Enterprise Data Standard

UNIVERSITY OF SOUTH CAROLINA
Chief Data Officer

Data Standard Info	**Data-related Terminology**			
	Reference ID	EDS 1.00	**Date**	02/09/2018
	Asset Classification < this document >	**Public Information**		
	Data Classification < related data element(s)>	n/a		
	MDM / RDM	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 06/30/2016.		

Data Standard Overview

Purpose and Use
This document standardizes terminology used in USC's data-related activities and offers definitions or explanations of each. This dictionary disambiguates items that are often (sometimes mistakenly) used interchangeably. This standard helps USC data users adopt a common vocabulary to ensure their usage is consistent and their intent is understood.

Conventions
While it is not always imperative or required that users be especially precise, this content may assist users in communicating effectively and efficiently.

Required Actions & Procedures
none

Justifications
n/a

Acknowledged Stakeholders

Stakeholder	RACI	Status
DAAC	Consulted	Complete 06/20/2016
Chief Data Officer	Responsible	n/a
UTS Data Warehouse	Informed	Pending 07/06/2016

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	Unanimous consent	06/30/2016

	Chief Data Officer	M. Kelly	06/20/2016
Change Log	Date	Comments	
	02/09/2018	Minor revisions to enable hyperlinks in this and related documents including University policies (M. Kelly)	
	06/20/2016	New standard drafted by M. Kelly for comment and revision by Data Administration Advisory Committee members and other stakeholders. Please contact Mike Kelly with questions or changes.	
See also	Appendix 1: Catalog of Data-related Terminology The following university Policies: • UNIV 1.51 Data and Information Governance • UNIV 1.52 Responsible Use of Data, Technology, and User Credentials The following Charter is in *draft* status: • Charter – Data Quality and Integrity Assurance Program		

Appendix 1. Catalog of Data-related Terminology

This catalog, grouped by topic areas, standardizes terminology used in USC's data-related activities and offers definitions or explanations of each. This dictionary disambiguates items that are often (sometimes mistakenly) used interchangeably.

Please note that the terminology is defined as used at USC and would not necessarily be defined the same everywhere.

Ideally, this catalog helps users to adopt a common vocabulary that will ensure their usage is consistent and their intent is understood.

Table of Contents

USC-specific Terminology

USC Data & Information Governance Framework & Programs 4
USC Data Governance Roles 5

General Use Terminology

Access and Permissions 6
Big Data and Analytics 7
Data Categories 8
Data Quality & Integrity 9
General Data Terminology 10
Inspection & Confirmation of Correctness / Expected Results or Values 11
Relationships between Data & Systems 12

USC Data & Information Governance Framework & Programs

USC's data governance framework is comprised of five programs under the direction of the Chief Data Officer; the named programs and their descriptions and purview are as follows.

-per University Policy UNIV 1.51

Program	Content, Concerns, and Initiatives
Data Quality & Integrity Assurance (DQIA)	program to continuously improve the quality of data, resulting in data that has integrity throughout its lifecycle. General indicators of data quality include: accessibility, accuracy, completeness, consistency, currency, definition, and timeliness.
Data Standards	conventions and services (such as address cleaning) that Data Stewards have adopted for select data elements, to ensure their appropriate and consistent use across the university system and to facilitate integrations.
Data Stewardship	assures coordinated and purposeful activities, services, procedures, and compliance with federal and state law and regulations, as well as university policy, with regard to data and information. This program, and the council which controls it, directs activities and resources for the other programs.
Identity & Access Management (IAM)	creates enabling structures and services that ensure purposeful, careful, and expert management of identities, access, and permissions to data and information systems, in ways that fulfill the critical business requirements of the university's functional and operational units.
Reporting, Analytics, and Decision Support (RADS)	identifies and promotes coordinated and purposeful activities, services, resources, and procedures to enhance the availability of data and improves the university's reporting and data driven decision-making capacities.

USC Data Governance Roles

USC has established the roles below in support of data governance. Full definitions are provided in University policy UNIV 1.51.

Role	Purview
Chief Data Officer (CDO)	coordinates and collaborates with stakeholders of university data to implement, administer, and continuously improve data governance programs.
Constituents	persons and entities that have a relationship to any organizational unit of the university system, including but not limited to: students, employees, and other affiliates.
Data and Information Strategy Council (DISC)	presidentially-appointed council with ultimate oversight of and decision-making authority for data and information governance, including but not limited to establishing goals and key performance indicators that align data governance to the university system's strategic priorities and initiatives.
Data Custodians	personnel who maintain hardware, information systems/databases, applications, backup systems, and networks through which data is transmitted, processed, and stored. Custodians may be university personnel or personnel/service providers under agreement or contract with the university.
Data Domains	see EDS1.05; discrete, definitive, and defined list of the university's major functional areas, lines of business (LOB), and subject matter into which data may be grouped. Each domain has a Domain Governor.
Data Steward	individuals who oversee the capture, maintenance, storage, use, and dissemination of University Data and information for a particular function or operation; they may be considered System Owners for stores and systems they purchase, operate, or contract with a third party/service provider to operate and/or host.
Data Trustee	individuals who have strategic planning and policy-making responsibilities with implications for University Data. Data Trustees designate Data Stewards for the organizational units under their care.
Domain Governor	see EDS 1.05; individual designated by a University Policy Responsible Officer to provide guidance, recommendations, and consultation to Data Stewards of the subdomains.
End User	any person or system that accesses university assets including data and information systems.

Access and Permissions

The following terms describe what a user needs, or needs to do, in order to do his or her job, as well as the processes involved in meeting those needs. These terms are *listed in a relatively logical progression*: identify the person first, then assign credentials, request access to be approved an appropriate party, authorization to use resources by specifying the permissions assigned to the identified user, and the user then authenticating his/her user credentials to gain access.

Terminology	Definition and Usage
Identification	ensures the correct and accurate end user or person and his/her associated credentials
User Credentials	accounts, email address, network username, and other forms of identification, often paired with a password, that are issued or assigned to a person
Access Request	the formal request for a user to be authorized for use of or access to a resource and the particular permissions being requested for that user on that resource
Access	*n.* authorization to view or use a system or data store *v.* the act of gaining entry to or use of a system or data store
Authorization	the formal approval by an appropriate authority (such as a manager or data steward) for a user to be granted access or use of a system or data store
Permissions *Acceptable synonyms:* - entitlements - rights	the actions, activities, and views to which a user is authorized for a particular resource. The gold standard is **Role-based Permissions** that are derived from and based upon the intersection of a user's job duties and the department/unit to which he/she belongs. Note: although USC promotes the Principle of Least Privilege, some systems do not enable administrators to limit or constrain access due to design limitations; in such cases users may be granted access or permissions to more than they truly need.
Authentication	passing of a user's credentials through a university system to ensure identification and to enable entry to a resource where he/she will exercise permissions

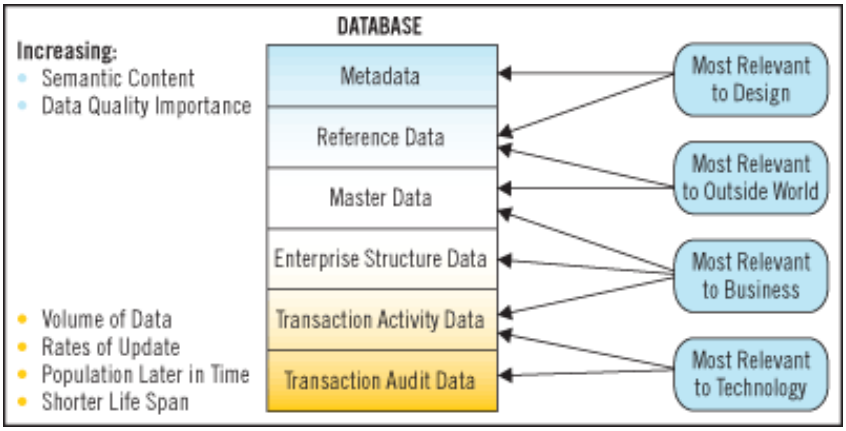
Big Data and Analytics

The definitions of the five Vs listed below are adapted from https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know

Terminology	Definition and Usage
Analytics	advanced analysis of data; similar to reporting except that underlying questions tend to be complex, the data is typically characterized by the "5 Vs," the statistical skills and techniques are more innovative, and the outputs are more sophisticated (e.g. interactive dashboards). At least 6 categories of analytics are used in higher education.
Descriptive Analytics	describes what happened (most similar to reporting)
Diagnostic Analytics	describes why something happened
Discovery Analytics	reveals previously unknown relationships in data
Learning Analytics	explores the relationships among instructional/pedagogical techniques, environmental factors, and student outcomes
Predictive Analytics	describes what is likely to happen
Prescriptive Analytics	describes what should happen or what should be done
Big Data	data, or data stores, that are characterized by the "5 Vs": Value, Variety, Velocity, Veracity, and Volume
Value	the costs and benefits of data that the organization possesses or may acquire
Variety	describes the varied types of data that may be available, including structured, unstructured, semi-structured, and relational data; may also refer to the variety of subjects the data pertains to, or the known values about those subjects
Velocity	the speed at which new data is generated, new values are produced, and/or the speed at which data moves through systems
Veracity	describes the trustworthiness of the data; with Big Data, quality and accuracy are often considered less controlled or controllable
Volume	refers to the vast quantity of data that has been generated and continues to be generated

Data Categories

The data categories described below are adapted from a touchstone article from information-management.com in 2006 entitled “Master Data versus Reference Data”, by Malcolm Chisholm. Accessed 07/06/2016 at http://www.information-management.com/issues/20060401/1051002-1.html



Items are presented by primary importance first, then alphabetically.

Terminology	Definition and Usage
Master Data	the cohesive, authoritative record of information about parties, subjects, persons, entities, products, services and offerings of and specific to the university or its organizational units (e.g. the person record as well as the courses for which they enrolled). The same instance may be known by many different names or IDs, all of which are traceable to the same instance.
Reference Data	any kind of data that is used solely to categorize other data found in a database, or solely for relating data in a database to information beyond USC (e.g. sex, country, states); generally change infrequently.
Metadata	data about data: definition, attributes, and qualities
Enterprise Structure Data	data that represents the structure of the enterprise, particularly for reporting business activity by responsibility. It includes things such as organizational structure and charts of accounts. Enterprise structure data is often a problem because when it changes it becomes difficult to do historical reporting
Transaction Activity Data	represents the transactions that operational systems are designed to automate. It is the traditional focus of IT, including things such as orders, sales and trades
Transaction Audit Data	data that tracks the progress of an individual transaction, such as Web logs and database logs

Data Quality & Integrity

Terminology	Definition and Usage
Data Integrity	the validity of data across its lifecycle, measurable by the combined characteristics of data quality; it is the opposite of data corruption
Data Quality	eight characteristics of data that are measurable indicators and serve as the foundation for data integrity: accessibility, accuracy, completeness, consistency, currency, definition, relevance, and timeliness (For more complete definition, please see Charter for Data Quality & Integrity Program {link forthcoming})
Accessibility	data is available for its intended use by a user (not blocked or hidden)
Accuracy	errors are not present in the data and the degree to which precision and detail align with data definitions or standards
Completeness	values are present in a data element for unit records that should contain them
Consistency	standards are applied and enforced for data date elements and their values, especially for structured data. May also refer to values for a given data element being equivalent (the same) when they are present in multiple information systems and/or data repositories
Currency	values are up to date, or, values are clearly tied to a specific and relevant time and date
Definition	a data element's intended meaning, actual meaning, and/or usage is clearly established and communicated
Timeliness	values are provisioned at the point in time when they are needed to fulfill operational or reporting needs

Note: literature and practice show that there are many definitions, metrics, and indicators of data quality; the terminology and definitions above reflect usage at the University of South Carolina; they are subject to change.

General Data Terminology

Terminology	Definition and Usage
Cognos	the application layer that enables and supports the USC Data Warehouse.
Data	information that is known and/or values that describe characteristics or quantities of a being, object, transaction, or event. Data may be used interchangeably with the terms content and information.
Data Definition	*brief:* describes what a data element is and what it means; system-specific. *extensive:* Information about a data element ensure its intended meaning, actual meaning, and/or usage are documented and made available to appropriate persons. The content of a definition follows a fairly consistent template, although the volume of detail needed varies according to importance or complexity of the data element it pertains to.
Data Element	denotes a discrete and purposeful, often single, point of information; Data Elements is the preferred term, but such items may also be known as a field, column, variable, or object.
Data Standard	*brief:* documented agreement about data and information across/between systems or org units. *extensive:* Written guidance that provides for consistent entry, handling, processing, monitoring, updating, and use of select data elements that are exchanged between information systems, agencies, or organizational units. The content of a data standard is highly adaptable to the data element(s) it pertains to; a data standard may call for processing postal addresses through a cleansing program prior to import to the ERP, or it may be a list of disallowed characters for an information system.
Data Warehouse	the university's central data repository, administered by UTS/IBM.
Historized Data *or historize, historization, history data*	data that can be rolled back to specific date to see what the values were at a particular point in time, and often enables comparison between two dates. Note: historization is only available for select Student-related data tables (and all elements contained on those tables) going back to August 12, 2014; generally person demo and address data is excluded. Each Historized tables were started on different dates. This term is unique to USC.
Operational Data Store *or ODS*	an interim repository of data that has been extracted and possibly integrated from information systems or data sources, and aligned to business rules. It sits between the source system (e.g. Banner) and the Data Warehouse.
Resource Account	a user account established for either non-persons (or, rarely, a strictly limited group of persons), to enable it/them to perform certain activities in a system

Terminology	Definition and Usage
Static Data	data that is frozen in a data store; the data is what is was at a specific date in time.
Unit Record	the data elements, or a subset of elements, that belong to a particular person, vendor, asset, etc. In an information system about students, each individual and unique student would normally have one cohesive unit record; in a spreadsheet view this is often seen as a single row of information for a given person. In relational databases, the illustration of relationships may be much more complicated
Value	the content that is present in a data element for a given unit record (person or object).

Inspection & Confirmation of Correctness / Expected Results or Values

Although the phrases below are relatively interchangeable in common use, "validate" is the preferred term.

Terminology	Definition and Usage
Validate *or* validating, validation, validated	establishing correctness of a deliverable or outcome (e.g. data value, report, etc.) by means of thorough testing and/or confirmation process; ensure factual proof.
Verify *or* verifying, verification, verified	establishing correspondence of actual facts or details with those proposed or guessed at.
Vet *or* vetting, vetted	to check (something) carefully to make sure it is acceptable.

Relationships between Data & Systems

Terminology	Definition and Usage
Export	pushing a set of data elements or unit records out of a system for use elsewhere
Extract	pulling a set of data elements or unit records out of a system for use outside the system
Feed (data feed)	an activity that provides data or unit records from one system for use in another system or business process. The data set is typically an export from a sending system that is 'fed' as an import into a receiving system
Flat file	a data source in which all data is contained in a single file (generally a spreadsheet or delimited .txt file), and often with one row of data per unit (person or object); distinguished from *relational database* (below)
Import	pulling a set of data elements or unit records into a system for use by that system
Integration	two or more systems combined to work as one solution; generally implies real-time access to database tables via programmed application interfaces. Ex. TouchNet integration with Banner
Interface	two systems share information by means of a bridge where the sources of information and even the programming language may be different; the interface may be one-way or bi-directional
Relational database	a data source in which data is distributed among numerous files, with a common primary key (identifier) linking the units on rows; columns contain the values associated with the units and may use additional keys to link codes to values from reference tables; distinguished from *flat file* (above)
UC4	a scheduling service that can initiate configured jobs or series of jobs to execute as planned. Most interfaces are automated via UC4 jobs. Ex. cleaning postal addresses as they move from one system into another