Enterprise Data Standard

UNIVERSITY OF SOUTH CAROLINA
Chief Data Officer

Data Standard Info	**Data Definitions & Data Dictionaries**			
	Reference ID	EDS 1.02	**Date**	06/21/2016
	Asset Classification <this document>	**Public Information**		
	Data Classification <related data element(s)>	n/a		
	Steward of this Standard	Chief Data Officer		
	Contact	Mike Kelly	kellymc2@mailbox.sc.edu 803-777-5230	
	Status	**APPROVED.** Approved by unanimous consent of Data Administration Advisory Committee on 06/13/2016.		
Data Standard Overview	The University of South Carolina, through its campuses, divisions, and organization units, establishes definitions for data elements that (1) are required for mandatory reporting, (2) support key institutional metrics, or (3) are otherwise essential to operations and services, including Enterprise Resource Planning and information systems. Data definitions and data dictionaries are encouraged for all data stores containing data and information important to the university as a whole or any particular organizational unit.			
	Purpose and Use This standard offers guidance about when data definitions and dictionaries may be required or recommended, as well as suggested content and available resources. **Required Actions & Procedures** Data definitions must be established for key data elements in enterprise information systems and data stores. Procedures are outlined below. **Justifications** The Banner Student Information System (OneCarolina) Post-Implementation Review Audit Report, dated 12/04/2015, found that “standard data definitions were not established during implementation, and still do not exist for most data elements within Banner. This was one of the Banner project objectives and is a leading practice.” Audit & Advisory Services recommended: “A data dictionary with clear definitions will facilitate efficient and accurate reporting. We recommend leveraging the new position of Chief Data officer (CDO) to accelerate efforts in developing a data dictionary for addressing undefined data elements in Banner. For future system implementation projects, the CDO should be included to facilitate the assessment, identification, and			

	implementation of data dictionary strategies. A data dictionary may save time, effort, and resources related to fixing erroneous reports and data."
Definitions	Data Element: a discrete and purposeful, often single, point of information; also known as a field, column, variable, or object. Data Definition: a statement of the precise meaning of a data element. Data Dictionary: a compendium of data definitions for multiple data elements that exist within a given data store.
Also known as	The following related terms are not recommended: data glossary; code book.
Disambiguation	n/a
Caveats and Exceptions	This standard is intended to provide definition and dictionary guidance for data stores in the university's administrative domains. While the guidance may may serve as best practice for researchers and Principal Investigators, such personnel should consult their funding or award source for requirements.
Revision	Requests for revision, additions, or other changes and suggestions may be submitted at any time to the Contact listed above. Following initial approval and adoption, the Contact may make non-substantive changes at any time; substantive changes will require approval by the appropriate group (e.g. DAAC, Data Stewardship Council, or Data Standards Committee).

Approval Log

Operational/Functional Area	Authorization	Date
Data Administration Advisory Committee	All	06/13/2016
Audit & Advisory Services (informed)	K. Dwiggins	02/24/2016
UTS Project Management Office	M. Branch-Frappier	02/24/2016
Chief Data Officer (Steward of this Standard)	M. Kelly	02/26/2016

Change Log

Date	Comments
06/21/2016	Post-approval, M. Kelly reformatted content to conform to Enterprise Data Standard template prior to publication.
03/15/2016	Proposed guidelines distributed to DAAC members for review, comment, and changes.
02/05/2016	New Standard drafted by M. Kelly and circulated to UTS PMO for review, comment, and changes.

See also	Guidelines for Data Definitions & Data Dictionaries (below)

Guidelines for Data Definitions & Data Dictionaries

Introduction

These guidelines introduce Data Stewards and Data Users to the concepts, importance, and potential parameters for data definitions and data dictionaries. Definitions of data elements may exist in the absence of a full data dictionary, but generally any information system or data store should offer a dictionary that provides clarity about its data elements.

I. Principles

Definitions and the dictionaries that contain them should conform to these principles:

A. Definitions serve to document the existence, meaning, use, and parameters of data elements.

B. Definitions should be made available to end users of a data store. In general, they should be broadly available to end users, including those who enter data into the data store and those who use or analyze the data.

C. When a data definition/dictionary is provided, end users are responsible for actively using them to correctly enter, select, and analyze data elements. When a data definition/dictionary is *not* provided, the Data Steward is responsible for any improper entry or use.

D. A definition/dictionary Contact Person should be clearly designated; users should contact this person for further clarification or to provide correction or other feedback concerning a data element

E. Each dictionary is a data asset and must be assigned a Data Classification, potentially limiting its availability (some dictionaries should not be publically accessible).

F. Definitions must be actively maintained, and changes must be communicated to known users and consumers. This is essential to the principles of change management in order to ensure that changes made at one level do not have inadvertent consequences on downstream systems.

II. Basic Guidelines

A. Purpose: Document and Share Knowledge

In its 2004 publication, the International Standards Organization (ISO) stated that "the increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data recorded in databases. One of the prerequisites for a correct and proper use and interpretation of data is that *both users and owners of data have a common understanding of the meaning and descriptive characteristics (e.g., representation) of that data*. To guarantee this shared view, a number of basic attributes has to be defined[1]." This is the essence of data definitions.

As USC offices and campuses rely heavily on each other for the entry, maintenance, and sharing of data, it is imperative to ensure we have a common understanding. Definitions and dictionaries document knowledge about: (a) what data exist; (b) who is responsible for the data; (c) data classification and other constraints, such as permissible values; (d) the intended and actual use of a data element; and (e) disambiguation from similar terminology and relationship to other data elements.

[1] International Standards Organization, 2004. Information Technology Parts 1-6 (2nd Edition) http://www.iso.org/

A data dictionary documents and shares knowledge; it is an enabling device that empowers end users to correctly enter, select, interpret, and analyze the data they access, without having to find someone to ask. Establishing a high quality data dictionary can partially relieve an office of the burden of responding to users' need for information about data; dictionaries enable self-service and improve efficiency.

B. Definition of Terms (from the forthcoming *Data Standards Program Charter*)

1) **Data Definition:** in simple terms, the degree to which a data element's intended meaning, actual meaning, and/or usage is clearly established and communicated. Depending on the particular use and purpose, a given data definition may be succinct and limited or it may be expansive. A succinct definition most often clarifies the intended use or meaning of a data element; it may be considered 'minimally adequate' to ensure proper selection and use of the data element. More expansive definitions may incorporate the universe of details about a given data element, a compilation of the element's intended meaning, purpose, and metadata. A single data element may have multiple definitions, each contextualized by function, operational area, purpose, or information system.

2) **Data Dictionary:** a compilation of data definitions for multiple data elements contained in, exchanged with, or transformed via an information system, data store, or report. A data dictionary is a companion guide for users to ensure consistent and appropriate entry, maintenance, interpretation, and use of data elements. Data dictionary is the standard terminology for such a compilation; although the term "data glossary" may be used semi-interchangeably by other organizations, the word glossary is defined as a list of terminology and as such is not equivalent to a dictionary.

C. Requirements

1) **Who is accountable for creating a data definition or dictionary?**
The individual who establishes a data element is also accountable for establishing the data definition; the same is true for the dictionaries that contain definitions. Most often the accountable person is a manager of an organization unit. We generally call these individuals "Data Stewards."

When another individual or unit receives a data element (or elements) and transforms it in any manner (e.g. calculations, joins, apply logic, trim digits, alter punctuation, etc.) then a new data element has been created; the recipient who transformed the data element becomes the Data Steward of the new data element and bears accountability for it, including the definition.

Certain enterprise systems have officially designated Data Stewards; these individuals have especially broad authority, responsibility, and discretion in the management of data in the systems indicated. Anyone receiving data from these mission-critical systems, and their companion data stores presented in Data Warehouse, must conform to the established definitions and dictionaries and not create uncertainty in duplicated terminology, meaning, or representation of the data.

Table 1: Enterprise Systems Data Stewards

Enterprise System (ERP)	Data Steward
Banner Student	University Registrar
Banner Financial Aid	Director of Financial Aid and Scholarships
PeopleSoft HCM	Director of Human Resources
PeopleSoft Finance	Controller
Oracle Identity Manager	Deputy Chief Information Officer
University Fact Books, OIRA Data Warehouse and Common Data Set	Executive Director, Office of Institutional Research, Assessment, and Analytics

2) When is a data element Definition required or recommended?

Generally, any data element should have a definition. The accountable manager (Data Steward) must exercise judgment to decide whether, when, and how to apply these requirements.

Must. A data element must have a definition if it is essential to a key institutional metric or mandatory reporting requirement. With very limited exception, each data element in an ERP system must have a definition.

Should. Under normal circumstances:

(1) Definitions should be established for any data element that is exchanged with an ERP system, whether unmodified or with transformation.
(2) Organizational units should define data elements that are essential to their mission, purpose, operations, internal reporting, and external accountability.
(3) Organizational units should define all data elements that are exchanged or shared with any other organizational unit or external entity.

3) When is a Data Dictionary required or recommended?

Generally, a data dictionary should exist for each data store (as a grouping of data elements) for which one or more data definitions have been created and approved.

Must. A data dictionary is required for all Enterprise Resource Projects (ERP), including data warehouse, Banner, PeopleSoft, and Oracle Identity Manager. Efforts to establish dictionaries must recognize that some data elements are more important than others, and that some systems require a period of use to properly understand and articulate the definitions of certain data elements. The potential need for a data dictionary, or changes to an existing one, should be a key consideration with any ERP system implementation, upgrade, or update. *NOTE: for ERP systems and data stores in existence prior to this Standard, every effort should be made to conform to the spirit and intent of this Standard as quickly as feasible.*

Should. For all information systems and databases that serve a unit or office of the university, sometimes known as Auxiliary Information Systems, a data dictionary should be compiled throughout the build and/or implementation processes. When such systems transact data with an ERP, or when the data in them is shared with other units, a dictionary is a practical requirement. *NOTE: for information systems and databases in*

existence prior to this Standard, every effort should be made to conform to the spirit and intent of this Standard as quickly as feasible.

4) What are the maintenance and notification requirements for data definitions and data dictionaries?

Data Stewards are responsible for the maintenance of data dictionaries and the data element definitions that comprise them. Data Stewards are also responsible for proactively notifying all data users and consumers of their availability as well as significant or meaningful changes to data element content and/or definitions over time.

Data definitions and/or dictionaries should:

a) be reviewed for currency at least annually
b) be modified whenever needed
c) be shared with individuals who are granted access and permissions to systems
d) be posted as an online, searchable resource
e) provide a mechanism for data users and consumers to be actively notified of changes
f) provide a mechanism for data users and consumers to provide feedback, make inquiries, and receive clarification about definitions

5) What are the publication and access requirements for definitions and dictionaries?

First and foremost, definitions and dictionaries are data assets, and must therefore have a Data Classification applied to them as described in policy UNIV 1.51 (Data & Information Governance, approval pending), generally based upon the Data Classification of the data elements or system described in the dictionary.

Considering the asset classification, Data Stewards are responsible for deciding appropriate publication and access criteria for a given dictionary. The Chief Data Officer (CDO) website or appropriate SharePoint site are the preferred host locations for enterprise dictionaries, unless there is a compelling reason to post a dictionary elsewhere. The intent is to make the CDO site a familiar and central go-to source for dictionaries; SharePoint sites may be used when the content of definitions or dictionaries requires access restrictions.

A couple illustrations may help clarify publication requirements and constraints:

a) Some data definitions may be ‘secret sauce recipes’ – calculations or transformations of one or more data elements that are proprietary, or business decision drivers, and/or which could present a potential risk of data controversy/public relations; in such cases, it is reasonable for the Data Steward to limit access. Even so the data element needs definition (perhaps part Public Information, part Restricted or Confidential). Instead of embedding the calculation or transformation in the definition, the definition may name the custodian of the calculations or transformation who retains the documented methodology in a secure location. The dictionary may also reside behind an access-constraining service, such as SharePoint.

b) State of SC info privacy requirements (SCDIS-200) call for agencies to classify their data assets in addition to classifying their data elements. Each data dictionary is an asset and must be classified. *Why?* The content requirements for definitions described below indicate that the source system data elements should be catalogued in the definition. While lineage and traceability is vitally important to understanding

data, publicly exposing some data table and field names could serve as an expedient blueprint for a hacker to get at select data points. Stewards must guard against that possibility. So, definition content requirements may need to be modified or waived for particularly sensitive data elements.

6) What are the content requirements for a data definition?

Effort and detail in defining data elements should be roughly matched to the importance of the data element; the more important a data element is to the operation of a unit or the enterprise, the more detail should generally be documented about it. The complexity of a given data element also has bearing on how extensive its definition is. As a rule of thumb, for less critical information systems and data stores, a minimally adequate approach may be sufficient. As the number of users increases, or the importance of the data increases, more effort should be invested. Below are general content guidelines for definitions which may be adapted on a case-by-case basis.

a) Minimally-adequate definitions should contain:

1. Unique ID – alphanumeric identifier (provides tracking in cases where Data Element Name may be changed)
2. Data Element Name – the name of the term or element (must be distinct and unique)
3. Basic Meaning – a brief definition or readily understandable meaning
4. Interpretation & Usage Notes – any other info enabling users to accurately select needed elements
5. Data Classification – governs security & access controls; 4-tier schema
6. Domain – name/location of the data store or org unit to which the data element belongs
7. Data Source Info – info about the database, table, and field
8. Integrations & Dependencies – list of data stores where the element is distributed, if any (AKA upstream- or downstream-systems)
9. Calculations and Transformations – describes any logic applied to the data element or values (e.g. number or decimal places in a GPA)
10. Data Type – describes the values the data element may contain (e.g. short text, long text, numeric, date or date-time, currency, etc.)
11. Effective-Dated Info – describes whether records may contain multiple values, each attached to a date. Effective-dated data can present special challenges for users to select the value in place at a desired point in time
12. Data Steward – manager accountable for the data element (should be the system owner and same person who approves the definition)
13. Contact Person and Contact Information – individual the end users may contact for clarification
14. Modified Date – date the definition was most recently established, modified, and/or approved
15. Modified By – person who most recently modified the definition

b) Extended definitions may contain:

1. All components of (a) Minimally-adequate definitions
2. Extended or exhaustive meaning
3. List of similar data elements and disambiguation guidance
4. List or range or description of valid/permissible values
5. For validated data elements, list of codes and descriptions OR naming the validation table and its location

6. Extended Interpretation & Usage Notes – providing instruction on selecting correct value (usually required if effective-dated rows of values exist for a given data element)
7. Log of Changes to the definition – each time a change occurs to the data element or permissible values (what changed, when it changed, and who authorized)
8. Campus or Entity – name(s) of campuses for which the data element has use
9. External Requirements – if a data element exists to meet external requirements, the requiring entity and/or mechanism should be identified (e.g. reports to CHE, IPEDS, or NCAA Title IX)

c) Optimized definitions may contain some or all of the following:

1. All components of (i.) Minimally adequate definitions
2. Appropriate components of (ii.) Extended definitions
3. Technical Source Details
   - Exhaustive list of system modules or screen names where data element is found
   - Designation of module/screen which is the authoritative source of the 'true' value and/or the location where change of value can be effected
   - System table(s)
   - Technical field name for the data element
   - Any known technical constraints on the data element and/or its values
4. Data Warehouse location details
   - Data element name
   - Packages and tables containing the data element
   - Known calculations or transformations (if any)

III. Resources for Establishing Data Definitions and Data Dictionaries

To fulfill their purposes, data definitions and data dictionaries must be made available in a timely manner to those who require or would benefit from the knowledge they contain. Need for definitions and dictionaries should be a factor in procuring, building, or otherwise implementing any information system or data store.
The Chief Data Officer serves as an internal resource for determining when a definition/dictionary is needed and for constructing data definitions and data dictionaries for a variety of purposes across the USC system, including:

A. Data Projects & Information Systems

Any organizational unit purchasing or creating an information system or data store is responsible for considering the need for a data dictionary (if the activity is a formal 'Project' please see item C below). Wherever practical, it a best practice to establish clear meanings of data elements. This is especially true when the data from the source will be shared or reported, internally or externally. The unit manager, as a Data Steward, determines whether definitions or dictionaries are required, and likewise determines the level of effort, content, and availability requirements. The Chief Data Officer serves as a resource for the unit manager in making these decisions.

Special consideration should be given to any new or revised data integrations, data interfaces, or data feeds to/from the system, resulting from the project. This is especially true for data elements exchanged with enterprise systems (ERPs including Banner, PeopleSoft, OIM, etc.). Coordination with Enterprise System Data Stewards may be critically necessary or highly advisable.

B. Existing Data Stores and Information Systems

Information systems and data stores existing prior to this Standard (March 2016) should be reviewed as quickly as feasible by their Data Stewards to assess the need or benefit of establishing data definitions and/or a data dictionary.

First priority should be given to stores and systems known to have data integrations, data interfaces, or data feeds to enterprise systems or involve external sharing. Second priority should be given to stores and systems that: (a) contain data of a sensitive nature (that is, Data Classification of Confidential or Restricted); or (b) share data between organizational units. Lowest priority should be given to stores or systems that serve a discrete/limited purpose or single organizational unit.

C. Formal PMO Data Projects & Information Systems

Certain data and information initiatives are accepted for management by organizations known as Project Management Offices (PMO). The university PMO methodology includes consideration of need for a data dictionary in the project charter/scope (effective March 2016). The Chief Data Officer serves as a resource for the Project Sponsor in this consideration and throughout the project lifecycle.

If a data dictionary is needed, the Project Sponsor will work with the Project Manager (PM) to ensure the scope, work breakdown structure, resource assignments, timeline, and deliverables address the need. Projects will also clarify who the System Owner will be; in most cases this person also holds the role of Data Steward.

In general, functional project team members will be responsible for the following under the direction of the Data Steward: produce definitions and dictionaries; communicate availability and access to stakeholders; and, perpetual maintenance of definitions and dictionaries after transition to operations (go-live).

Special consideration should be given to any new or revised data integrations, data interfaces, or data feeds to/from the system, resulting from the project. This is especially true for data elements exchanged with enterprise systems (ERPs including Banner, PeopleSoft, OIM, etc.). Coordination with Enterprise System Data Stewards (see *Table 1*) may be critically necessary or highly advisable.

D. Research Data & Information Systems – *see also RSCH 1.05 Data Access and Retention*

Unless otherwise noted, Principle Investigators (PIs) essentially serve as Data Stewards of information systems and data stores related to their initiatives. In many cases, research funding requires data to be collected, maintained, reported, or shared according to specific parameters. PIs must consider any such requirements and other benefits of establishing a data dictionary and/or research codebook, which may include adopting an external model. PIs may consult the Chief Data Officer for additional guidance.