

APPENDIX D. GLOSSARY³³

3-sigma (3 σ or 3s): Three standard deviations used to describe a level of quality in which three standard deviations of the population fall within the upper and lower control limits of quality with a shift of the process mean of 1.5 standard deviations, and in which the defect rate approaches 6.681%, allowing no more than 66,810 defects per million parts. ^{LPE}

6-sigma (6 σ or 6s): Six standard deviations used to describe a level of quality in which six standard deviations of the population fall within the upper and lower control limits of quality with a shift of the process mean of 1.5 standard deviations, and in which the defect rate approaches zero, allowing no more than 3.4 defects per million parts. ^{LPE}

Accessibility: the degree to which the knowledge worker or end customer is able to access or get the information they need. ^{LPE}

Accuracy to reality: A characteristic of information quality measuring the degree to which a data value (or set of data values) correctly represents the attributes of the real-world object or event. ^{LPE}

Accuracy to surrogate source: A measure of the degree to which data agrees with an original, acknowledged authoritative source of data about a real world object or event, such as a form, document, or unaltered electronic data received from outside the organization. ^{LPE}

Atomic level: Defines attributes that contain a single fact. For instance, "Full Name" is not an atomic level attribute because it can be split into at least two distinct pieces of information: "First Name" and "Last Name."

Automated information quality assessment: Information quality inspection using software tools to analyze data for business rule conformance. Automated tools can assess that a data element content is valid (adheres to business rules) for most business rules, and they can determine consistency across files or databases, referential integrity and other mechanical aspects of information quality. However, they may not automate assessment of some very complex business rules and they *cannot* evaluate accuracy. ^{LPE}

Business concept: A person, place, thing, event or idea that is relevant to the business and for which the enterprise collects, stores, and applies information. Procedural note: for business concepts to be properly used and managed, they must be clearly understood;

this requires that they be concisely defined using rigorous declarative language (as opposed to procedural language).

Business information steward: The person or group that manages the development, approval, and use of data within a specified functional area, ensuring that it can be used to satisfy business data requirements throughout the organization.

Business rule: A statement expressing a policy or condition that governs business actions and establishes data integrity guidelines. ^{LPE}

CASE: Acronym for *Computer-Aided Systems (or Software) Engineering*. The application of automated technologies to business and information modeling and systems (or software) engineering. ^{LPE}

(HUD's) Common Data Element Correction

Method: A method for correcting data, developed for HUD, based upon accepted industry standards and incorporating project management and total quality management principles. Now replaced with HUD's Total Information Quality Management methodology. See *Total Information Quality Management*.

Common term: A standard English word used by HUD as defined in a commercial dictionary (for instance "Enterprise is a unit of economic organization or activity, esp.: a business organization"). ^{MWD}

Completeness: A characteristic of information quality measuring the degree to which all required data is known. (1) *Fact* completeness is a measure of data definition quality expressed as a percentage of the attributes about an entity type that need to be known to assure that they are defined in the model and implemented in a database. For example, "80 percent of the attributes required to be known about customers have fields in a database to store the attribute values." (2) *Value* completeness is the first measure of data content quality expressed as a percentage of the columns or fields of a table or file that should have values in them, in fact do so. For example, "95 percent of the columns for the customer table have a value in them." Also referred to as *Coverage*. (3) *Occurrence* completeness is the second measure of the data content quality expressed as a percentage of the rows or records of a table or file that should be present in them. For example, "95 percent of the households in which HUD needs to know about, have a record (row) in the household table." ^{LPE}

Concurrency: A characteristic of information quality measuring the degree to which the timing of equivalence of data is stored in redundant or

distributed database files. The measure data concurrency may describe the minimum, maximum, and average information float time from when data is available in one data source and when it becomes available in another data source. Or it may consist of the relative percent of data from a data source that is propagated to the target within a specified time frame (also see *Information float*).^{LPE}

Consistency: A measure of information quality expressed as the degree to which a set of data is equivalent in redundant or distributed databases.^{LPE}

Contextual clarity: the degree to which information presentation enables the knowledge worker or end customer to understand the meaning of the data and avoid misinterpretation.^{LPE}

Controllable mission-critical information: Controllable means that HUD has control over information or data content because it is collected following HUD standards (e.g., housing authority filings) or produced within HUD. Non-controllable mission-critical information is information acquired by HUD from sources that cannot be controlled by HUD such as survey information from an external source (e.g., homeownership rates from the Bureau of Census) or information resulting from HUD actions such as surveys from small samples with large margin of error, non-respondents impacting the representativeness of the sample or inaccurate responses from respondents.

Data: The representation of facts. Data can represent facts in many mediums or forms including digital, textual, numerical, or graphical. The raw material from which information can be produced when it is put in context that gives it meaning.^{LPE}

Data correction: See *Information Product Improvement*.

Data definition: The process of analyzing, documenting, reviewing, and approving unique names, definitions, characteristics and representations of data according to established procedures and conventions and standards.

Data dictionary: A repository of information (metadata) defining and describing the data resource. A repository containing metadata. An *active* data dictionary, such as a catalog, is one that is capable of interacting with and controlling the environment about which it stores information or metadata. An *integrated* data dictionary is one that is capable of controlling the data and process environments. A *passive* data dictionary is one that is capable of storing metadata or data about the data resource, but is not capable of interacting with or controlling the

computerized environment external to the data dictionary. See also *Repository*.^{LPE}

Data element: The smallest unit of named data that has meaning to a knowledge worker. A data element is the implementation of an attribute. Synonymous with data item and *field*.^{LPE}

Data improvement: See *Information Product Improvement*.

Data intermediary: a role in which individuals transform data from one form, not created by them, to another form (e.g., data entry technicians).^{LPE}

Data quality: See *Information quality*.

Data reengineering: The process of analyzing, standardizing, and transforming data from un-architected or non-standardized files or databases into enterprise-standardized information architecture (definition and architecture).^{LPE}

Data standardization: See *Data Definition*.

Defect: An item that does not conform to its quality standard or customer expectation.^{LPE}

Definition conformance: the degree of consistency of the meaning of the actual data values with its data definition.^{LPE}

Dependency rules: The restrictions and requirements imposed upon the valid data values of a data element by the data value of another data element. Dependency rules are revealed in the business rules. Examples of dependency rules include

- An Order without a Customer Name is not valid.
- If Employee Marital Status is 'Married'.
- Employee Spouse information must be present.
- An Employee Termination Date is not valid for an Active Employee.
- When an Order is 'Shipped', the Order Shipping Date must be captured.

Derivation integrity: the correctness with which derived data is calculated from its base data.

Derived data: Data that is created or calculated from other data within the database or system.^{LPE}

Dissemination: to spread abroad as if sowing seed (to plant seed for growth especially by scattering; e.g., disseminating ideas); to disperse throughout.^{MWD}

Dissemination of information: (In the context of information dissemination by federal agencies) Agency initiated or sponsored distribution of information to the public (see 5 CFR 1320.3(d))

(definition of "Conduct or Sponsor"). Dissemination does not include distribution limited to government employees or agency contractors or grantees; intra- or inter-agency use or sharing of government information; and responses to requests for agency records under the Freedom of Information Act, the Privacy Act, the Federal Advisory Committee Act or other similar law. This definition also does not include distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes. ^{OMB}

Domain: (1) Set or range of valid values for a given attribute or field, or the specification of business rules for determining the valid values. (2) The area or field of reference of an application or problem set. ^{LPE}

External partner: These are individuals and organizations that provide to and/or receive from HUD services and/or information regarding housing and/or urban development. They include state and local governments, other federal agencies, housing authorities, and public service organizations.

Fact: the quality of being actual; something that has actual existence; an actual occurrence; a deed. ^{MWD}

Format consistency: The use of a standard format for storage of a data element that has several format options. For example, Social Security Number may be stored as the numeric "123456789" or as the character "123-45-6789". The use of a uniform format facilitates the comparison of data across databases.

Influential information: is scientific, financial, or statistical information that the U. S. Government Agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions. ^{OMB}

Information (1): the communication or reception of knowledge or intelligence; knowledge obtained from investigation, study, or instruction; intelligence; news; facts, data; the attribute inherent in and communicated by one of two or more alternative sequences or arrangements of something (as nucleotides in DNA or binary digits in a computer program) that produce specific effects; a signal or character (as in a communication system or computer) representing data; something (as a message, experimental data, or a picture) which justifies change in a construct (as a plan or theory) that represents physical or mental experience or another construct; a quantitative measure of the content of information –*specifically* a numerical quantity that measures the uncertainty in the outcome of an experiment to be performed. ^{MWD}

Information (2): (In the context of business and government use; disseminated or not; this is the definition used in this Handbook) Data in context. The meaning given to data or the interpretation of data based on its context. It is the finished product as a result of the interpretation of data. ^{LPE}

Information (3): (In the context of information dissemination by federal agencies) Any communication or representation of knowledge such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms. This definition includes information that an agency *disseminates* from a web page, but does not include the provision of hyperlinks to information that others disseminate. This definition does not include opinions, where the agency's presentation makes it clear that what is being offered is someone's opinion rather than fact or the agency's views. ^{OMB}

Information architecture: A "blueprint" of an enterprise expressed in terms of a business process model, showing what the enterprise does; an enterprise information model, showing what information resources are required; and a business information model, showing the relationships of the processes and information. ^{LPE}

Information dissemination: see *Dissemination of information*.

Information float: The length of the delay in the time a fact becomes known in an organization to the time in which an interested knowledge worker is able to know that fact. Information float has two components: Manual float is the length of the delay in the time a fact becomes known to when it is first captured electronically in a potentially sharable database. Electronic float is the length in time from when a fact is captured in its electronic form in a potentially sharable database, to the time it is "moved" to a database that makes it accessible to an interested knowledge worker. ^{LPE}

Information group: A relatively small and cohesive collection of information, consisting of 20–50 data elements and related entity types, grouped around a single subject or subset of a major subject. An information group will generally have one or more subject matter experts and several business roles that use the information. ^{LPE}

Information producer: The role of individuals in which they originate, capture, create, or maintain data or knowledge as a part of their job function or as part of the process they perform. Information producers create the actual information content and are accountable for its accuracy and completeness to

meet all information stakeholders' needs. See also *Data intermediary*.^{LPE}

Information product improvement: The process of data correction, reengineering, and transformation required to improve existing defective data up to an acceptable level of quality. This can be achieved through manual correction (by inspection or verification), manual or automated completion, filtering, merging, decoding, and translating. This is one component of *information scrap and rework*. See also *Data reengineering*. Information product improvement is *reactive* information quality.^{LPE}

Information quality: The degree to which information consistently meets the requirements and expectations of the knowledge workers in performing their jobs.^{LPE}

Information quality (assessed level): The measurement of actual quality of a set of information against its required quality characteristics.^{LPE}

Information quality (desired level): The level of data quality required to support the business needs of all information consumers.^{LPE}

Information quality assessment: The random sampling of a collection of data and testing it against its valid data values to determine its accuracy and reliability. Also called *data quality assessment* or *data audit*.^{LPE}

Information stakeholder: Any individual who has an interest in and dependence on a set of data or information. Stakeholders may include information producers, knowledge workers, external customers, regulatory bodies, and various information systems roles such as database designers, application developers, and maintenance personnel.^{LPE}

Information steward: There are seven business roles in information stewardship and nine information systems roles in information stewardship. See *Business information steward*.^{LPE}

Information value / cost chain: The end-to-end set, beginning with suppliers and ending with customers, of processes and data stores, electronic and otherwise, involved in creating, updating, interfacing, and propagating data of a type from its origination to its ultimate data store, including independent data entry processes, if any.^{LPE}

Integrity: The security of information; protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.^{OMB}

Knowledge worker: The role of individuals in which they use information in any form as part of their job

function or in the course of performing a process, whether operational or strategic. Also referred to as an *information consumer* or *customer*. Accountable for work results created as a result of the use of information and for adhering to any policies governing the security, privacy, and confidentiality of the information used. The term *knowledge worker* was created by and has been used consistently by Peter Drucker since as early as 1973 to describe in general all "workers" in the Information Age organization.^{LPE}

Metadata: A term used to mean data that describes or specifies other data. The term *metadata* is used to define all of the characteristics that need to be known about data in order to build databases and applications and to support knowledge workers and information producers.^{LPE}

Mission-critical information: Is information considered fundamental for HUD to conduct business, or information frequently used by the Department, particularly financial information, key to the Department's integrity and accountability, and information used to support Annual Performance Plan reports. Program Areas, the Office of the Chief Financial Officer (OCFO), the Office of the Chief Information Officer (OCIO), the Deputy Secretary or the Secretary will categorize information as mission-critical. Mission-critical information will be managed using the TIQM[®] approach to enable HUD to achieve expected levels of information quality necessary to serve its constituents properly (also, see controllable mission-critical information).

Non-duplication: A characteristic of information quality measuring the degree to which there are no redundant occurrences of data.^{LPE}

Objectivity: The state whereby disseminated information is being presented in an accurate, clear, complete, and unbiased manner. This involves whether the information is presented within a proper context. Sometimes, in disseminating certain types of information to the public, other information must also be disseminated in order to ensure an accurate, clear, complete, and unbiased presentation. Also, the agency needs to identify the sources of the disseminated information (to the extent possible, consistent with confidentiality protections) and, in a scientific, financial, or statistical context, the supporting data and models, so that the public can assess for itself whether there may be some reason to question the objectivity of the sources. Where appropriate, data should have full, accurate, transparent documentation, and error sources affecting data quality should be identified and disclosed to information consumers.^{OMB}

Physical Information Quality Assessment:

Physical assessments compare data values to the real-world objects and events that the data represents in order to confirm that the values are accurate. This type of testing is more time and labor intensive than automated testing, but is a necessity for confirming the accuracy of data. Physical assessments are usually complementary and must be consistent with and complementary to the corresponding automated assessment. ^{LPE}

Precision: the degree to which data is known to the right level of granularity (e.g., the right number of decimal digits right of the decimal point, time to the hour or the half-hour or the minute, or the square footage of a building is known to within one square foot as opposed to the nearest 100s of feet). ^{LPE}

Primary key: The attribute(s) that are used to uniquely identify a specific occurrence of an entity, relation, or file. A primary key that consists of more than one attribute is called a *composite* (or *concatenated*) primary key. ^{LPE}

Primary key uniqueness: The prerequisite of a primary key to identify a single entity, row in a database, or occurrence in a file.

Process owner: The person responsible for the process definition and/or process execution. The process owner is the managerial information steward for the data created or updated by the process, and is accountable for process performance integrity and the quality of information produced. ^{LPE}

Quality standard: A mandated or required quality goal, reliability level, or quality model to be met and maintained. ^{LPE}

Ranges, reasonability tests: General tests applied to information to determine if the value is correct. For example:

- A test for Birth Date on a Drivers License Application might be that the resulting age of the applicant be between 16 and 120.
- A range for Patient Temperature might be 80-110 degrees, while the range for Room Temperature might be -20 to 120 degrees.

Record of origin: The first electronic file in which an occurrence of an entity type is created. ^{LPE}

Record of reference: The single, authoritative database file for a collection of fields for occurrences of an entity type. This file represents the most reliable source of operational data for these attributes or fields. In a fragmented data environment, a single occurrence may have different collections of fields whose record of reference is in different files. ^{LPE}

Referential integrity: Integrity constraints that govern the relationship of an occurrence of one entity type or file to one or more occurrences of another entity type or file, such as the relationship of a customer to the orders that customer may place. Referential integrity defines constraints for creating, updating, or deleting occurrences of either or both files. ^{LPE}

Relationship Validity: The degree to which related data conforms to the associative business rules. ^{LPE}

Repository: A database for storing information about objects of interest to the enterprise, especially those required in all phases of database and application development. A repository can contain all objects related to the building of systems including code, objects, pictures, definitions. The repository acts as a basis for documentation and code generation specifications that will be used further in the systems development life cycle. Also referred to as *design dictionary*, *encyclopedia*, *object-oriented dictionary*, and *knowledge base*. ^{LPE}

Rightness or fact completeness: The degree to which the information presented is the right kind and has the right quality to support a given process or decision. ^{LPE}

Scalability: The ability to scale to support larger or smaller volumes of data and more or less knowledge workers. The ability to increase or decrease size or capability in cost-effective increments with minimal impact on the unit cost of business and the procurement of additional services.

Surrogate source: a document, form, application, or other paper copy of the information from which the data was originally entered. Also, an electronic copy of the data generated outside the organization that is known to be accurate. ^{LPE}

Timeliness: A characteristic of information quality measuring the degree to which data is available when knowledge workers or processes require it. ^{LPE}

(HUD's) Total Information Quality Management (TIQM[®]): Techniques, methods and management principles that provide for continuous improvement to the information processes of an enterprise. A management approach used by HUD, based upon accepted industry standards and incorporating project management and total quality management principles. It replaced HUD's Common Data Element Correction Method.

Usability: the degree to which the information presentation is directly and efficiently applicable for its purpose. ^{LPE}

User: A term used by many to refer to the role of people in information technology, computer systems, or data. The term is inappropriate to describe the role of information producers and knowledge workers who perform the value work of the enterprise, and for whom information technology should enable them to transform their work and who depend on information to perform their work. With respect to information technology, applications, and data, the role of business personnel is that of information producers and knowledge workers. The term knowledge worker was created by and has been used consistently by Peter Drucker since as early as 1973 to describe in general all "workers" in the Information-Age organization. The relationship of business personnel to information systems personnel is not as "users," but as *partners* who together solve the information and work problems of the enterprise. ^{LPE}

Utility: The usefulness of the information to its intended consumers, including the public. In assessing the usefulness of information that the agency disseminates to the public, the agency needs to consider the uses of the information not only from the perspective of the agency but also from the perspective of the public. As a result, when transparency of information is relevant for assessing the information's usefulness from the public's perspective, the agency must take care to ensure that transparency has been addressed in its review of the information. ^{OMB}

Validation (of Performance Data): is the assessment of whether the data are appropriate for the performance measure. Also, validation indicates the appropriateness of ... performance measures in relation to ... goals and objectives. ^{GAO}

Validity: A characteristic of information quality measuring the degree to which the data conforms to defined business rules. Validity is not synonymous with *accuracy*, which means the values are the correct values. A value may be a valid value, but still be incorrect. For example, a customer date of first service can be a *valid* date (within the correct range) and yet not be an *accurate* date. ^{LPE}

Verification (of Performance Data): The assessment of [performance] data completeness, accuracy, and consistency, timeliness, and related quality control practices. ^{GAO}

Zero defects: A state of quality characterized by defect-free products or 6-Sigma level quality. See *6 Sigma*. ^{LPE}