ihtsdo Leading healthcare terminology, worldwide

SNOMED CT Quality Evaluation

Part I : Descriptive Release Statistics

Date: 20160219
Version: 1.00
Status: Draft

Amendment History

Version	Date	Editor	Comments
0.8	20151008	Monique Van Berkum (MVB)	Initial Analysis and preliminary queries
0.9	20151201	Guillermo Reynoso (GR)	Document reorganization
0.95	20160212	MVB/GR	Addition of reporting examples
1.00	20160219	MVB/GR	Updates to Tables

Reviewers

Version	Date	Approver	Comments

Table of Contents

1 Introduction

1.1 Background

This document describes general descriptive release statistics to be used as part of an overall Quality Assurance process for the SNOMED CT® International Release. The objective of the general descriptive release statistics and an accompanying set of content patterns recommended for periodic monitoring (described in a complimentary document) is to provide information to assist the IHTSDO in:

- Providing metrics on temporal trends as a measurement of progress
- Ensuring that newly added content meets quality assurance criteria
- Ensuring that changes to existing content meet quality assurance criteria
- Identifying and prioritizing areas of existing content which do not currently meet quality assurance criteria

1.2 Purpose

The purpose of this document is to propose an initial set of descriptive measures and statistics, obtained through automated queries, with the objective of providing information on content changes and temporal trends as a measurement of progress. The gathered data will also assist in comparing consecutive SNOMED CT International Releases and trends across multiple releases.

1.3 Scope

This document describes:

- A set of proposed measures and statistics to be computed from SNOMED CT releases

A complementary document will describe a set of content patterns recommended for periodic monitoring and evaluation of conformance with content management, quality assurance processes and technical specifications.

1.4 Audience

The audience for this document includes IHTSDO Content staff and the Community of Practice.

2 General Release Statistics

2.1 General release statistics for all of SNOMED CT

General release statistics provided for all of SNOMED CT include:

- Total number of active Concepts
- Percentage of Concepts which are sufficiently defined (SD) (see Section 3 for explanation of SD)
- Total number of active Descriptions
- Total number of active Relationships

2.2 General release statistics by hierarchy

General release statistics provided for each hierarchy include:

- Total number of active Concepts
- Percentage of active Concepts in each hierarchy as a percentage of all active Concepts in the release
- Percentage of Concepts in each hierarchy which are sufficiently defined (SD)
- Total number of active Descriptions
- Total number of active Relationships

2.3 'General Release Statistics' Table – Example

General Release Statistics

Hierarchy	Concepts	% of SCT	% SD (*)	Descriptions	Relationships
SNOMED CT Concept (SNOMED RT+CTV3)	319446	100.00%	23.90%	838727	2527103
Body structure (body structure)	31039	9.71%	2.56%	90639	92445
Clinical finding (finding)	103911	32.52%	43.63%	276111	405890
Environment or geographical location (environment / location)	1815	0.56%	0.00%	4119	1846
Event (event)	3611	1.13%	1.74%	8020	3804
Observable entity (observable entity)	8469	2.65%	0.00%	21787	8936
Organism (organism)	33535	10.49%	0.00%	91909	35034
Pharmaceutical / biologic product (product)	17291	5.41%	14.69%	42684	56249
Physical force (physical force)	170	0.05%	0.00%	441	181
Physical object (physical object)	14625	4.57%	0.00%	32728	17643
Procedure (procedure)	55423	17.34%	42.21%	142988	261415
Qualifier value (qualifier value)	9393	2.94%	0.00%	23843	10166
Record artifact (record artifact)	228	0.07%	0.00%	513	230
SNOMED CT Model Component (metadata)	1553	0.48%	0.00%	3522	1568
Situation with explicit context (situation)	4251	1.33%	68.73%	10843	21697
Social context (social concept)	4720	1.47%	0.14%	10621	5505
Special concept (special concept)	648	0.20%	0.00%	1541	648
Specimen (specimen)	1634	0.51%	78.82%	3633	6050
Staging and scales (staging scale)	1360	0.42%	0.00%	3810	1361
Substance (substance)	25761	8.06%	0.00%	68946	31836

(*) Percentage of existing concepts in hierarchy which are Sufficiently Defined (SD). For the root concept the value includes all SNOMED CT.

(Example based on the January 2016 International Release)

3 Statistics on New Concepts

Factors affecting the number of new Concepts in the release include but are not limited to:

1. Ongoing content projects – influenced by:
    a. Prioritization
    b. Editorial Policy development
    c. Project timelines
2. Request submissions (single and batch) – influenced by:
    a. Submission timing with respect to the release cycle
    b. Submission appropriateness for the International Release
    c. Submission complexity
    d. Submission quality
    e. State of existing content in the hierarchy where the submission is added
    f. Ongoing changes to Editorial Policy

New concepts can be added with a definition status of sufficiently defined (SD) or primitive (P). A concept is defined by a set of relationships to other concepts. If the resulting definition is sufficient to distinguish the concept from its parents and siblings in the subtype hierarchy, the concept is considered to be sufficiently defined. If the definition is not sufficient to distinguish the concept from its parents and siblings, the concept is considered to be primitive (SNOMED CT® Technical Implementation Guide January 2015 International Release Section 4.2.1.2).

Factors affecting the definition status of new concepts include but are not limited to:

1. The Concept Model only allows concepts in some hierarchies to be sufficiently defined.
2. For hierarchies which allow sufficiently defined concepts, the Concept Model may not be expressive enough to support sufficiently defining some concepts.
3. Some concepts may be unable to be sufficiently defined within the clinical scope of SNOMED CT even with enhancements to the Concept Model.

3.1 Statistics on new Concepts for all of SNOMED CT

Statistics provided for new Concepts for all of SNOMED CT include:

- Number of new Concepts
- New Concepts as a percentage of all Concepts in the release
- Percentage of new Concepts which are sufficiently defined (SD)
- Percentage of new Concepts which are primitive (P)

3.2 Statistics on new Concepts by hierarchy

Statistics provided for new Concepts in each hierarchy include:

- Number of new Concepts
- New Concepts as a percentage of each hierarchy

- New Concepts in each hierarchy as a percentage of all new Concepts in the release
- Percentage of new Concepts per hierarchy which are added as sufficiently defined (SD)
- Percentage of new Concepts per hierarchy which are added as primitive (P)

3.3 'New Concepts Statistics' Table – Example

New Concepts Statistics

Hierarchy	New Concepts	% New in Hierarchy	% of All New in Release	% Added as SD (*)	% Added as P (*)
SNOMED CT Concept (SNOMED RT+CTV3)	3102	0.73%	100.00%	43.32%	56.68%
Body structure (body structure)	245	0.78%	7.89%	0.00%	100.00%
Clinical finding (finding)	957	0.92%	30.85%	72.30%	27.70%
Environment or geographical location (environment / location)	1	0.05%	0.03%	0.00%	100.00%
Event (event)	6	0.16%	0.19%	33.33%	66.67%
Observable entity (observable entity)	61	0.72%	1.96%	0.00%	100.00%
Organism (organism)	541	1.61%	17.44%	0.00%	100.00%
Pharmaceutical / biologic product (product)	425	2.45%	13.70%	88.23%	11.77%
Physical force (physical force)	0	0.00%	0.00%	0.00%	100.00%
Physical object (physical object)	225	1.53%	7.25%	0.00%	100.00%
Procedure (procedure)	334	0.60%	10.76%	49.40%	50.60%
Qualifier value (qualifier value)	6	0.06%	0.19%	0.00%	100.00%
Record artifact (record artifact)	4	1.74%	0.12%	0.00%	100.00%
SNOMED CT Model Component (metadata)	7	0.45%	0.22%	0.00%	100.00%
Situation with explicit context (situation)	102	2.39%	3.28%	87.25%	12.75%
Social context (social concept)	10	0.21%	0.32%	70.00%	30.00%
Special concept (special concept)	0	0.00%	0.00%	0.00%	100.00%
Specimen (specimen)	15	0.91%	0.48%	93.33%	6.67%
Staging and scales (staging scale)	21	1.54%	0.67%	0.00%	100.00%
Substance (substance)	142	0.55%	4.57%	0.00%	100.00%

(*) SD = Sufficiently defined
P = Primitive

(Example based on the January 2016 International Release)

4 Statistics on Concepts Inactivated During the Release

The statistics provided are only for concepts which were inactivated during the release. They are not statistics for all concepts that have ever been inactivated in SNOMED CT.

4.1 Statistics for SNOMED CT for Concepts inactivated during the release

Statistics provided for SNOMED CT for Concepts inactivated during the release include:

- Number of Concepts in SNOMED CT inactivated during the release
    - Percentage of Concepts in SNOMED CT inactivated during the release
    - Percentage of Concepts in SNOMED CT inactivated during the release in each inactivation category (e.g., ambiguous, duplicate, outdated etc.)

4.2 Statistics by hierarchy for Concepts inactivated during the release

Statistics provided by hierarchy for Concepts inactivated during the release include:

- Number of Concepts in the hierarchy inactivated during the release
    - Percentage of Concepts in the hierarchy inactivated during the release (inactivated Concepts in hierarchy/total Concepts in hierarchy)
    - Concepts inactivated in the hierarchy during the release as a percentage of all Concepts inactivated in SNOMED CT during the release
- Percentage of Concepts in the hierarchy inactivated during the release by inactivation category (e.g., ambiguous, duplicate, outdated etc.)

4.3 'Statistics on Concepts Inactivated During the Release' Table – Example

Statistics on Concepts Inactivated During the Release

Hierarchy	Inactivated Concepts	% of Hierarchy Inactivated	% of All Inactivated	% as Ambiguous	% as Duplicate	% as Erroneous	% as Not Stated	% as Outdated	% as Moved	% as Limited
SNOMED CT Concept (SNOMED RT+CTV3)	726	0.17%	100.00%	51.23%	21.76%	17.90%	0.13%	3.71%	5.23%	0.00%
Body structure (body structure)	77	0.24%	10.60%	98.70%	0.00%	1.29%	0.00%	0.00%	0.00%	0.00%
Clinical finding (finding)	277	0.26%	38.15%	68.59%	12.27%	5.05%	0.00%	1.44%	12.63%	0.00%
Environment or geographical location (environment / location)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Event (event)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Observable entity (observable entity)	1	0.01%	0.13%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
Organism (organism)	41	0.12%	5.64%	12.19%	73.17%	9.75%	0.00%	4.87%	0.00%	0.00%
Pharmaceutical / biologic product (product)	65	0.37%	8.95%	36.92%	4.61%	26.15%	0.00%	32.30%	0.00%	0.00%
Physical force (physical force)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Physical object (physical object)	3	0.02%	0.41%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Procedure (procedure)	38	0.06%	5.23%	42.10%	36.84%	18.42%	0.00%	0.00%	2.63%	0.00%
Qualifier value (qualifier value)	1	0.01%	0.13%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Record artifact (record artifact)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SNOMED CT Model Component (metadata)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Situation with explicit context (situation)	13	0.30%	1.79%	92.30%	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%
Social context (social concept)	1	0.02%	0.13%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Special concept (special concept)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specimen (specimen)	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Staging and scales (staging scale)	1	0.07%	0.13%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Substance (substance)	208	0.80%	28.65%	22.59%	35.09%	41.34%	0.00%	0.00%	0.96%	0.00%

(Example based on the January 2016 International Release)

5 Statistics on Changes to Existing Active Concepts

Statistics provided on existing active Concepts (active in the prior release and active in the current release) include information on changes to:

1. **Concept definitions (including):**
   - Stated concept definition changes
   - Inferred concept definition changes

   Changes to concept definitions include changes to |Is a| Relationships to proximal parent concepts and to Defining Relationships (attribute-value pairs).

   Stated changes to the concept definition are changes made manually by content editors to the set of Relationships that are the defining characteristics of a Concept.

   Inferred changes to the concept definition are changes made after the description logic classifier has applied a set of logical rules to the set of Relationships that are the stated concept definition. The inferred concept definition is the definition included in the standard SNOMED CT distribution relationship file.

2. **Concept definition status (including):**
   - From Primitive to Sufficiently Defined (P to SD)
   - From Sufficiently Defined to Primitive (SD to P)

   Concept definition status changes are made manually by content editors after having reviewed or edited a concept definition.

   A Concept is defined by a set of Relationships to other Concepts. If the resulting definition is sufficient to distinguish the Concept from its parents and sibling Concepts in the subtype hierarchy, the Concept is considered to be sufficiently defined. If the definition is not sufficient to distinguish the Concept from its parents and siblings, the Concept is considered to be primitive. (SNOMED CT® Technical Implementation Guide January 2015 International Release Section 4.2.1.2).

3. **Descriptions (including):**
   - New Descriptions
   - Retired Descriptions (e.g., inappropriate, outdated, etc.)

4. **Fully Specified Names (FSN)**

   Statistics on changes to FSNs are provided separate from changes to other Descriptions because the FSN represents the meaning of the Concept code. Policy regarding changes to the FSN of an existing Concept is stricter than policy regarding changes to Synonyms or Preferred Terms for an existing concept. A change to an FSN requires retiring the Concept except in cases where the Fully Specified Name (FSN) undergoes a minor change (as permitted in the Editorial Guide Section on “Minor changes in the Fully Specified Name”).

5.1 Statistics on changes to existing active Concepts for all of SNOMED CT

The statistics provided for changes to existing active Concepts (active in the prior release and active in the current release) for all of SNOMED CT include:

- Number of Concepts in SNOMED CT that underwent any change (as described by 1-4 above)
  - Percentage of Concepts in SNOMED CT that underwent any change

- Number of Concepts in SNOMED CT with any stated change to the concept definition
    - Percentage of Concepts in SNOMED CT with any stated change to the concept definition
- Number of Concepts in SNOMED CT with any inferred change to the concept definition
    - Percentage of Concepts in SNOMED CT with any inferred change to the concept definition
- Number of Concepts in SNOMED CT with a definition status change from P to SD
    - Percentage of Concepts in SNOMED CT with a definition status change from P to SD
- Number of Concepts in SNOMED CT with a definition status change from SD to P
    - Percentage of Concepts in SNOMED CT with a definition status change from SD to P
- Number of Concepts in SNOMED CT with any Description change
    - Percentage of Concepts in SNOMED CT with any Description change
- Number of Concepts in SNOMED CT with a FSN change
    - Percentage of Concepts in SNOMED CT with a FSN change

5.2 Statistics on changes to existing active Concepts by hierarchy

The statistics provided for changes to existing active Concepts (active in the prior release and active in the current release) by hierarchy include:

- Number of Concepts in the hierarchy that underwent any change (as described by 1-4 above)
    - Percentage of Concepts in the hierarchy that underwent any change
- Number of Concepts in the hierarchy with any stated change to the concept definition
    - Percentage of Concepts in the hierarchy with any stated change to the concept definition
- Number of Concepts in the hierarchy with any inferred change to the concept definition
    - Percentage of Concepts in the hierarchy with any inferred change to the concept definition
- Number of Concepts in the hierarchy with a definition status change from P to SD
    - Percentage of Concepts in the hierarchy with a definition status change from P to SD
- Number of Concepts in the hierarchy with a definition status change from SD to P
    - Percentage of Concepts in the hierarchy with a definition status change from SD to P
- Number of Concepts in the hierarchy with any Description change
    - Percentage of Concepts in the hierarchy with any Description change
- Number of Concepts in the hierarchy with a FSN change
    - Percentage of Concepts in the hierarchy with a FSN change

5.3 'Changes to Existing Concepts - Counts by Change Type' Table – Example

Changes to Existing Concepts - Counts by Change Type (*)

Hierarchy	Any	Stated Defn	Inferred Defn	P to SD	SD to P	Description Change	FSN Change
SNOMED CT Concept (SNOMED RT+CTV3)	20525	4654	16977	1373	138	4154	3339
Body structure (body structure)	238	165	226	1	0	19	11
Clinical finding (finding)	15210	2400	13070	437	117	3332	3187
Environment or geographical location (environment / location)	0	0	0	0	0	0	0
Event (event)	24	18	21	1	0	1	1
Observable entity (observable entity)	15	3	3	0	0	13	9
Organism (organism)	143	98	88	0	0	96	41
Pharmaceutical / biologic product (product)	3177	1208	2050	887	2	535	50
Physical force (physical force)	0	0	0	0	0	0	0
Physical object (physical object)	17	12	12	0	0	6	3
Procedure (procedure)	1056	274	933	36	15	65	16
Qualifier value (qualifier value)	6	3	3	0	0	3	0
Record artifact (record artifact)	0	0	0	0	0	0	0
SNOMED CT Model Component (metadata)	1	0	0	0	0	1	0
Situation with explicit context (situation)	113	34	103	10	3	9	3
Social context (social concept)	1	1	1	0	0	0	0
Special concept (special concept)	0	0	0	0	0	0	0
Specimen (specimen)	30	2	30	1	0	0	0
Staging and scales (staging scale)	1	0	0	0	0	1	0
Substance (substance)	492	436	437	0	1	72	18

(*) Stated Defn = Change to stated concept definition
Inferred Defn = Change to inferred concept definition
P to SD = Change in concept definition status from Primitive to Sufficiently Defined
SD to P = Change in concept definition status Sufficiently Defined to Primitive

(Example based on the January 2016 International Release)

5.4 'Changes to Existing Concepts - Percentage of Concepts with Each Change Type' Table – Example

The first row provides the percentage of Concepts in all of SNOMED CT that underwent each change type. The remainder of the rows provide the percentage of Concepts in each hierarchy that underwent each change type.

Changes to Existing Concepts - Percentage of Concepts with each Change Type (*)

Hierarchy	Any	Stated Defn	Inferred Defn	P to SD	SD to P	Description Change	FSN Change
SNOMED CT Concept (SNOMED RT+CTV3)	4.83%	1.09%	3.99%	0.32%	0.03%	0.97%	0.78%
Body structure (body structure)	0.76%	0.53%	0.72%	0.00%	0.00%	0.06%	0.03%
Clinical finding (finding)	14.63%	2.30%	12.57%	0.42%	0.11%	3.20%	3.06%
Environment or geographical location (environment / location)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Event (event)	0.66%	0.49%	0.58%	0.02%	0.00%	0.02%	0.02%
Observable entity (observable entity)	0.17%	0.03%	0.03%	0.00%	0.00%	0.15%	0.10%
Organism (organism)	0.42%	0.29%	0.26%	0.00%	0.00%	0.28%	0.12%
Pharmaceutical / biologic product (product)	18.37%	6.98%	11.85%	5.12%	0.01%	3.09%	0.28%
Physical force (physical force)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Physical object (physical object)	0.11%	0.08%	0.08%	0.00%	0.00%	0.04%	0.02%
Procedure (procedure)	1.90%	0.49%	1.68%	0.06%	0.02%	0.11%	0.02%
Qualifier value (qualifier value)	0.06%	0.03%	0.03%	0.00%	0.00%	0.03%	0.00%
Record artifact (record artifact)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SNOMED CT Model Component (metadata)	0.06%	0.00%	0.00%	0.00%	0.00%	0.06%	0.00%
Situation with explicit context (situation)	2.65%	0.79%	2.42%	0.23%	0.07%	0.21%	0.07%
Social context (social concept)	0.02%	0.02%	0.02%	0.00%	0.00%	0.00%	0.00%
Special concept (special concept)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Specimen (specimen)	1.83%	0.12%	1.83%	0.06%	0.00%	0.00%	0.00%
Staging and scales (staging scale)	0.07%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%
Substance (substance)	1.90%	1.69%	1.69%	0.00%	0.00%	0.27%	0.06%

(*) Stated Defn = Change to stated concept definition
Inferred Defn = Change to inferred concept definition
P to SD = Change in concept definition status from Primitive to Sufficiently Defined
SD to P = Change in concept definition status Sufficiently Defined to Primitive

(Example based on the January 2016 International Release)