Overview

In the mid 90s the vision that guided the work of Data Administration was to make our Institutional data accessible via a Warehouse, so that we could use it better and improve it's quality. Today much of that vision has been achieved, most information collected centrally is available in the Warehouse and is refreshed on a nightly basis. However, many in the MIT Community find that information is hard to understand and use. Now is the right time to recalibrate and set new goals in order to bring MIT's management and use of data to a higher level, that will support the expectations of a Digital MIT.

The rallying cry has been "Make the Data accessible"! Now it is time to set our sites at a more challenging and complex set of goals and look at how data is collected and managed as well.

Since using and managing information involves the entire MIT community, it is important that we agree on a set of Goals and Principles that will guide this work.

Vision

- Using information is natural and intuitive for everyone
- MIT spends more time making use of data than collecting, correcting and managing it
- All systems have automated reconciliation and integrity checking processes
- We understand and document, why we're collecting information, how it will be used, how it is going to be maintained, and who is responsible, before deciding to build any system
- All information to support important decisions is easily obtained
- Members of the MIT Community have a shared understanding of their roles and responsibilities concerning data
- All data is considered sharable unless there is a documented business policy or government regulation

Goals & Principles

- Data should be Complete, Accurate and Timely
- People can easily get the information needed to do their job
- People understand the information they need to use
- Information is easy to update
- Information is only entered once
- Information and Reporting requirements are understood as systems are designed
- All enterprise data should have a **Custodian** identified who is responsible for it's quality
Measures & Concrete Goals
The high level Goals identified above lead to more concrete goals and measures:

- All enterprise information is available in the Warehouse
- Data definitions exist for all information in the Warehouse
- Data Models exist for each Business area
- Models are produced as part of development process
- There are documented processes for resolving data issues
- Flows of information are streamlined
  - Information is collected at the source
  - Number of steps in propagation are reduced
  - Reconciliation and integrity checks of data are automatic

Major Work Areas
Achieving these goals will require work in three main areas:

1. Improving our understanding and use of data
2. Improving our Development Practices
3. Establishing Roles, Responsibilities and Processes concerning Data

Improving the understanding and use of data

**Current State**
- MIT is not always making use of the data it has
- The MIT Community is not fully aware of the data resources available
- Data Definitions are hard to find, and not always accurate or up to date
- Lack of understanding of the structure of information
- Lack of understanding of the flows of information

**End State**
- The community knows how to find the information they need
- The community knows how to get data issues resolved
- There are data definitions available to the MIT Community for all centrally managed data

**Actions**
- Teach people how to access, understand and use information
- Improve the existing data definitions
- Create documentation about areas that are hardest to understand
- Publish models of data
  - Enterprise view
  - Department view
  - Subject views and flows
Improving our Development Practices

There are two primary areas where Data Administration can impact our development. The first is gathering requirements and producing a conceptual model of the domain, so that systems being built are well integrated with existing data. The second is in improving practices of designing and maintaining the Database itself.

**Current State**
- There are no standard Data Analysis techniques used in Projects
- There is no standard process for identifying where data might already exist
- Lack of complete understanding of who enters what data, and who uses what data
- System and Data Modeling is often limited to the central portion of a business function
- Data integrity is often unenforced in transactional systems  
  - leading to inaccurate information  
  - The system is not designed with reuse of information in mind
- Existing data structures are poorly documented  
  - New structures are created rather than using existing ones due to lack of understanding
- Reporting needs are often done in the transactional system  
  - Adding extra cost to development
- Database Administration (really DBMS Administration) is largely focused on managing space, tuning and performing upgrades, rather than reviewing the structure and naming of proposed changes.  
  - Field names aren't consistent leading to extra developer effort

**End State**
- Conceptual Data Models and Business Process Models are produced and reviewed before implementation
- All Data Definitions are published as new systems are rolled out
- All developers can easily discover what data exists and what the System of Record is
- Reporting functionality is always included in new Projects
- Reporting is done using standard reporting tools against the Warehouse and embedded as needed in applications

**Requirements and Design Activities**
Work with development teams on the following activities;
- Define a scorecard for assessing Data Management practices
- Construct a conceptual data model prior to implementation
- Developing Business Process models
- Identify reporting requirements, and determine how they will be met
- Identify any existing or related data, and define integration keys

**Database Design**
Work with database and development teams on the following activities
- Draft database guidelines for creating, altering and documenting schemas;
Establishing Roles, Responsibilities and Processes concerning Data

This work requires collaboration amongst the wider community.

Current State

• There are no established guidelines for Data Management Practices
• There is not a consensus on how data should be managed and accessed
• Data is often reconciled manually

Actions

• Get broad agreement and publish the goals of Data Administration
• Collectively establish goals and guidelines for Data Management Practices
  o What are appropriate uses of data
  o Data access policies
  o Policies around transfer of data to a third party
• Establish a mechanism for reporting and prioritizing data issues
• Establish a set of measures to assess data practices
• Publish and socialize data guidelines
• Publish the Roles and Responsibilities for
  o Data Custodians
  o Data Consumers
  o Projects Teams
• Establish a Role for Data Administration in each area (HR, Facilities etc)