DATA INTEROPERABILITY STANDARDS, 2016

NATIONAL INFORMATION TECHNOLOGY DEVELOPMENT AGENCY (NITDA)

NATIONAL INFORMATION TECHNOLOGY DEVELOPMENT AGENCY ACT 2007

DATA INTEROPERABILITY STANDARDS, 2016

ARRANGEMENT OF STANDARDS

The Government of the Federal Republic of Nigeria

NATIONAL INFORMATAION TECHNOLOGY DEVELOPMENT AGENCY (NITDA)

DATA INTEROPERABILITY STANDARDS, 2016

In exercise of the powers conferred on it by section 6 of the National Information Technology Development Agency Act of 2007, NITDA in consultation with key stakeholders hereby issues the following minimum Standard for Data Interoperability.

Powers of NITDA

[................2016]

Commencement.

WHEREAS:

Preamble

Semantic interoperability refers to the ability of information systems to exchange information on the basis of shared, pre-established and negotiated meanings of terms and expressions; and is needed in order to make other types of interoperability work (syntactic, technical, organizational, cross-cultural etc.);

Semantic interoperability entails a co-ordination of meaning; and more importantly, the acceptability of the precise meaning of exchanged information by any other application not initially developed for this purpose;

There is a key challenge associated with interoperability which is the ability to understand the context and attach a meaning to it so that users get appropriate results;

In addressing this challenge, NITDA has followed global standards on interoperability for data sharing and drawn from the experience of Interoperability Frameworks of other countries including the Estonian IT Interoperability Framework, Dutch

Interoperability Framework, Australian Information Interoperability Framework, US Data Reference Model, EU Content Interoperability Framework, UK eGovernment Metadata Standards and some other Government Semantic Interoperability Frameworks (GSIFs), ISO/IEC 19794 Information Technology Biometric Interchange Formats – Part 1 – 7, ISO/IEC19785 -1:2006, ISO/IEC 10918 1- 4 as well as standards from bodies like the OMG, IEC, W3C, ITU-T, IETF, OASIS; and

With its wealth of experience and in the exercise of its mandate, NITDA hereby issues these Standards to achieve data interoperability in the Public Sector and to facilitate data sharing and reuse.

PART ONE – PRELIMINARIES

1. These Standards may be cited as Standards for Data Interoperability. Short title

2. The various TERMS used herein, shall except where the context permit, be as defined in Appendix (1) of this Standards. Definitions

3. The persons and bodies to which these Standards shall apply include: Application

(a) Information System Developers, (e.g. architects, analysts, designers, programmers, record-keepers, project managers etc),

(b) Database Owners, Users and Integrators of operations as well as Contracting Entities, Information System Owners, System Auditors etc in both the Public and Private Sectors;

(c) Government agencies directly involved in data acquisition; solution architecture, data design & implementation.

4. These Standards are issued for the purpose of: Objectives

(a) Providing standards to improve data sharing by defining a number of important, universally understandable concepts and standards between data-sharing communities without requiring complex mediations.

(b) Ensuring minimum level of interoperability within and between government agencies.

5. The Scope covers standards for data interoperability, data standardization, naming standards, standards for data element definition. Scope

PART TWO – BACKGROUND- SEMANTIC INTEROPERABILITY

6. (1) Semantic interoperability includes- Semantic Interoperability

(a) the ability of organizations to understand exchanged data in a similar way;

(b) the ability of software systems to make adequate use of data received from other software systems;

(c) how the elements of the data structures exchanged are related to real world objects, relations and events;

(d) exchange of information about the context of data, i.e. relations, operations and functioning in general and

(e) exchange of metadata between organizations/agencies

(2) Semantic Interoperability is achieved when:

(a) Data exchange partners have a shared understanding of the meaning of shared data;

(b) Data exchanges adhere to the shared understanding; and

(c) Data is exchanged without misinterpretations.

7. Semantic Interoperability addresses three (3) major areas of incompatibilities: Major Incompatibilities

(a) Domain level incompatibilities

(b) Entity level incompatibilities

(c) Abstraction level incompatibilities

8. Domain level incompatibilities arises where semantically similar *attributes* are modeled using different descriptions and these include Naming, Data Representation and Data Scaling conflicts- Domain level incompatibility

(a) Naming conflicts or Labeling conflicts – these arise when-

(I) different concepts described by the same word or

(II) the same attribute name have different meanings (homonyms) and multiple alternative words that describe the same concepts or

(III) different attributes names referring to the same thing (synonyms) or

(IV) The uncoordinated assignment of names in a database schema[1].

(b) Representation conflicts or Data Value conflicts – these arise from representing values in different ways – using different representations for the same value type, for instance sex values represented as "M" and "F" in one system, while being represented as "male" and "female" in another;

(c) Spatial Domain Conflict – these arise when there are different legal implications for the same piece of data. For instance "blood group" information as part of personal details is allowed in one context, while it violates privacy rights in another.

9. Entity Level Incompatibilities or heterogeneities arise when semantically similar entities are modeled using different descriptions and these include measurement conflict, cofounding conflict, Schema Isomorphism conflicts and structural conflicts:

Entity level incompatibilities

(a) Measurement conflicts or Data Precision conflicts - arise from data being represented in different units or scales – differences in the units or scales of values. For instance one system represents length in "meters" and the other in represents it as "feet";

(b) Confounding conflicts may result from having different shades of meanings assigned to a single concept. For example, the price of a stock may be the "latest closing price" or "latest trade price" depending on the exchange or the time of day. Confounding conflicts could arise when actual meaning of information is dependent on the content or value of another

[1] Examples of homonyms and synonyms abound: an attribute name such as "name" is frequently used in referring to different things in different scenarios (e.g., name of a person and name of a company); on the other hand, an item such as "employee name" is frequently given a variety of attribute names such as "empname" or "employee-name", "staffname" and so on. Another example is "family name" which is variously referred to as "Lastname", "Surname" in different systems;

data;

(c) Schema Isomorphism Conflicts – these arise when same concept is described with different set of attributes;

(d) Structural conflicts - these arise when the same piece of information is modeled as a relation name, an attribute name or value in a table.

10. Abstraction level incompatibilities arise when two semantically similar entities or attributes are represented at different levels of abstraction and these include Generalization, Aggregation and Attribute Entity conflicts:

Abstraction Level incompatibilities

(a) Integrity conflicts – data considered correct in one context that violates integrity constraints in another, for instance a citizen may be permitted to have more than one address in one context and only allowed one address in another;

(b) Generalization conflicts – data held in one context may be a subset or superset of another, therefore data with mandatory attributes in one context may be invalid in another;

(c) Computational conflicts – these conflicts arise from alternative ways of computing data.

(d) Aggregation conflicts – data stored in one context is defined collectively in another (or vice versa). For instance individual student records in one system are grouped by ages in another system.

(e) Granularity conflicts – these occur when data are reported at different levels of abstraction or granularity. Consider, for instance, the following granularity conflict: in “Database A”, a student's grade is represented as a letter grade “A, B or C”; while in “Database B”, the corresponding information is represented as points in the range “0 to 100”.

11. The benefits of semantic interoperability include-

Benefits/ motivation

(a) The quality of data will improve as data from different sources can be seamlessly integrated. There will be less errors and inaccuracies upon using data and making decisions (related to misinterpretation of data as well as discrepancies arising from duplication of data);

(b) Less investment in the production (acquisition) of data will be needed since it will be easier to reuse data;

(c) Parties will have to spend less time on integrating the information systems of different organizations;

(d) Sustainability of the application or information system will increase and it will be easier to make further developments. The knowledge base related to information systems will be preserved also after key persons have left the agency or organization.

PART THREE – DATA INTEROPERABILITY

Data interoperability

12. Data interoperability includes-

(a) The ability to correctly interpret data across different systems or organizational boundaries.

(b) Facilitating a common understanding of the data meaning and usage between systems and across agencies – providing clarity in plain English or familiar business language.

Implementing Data interoperability

13. (1) Implementing data interoperability requires achieving both data integration and data exchange as well as enabling effective use of the data that becomes available.

(2) Each of these three tasks involves some type of standards and guidelines in the way data is captured and consumed between disparate systems.

(3) Data semantics irregularities are most commonly evidenced through differences in:

(a) data names;

(b) data types;

(c) data lengths; and

(d) data structures

Principles relating to data interoperability

14. Standards & Principle Statement Relating to Data Interoperability include:

(a) Use XML and XML Schema for Data Interoperability;

(b) Use XMI for exchange of all business information and information system modelling;

(c) Use XSL for data transformation;

(d) Ensure XML products are written so as to comply with the recommendations of the World Wide Web Consortium (W3C); and

(e) Where necessary base the work on the draft W3C standards but avoid the use of any product specific XML extensions that are not being considered for open standardization with the W3C.

15. Recommended Standards and specifications for Data Interoperability and Transformation are as stipulated in Appendix 2; while Standards for biometric interchange is as stipulated in Appendix 3. Appendix 2 and 3

PART FOUR – STANDARD FOR DATA ELEMENT DEFINITIONS

16. (1) Data element definitions shall be the primary vehicle for conveying the meaning of data. Guidelines for Data Definition

(2) Data element definitions shall, in addition to being precise and unambiguous, ensure a seamless exchange of data.

(3) When two or more parties exchange data, the meaning of that data shall be explicit in an agreement.

(4) Definitions shall be written to facilitate understanding by any user and by recipients of shared data.

17. Some of the benefits of standardizing data definition include: Benefits of standardizing data definition

(a) Standardize structure and contents of metadata registries;

(b) Make metadata collections accessible, searchable by semantic content;

(c) Support understanding and reuse of data standards;

(d) Promote use of standards for greater interoperability.

18. (1) Requirements: Guidelines for standardizing

(a) A data definition shall – data definition

- (I) be stated in the singular;
- (II) state what the concept is, not only what it is not;
- (III) be stated as a descriptive phrase or sentence(s);
- (IV) contain only commonly understood abbreviations (i.e. understood also outside the given domain);
- (V) be expressed without embedding definitions of other data or underlying concepts.

(2) Recommendations:

(b) A data definitions should-

- (I) state the essential meaning of the concept;
- (II) be precise and unambiguous;
- (III) be concise;
- (IV) be able to stand alone;
- (V) be expressed without embedding rationale, functional usage, or procedural information;
- (VI) avoid circular reasoning;
- VII. use the same terminology and consistent logical structure for related definitions;

19. The explanations for requirements and recommendations in Article 18 above are contained in Appendix 4 below. Appendix 4

PART FIVE – DATA NAMING STANDARDS

20. (1) A naming standard is a collection of rules, which, when applied to data elements, results in a set of data objects named in a logical and standardized way. Naming standards

(2) These names convey some information about the named objects; an element name, for example, indicates the set of possible valid values (its data domain), and its usage.

(3) Having naming standards help to achieve efficient use and reuse of data through better understanding of what data is available across government agencies.

(4) Standardized names will enhance interoperability within and amongst government agencies.

21. Data Element Naming Standards will help:

Usefulness of Data Element

(a) foster a common understanding of data across the Ministries, Departments and Agencies (MDAs) of Government,

(b) promote data sharing across systems and among data users by providing uniformity in how data is defined and named;

(c) facilitate the discovery of data which can be shared across MDAs; and

(d) provide uniformity which enables the control and management of data across MDAs.

22. Standards Used includes- Kinds of standards

(a) Data element definitions – English phrase (or phrases) which describe a data element. It consists of Prime Words, Optional Qualifiers, and Class Words.

(b) Logical Data Element Names - uniquely identifies a data element within an agency's data resource. These are derived from the data element definitions.

(c) Physical Data Element Names – required by the operating software (database) to uniquely identify the data and manage it. Developed from the logical data element names

23. Components of Data Element Names includes- Components of Data Elements

(a) Prime Word – It is used to describe the subject area of data and represents the business portion of the name. It gives the general grouping or context. It might be the entity name, e.g. Person

(b) Class Word – It is used to describe the major classifications or types of data associated with data elements. For each data element defined and named, one Class Word becomes part of the data element definition and associated logical and physical names. It identifies the general function or purpose of the data, e.g. DATE is used to give a point in time, and NAME is used to identify or classify data. It is usually the last word in the name. There are four categories of Class Words:
   (I) Chronology – indicate a point in time, span of time.
   (II) Measurement – have dimension, capacity, amount or duration.
   (III) Identification – distinguish a person, place or thing.
   (IV) Text – identify more free form or narrative data

(c) Qualifier – It is used to further define and distinguish the prime and class words. Qualifiers should be used as necessary to fully identify the data element, and should be listed in an order that has logical English meaning. It is used to expand the Class Word when that does not say enough about the type of data. It is called a Qualifier because it qualifies the Class Word. It may precede or follow the Class Word.

(d) Modifiers. They are a number of words after the Prime Word and before the Class Word to provide the rest of the meaning. They are not necessary or compulsory in naming a data element. Therefore, this Standard does not include its usage.

General Rules of Data Element Definition

24. General Rules of Data Element Definitions are:

(1) Data Element Definitions which shall:
   (a) include the classification of the data;
   (b) identify the prime word;
   (c) be clear and concise;
   (d) define what data is, not what it is not; and
   (e) should note if this data element is derived from others

(2) Logical Data Element Names which shall:
   (a) be of the format prime word (required) – qualifier (optional) – class (required);
   (b) not use abbreviations unless required to meet a name length limitation in the tool;
   (c) nouns are singular and verbs are in present tense.

(3) Physical Data Element Names which shall:

   (a) be of the format prime word (required) – qualifier (optional) – class (required),
   (b) use approved two or three character abbreviations – (approved by the agency/community)
   (c) contain syntax mandated by database or operating software

(4) Approved abbreviations which shall:

   (a) make sense to users;
   (b) as a general rule, should only be created for words with five or more characters in length;
   (c) be associated with only one abbreviated word or forms of that word (i.e. ALLOC for allocate, allocated and allocation),
   (d) not be an English word themselves (for example: do not use CLASS for classification);
   (e) not be confused with other standards abbreviations (e.g. TM is Trademark, not for tax money);
   (f) start with the letter of the word being abbreviated (e.g. use

EXCPT not XCPT for exception);

(g) if one abbreviation is used within a name, be abbreviated if possible within the above constraints (e.g. UNIV_ID_NBR is better than UNIVERSITY_ID_NUMBER).

Tabulated illustration of Logical and Physical data element

25. Logical data elements names and physical data element names in 24(2) and (3) above are illustrated in tables A and B respectively, in Appendix 5.

Class words

26. (1) Class words will be used as suffixes to column names in a consistent manner to simplify the task of understanding the underlying data type of a data column.

(2) All column names will be Pascal cased (First letter of each word will be in upper case followed by Lower case letters to complete each word).

(3) There shall be limit to the number of data Class words to prohibit the Class words from becoming cryptic in nature and creating undue complexity in data naming standards.

(4) To avoid compromising or using class word in an inappropriate manner, the list of acceptable Class words shall be populated with a complete collection.

(5) The table in Appendix 6 contains the list of draft Class words that will be used in naming attributes to ensure uniformity across the board.

PART SIX- DATA STANDARDIZATION

Data Standardization

27. (1) Data Standardization is the first step to ensure that data is able to be shared across agencies of government. This establishes trustworthy data for use by other applications across MDAs.

(2) Data standardization may, where practicable, be performed during data entry; or where not practicable, a comprehensive back-end process is necessary to eliminate any inconsistencies in the data.

(3) Data standards ensure that the terms exchanged between IT systems and their components are unambiguously defined. Such standards can e.g. define whether an address field contains street name as well as number, or whether the number is to be located in a separate field. Such aspects are usually

not specified in the technical standards, but they are important in order to ensure an efficient and problem free exchange of information with IT systems, that fully understand each other.

(4) Standardizing data helps make the source data internally consistent; that is, each data type has the same kind of content and format. Common examples of data elements that can be identified are name, address, city, state, and post code.

(5) A valuable approach is the promotion of common syntax and semantics for the few things we can all agree on.

(6) At the most basic level, data standards are about the standardization of data elements:
   (a) defining what to collect,
   (b) deciding how to represent what is collected (by designating data types or terminologies), and
   (c) determining how to encode the data for transmission.

28. The purposes of data standardization are to promote the following: Purpose of data standardization

   (a) Standard description of data;
   (b) Common understanding of data across organizational elements and between organizations;
   (c) Re-use and standardization of data over time, space, and applications;
   (d) Harmonization and standardization of data within an organization/agency and across organizations/agencies;
   (e) Management of the components of data;
   (f) Re-use of the components of data.

29. Poor integration of data from different sources occur often where there are different rubrics for the same fields, such as “address”, “street address” or “local address”; or “postal code” and “zip code” Poor integration

30. Name standardization includes- Name standardization

   (a) the process in which a given name is broken down into its constituent parts.

   (b) the ability to recognize the many different name components in many different formats and patterns and then the ability to extract the corresponding strings and rearrange them into a format suitable for subsequent interoperability and integration processes.

31. A standard name format should consist of the following components: Content of a standard name format

(a) Title or Prefix
(b) First Name
(c) Middle name
(d) Last Name or Surname
(e) Suffix

32. For prefix or title to a full name on all official transactions/documentations, the following applies: Prefix or title

(a) All prefixes or titles are entered into the Prefix/Title field;
(b) Minimum field length: 2 characters;
(c) Maximum field length: 20 characters;
(d) This field may be left blank, but it is preferred to use prefixes for records of individuals;
(e) All information is to be entered using uppercase/lowercase letters. Standard capitalization rules should be used. Punctuation is to be used in prefixes. Standard abbreviations are preferred, but full titles may sometimes be used in special circumstances;
(f) The preference is to use the abbreviated form of the prefix;
(g) Do not enter prefixes in any other fields;
(h) Commas, slashes and the pound sign (#) are never used in this field;
(i) If no prefix is provided, the default is Mr or Ms.

33. For First Name which is any given/chosen name by which a person is known or designated on all official transactions/documentations; the following applies- First name

(a) All information is to be entered using uppercase and lowercase letters. Never use all uppercase or all lowercase letters;
(b) Minimum field length: 2 characters;
(c) Maximum field length: 60 characters;
(d) Hyphens or apostrophes may be used;
(e) The period (.) is allowed;
(f) In those cases where a single character or initial is designated as the first name and followed by a middle name, place the single character in the first name field and the middle name in the middle name field;
(g) Spaces are permitted for double names (e.g., Mary Ann);
(h) Single Character First Names should be entered with no period;

Example: B. Steven Emeka
enter B in First Name,
Steven in Middle Name field and
Emeka in the Last name field

(i) Do NOT use titles, prefixes and suffixes in the first name field;
(j) Commas, slashes and the pound sign (#) are never used in this field.

34. For middle name or initial by which a person is known or designated on all official transactions/documentations and the following applies: Middle Name

(a) All information is to be entered using uppercase/lowercase letters. Never use all uppercase or lowercase letters. If no middle name or middle initial exists, leave the field blank;
(b) Minimum Field Length: 2 characters;
(c) Maximum field length: 60 characters;
(d) Hyphens may be used to separate double names;
(e) Apostrophes may be used. Example: O'Deji;
(f) Spaces are permitted between multiple names. Example: Mary Ann;
(g) The period is allowed;
(h) Enter the entire middle name if available for identification purposes;
(i) Do NOT use title or prefixes and suffixes in the middle name field;
(j) Single Character Middle Names should be entered without a period;
(k) Commas, slashes and the pound sign (#) should not be used in this field;

35. For Last Name or Surname which refers to Non-chosen/inherited/married name by which a person is known or designated on all official transactions/Documentations; and the following applies: Last name

(a) Minimum Field Length: 2 characters;
(b) Maximum Field Length: 60 characters;
(c) Hyphens may be used to separate double last names as indicated e.g. Kehinde-Phillips, Mobolaji-Johnson etc. If there are two last names that are not hyphenated (as specified by the individual), the two names should both be entered in the “Last name” field. Spaces are used for two last names that are not hyphenated e.g. Salisu Ibrahim
(d) Example: Mary-Ann Salisu Ibrahim ---

'Salisu Ibrahim' is entered in the last name field. 'Mary-Ann' is entered in the first name field. No middle name is entered;

(e) Commas, slashes and the pound sign (#) should not be used in the last name field;

(f) The period is used in names that are written as abbreviations e.g. St. Brown.

36. For Suffix which refers to a full name on all official transactions/documentations and the following applies

Suffix

(a) Suffixes are entered only into the Suffix field.

(b) Minimum Field Length: 2 characters;

(c) Maximum Field Length: 20 characters;

(d) If there is no suffix, the field should be left blank;

(e) All information is to be entered using uppercase/lowercase letters. Standard capitalization rules should be used. Punctuation is used in some suffixes. Acronyms used to indicate degrees, medical certifications or to indicate religious orders generally don't have periods;

(f) Both a prefix and suffix are used when there is an inherited suffix (i.e., Jr., III);

(g) A suffix indicating academic or medical degree is not used when the prefix is entered.
Example: Do not enter Dr. Baba Abubakar, PhD

(h) Slashes and the pound sign (#) are never used in this field.

37. The Standards provided in Articles 32 to 36 above are applicable only to new systems, applications and databases. For existing data, truncation rules will be defined in line with the requirements and the rules of the operating DBs to guide exchange of data.

Application to new systems

38. Address standardization is-

Address standardization

(a) Critical for data integration and interoperability across government agencies.

(b) An important aspect with regards to citizen service processes, entity identity resolution processes, fraud detection, along with a number of other processes.

(c) To separate the address elements into different fields rather than grouping all this data into a single field.

39. A standard Address format should have the following components:

Components of address format

(a) Address Type 2
(b) Address 1 (Street Address)
(c) Address 2 (Street Address Cont'd)
(d) City
(e) LGA
(f) State
(g) Postcode

40. For Address (Street) Fields, the following applies:

Address

(a) All information should be typed in upper/lower case format, i.e., not all caps or all lower;
(b) Special Characters and the period (.) cannot be used when entering the number portion of an address. The following is an example:
Incorrect: 33. Oran Street. Use instead 33 Oran Street.

41. For City/LGA/State/Post code, the following applies:

City/LGA/State/Post code fields

(a) These fields should be pre-populated as a matter of rule so that consistency and accuracy can be maintained. Where postcode is not available, the City, LGA and State fields should be pre-populated.

(b) Defining standards abbreviation forms for States and LGAs
(c) Adoption by NIPOST, 2-3 letter abbreviation, for States and LGAs in Nigeria to ensure uniformity.

42. On date display format, the international standard date notation is YYYY-MM-DD. Where "YYYY" is the year in the usual Gregorian calendar, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31. For example, the fourth day of February in the year 1995 is written in the standard notation as: "1995-02-04"

ISO date display format

43. Other commonly used notations are "2/4/95", "4/2/95", "95/2/4", "4.2.1995", "04-FEB-1995", "4-February-1995", and many more. The first two examples are dangerous and confusing, because both are used quite often in the U.S. and in the UK as well as here in Nigeria and both cannot be easily distinguished; it is unclear whether 2/4/95 means 1995-04-02 or 1995-02-04.

Commonly used date display format

44. Advantages of the ISO 8601 standard date notation compared to other commonly used variants:

Advantages of ISO over others

(a) easily readable and writeable by software (no 'JAN', 'FEB', ... table necessary);

(b) easily comparable and sortable with a trivial string comparison language independent;
(c) cannot be confused with other popular date notations;
(d) consistency with the common 24h time notation system, where the larger units (hours) are also written in front of the smaller ones (minutes and seconds);
(e) strings containing a date followed by a time are also easily comparable and sortable (e.g. “1995-02-04 22:45:00”);
(f) the notation is short and has constant length, which makes both keyboard data entry and table layout easier.
PROVIDED THAT-
(I) a 4-digit year representation avoids overflow problems after 2099-12-31;
(II) ISO 8601 is only specifying numeric notations and does not cover dates and times where words are used in the representation. It is not intended as a replacement for language-dependent worded date notations such as “24. December 2001” or “February 4, 1995” ISO 8601 should however be used to replace notations such as “2/4/95” and “9.30 p.m.”.

45. These formats may be used in view of the widely used formats in Nigeria: Optional formats

(a) Date fields will have a default format for data entry of DD-MM-YYYY or as agreed by Committee.

(b) Date fields will have a format of DD/MM/YYYY for display only fields, or DD/MM/YY as agreed, PROVIDED that ISO date notations are: YYYY-MM-DD and YY-MM-DD.

PART SEVEN: ENFORCEMENT AND AMENDMENT

46. Any breach of these standards shall be construed as a breach of the provisions of the National Information Technology Development Agency Act of 2007 and the guidelines issued under that Act. Breach

47. The enforcement of these Regulations shall be by NITDA. Enforcement

48. NITDA shall amend or review these Standards periodically or as the need arises in consultation with stakeholders. In reviewing them, the Agency shall be guided, among other considerations, by global trends and practices in Information Technology and the developmental aspirations of Nigeria. Amendment

Appendix 1: Terms and Definitions [Article 2]

TERM	DEFINITION
Data Element	The smallest unit of a data structure, e.g. column of a table; a dataset separated by XML tags. Also called “data unit”.
Data Element	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
Data Element Name	A single or multi-word designation used as the primary means of identification of data elements for humans.
Data Structure	A physical or logical relationship among units of data and the data themselves. [EVS-ISO 2382 15.03.01]
Database	Structured collection of data. A collection of data organised according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas. [EVS-ISO 2382 01.08.05] A data structure for accepting, storing and providing on demand data for multiple independent users. [EVS-ISO 2382 04.07.15]
Definition	A representation of a concept by a descriptive statement which serves to differentiate it from related concepts.
Domain	A field of special knowledge. [EVS-ISO 1087-1:2002 and EVS-ISO 5127 1.1.1.07]
Domain Expert	A specialist in a field of special knowledge.
Domain Glossary	An alphabetical list of terms with definitions. A glossary also commonly contains an explanation of words, concepts or terms that are usually listed in alphabetical order. A domain glossary is an alphabetical list of terms and their definitions in a particular field or subject matter.
End User	A person/role using an operation and usually also the semantic description of this operation.
ERD	Entity Relationship Diagram. A model of relationship between database and tables.

TERM	DEFINITION
Data Element	The smallest unit of a data structure, e.g. column of a table; a dataset separated by XML tags. Also called “data unit”.
Data Element	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
Glossary	An alphabetical or systematic list of all words in a language or words of certain category, including explanations in that language or translations into one or more languages. [EVS-ISO 5127 2.2.1.16]
GML	Geospatial Markup Language
Human-readable description	A description in a form readable to both developers and domain experts, disclosing the meaning of data (semantics).
IDABC	**I**nteroperable **D**elivery of European eGovernment Services to public **A**dministrations, **B**usinesses and **C**itizens.
IDEF	Integration Definition. IDEF is a family of modeling languages in the field of systems and software engineering.
IETF	Internet Engineering Task Force
ITU-T	International Telecommunication Union for Telecommunication Standards
Machine-readable description	A description in a form readable to software systems and conveying the data without changing their meaning (semantics).
Name	A term used for referring to a domain concept (human-readable)..
NITDA	National Information Technology Development Agency
OASIS XCBF	OASIS XML Common Biometric Format
Object Class	A set of objects classified or grouped on the basis of a common property.
Object Status	A status experienced during the object’s life cycle.
OMG	Object Management Group. A consortium which, among other things, develops the UML standard. See http://www.omg.org/.
Ontology	For the purposes of this Guide, a way of describing a domain both in machine- and human-readable form. Includes a

TERM	DEFINITION
Data Element	The smallest unit of a data structure, e.g. column of a table; a dataset separated by XML tags. Also called “data unit”.
Data Element	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
	glossary of concepts (terms describing and presenting concepts) and relations (e.g. relation of succession) between them used in the particular domain. Ontology is a set of concepts – such as things, events and relations that are specified in some way in order to create an agreed-upon vocabulary for exchanging information. Therefore, ontology describes semantics of exchanged information. Ontology comprises two types of elements: Class – an abstraction mechanism used for classification; and Property - assertion of facts about classes.
Ontology Description Language	For the purposes of these instructions, a way of describing a domain both in machine- and human-readable form. Includes a glossary of concepts (terms describing and presenting concepts) and relations (e.g. relation of succession) between them used in the particular domain.
OWL	Ontology Web Language. A markup language for presenting ontologies in the WWW. See http://www.w3.org/TR/owl-features/. More powerful while also more complicated than RDFS.
Process	A certain operation or sequence of operations of an organisation necessary for fulfilling a task or providing a service. Also called “operational process”, “business process”.
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema. A markup language for presenting ontologies in the WWW. See also “OWL” and http://www.w3.org/TR/rdf-schema/.
Recommendation	A requirement that is not mandatory
RSS	Really Simple Syndication

TERM	DEFINITION
Data Element	The smallest unit of a data structure, e.g. column of a table; a dataset separated by XML tags. Also called "data unit".
Data Element	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
Rule	A mandatory requirement to be followed.
SAML	Security Assertion Markup Language
SA-WSDL	Semantic Annotations for WSDL and XML Schema. A W3C recommendation for the semantic description of web services and data structures. http://www.w3.org/TR/sawsdl/ had the status of W3C Candidate Recommendation as at 25 February 2007.
Semantic Web Standards	The semantic web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.
UML	Unified Modelling Language. A language for modelling and visual presentation of software systems and business domains. See also http://www.uml.org/.
Value Domain	A set of permissible values. **Context:** in the context of ISO/IEC 11179, a domain is a set of possible data values of an attribute. A data value is an element of a value domain.
W3C	World Wide Web Consortium. The main organisation developing web standards. See http://www.w3.org/. W3C standards include XML, RDF and OWL.
WSDL	Web Services Description Language. An XML-based language for describing web services. WSDL 2.0 (http://www.w3.org/TR/wsdl20-primer/) had the status of W3C Candidate Recommendation as at January 2007. See also http://www.w3schools.com/wsdl/default.asp.
XMI	XML Metadata Interchange. A standard allowing, among other things, exchange of UML models between parties in a standard way. XMI is one of OMG's standards.

TERM	DEFINITION
Data Element	The smallest unit of a data structure, e.g. column of a table; a dataset separated by XML tags. Also called “data unit”.
Data Element	A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
	See also http://www.omg.org/technology/documents/formal/xmi.htm .
XSL	Extensible Stylesheet Language

Appendix 2- The Recommended Standards & Specifications for Data Interoperability & Transformation [article 15]

Table 1.0

Component	Standard	Standard Body
Metadata/MetaLanguage	XML (Extensible Markup Language)	W3C
XML Metadata Definition	XML-Schema RelaxNG	W3C OASIS/ISO
XML Data Transformation	XSL (Extensible Stylesheet Language)	W3C
XML Data Query	Xpath	W3C
XML Signature	XML DSIG	W3C
XML Security Mark-up	SAML v2.0 (Security Assertion Markup Language)	OASIS
Public Key Infrastructure	X509v3 (SSL and TSL)	ITU-T
Model exchange	XMI (XML Metadata Interchange)	OMG

Appendix 3 - Standards for Biometric Interchange [article 15]

Table 2.0

Component	Standard	Standard Body
Secure XML Enconding for exchanging Biometric data	OASIS XCBF 1.1 Specification Secure XML encodings for the patron formats specified in CBEFF (Common Biometric Exchange File Format (NISTRI 6529).	OASIS
Data Element Specification	ISO/IEC 19785-1:2006 Information Technology – Common Biometric Exchange Formats Framework – Part 1: Data Element Specification	ISO/IEC
Interchange Format Framework	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 1: Framework	ISO/IEC
Interchange Format Framework for finger minutiae data	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 2: Finger minutiae data	ISO/IEC
Interchange Format Framework for finger pattern spectral	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 3: Finger pattern spectral	ISO/IEC
Interchange Format Framework for finger image data	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 4: Finger image data	ISO/IEC
Interchange Format Framework for Face image data	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 5: Face image data	ISO/IEC
Interchange Format Framework for Signature/sign behaviour data	ISO/IEC 19794: Information Technology Biometric data interchange formats – Part 7: Signature/sign behaviour data	ISO/IEC
Graphical/still image information exchange specification	ISO/IEC 10918-1:1994 Information Technology – Digital compression and coding of continuous - tone still images: Requirements and Guidelines ISO/IEC 10918-1: 1994/CD Cor 1 ISO/IEC 10918-2: 1995 Information Technology - Digital compression and coding of continuous-tone still images: Compliance testing ISO/IEC 10918-3: 1997 Information Technology - Digital compression and coding of continuous-tone still images: Extensions ISO/IEC 10918-3: 1997/Amd 1: 1999 ISO/IEC 10918-4: 1999 Information Technology - Digital compression and coding	ISO/IEC

Component	Standard	Standard Body
	of continuous-tone still images: Registration of JPEG profiles, SPIFF profiles, SPIFF tags, SPIFF compression types and Registration Authorities (REGAUT)	

NOTE: The standards provided in Section 4.2 above apply to exchange of biometric data only and not how the data is captured. However, data sharing and exchange will be governed by the provisions of Data Protection Act or any relevant legislation.

Appendix 4 - Explanations for Requirements and Recommendations [Article 19]

1.1.1 Requirements

To facilitate understanding of the requirements for construction of well-formed data definitions, explanations and examples are provided below. Each requirement is followed by a short explanation of its meaning. Examples are given to support the explanations. In all cases, a good example is provided to exemplify the explanation. When deemed beneficial, a poor, but commonly used example is given to show how a definition should NOT be constructed.
Note that the examples below are definitions for data elements and these definitions are illustrative.

1.1.1.1 Requirement 1

A data definition shall: be stated in the singular

EXPLANATION - The concept expressed by the data definition shall be expressed in the singular. (An exception is made if the concept itself is plural.)

EXAMPLE - “Article Number”

Good definition: A reference number that identifies an article.
Poor definition: Reference number identifying articles.

REASON - The poor definition uses the plural word “articles,” which is ambiguous, since it could imply that an “article number” refers to more than one article.

1.1.1.2 Requirement 2

A data definition shall: state what the concept is, not only what it is not

EXPLANATION - When constructing definitions, the concept cannot be defined exclusively by stating what the concept is not.

EXAMPLE - “Freight Cost Amount”

Good definition: Cost amount incurred by a shipper in moving goods from one place to another.
Poor definition: Costs which are not related to packing, documentation, loading, unloading, and insurance.

REASON - The poor definition does not specify what is included in the meaning of the data.

1.1.1.3 Requirement 3

A data definition shall: be stated as a descriptive phrase or sentence(s) (in most languages)

EXPLANATION - A phrase is necessary (in most languages) to form a precise definition that includes the essential characteristics of the concept. Simply stating one or more synonym(s) is insufficient. Simply restating the words of the name in a different order is insufficient. If more than a descriptive phrase is needed, use complete, grammatically correct sentences.

EXAMPLE - “Agent Name”

Good definition: Name of party authorized to act on behalf of another party.
Poor definition: Representative.

REASON - “Representative” is a near-synonym of the data element name, which is not adequate for a definition.

1.1.1.4 Requirement 4

A data definition shall: contain only commonly understood abbreviations

EXPLANATION - Understanding the meaning of an abbreviation, including acronyms and initialisms, is usually confined to a certain environment. In other environments the same abbreviation can cause misinterpretation or confusion. Therefore, to avoid ambiguity, full words, not abbreviations, shall be used in the definition.
Exceptions to this requirement may be made if an abbreviation is commonly understood such as “i.e.” and “e.g.” or if an abbreviation is more readily understood than the full form of a complex term and has been adopted as a term in its own right such as “radar” standing for “radio detecting and ranging.”
All acronyms must be expanded on the first occurrence.

EXAMPLE 1 - “Tide Height”

Good definition: The vertical distance from mean sea level (MSL) to a specific tide level.
Poor definition: The vertical distance from MSL to a specific tide level.

REASON - The poor definition is unclear because the acronym, MSL, is not commonly understood and some users may need to refer to other sources to determine what it represents. Without the full word, finding the term in a glossary may be difficult or impossible.

EXAMPLE 2 - “Unit of Density Measurement”

Good definition: The unit employed in measuring the concentration of matter in terms of mass per unit (m.p.u.) volume (e.g., pound per cubic foot; kilogram per cubic meter).
Poor definition: The unit employed in measuring the concentration of matter in terms of m.p.u. volume (e.g., pound per cubic foot; kilogram per cubic meter).
REASON - m.p.u. is not a common abbreviation, and its meaning may not be understood by some users. The abbreviation should be expanded to full words.

1.1.1.5 Requirement 5

A data definition shall: be expressed without embedding definitions of other data or underlying concepts

EXPLANATION - As shown in the following example, the definition of a second data element or related concept should not appear in the definition proper of the primary data element. Definitions of terms should be provided in an associated glossary. If the second definition is necessary, it may be attached by a note at the end of the primary definition's main text or as a separate entry in the dictionary. Related definitions can be accessed through relational attributes (e.g., cross-reference).

EXAMPLE 1- “Sample Type Code”

Good definition: A code identifying the kind of sample.
Poor definition: A code identifying the kind of sample collected. A sample is a small specimen taken for testing. It can be either an actual sample for testing, or a quality control surrogate sample. A quality control sample is a surrogate sample taken to verify results of actual samples.

REASON - The poor definition contains two extraneous definitions embedded in it. They are definitions of “sample” and of “quality control sample.”

EXAMPLE 2 - "Issuing Bank Documentary Credit Number"

Good definition: Reference number assigned by issuing bank to a documentary credit.
Poor definition: Reference number assigned by issuing bank to a documentary credit. A documentary credit is a document in which a bank states that it has issued a documentary credit under which the beneficiary is to obtain payment, acceptance, or negotiation on compliance with certain terms and conditions and against presentation of stipulated documents and such drafts as may be specified.

REASON - The poor definition contains a concept definition, which should be included in a glossary.

1.1.2 Recommendations

1.1.2.1 Recommendation 1

A data definition should: state the essential meaning of the concept

EXPLANATION - All primary characteristics of the concept represented should appear in the definition at the relevant level of specificity for the context. The inclusion of non-essential characteristics should be avoided. The level of detail necessary is dependent upon the needs of the system user and environment.

EXAMPLE 1 - “Consignment Loading Sequence Number” (Intended context: any form of transportation)
Good definition: A number indicating the sequence in which consignments are loaded in a means of transport or piece of transport equipment.
Poor definition: A number indicating the sequence in which consignments are loaded in a truck.

REASON - In the intended context, consignments can be transported by various transportation modes, e.g., trucks, vessels or freight trains. Consignments are not limited to trucks for transport.

EXAMPLE 2 - “Invoice Amount”

Good definition: Total sum charged on an invoice.
Poor definition: The total sum of all chargeable items mentioned on an invoice, taking into account deductions on the one hand, such as allowances and discounts, and additions on the other hand, such as charges for insurance, transport, handling, etc.

REASON - The poor definition includes extraneous material.

1.1.2.2 Recommendation 2

A data definition should: be precise and unambiguous

EXPLANATION -The exact meaning and interpretation of the defined concept should be apparent from the definition. A definition should be clear enough to allow only one possible interpretation.

EXAMPLE - “Shipment Receipt Date”

Good definition: Date on which a shipment is received by the receiving party.
Poor definition: Date on which a specific shipment is delivered.

REASON - The poor definition does not specify what determines a “delivery.” “Delivery” could be understood as either the act of unloading a product at the intended destination or the point at which the intended customer actually obtains the product. It is possible that the intended customer never receives the product that has been unloaded at his site or the customer may receive the product days after it was unloaded at the site.

1.1.2.3 Recommendation 3

A data definition should: be concise

EXPLANATION - The definition should be brief and comprehensive. Extraneous qualifying phrases such as “for the purpose of this metadata registry,” “terms to be described,” shall be avoided.

EXAMPLE - “Character Set Name”

Good definition: The name given to the set of phonetic or ideographic symbols in which data is encoded.
Poor definition: The name given to the set of phonetic or ideographic symbols in which data is encoded, for the purpose of this metadata registry, or, as used elsewhere, the capability of systems hardware and software to process data encoded in one or more scripts.

REASON - In the poor definition, all the phrases after “...data is encoded” are extraneous qualifying phrases.

1.1.2.4 Recommendation 4

A data definition should: be able to stand alone

EXPLANATION - The meaning of the concept should be apparent from the definition. Additional explanations or references should not be necessary for understanding the meaning of the definition.

EXAMPLE - “School Location City Name”

Good definition: Name of the city where a school is situated.
Poor definition: See “school site”.

REASON - The poor definition does not stand alone, it requires the aid of a second definition (school site) to understand the meaning of the first.

1.1.2.5 Recommendation 5

A data definition should: be expressed without embedding rationale, functional usage, domain information, or procedural information

EXPLANATION - Although they are often necessary, such statements do not belong in the definition proper because they contain information extraneous to the definition. If deemed useful, such expressions may be placed in other metadata attributes (see ISO/IEC 11179-3). It is, however, permissible to add examples after the definition.

1) The rationale for a given definition should not be included as part of the definition (e.g. if a data element uses miles instead of kilometers, the reason should not be indicated in the definition).
2) Functional usage such as: “this data element should not be used for ...” should not be included in the definition proper.
3) Remarks about procedural aspects. For example, “This data element is used in conjunction with data element 'xxx'”, should not appear in the definition; instead use “Related data reference” and “Type of relationship” as specified in ISO/IEC 11179-3.

EXAMPLE - “Data Field Label”

Good definition: Identification of a field in an index, thesaurus, query, database, etc.
Poor definition: Identification of a field in an index, thesaurus, query, database, etc., which is provided for units of information such as abstracts, columns within tables.

REASON - The poor definition contains remarks about functional usage. This information starting with “which is provided for...” must be excluded from the definition and placed in another attribute, if it is necessary information.

1.1.2.6 Recommendation 6

A data definition should: avoid circular reasoning

EXPLANATION - Two definitions shall not be defined in terms of each other. A definition should not use another concept's definition as its definition. This results in a situation where a concept is defined with the aid of another concept that is, in turn, defined with the aid of the given concept.

EXAMPLE - two data elements with poor definitions:

1) Employee ID Number - Number assigned to an employee.
2) Employee - Person corresponding to the employee ID number.

REASON - Each definition refers to the other for its meaning. The meaning is not given in either definition.

1.1.2.7 Recommendation 7

A data definition should: use the same terminology and consistent logical structure for related definitions

EXPLANATION - A common terminology and syntax should be used for similar or associated definitions.

EXAMPLE - The following example illustrates this idea. Both definitions pertain to related concepts and therefore have the same logical structure and similar terminology.

1) "Goods Dispatch Date" - Date on which goods were dispatched by a given party.
2) "Goods Receipt Date" - Date on which goods were received by a given party.

REASON - Using the same terminology and syntax facilitates understanding. Otherwise, users wonder whether some difference is implied by use of synonymous terms and variable syntax.

1.1.2.8 Recommendation 8

A data definition should: be appropriate for the type of metadata item being defined

EXPLANATION – Different types of metadata item in a metadata registry (e.g. data element concept, data element, conceptual domain, value domain) each play a different role and this should be reflected in the definitions.

EXAMPLE –

Data element concept: "Job Grade Maximum Salary Amount"
Definition: The maximum salary permitted for the associated job grade.
Note: The data element concept makes no reference to a specific value domain.
Conceptual Domain: "Monetary amount"
Definition: An amount that may be expressed in a unit of currency.
Note: The definition refers to a "dimensionality" of currency, but not to a specific currency.

Data element 1": "European Job Grade Maximum Salary Amount"
Definition: The maximum salary permitted for the associated job grade expressed in Euros.

Data element 2": "U.S. Job Grade Maximum Salary Amount"
Definition: The maximum salary permitted for the associated job grade expressed in US dollars.

Data element 3": "Nigerian Job Grade Maximum Salary Amount"
Definition: The maximum salary permitted for the associated job grade expressed in Naira.

Note: Data element definitions may refer to explicit values domains, since this may be all that distinguishes two data elements.

Appendix 5 [article 25]

Table A- logical data element names

Prime Word	Qualifier (or Secondary Prime Word)	Class Word	Logical Name
Account		Balance	Account Balance
Employee	Salary	Amount	Employee Salary Amount
Student	Last	Name	Student Last Name

Table B- physical data element

Logical Name	DB 1	DB 2
Account Balance	ACCT_BAL	ACCT-BAL
Employee Salary Amount	EMP_SAL_AMT	EMP-SAL-AMT
Student Last Name	STU_LST_NAME	STU-LST-NAME

Appendix 6 [article 26(5)]

Note that the Domain column fields identified as “Text” can be any domain of text data, char, varchar, nvarchar or other. Also numeric is any number domain as int, double and other.

Table 5.0

Suffix Classwords	Description	Domain	Total Length	Decimal Places
Address	Descriptive text used to denote a place where a person or organization may be communicated with (i.e. PO Box), or a physical location (i.e. Street Address).	Text, nvarchar	60	
Amount	Most values of the number data type expressed as two decimal places which are meant to define currency. For special accounting purposes four digits to the right of the decimal point are acceptable.	Number	15	4
BLOB	A subset of Blobs that are non-imaged data. This includes data that is compressed, zipped, or otherwise encrypted	Binary	Variable	
Code	A numeric or character value that identifies classifications or categories of a member of a set of like values. A code does not include the description of the code value rather a simple abbreviation that stands for that description.	Number or Text as is required	Integer or variable not exceeding 9 bytes	
Constant	Data which does not change value over time or in different circumstances or uses.	Text or Number	Numeric or Variable not exceeding 8 bytes	
Date	A unit of time expressed in months, days, and years, used to designate a specific 24-hour period	Date	8	
Datetime	A specific instance of time that includes date and time components	Dependent on the DBMS used		
Decimal	A numeric representation of data that is not normally considered a quantity and represented in float decimal or numeric notation with or without significant digits to the right of the decimal point	Number preferred Or Float if no other options is available	Variable	Variable
Description	Data having undefined, freeform, unstructured, or unformatted content and is not an Address or Name	Text, nvarchar	Variable	

Suffix Classwords	Description	Domain	Total Length	Decimal Places
Flag	A bit or series of bits with two stable states. A binary condition permitting only two values (i.e. True/False, Yes/No, Pass/Fail)	Char(1) or bit	1	
Hash	A resulting hash from Secure Hash Algorithm-256 (SHA256) or higher when available	Text, nvarchar	Variable	
ID	Either a numeric value that implies sequence; or a computer generated serial identifier used to generate primary keys in a database to maintain referential integrity	Number	9	
Image	A subset of Binary Large Objects, Blobs, that represents a digitized or scanned image or document. (This includes PDF's, bitmaps, jpegs and other image forms and document types	Binary	Variable	
Name	A word or phrase that constitutes the distinctive designation of a person, place, thing, or concept	Text, nvarchar	Variable	
ShortName	A word or phrase that constitutes the distinctive designation of a person, place, thing, or concept in a shortened string. Acronyms, codes, etc. will be valid uses of this field	Text, nvarchar	Variable	
Number	A combination of letters and/or numbers used to uniquely identify an occurrence of something. (i.e. Social Security Number, Vehicle Identification Number). Special characters used as separators would be excluded from occurrences of attributes or fields in this class. Rather, display formats would achieve this effect. For example, Social Security Number would be 9 digits without the 2 „-‟separators	Number or Text as is required	Variable	
Percent	A number that represents the ratio between two values that have the same unit of measure multiplied by 100 (A rate times 100)	Number	5	2
Quantity	A number of non-monetary units expressed in conjunction with a unit of measure	Number	Dependent on associated Measurement and Unit	

Suffix Classwords	Description	Domain	Total Length	Decimal Places
Rate	A quantity or amount measured with respect to another measured quantity or amount (i.e. naira/hour, miles/gallon, etc.)	Number	9	4
XML	A valid XML Document, which could be XML, XSLT, Schema, or other well-formed XML document type	XML	Variable	