

## Data Quality Management Overview

SFA Data Quality Management provides the framework, guidance and support for managing data as an enterprise asset. Its purpose is to assess and recommend techniques and technologies that promote sound data management practices. The objective of the Data Quality Management function is to ensure data is shareable across the enterprise and that over time, data quality is improved and data is minimally redundant. The vision for data quality at SFA is:

*“Knowledge, acquired from multiple sources, about the information assets of SFA is integrated and made available to both the business and technical communities via a common enterprise capability. Data and its underlying descriptive knowledge, once acquired, is accurate, timely and easily reused.”*

The function seeks to ensure the value SFA receives from its data assets are maximized by increasing the quality of data and the degree of data sharing. Increased data quality is essential to modernizing and improving the systems at SFA. Data quality is equally essential to plan, operate, optimize and monitor the many activities undertaken by the channels.

When properly applied, the Data Management standards and procedures preset in this document will resolve many of the problems created by the traditional environment and derive the following benefits:

### *Benefits of Data Quality Management*

- Information technology will support improved operational methods within the department.
- The information delivery environment will have a greater level of stability.
- Long-term costs to maintain legacy data will be reduced.
- Business methods through improved Data Management will be achieved.
- The quality of data documentation will be enhanced through data standards and procedures.
- Data usage consistency and efficiency will be obtained.
- Improved flexibility and modularity of data structures will result.
- Reduced and controlled data redundancy will be achieved.
- Increased compatibility of data among data structures will result.

## I. Data Quality Management Objectives

The objectives of the Data Quality Management function are:

### *Economic Commitment*

To make a maximum contribution to the competitive survival, profitability and growth of SFA through the effective management and use of SFA data resources.

### *Service Commitment*

To provide SFA with the planning, policy, guidance and products to ensure that data quality is achieved.

### *Functional Commitment*

To provide advice, products, services and support such that:

- Data quality is maximized,
- Enterprise data resources are integrated and shared in support of enterprise business objectives,
- Proper data planning, analysis and design techniques are utilized as an integral part of automating the enterprises business,
- Data and related descriptive information (meta-data) is optimally reused.

### *Client Commitment*

All business channels and their technical organizational counterparts.

### *Scope Commitment*

All data and related descriptive information (meta-data) whether implemented through manual or automated means.

## II Products and Services

The current/future products and services provided by the Data Quality Management function of the OCIO It Management department are as follows:

- ✓ *Data Quality Management Policies, Procedures and Standards.* This product includes the policies, procedures and standards that are intended to facilitate meeting the objectives stated above.
- ✓ *Metadata Repository.* This product provides a standard, automated mechanism to capture, organize and integrate descriptive information about SFA in a central location and to make this available to SFA personnel at the desktop.
- ✓ *Data Standardization Products and Services.* Products includes the conceptual enterprise data model (CEDM) and mappings of legacy data to the CEDM. Services include assisting the channels to identify data stewards, common security levels, and authoritative sources for data.
- ✓ *Legacy/Operational System Data Dictionaries.* This service includes the conversion of legacy system data dictionary/data models into a standardized tool and format; and the maintenance of these models over time.
- ✓ *Project Support to IPTs.* This service includes “mentoring” IPTs on data quality, use of standardized data, and legacy mappings.
- ✓ *Support to Systems Development Life Cycle (SDLC) Process Definition.* This service includes the definition of data artifacts needed to support the SDLC.

Taken collectively the SFA Data Quality Management framework can be depicted as seen in Figure 1 below.

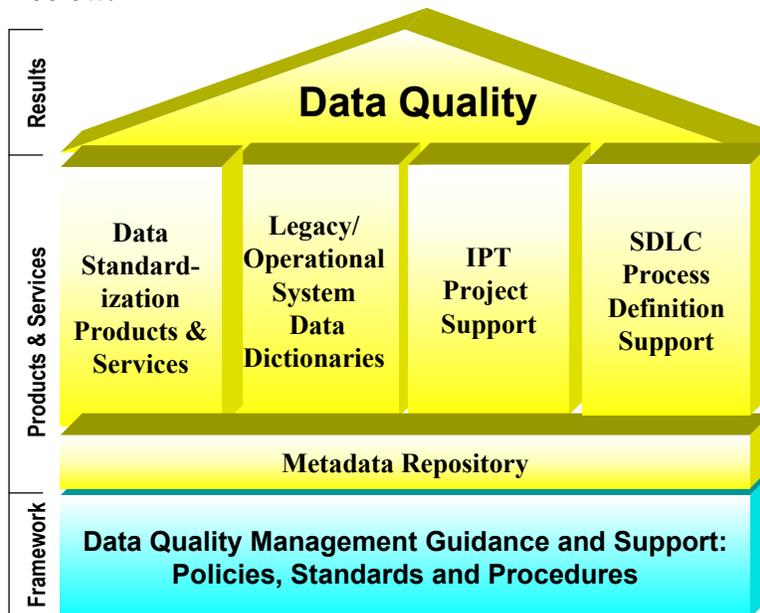


Figure 1- Data Quality Management Environment

### III. Data Quality Management Concepts

This section describes some of the key concepts underlying the SFA Data Quality Management function. As this document evolves over time additional concepts will be included. Contained in this section you will find an explanation of the following concepts:

- Data Stewardship
- Data Standardization

#### III.1 Data Stewardship

Data quality can only be achieved when clearly defined roles and responsibilities for data management are defined. The data stewardship concept addresses these management needs. An enterprise data stewardship program is usually established to formalize and manage the roles and responsibilities adopted by an organization. Currently, SFA has not adopted a formal enterprise data stewardship program. The CIO/Data Quality Management team anticipates that this will occur in the near future.

#### III.2 Data Stewardship Overview

Data stewardship may be defined as:

*An enterprise program that formalizes data accountability in order to facilitate reusability, accessibility, and sharing of quality information across the organization.*

Stewardship is an important concept in building shareable databases. Stewardship literally means managing another's property or resources and this concept applies directly to an SFA's data. It is critical that data is managed such that it is consistent, accessible, credible, usable, and secure.

The objectives of a data stewardship program are to:

- Ensure the accountability for the data assets as a key enterprise resource
- Ensure that enterprise data is reliable, consistent with data quality objectives, and usable for both internal and external consumption
- Facilitate reusability, accessibility, and sharing of the data assets across the enterprise
- Facilitate change management for data format and content across the enterprise
- Facilitate data quality initiatives throughout information systems

#### Data Stewardship Benefits:

- Improved data quality as a result of business accountability for data content
- As a result of business ownership of data definition, increased business communication and understanding through data as a common business language
- Greater alignment of information requirements with business needs - better able to respond to business drivers
- Reduced data costs through elimination of non-shared or redundant databases, interfaces and applications
- Cost avoidance by taking advantage of object reusability and information sharing
- Consistent data definition and rules ensures maximum data value, common data usage and consistent answer sets
- Provides a data conflict resolution mechanism
- Reduced data warehouse development costs through improved data quality as well as documentation of business rules.
- Documentation of Stewards and stewarded objects facilitates data change impact analysis

### III.3 Stewardship Roles and Responsibilities

All participants of the application development process have stewardship roles. Attributes, their definitions, and the applicable business rules are usually stewarded by end-users because they are the source of data and they best understand the data's uses. Other stewards maintain data models (DA), databases (DBA), shared accesses to data (Data Security), automated data processes (Developers), and computer facilities (Operations). Provided below are a definition of the roles, responsibilities and selection criteria for a data stewardship program.

#### ***Strategic Steward***

A strategic steward is the business organization responsible for setting the strategic direction towards which his organization is to move. This would most likely be within a specific channel and at a Director or General Manager. This individual has responsibility for overseeing the development of the requirements necessary to the support of the business in one or more Subject Areas to include security, usage, access, retention, and archiving across the enterprise. The responsibilities of the Strategic Steward are to:

- Set information policy and business direction for one or more Subject Area.
- Determine usage of the Subject Area information across and outside of the enterprise.
- Establish security guidelines.
- Determine overall Subject Area data quality objectives.
- Provide signature authority on Data Quality Plans.

The selection criteria for a candidate Strategic Steward are a business person that:

- Has policy and decision making authority with respect to one or more Subject Areas
- Has budget authority for the Subject Areas/Functions
- Is instrumental in defining the scope and high-level requirements/rules of one or more Subject Areas

### ***Business Steward***

The Business Steward is the business organization responsible for defining the data and business rules relating to the data, within one or more Subject Areas across the enterprise. The business steward insures that once identified and developed no change occurs that might impact what is within its purview. The business steward takes overall responsibility to insure the quality, integrity, and timeliness of the information being used by the business. The responsibilities of the Business Steward are:

- Define the business rules for internal and external data
- Provide clear concise entity and attribute definitions that describe the business content of the data item
- Provide standard calculation and summarization definitions
- Define integrity constraints
- Define attribute domains
- Define business data access/security/archival/currency requirements down to attribute level
- Define business backup and recovery policies and procedures
- Ensure the quality of data entered by data providers
- Provide Data Quality goals based on defined objectives
- Review/develop the Data Quality Plan
- Receives notification of and provides resolution for data quality issues
- Sign off on system interface design, involving data in his area
- Approval authority on data content/format or data classification change
- Effectively communicate data changes to all subscribers
- Meet data subscriber data requirements

The criteria for selecting a candidate Business Steward are:

- Detailed understanding of one or more subject areas excellent communication skills
- Respected for overall knowledge of the enterprise data
- Basic understanding of data modeling and databases
- Good technical writing skills
- Good Facilitation skills

### ***Technical Steward***

The Technical Steward is the IT organization responsible for the integrity of the data within the architecture that supports the Subject Areas in a given information system. The Technical Steward ensures data implementations comply with the direction set forth by the Business Steward. The Technical Steward assumes the responsibility for the identification, selection, implementation, and maintenance of the necessary physical environment and the components necessary to enable the data and processes that support the business requirements and therefore the business. The responsibilities of the Technical Steward are to:

- Ensure the system architecture supports data quality and the business rules as defined by the business steward

- Ensure the operational availability of data in line with the business requirements
- Provide data to other applications in a usable format
- Manage data redundancy to ensure data consistency and integrity
- Ensure information can be moved to the enterprise repository

The criteria for selecting a candidate Technical Steward are:

- Knowledgeable of overall architecture within one or more technical environments
- Knowledgeable of data from an enterprise perspective
- Technical architecture decision maker
- Member of the Information Technology organization

### ***Data Provider***

The Data Provider is the business, IT, or external organization responsible for creating or updating the data within one or more Subject Areas in a given information system. Again these are not stewards in the truest sense but they are stewards of that information, application, or enabling hardware that they provide. What is provided must be in a form or format acceptable to the business, and agreed upon most likely through a contract. If however the business chooses to alter in any way shape or form that which has been provided, then their liability ends and the business becomes the steward as is appropriate. The responsibilities of the Data Provider are to:

- Provide day to day entry/update of the Subject Area data within a given information system
- Ensure the timeliness of the data that is captured
- Notify business steward if there's a data quality issue - supply detailed information sufficient to permit understanding and diagnosis of problems
- Notify business steward if there is a change to or additional business requirements
- Review and follow procedures outlined in the Data Quality Plan

The criteria for selecting a candidate Data Provider are:

- Knowledgeable of the processes and business rules of the data
- Responsible for entering/updating data for one or more Subject Areas

### ***Data Subscriber***

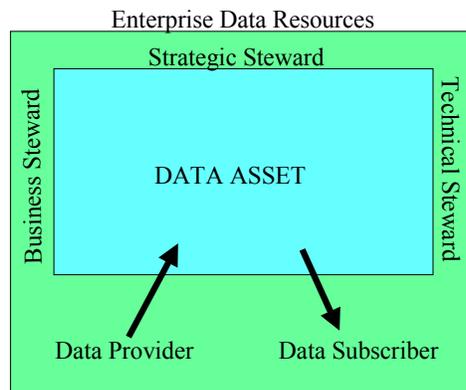
The business, IT, or external organization that uses the Subject Area data within a given information system as part of its business and is, therefore, affected by policy or content changes to that data. Although not a true steward by definition, the customer/subscriber has a responsibility to use the information provided in a means consistent with the wishes of the business. This may be either through contract, written consent, verbal consent, or implicit consent based on granted access. The customer/subscriber may also request the strategic steward to provide additional details, which may result in newly, defined requirements. The responsibilities of the Data Subscriber are to:

- Ensure data conforms to the requested requirements
- Responsible for the integrity of the data usage
- Notify business steward if there's a data quality issue - supply detailed information sufficient to permit understanding and diagnosis of problems

- Provide data and access requirements to business steward

**Organization Contact**

An Organization Contact is the representative/point of contact within an Organization who acts in one or more of the above capacities. The Strategic, Business, and Technical Stewards are the custodians of the enterprise data. They are responsible for managing its creation, usage and implementation. The Data Subscribers and Providers are also accountable for the enterprise data in that they must use the data according to the business rules set forth. In the case of the Data Provider they must provide consistently accurate data entry. The following diagram depicts data asset management in terms of the above stewardship roles:



**Figure 2- Data Management Environment**

**III.4 How Stewardship Affects Data**

In a shared data environment, end-user specifications of data access authorizations have a significant impact on how data is modeled. For example, if different business areas agree on sharing physical resources (databases), building the safeguards to protect each party's portion of the shared resource can be designed into the database. An alternative solution may be sharing the logical concepts but replicating the physical implementation of the design. Although data replication is discouraged, the whole idea behind allowing this type of replication is that it is managed, identified, and flagged for future resolution. The shared environment is a vital goal of the information architecture but along with it comes a host of business and technical considerations that have to be reconciled. This is an area where the practice of sound Data Quality Management principles can bring these viewpoints together.

### III.5 Stewardship Data Objects

Stewards are accountable for the following types of objects:

- Subject Area
- Entity
- Attribute
- Relationships

A Subject Area consists of data about a critical business area of the company, such as "Properties" or "Residents". An Entity is a person, place, thing, concept, or event about which the business wishes to store information. Entities are contained within Subject Areas. An Attribute is a piece of information that describes an Entity. For example, Resident Name and Resident Address may be attributes that describe the Entity "Resident".

Stewardship is, for the most part, assigned at the Subject Area level. All the entities and attributes within that area will be the responsibility of those stewards. There are situations where stewards can also be identified at the entity or attribute level. If stewardship needs to be assigned at the entity level, it will be those stewards who will be responsible for any entity business rules and information concerning the entity in question. If stewardship is assigned at the attribute level, it will be those stewards who are responsible for any attribute business rule information concerning the attribute in question.

When two entities are involved in an entity relationship, and they have different business stewards, it will be the responsibility of both stewards to validate the entity relationship and its detail information.

### III.6 Stewardship Maintenance

Business and technical units must select their own stewards who understand the business and technical processes and requirements as well as cross-functional uses of the data.

Stewardship can change over time due to:

1. Reorganization
2. Merger
3. Acquisition
4. People leaving/moving

Stewardship should be maintained at the organization or channel level. Those organizations will assign organization contacts to carry out the data stewardship functions. At this time a Data Stewardship Program is not implemented at SFA. When implemented, stewards and stewardship components will be maintained in the SFA Meta Data Repository and published on the Intranet for use by technical and business users. Additionally, OCIO/IT Management/Data Quality Management will be responsible for

communicating changes to data stewards. Procedures will be in place to manage these changes.

### III.7 Stewardship Communications

Data Providers/Stewards must be trained and made aware of downstream subscribers and the uses of their information products.

**Note:** No one organization owns the data since it's sharable across the enterprise and may be distributed outside of the SFA. The strategic information steward, however, is ultimately responsible for policy decisions and the distribution of the data across the enterprise as well as externally. The business steward is responsible for the integrity and security of this data based on the directions set by the strategic steward. Inter-project communication must be established in order to:

- Communicate business rule/requirements changes
- Communicate policy changes
- Communicate architecture changes

Vehicles of communication include:

- Repository look-up
- Notification to impacted stewards/subscribers/providers
- Impact analysis of data requirements/content changes, including changes to implementation plans

The business steward must establish policies and procedures for data subscribers to:

- Obtain access to data
- Submit their requirements for use of the data
- Restrictions in the use and distribution of the data

### III.7 Data Standardization

The Modernization Blueprint provides for more integrated systems. Integrated systems will require enhanced data sharing. Data standardization is the unifying influence that supports a shared data environment. The shared environment is a vital goal of the information architecture but along with it comes a host of business and technical considerations that have to be reconciled. This is an area where the practice of sound Data Quality Management principles can bring these viewpoints together.

Data standardization is important because it is easier to detect and reduce the occurrences of homonyms, synonyms, and aliases of data elements. This can be accomplished by following rigorous and comprehensive naming conventions. By doing so, the use of redundant or duplicate data elements in the enterprise data processing systems can be minimized. This reduces the total number of data elements used in these systems and minimizes maintenance to duplicate data fields.

Naming standards make a data name more descriptive and self-documenting. A meaningful, unique name provides information to the user of that data. The data name itself is the first source of information about its purpose and content.

Meaningful, descriptive definitions for all data elements are essential. Without them, data processing systems are plagued with redundant and inconsistent data. If data naming standards are not enforced, there will be problems with data incompatibility during the development and operation of systems.

Data elements are defined and documented in a variety of media throughout the organization. These include program source code, data models, process models, functional specifications, and business rules. With Data Quality Management standards enforced it is possible to improve the consistency of data references in each of these areas.

As a matter of clarification, certain tools or methodologies have specific meanings for certain words. In the Information Engineering (IE) methodology, attribute is synonymous with data element. In other environments, data attributes mean the definitive characteristics of a data element such as its length, description, format, and edit properties. Throughout this document the term “attribute” will be used in the context of the latter. Another distinction that needs to be made is the difference between a standard and a guideline. A standard is a prescribed method of doing things. A standard is an enforceable rule. A guideline is a suggested method for a course of action. A guideline serves to assist in the compliance with defined standards.

### III.8 Conclusion

The Data Quality Management function is a responsibility of everyone involved in the specification, definition, construction, operation, and usage of information systems. The principles, standards, and procedures included in this document along with a defined Solution Development Lifecycle (SDLC) Methodology form a foundation to guide the development and management of accessible, credible, and usable information delivery applications.