Issues from Metadata Glossary team

Metadata glossary team reviewed definition and explanatory text of GSIM (v1.2) information objects and made following comments for next GSIM revision team to consider.

Some writing convention applied to definition include:

- Spelling: UK English will be used (example: organisation instead of organization)
- No leading articles in definitions: All starting "A", "An", "The", etc. will be removed. Definition of term should be able to substitute grammatically; starting with article/using sentence will not allow this
- Definitions will not start with the concept to be defined (e.g. "A Classification Family is...")
- Definitions will start with lowercase and have no ending dot

Base Group

- **Agent**: (Meeting 16 July 2019) 1) we found "some form of responsibility" is vague, we understand that it was borrowed from prov ontology (during GSIM revision). Can this be made clearer in explanatory text?; 2) can Agent be non-Individual or non-Organisation, e.g. programme in automatized process (in this case, "some form of responsibility" does not make sense)?; 3) consider adding reference to prov ontology as source of definition in explanatory text (see how it is done for Agent In Role)
- **Change Event**: (Meeting 16 July 2019) the way definition is written does not conform with writing convention. We suggest to change definition to "indication of the occurrence of a change to an Identifiable Artefact"
- **Change Event Tuple**: (Meeting 6 August 2019) The second sentence should be moved to the explanatory text
- **Identifiable Artefact**: (Meeting 6 August 2019) There is a similar definition of this concept in the SDMX Glossary. What should be our philosophy? Keep the two glossaries totally distinct or should we refer here to the SDMX Glossary (e.g. in the explanatory text)? We could also adopt the SDMX definition and push the current definition to the "explanatory text" section; Create GSIM Issue Need definition & explanatory text of artefact to clarify the definition & provide reference to SDMX definition
- **Maintainer**: (Meeting 6 August 2019) Should we mention here the concept of "Maintenance Agency" from the SDMX Glossary?; to move example (e.g. ...) in the definition to the explanatory text
- **Owner**: (Meeting 6 August 2019) Proposal: second sentence to be moved to "Explanatory text"; to add 'authority' to include working groups, task forces etc. (i.e. instead of "A statistical office or other authority ..."); "statistical office, authority or other organisation..."
- **Role**: (Meeting 6 August 2019) to change responsible function to responsibility function as functions can't be responsible

Business Group

- **Process Design**: (Meeting 3 September 2019) the current definition does not describe essence of process design. It should be reformulated. Closer approximation: "description of the arrangement of steps needed to perform a business function"
- **Process Execution Log**: (Meeting 3 September 2019) in definition 1) second sentence should be in explanatory text; 2) proposed to change to "recording of a multi-set of process traces, i.e. a list of events generated by process instances"
- **Process Input**: (Meeting 3 September 2019) we propose to change definition to "any instance of an information object supplied to a Process Step Instance at the time its execution is initiated" (removing "which is")
- **Process Step Instance**: (Meeting 3 September 2019) we propose following 1) definition: "executed step in a Business Process specifying the actual inputs to and outputs from an occurrence of a Process Step"; 2) explanatory text: change "For this reason, each Process Step Instance details the inputs and outputs for that instance of the implementation of the Process Step." (remove "of" and replace "implementing")
- **Transformable Input**: (Meeting 3 September 2019) we propose to move the second sentence with change ("...") to explanatory text
- **Business Case**: (Meeting 24 September 2019) we propose to move the second sentence to the explanatory text
- **Business Function**: (Meeting 24 September 2019) we propose to change definition to "activities undertaken by a statistical organisation to achieve its one or more objectives"
- **Information Request**: (Meeting 24 September 2019) we propose to change definition to "Statistical Need that is a request for new information for a particular purpose"
- **Process Pattern**: (Meeting 24 September 2019) we propose to change definition to "named set of Process Designs that is highlighted for possible reuse"
- **Statistical Need**: (Meeting 24 September 2019) in definition, we propose 1) to move the second and third sentences to explanatory text (also fix error "Environmental Change", it should be "Environment Change"); 2) to change first sentence to "requirement, request, or other notification that will be considered by a statistical organisation"
- **Statistical Program**: (Meeting 24 September 2019) we propose to change definition to "set of activities, which may be repeated, that describe the purpose and context of a set of Business Processes within the relevant Statistical Programme Cycles"
- **Statistical Support Program**: (Meeting 24 September 2019) we are concerned about how the definition is written (e.g. definition written in negative and doesn’t really make sense – i.e., “not related to post-design production” but “necessary to support production”? if it “supports”, how can it not then be “related”? ) Can you reformulate it?

Concept Group

- **Classification Family**: (Meeting 15 October 2019) we propose to change definition to "group of Classification Series based on a common Concept (e.g. economic activity)"
- **Classification Index**: (Meeting 15 October 2019) we propose to move the second sentence to the explanatory text
- **Classification Index Entry**: (Meeting 15 October 2019) we propose 1) to update the definition as "word or short text (e.g. locality, product, economic activity, or occupational title) describing a type of object/unit or object property to which a Classification Item applies, together with the code of the corresponding Classification Item"; 2) to move the second and the third sentences to the explanatory text
• **Classification Item**: (Meeting 15 October 2019) we propose 1) to update the definition as "Category, at a certain Level within a Statistical Classification, whose content and borders are clearly defined" by combining the first and second sentence; 2) to move the third sentence to the explanatory text; 3) add following text to the explanatory text: "Categories are used to create sub-populations and must be mutually exclusive when contained in a Statistical Classification."

• **Classification Series (Meeting 15 October 2019)**: Second sentence in the definition should be moved to the "Explanatory Text" (Meeting 15 October 2019) we propose to add in the Explanatory Text a short text explaining what Code Lists are used for, e.g. "A Code List provides a predefined set of permissible values for an Enumerated Value Domain"

• **Correspondence Table**: (Meeting 15 October 2019) we propose 1) to update the definition as "set of Maps between the Classification Items of two Statistical Classifications"; 2) to move the second and the third sentences to the explanatory text

• **Datum**: (Meeting 7 Nov. 2019) "value" is not a definition, it's a synonym. Please consider revising (the team did not reach an agreement on the proposal)

• **Described Conceptual Domain**: (Meeting 7 Nov. 2019) To add in the explanatory text "Described Conceptual Domain is a synonym for non-enumerated conceptual domain (source: GSIM)"

• **Description Value Domain**: (Meeting 7 Nov. 2019) To add in the explanatory text "Description Value Domain is a synonym for non-enumerated value domain (source: GSIM)"

• **Designation**: (Meeting 7 Nov. 2019) To switch definition and explanatory text.

• **Instance Variable**: (Meeting 7 Nov. 2019) To move second sentence to explanatory text. The examples are US-centric, maybe make them more neutral?

• **Level**: (Meeting 7 Nov. 2019)
  1. Move current definition to explanatory text;
  2. Change text in explanatory text "A linear classification has only one Level " to "A linear (of flat) classification has only one Level
  3. Consider revising the definition. Here are candidates the glossary team considered but didn't reach agreement (i. position of a Category or a group of Categories within the hierarchy of a Statistical Classification (Source: United Nations, online Glossary of Classification Terms, last consulted on 8 June 2017); ii. identifiable position to which codes in a scheme of codes are related (SDMX Glossary 2018); iii. set of nodes in a statistical classification in which 1) each node in the set is the same number of arcs away from the root node in the hierarchy, and 2) the set is defined by a concept);
  4. Explanatory text should include more illustrative examples.

• Map (Meeting 7 Nov. 2019): 1) To move second and third sentences to explanatory text; 2) To fix typos under explanatory text: Node and Node Set should read Nodes and Node Sets

• **Measurement Type** (Meeting 7 Nov. 2019): 1) To consider revising the definition. It seems circular; 2) To move the second and third sentences to explanatory text

• **Measurement Unit (Meeting 7 Nov. 2019)**: To add in the explanatory text "is a definite magnitude of a quantity, defined and adopted by convention or by law, that is used as a standard for measurement of the same kind of quantity"

• **Population (Meeting 7 Nov. 2019)**: To make the examples less US-centric. E.g. 1. Adult persons in West Europe Region on 13 November 1956; 2. Computer companies in the World at the end of 2012; 3. Universities in Asia on 1 January 2011

• **Represented Variable (Meeting 26 Nov. 2019)**: To add in the beginning of the explanatory text "The measure applies to quantitative, categorical, and descriptive Variable"

• **Sentinel Value Domain (Meeting 26 Nov. 2019)**: 1) To change the definition as "Value Domain containing sentinel values, i.e. processing-related special values"; 2) To add in the beginning of the explanatory text "A sentinel value is one used for processing and with no subject matter content, such as missing or refused.

• **Statistical Classification (Meeting 17 Dec. 2019)**: Proposed a new definition: "hierarchically organised set of mutually exclusive and jointly exhaustive categories that share the same or similar characteristics, used for meaningfully grouping the objects or units in the population of interest"

• **Subject Field (Meeting 17 Dec. 2019)**: Proposed a new definition: "field of knowledge under which a set of Concepts and their Designations is used." (the definition does not need to depend on existence of Concept Systems. First sentence in the explanatory text better conveys the meaning of Subject Field"

• **Substantive Value Domain**: (Meeting 17 Dec. 2019) proposed 1) a new definition: "Value Domain containing substantive values, where a substantive value is subject matter related.", 2) add at the beginning of the explanatory text: "A substantive value is one representing subject matter content, such as <female> in a gender classification"

• **Unit (Meeting 17 Dec. 2019)**: Proposed a new definition from SDMX: "entity for which information is sought and for which statistics are ultimately compiled" (this is clearer and more specific to the business what we do); 2) to make the example less US-centric: "Individual person, i.e. John DOE, Erika MUSTERMANN, Albert DURAND"

• **Unit Type (Meeting 17 Dec. 2019)**: Proposed a new definition "class or group of Units based on a single characteristic" (note that we proposed a new definition for Unit)

• **Universal (Meeting 17 Dec. 2019)**: Proposed a new definition: "specialized Unit Type, but not by time or geography"; 2) to add to the explanatory text "In both cases, adding the condition of the university degree specializes persons, which is a Unit Type."

• **Variable (Meeting 17 Dec. 2019)**: Proposed a new definition "use of a Concept as a characteristic of Unit Type intended to be observed"

---

**Exchange Group**

- **Exchange Channel**: (Meeting 17 Dec. 2019) Proposed a new definition: "means or vehicles of exchanging information" ("means" seems vague)
- **Information Consumer**: (Meeting 17 Dec. 2019) Proposed a new definition: "Agent that consumes disseminated data"
- **Information Provider**: (Meeting 17 Dec. 14 Dec. 2019): Proposed a new definition: "Agent that provides collected information"
- **Instance Question**: (Meeting 17 Dec. 2019) To consider adding concrete examples. Explanation is esoteric.
- **Instance Question Block**: (Meeting 17 Dec. 2019): To consider adding concrete examples. Explanation is esoteric.
- **Instance Statement**: (Meeting 17 Dec. 2019): To consider adding concrete examples. Explanation is esoteric.
- **Output Specification**: (Meeting 17 Dec. 2019): Proposed a new definition: "outline or description of how Information Sets are presented and arranged for Information Consumer" (definition Consumer should not start with v)
- **Presentation**: (Meeting 17 Dec. 2019): Proposed a new definition: "manner through which data and referential metadata are arranged in a Product" (That way we don't use presented in the definition of Presentation)
- **Protocol**: (Meeting 17 Dec. 2019): Proposed a new definition: "method defining how information is communicated through an Exchange Channel" (word "mechanism" does not convey much information); 2) to replace "mechanism" in the explanatory text with "method"
- **Question**: (Meeting 17 Dec. 2019): Proposed a new definition: "text used to elicit a response for the Concept to be measured" (Question does not need to depend on existence of Concept Systems)
- **Question Block**: (Meeting 9 Jan. 2020): To change the first sentence of explanatory text from "A Question Block should be designed for reuse, as it can be used in multiple Questionnaires," to "A Question Block should ideally be designed for reuse"
- **Questionnaire**: (Meeting 9 Jan. 2020): 1) To add what CAPI stands for, namely "Computer-assisted personal interviewing" in the explanatory note; 2) The definitions of Questionnaire and Questionnaire Specification are very similar. It's not clear where designing of questionnaire and
implementation of questionnaire are. Also, Questionnaire Specification cannot exist without Questionnaire (composition black diamond) - is it correct? If Questionnaire was a design, it means you cannot design without an instrument - is this correct?

- **Questionnaire Component** (Meeting 9 Jan. 2020): 1) To start explanatory text with "A Questionnaire Component defines..."; 2) General comment about set of GSIM objects: include a real/concrete examples in the explanatory text, not just about GSIM modeling perspective (Questionnaire, Questionnaire Component, Questionnaire Specification, Question Block, Statement, and Questionnaire Logic; see examples given under Question Block explanatory text - something like this would be very helpful for readers; also see concrete examples given about Variable (Represented Variable/Instance Variable in GSIM e-training)

- **Questionnaire Logic** (Meeting 9 Jan. 2020): To change the definition to "regularly updated database of units and their properties that is designed for statistical purposes" (list is too colloquial)

- **Questionnaire Specification** (Meeting 9 Jan. 2020): See comment #2 under Questionnaire Component

- **Statement** (Meeting 9 Jan. 2020): "Statement" in the model specifically refers to instruction to the respondents, but the word "statement" can appear other parts in the question (e.g. Please indicate your level of agreement or disagreement with each of these statements...: a) I like to go to work every day; b) I hate going to work everyday). When it is used that way, it means what dictionary normally means. This distinction should also be explained in the explanatory not

- **Statistical Register** (Meeting 9 Jan. 2020): To change the definition to "regularly updated database of units and their properties that is designed for statistical purposes" (list is too colloquial)

### Structure Group

- **Data Point** (Meeting 9 Jan. 2020): Explanatory text seems limited. Usage of the word "tables" seems to convey a limited sense of how data might be organized. Propose "Field in a Data Structure which corresponds to, e.g., a cell in a table..."

- **Data Structure** (Meeting 9 Jan. 2020): To change the definition as "structure of an organised collection of data (Data Set)" (definition should not start with a verb nor contain the word "Structure")

- **Dimensional Data Point** (Meeting 28 Jan. 2020): Explanatory text is incomplete, the examples given for why multiple data points could appear in the same data point doesn’t account for using different measures, using values different time, i.e. times series, or values from different statistical program. The text needs to account for these.

- **Dimensional Data Set** (Meeting 28 Jan. 2020): Need explanatory text

- **Dimensional Data Structure** (Meeting 28 Jan. 2020): The definition can’t start with a verb, and it should focus on the structural description: identifiers, attribute, and measures. In explanatory text, add gender as an identifier component.

- **Information Resource** (Meeting 28 Jan. 2020): Put a hyphen between sub and classes in explanatory text. Concrete examples are needed here.

- **Logical Record** (Meeting 28 Jan. 2020): removed verb in def. and add comma in explanatory text

- **Record Relationship** (Meeting 20 Feb. 2020): 1) Definition should not start with a verb, remove "Describes"; 2) Second sentence in the definition should be moved to the explanatory text

- **Referential Metadata Content Item** (Meeting 20 Feb. 2020): Explanatory text is too abstract and model dependent. Suggest adding a concrete example

- **Referential Metadata Resource** (Meeting 20 Feb. 2020): Explanatory text is too abstract and model dependent. Suggest adding a concrete example

- **Referential Metadata Set** (Meeting 20 Feb. 2020): Explanatory text is too abstract and model dependent. Suggest adding a concrete example

- **Referential Metadata Structure** (Meeting 20 Feb. 2020): 1) Definition should not start with a verb, remove "Describes the"; 2) Remove "(Referential Metadata Set)"

- **Referential Metadata Subject** (Meeting 20 Feb. 2020): Definition begins with a verb. Remove “Identifies the”. Also the definition relies on the term itself, as a result, it’s circular. Consider to write a definition that does not depend on the term.

- **Referential Metadata Subject Item** (Meeting 20 Feb. 2020): Definition begins with a verb. Remove “Identifies the”. Also the definition relies on the term itself, as a result, it’s circular. Consider to write a definition that does not depend on the term.

- **Unit Data Point** (Meeting 20 Feb. 2020): Explanatory text seems wrong or incomplete. A GSIM implementation could document the same data point from several perspectives at the same time. The explanation of what differentiates data points in that data point doesn’t handle this case.

- **Unit Data Record** (Meeting 20 Feb. 2020): 1) Remove “Contains the”. Proposed to change from "Contains" to "Container for ...." 2) Explanatory text is incomplete. It is not clear if the example given fits the meaning of the term defined. 3) In the explanatory note, how can we infer that the age is calculated on the 1st of January 2012 ? Words "in years" after 1st of January 2012 should be removed. This seems a simple mistake.

- **Unit Data Set** (Meeting 20 Feb. 2020): The example is not clear. Pick a more direct example

- **Unit Data Structure** (Meeting 20 Feb. 2020): Definition should not start with a verb