basic person information for any person associated with the University of Maryland. Generally, basic person information is information about a person that is the same across all application systems – University Identification Number, names, birth date, addresses, email address, etc. Historically, each of the major application systems has stored this information separately causing data to become out of sync among systems. The UPS concept is to store basic person information in a singular database that is updated and accessed by each major application, thus avoiding separate storage in multiple databases.

The Office of Data Administration manages the synchronization of data that flows into UPS from major enterprise applications (PHR and SIS). PHR updates UPS tables real-time upon user entry of information via web-based screens. SIS does not update UPS real-time therefore batch programs are run to handle the synchronization. Note: UPS is under the Oracle database structure and SIS is under the Datacom structure. As major applications are replaced or upgraded, UPS will be updated real-time and redundant data will not be stored in each separate application. The long-term goal is to have UPS be the Person system for all applications.

**Directory**

The Enterprise Internet Services unit within the Office of Information Technology is working on a middleware initiative that provides a comprehensive framework for network based applications to do enterprise based authorization, authentication, and searches. The core piece of the middleware initiative is the University Directory. The Directory being implemented at the University of Maryland uses LDAP (Lightweight Directory Access Protocol) and is commonly referred to as the Directory. Working in partnership with the Enterprise Internet Services unit, ODA manages the scope of inclusion of data elements, data access requests for Directory attributes, and data quality. The Directory is not meant to replace core application systems or the University Data Warehouse. It provides the means for applications that need core person information to access data from a single source rather than having to go to multiple sources for the data.
The Directory contains certain information about the members of the university community (faculty, staff, students, and affiliates). Employees with current and future approved appointments in the PHR system are included. Students who have been admitted, or are currently enrolled are also included. Currently the directory contains College Park employees, students and affiliates only. The long-term goal is to add other institutions and entities that use College Park applications and services.

The following types of information can be found in the University Directory.

- Identifiers (e.g. Directory ID, U ID)
- Name
- Demographic information (e.g. gender, DOB)
- Contact information (e.g. email, phone, fax, cell, address)
- Person indicators – faculty, staff, student, alumni
- Confidentiality indicators
- Calendaring attributes
- Services attributes
- Termination Dates (student, employee, affiliate)
- Employee appointment (e.g. unit, title, organization, employment category)
- Student data (e.g. major, college, class standing [freshman, junior, senior], course student is registered for, number of credits per course, and instructor of course).

**Data Modeling**

An enterprise architecture comprises a set of pictorial representations (models) of the business (e.g. business functions, business data). Data models provide an enterprise-wide view of the data, help in understanding the organization, and avoid misusing or redundantly recreating processes or data. The Office of Data Administration maintains high-level business function models as well as detailed data warehouse business data models. The detailed data models depict the data by subject area and contain data elements associated with sub-processes within those areas.

**Data Warehouse**

The UM Data Warehouse is a collection of integrated institutional data for the purpose of performing analysis, producing ad hoc queries and reports, and maintaining data subsets. Data subsets have been extracted from various campus transactional databases, structured for query and analysis, and located on platforms that can be accessed using standard SQL query tools. The data warehouse consists of several different types of data: transactional, frozen (moment in time) and year-end. In order to achieve consistency of data, a core of defined data elements are identified for each data topic to be included in the warehouse, data standards and principles are applied to the data, derived data elements are developed, and data from transactional systems are loaded into a transactional data warehouse. Frozen data, for institutional reporting, are then taken as a snapshot (moment in time) of the transactional data and placed in a frozen data warehouse. Some data subsets, such as financial data, may be loaded with cumulative totals at the end of a fiscal year. The
Office of Data Administration manages the “current” data warehouse. The Office of Institutional Research and Planning manages the “frozen” data warehouse.

**Data Access**

Another critical aspect of the data architecture is having institutional data available as needed, when needed to facilitate business processes, manage data quality programs, perform analysis, explore trends, and perform institutional reporting. The Office of Data Administration facilitates directing requests from internal business customers to the appropriate office. Public Information Act requests for data are handled by the data steward for a subset and in consultation with the Legal Office, if appropriate.

**Source Application System Access**

For direct access to most source applications (e.g. PHR, SIS) for the purposes of creating transactions and updating records, or viewing data directly in the transaction system, the system data manager can be contacted. The OIT Helpdesk maintains a web page to assist users in obtaining more information about access to the major campus application systems ([http://www.helpdesk.umd.edu/topics/accounts/](http://www.helpdesk.umd.edu/topics/accounts/)). Appropriate approvals must be secured from the employee’s unit head and may also be required from the data steward for the system.

**Data Warehouse Access**

Client users access the “current” data warehouse, to build their own ad-hoc queries utilizing query tool software. The query tool allows them to access the data warehouse tables/views, select their query criteria, retrieve data and format reports & charts. Any SQL-based Oracle compatible query tool can be used with the UM data warehouse. The query tool supported by the campus is Brio Software’s Brio Intelligence Explorer that provides point-and-click access to the warehouse on Mac, PC and Unix platforms. Purchase of the Brio Intelligence Explorer client software is necessary.

Web users access the “current” data warehouse, and can utilize secured data and reports via the web for queries created by Brio client users and loaded to the Warehouse on the Web (WOW) server. Purchase of the Brio IntelligenceQuickview (for PC users) or Brio Intelligence Insight (for MAC users) web plugin software is necessary.

**Exported Data from Source Applications**

Due to new technologies and the desire to use information for analysis and planning there are ever increasing demands for exported data from source application systems:

- Central administrative applications need data from core application systems (e.g. the Directory).
• Departmental applications need data from core application systems (e.g. FM Work Control).
• Service Offices need data to manage their systems and ensure data quality.
• Departmental staff need data for inquiry, analysis and planning.
• The Office of Institutional Research and Planning needs data for institutional reporting.
• Individual faculty, staff, and students need data to confirm the information collected about them.

Timeliness of data is also a factor - static vs. real-time, daily vs. weekly, frozen vs. current. Data are initially generated in source applications systems and made accessible in those systems. Additionally, to enhance access for query purposes and to preserve performance in source application systems, data are exported and made available based on usage criteria.

**Data Quality**

Managing data quality is an important aspect of any organization’s data architecture. Data quality is the ability of data to meet business requirements. Components of data quality include accuracy, consistency, and completeness. Accuracy ensures that a data element is within its context (e.g. is the zip code correct for the address?). Consistency ensures that a data element is the same across databases. Completeness ensures that a data element is fully populated. Poor data quality can negatively impact the health of our institution through incorrect business transactions, the inability to strategically plan, and by jeopardizing customer relationships.

Achieving data quality involves initiating data quality programs, maintaining a commitment to managing quality and changing business processes, developing metrics to measure data quality, and changing the underlying business applications.

As stated in the Institutional Data Management Policy, data quality will be actively managed. Our data stewardship program imparts the leadership in this area, and also puts responsibility on users who input data into source application systems. Depending on the system, these users are departmental staff or service office staff.

Working in partnership with the Office of Institutional Research and Planning, a plan is being developed to identify specific data elements that are critical for institutional reporting and to develop processes for keeping data clean and in sync for those elements. Where appropriate, source system applications are revised to incorporate edit checks to ensure data quality. Maintenance queries that also provide edit checking are run on a regular basis and source system data corrected as appropriate.

**Processes**
Requesting, Changing or Making Units Obsolete

Departments that have changes to existing units or need to create new units initiate these transactions through ODA. The following processes are handled by ODA:

- ODA is the initial point of contact for the establishment and maintenance of new codes to represent units, name changes, and units to be made obsolete. Departments initiate the request using web forms that are located on the ODA web site. Note: the creation of new academic departments, centers and institutions must first receive the appropriate approvals through the Office of the Sr. VP & Provost for Academic Affairs.

- ODA works with the Office of the Sr. VP & Provost for Academic Affairs to ensure that the appropriate approvals have been received regarding the creation of new academic departments, centers and institutes.

- Notifications are sent to the Comptroller’s Office to establish, change or archive obsolete units represented in the Financial Records System (FRS).

- Notifications are sent to all service office personnel involved in database management to ensure that new units, changed names, and obsolete units are updated in the various application systems that are maintained throughout campus.

- ODA updates the PHR Unit table to set the foundation for the departmental creation of appointments.

- ODA creates and maintains the University Bridge that tracks and keeps history on all recognized organizational entities (some entities are not represented by codes) and the various campus codes that map to them – unit code, undergraduate major, graduate program, course prefix, obsolete FAS codes, and obsolete HRS codes. A listing of current organizational entities is available on the ODA web site.

Database Service

There are a number of other database system support activities provided by OIT-Technical Services and Support. These include:

- On a per course basis, instructor and student access of Oracle Relational Database Management System (RDBMS) for academic purposes.

- Installation and consultative support for Oracle server applications in Colleges and other units.

- Support for databases used by several specific applications, such as Web hosting services, the University of Maryland Certificate Authority, Optix, and OIT Financial Services.
Business Process Analysis

The Office of Data Administration provides business process analysis of processes that affect the core data infrastructure. Some examples of analyses include:

- Graduate Assistant Processing - Researched graduate assistant appointment and tracking process, recommended FTE change to match time worked, recommended changes to the methodology for issuing stipends for fellows, and facilitated single tracking of fellows. These recommended changes led to core changes in business practice.

- U ID-Identifier Synchronization – developing a plan for changing person identifiers from SSN to U ID. Identified issues associated with identifiers currently used.