SNOMED International

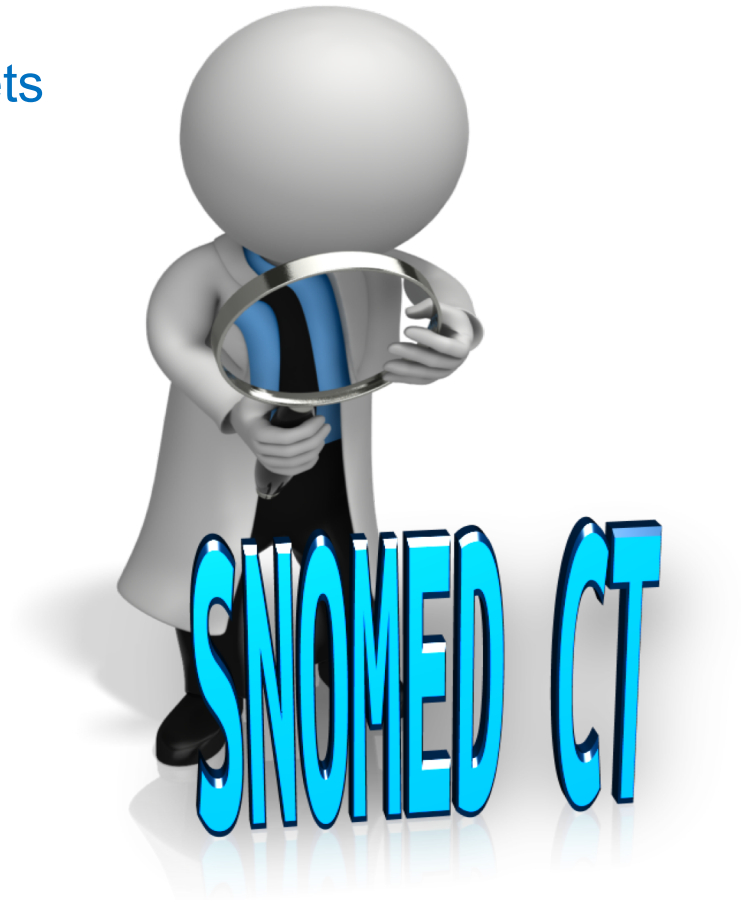
Practical Guide to SNOMED CT Reference Sets

Expo 2018 Tutorial

David Markwell and Anne Randorff Højen
SNOMED International

Overview

- Part 1. Introduction to reference sets
  - Purpose of reference sets
  - Subsets, value set and reference sets
  - Requirements
  - Reference set types and examples
- Part 2. Reference set life cycle
  - Overview of development process
  - Practical examples
  - Questions



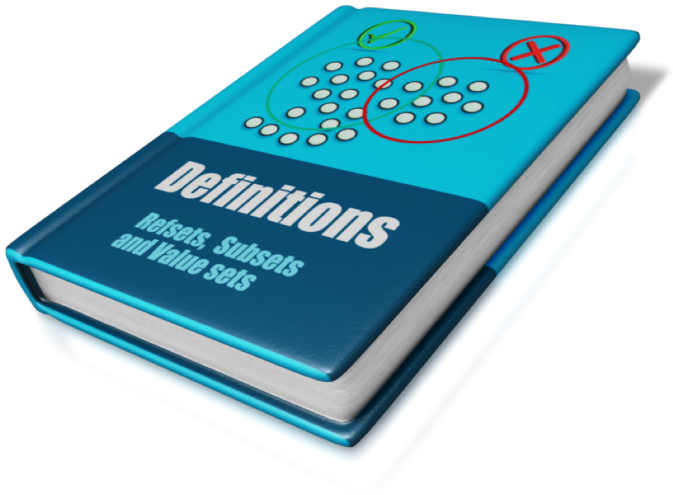
SNOMED International

Introduction to Reference Sets

The Purpose of Reference Sets

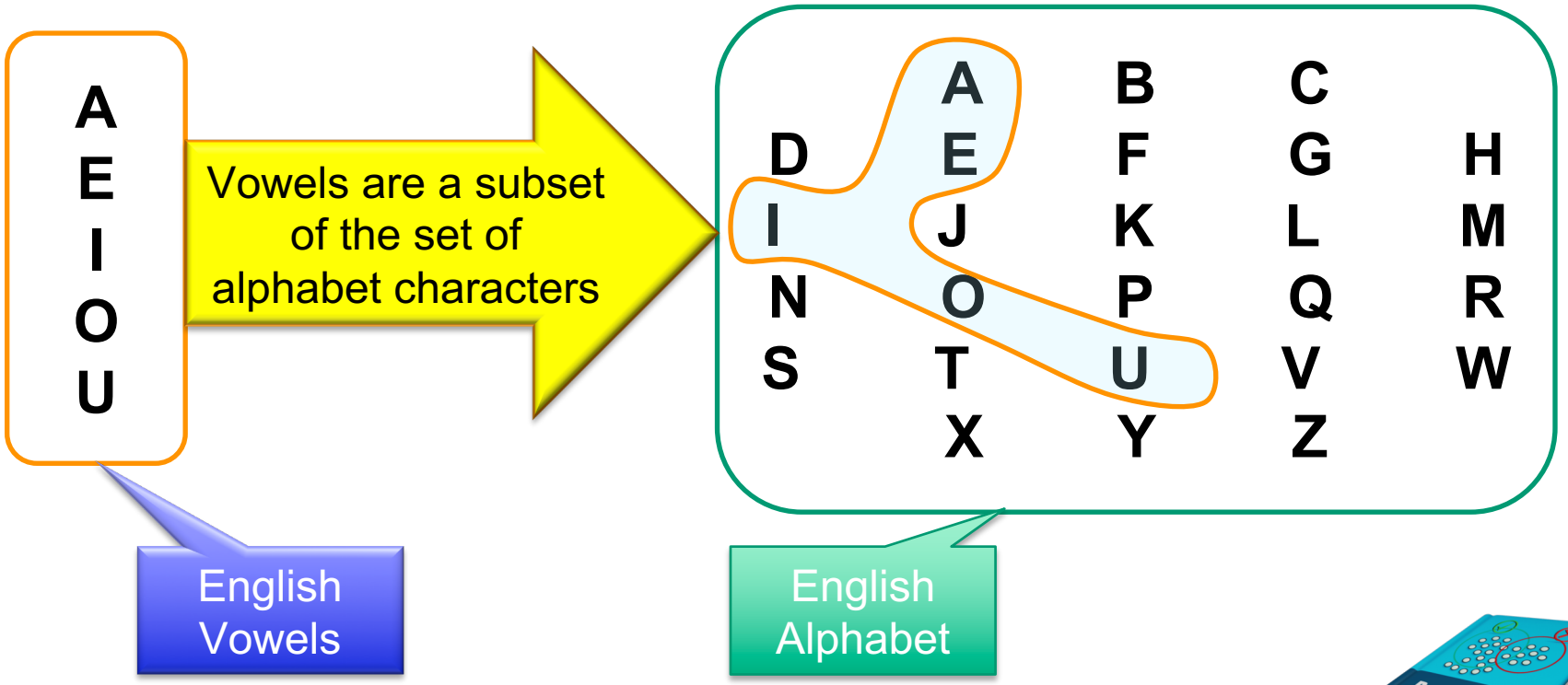
- To provide a standard was to enable customization and enhancement of SNOMED CT
  - Customization to meet the needs of different countries, languages, specialties and contexts
  - Enhancement by adding supporting data that extends the features and functionality of the terminology

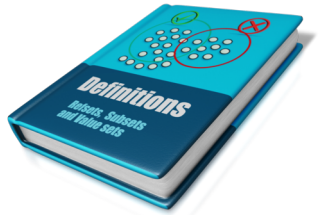
Definitions of Subset, Value Set and Reference Set



Definition of a Subset

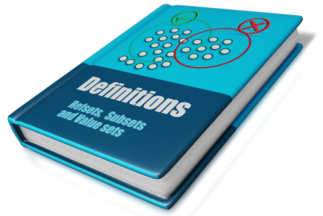
- A set whose members are all contained in another set
  - This is a general definition not specific to SNOMED CT
- Example





Definition of a Value Set

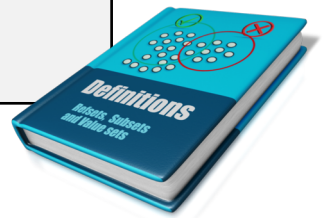
- A set of concept representations used to represent values in a particular data item
  - This definition is not specific to SNOMED CT
  - May include codes from different code systems, coded refinements or predefined text strings
- Example
  - A specification for a problem list might define a single value-set including:
    - SNOMED CT disorder concepts
    - SNOMED CT expressions that are subtypes of disorder
    - ICD-10 classification codes representing diseases



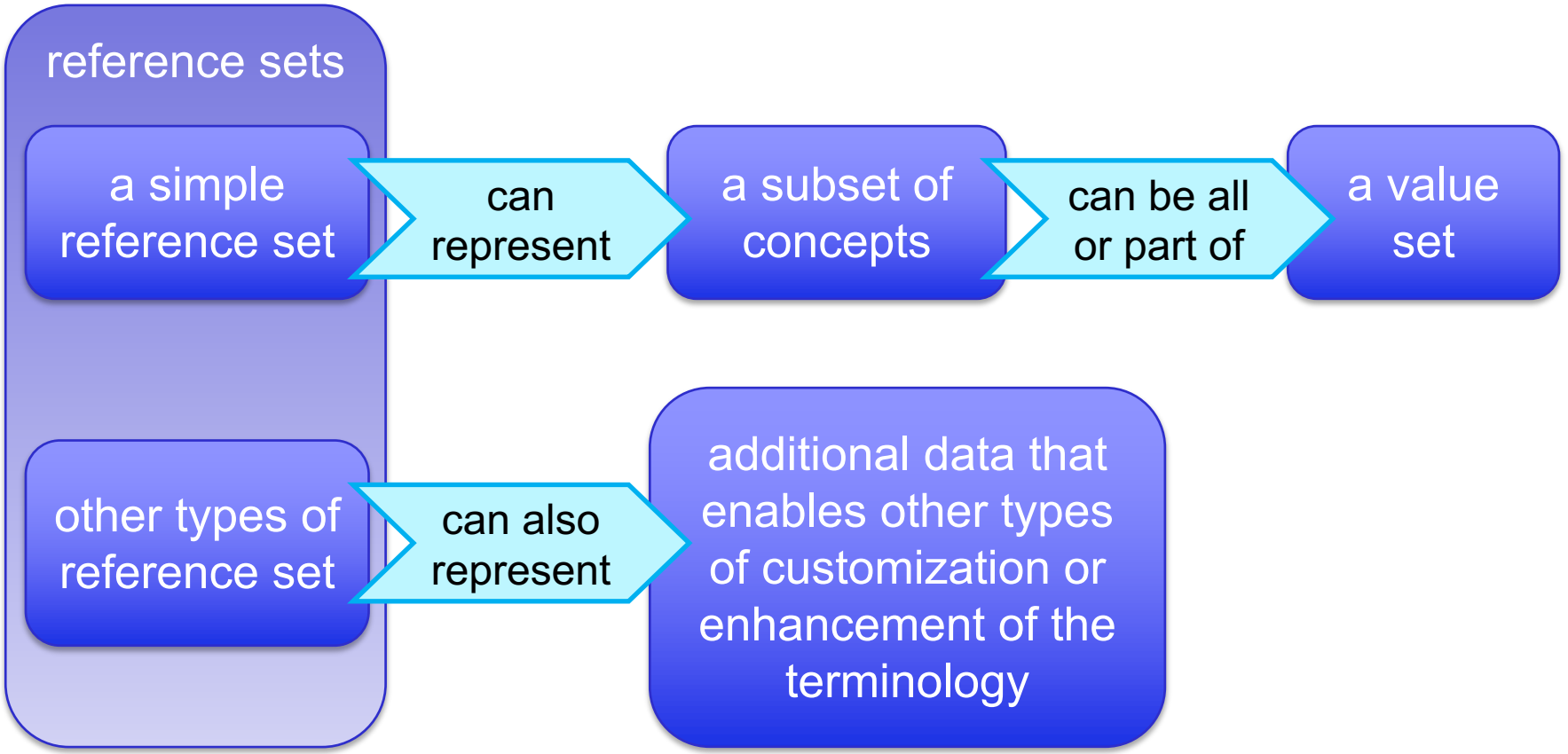
Definition of a Reference Set

- A standard SNOMED CT release file format
- A set of references to SNOMED CT components
- May include additional data for each referenced component

Identification and versioning and modularization information	**id**
	effectiveTime
	active
	moduleId
An identifier of the reference set	**refsetId**
Component reference	**referencedComponentId**
Additional data items defined to meet specific requirements	*<attribute-1 … attribute-n>*

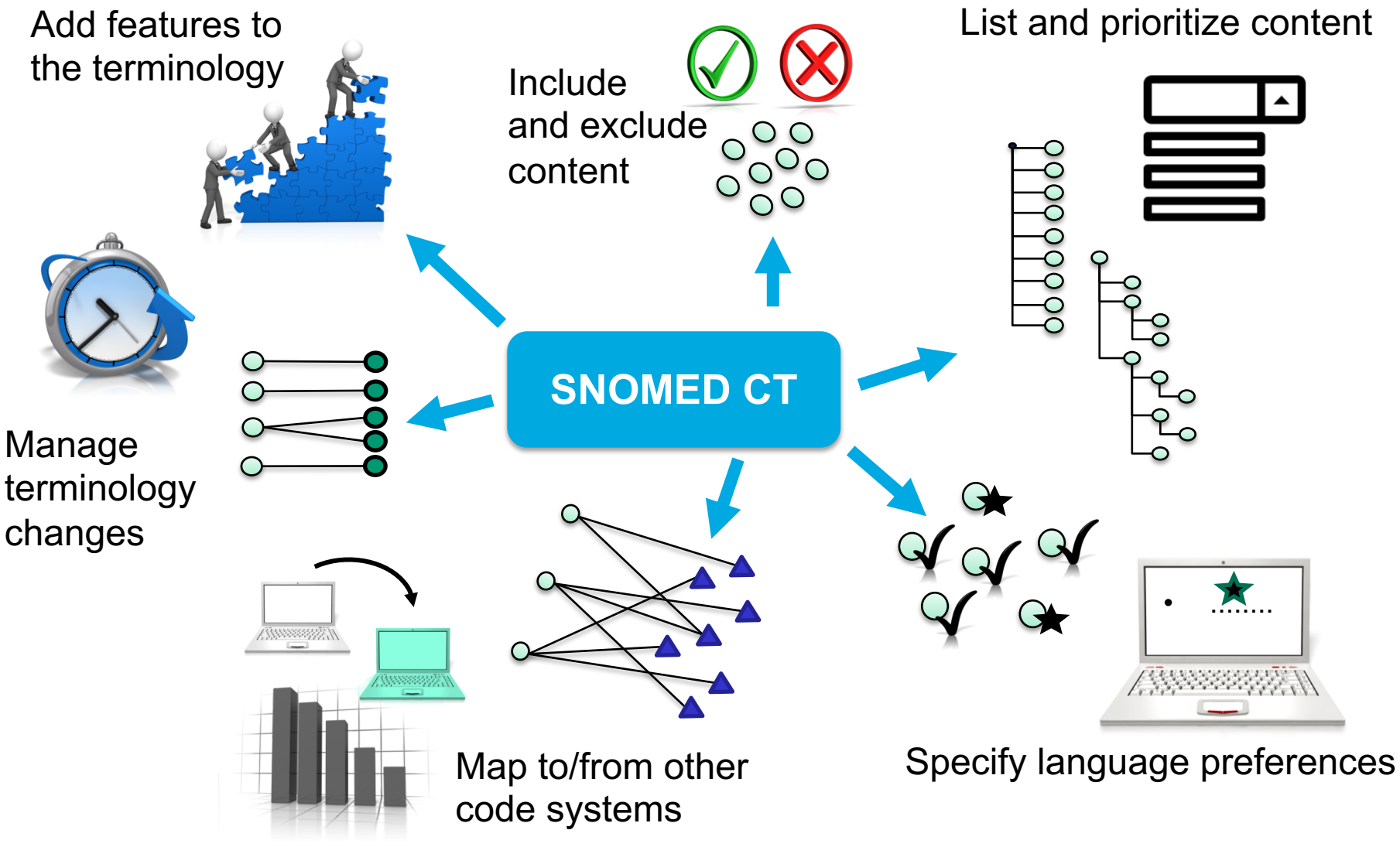


Subsets, Value Sets and Reference Sets – Connections



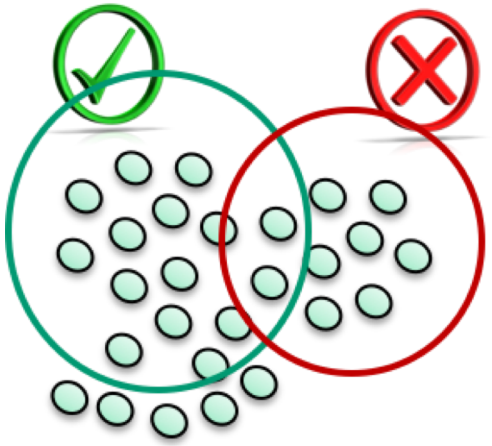
Requirements for Reference Sets

Requirements for Reference Sets



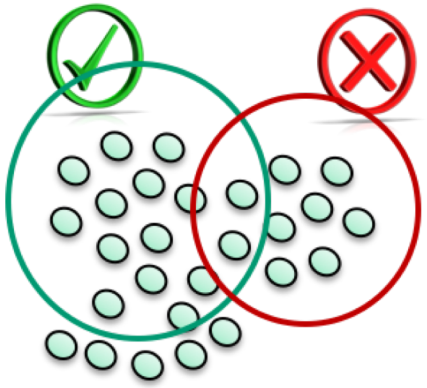
SNOMED International

Requirements for Including or Excluding Concepts and Descriptions



Practical Requirements for Subsets

- Searches and data entry
  - Restricting searches to a set of concepts or descriptions
  - Specifying descriptions to appear in a list of options
  - Constraining data entry to a specified set of concepts
- Information model and communications
  - Specifying value sets for particular data items
- Data retrieval and analysis
  - Specifying query criteria
- Other uses
  - Subsets can be used for any purpose that requires selective inclusion or exclusion of specified sets of components



Data Structure Requirements for Subsets

1. **A simple list of identifiers can represent a subset of SNOMED CT components**
   - For example:
     - 82272006
     - 6142004
     - 55604004
     - … etc. …

When a subset is represented by a SNOMED CT simple reference set:

- The list of component identifiers is represented by the **referencedComponentId**

Subset Represented by a Simple Reference Set

Subset

Reference Set

Id	Term
82272006	Common cold
6142004	Influenza
55604004	Avian influenza

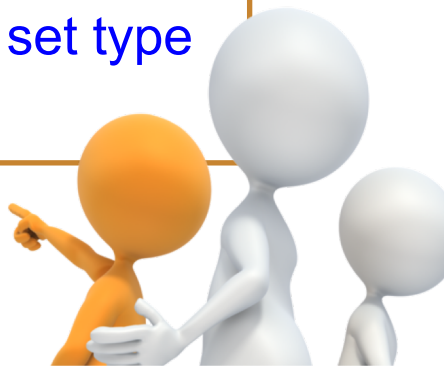
id	effective Time	active	moduleId	refsetId	referenced ComponentId
0000d6608041-005b-5108-b608-07891b210365	20140731	1	19999999103	49999999102	82272006
00008774f374-5f54-5234-8af1-8aca6e73b5c2	20140731	1	19999999103	49999999102	6142004
00006ac9-2f97-5ed4-b37e-84cb158c388d	20140731	1	19999999103	49999999102	55604004

Data Structure Requirements for Subsets

1. A simple list identifiers can represent a subset of SNOMED CT components
2. **For practical use, a subset needs to be identified and named so it can be referred to unambiguously**

When a subset is represented by a reference set:

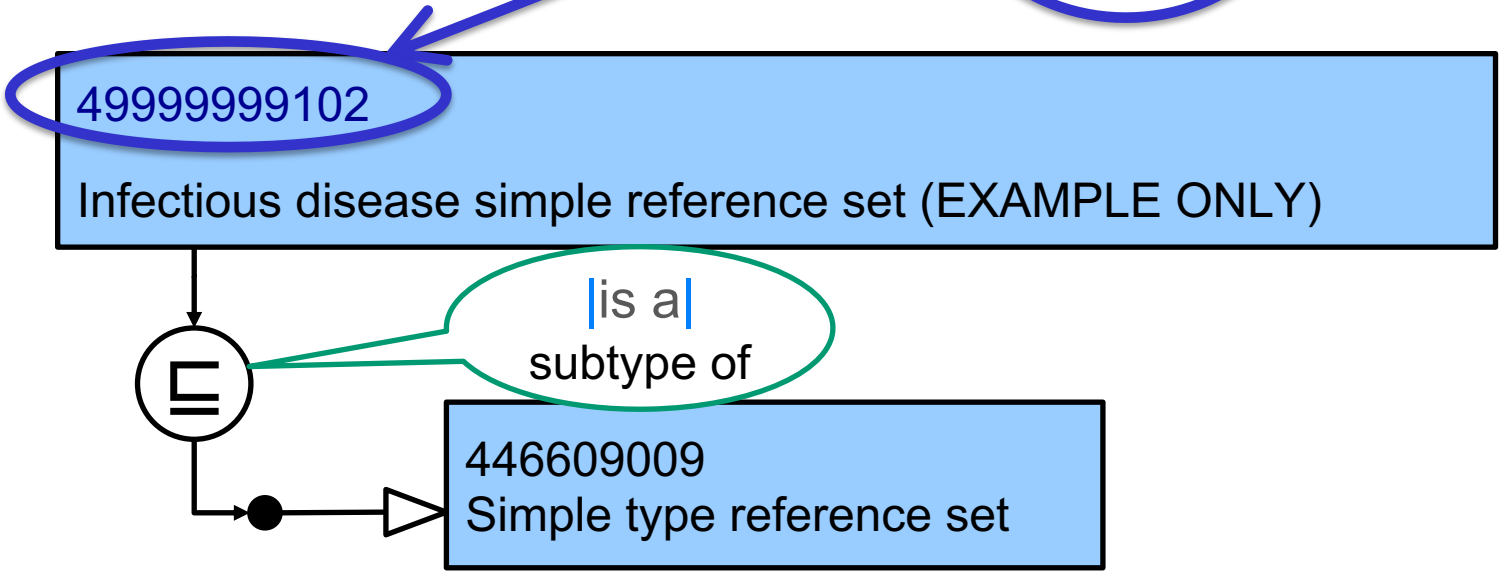
- The subset is identified by the **refsetId**
- The **refsetId** refers to a concept
- Descriptions of that concept name the reference set
- A relationship of that concept refers to the reference set type
  (in this case |is a| |simple type reference set|)



Subset Represented by a Simple Reference Set

Reference Set

id	effective Time	active	moduleId	refsetId	referenced ComponentId
0000d6608041-005b-5108-b608-07891b210365	20140731	1	19999999103	49999999102	82272006
00008774f374-5f54-5234-8af1-8aca6e73b5c2	20140731	1	19999999103	49999999102	6142004
00006ac9-2f97-5ed4-b37e-84cb158c388d	20140731	1	19999999103	49999999102	55604004



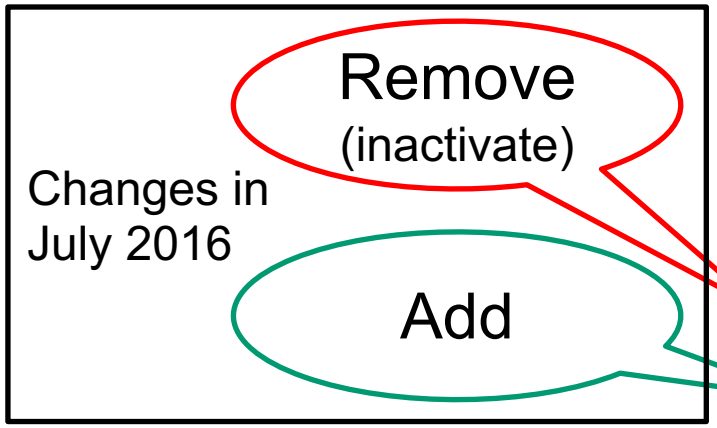
Data Structure Requirements for Subsets

1. A simple list identifiers can represent a subset of SNOMED CT components
2. For practical use, a subset needs to be identified and named so it can be referred to unambiguously
3. **Subset membership may need to change with future releases of SNOMED CT or due to evolving requirements for inclusion of different content**

When a subset is represented by a reference set:

- The standard SNOMED CT approach to versioning and modularization allows full tracking of changes.
- This uses the following columns
  - id, effectiveTime, active, moduleId

Subset Membership Changes Represented by a Simple Reference Set



Id	Term
82272006	Common cold
6142004	Influenza
~~55604004~~	~~Avian influenza~~
3928002	Zika fever

Reference Set

id	effective Time	active	moduleId	refsetId	referenced ComponentId
0000d6608041-005b-5108-b608-07891b210365	20140731	1	19999999103	49999999102	82272006
00008774f374-5f54-5234-8af1-8aca6e73b5c2	20140731	1	19999999103	49999999102	6142004
00006ac9-2f97-5ed4-b37e-84cb158c388d	20140731	1	19999999103	49999999102	55604004
00006ac9-2f97-5ed4-b37e-84cb158c388d	**20160731**	**0**	19999999103	**49999999102**	**55604004**
0000c5c4-50bb-5937-9141-1ec3e2d517e8	**20160731**	**1**	19999999103	**49999999102**	**3928002**

Data Structure Requirements for Subsets

1. A simple list identifiers can represent a subset of SNOMED CT components
2. For practical use, a subset needs to be identified and named so it can be referred to unambiguously
3. Subset membership may need to change with future releases of SNOMED CT or due to evolving requirements for inclusion of different content
4. **It may be useful to define the membership of a subset using rules rather than a list of identifiers**
   - **This is called an *intensional subset definition***

Not a misspelling !
Intensional is not the same as intentional

Extensional and Intensional Subset Definitions

- Extensional subset definitions
  - Subset membership defined by enumeration
    - Identifying each of the members individually

SNOMED CT supports extensional subset definitions:

- **Simple type reference sets** with members identified by component identifiers (referencedComponentId)

- Intensional subset definitions
  - Subset membership is defined by a set of rules
    - The rules are expressed as a query that computes the membership of a subset

SNOMED CT supports intensional subset definitions:

- **Query type reference sets** with rules represented as
- **SNOMED CT expression constraints**

Intensional Subset Definitions: Substrate and Expansion

Substrate

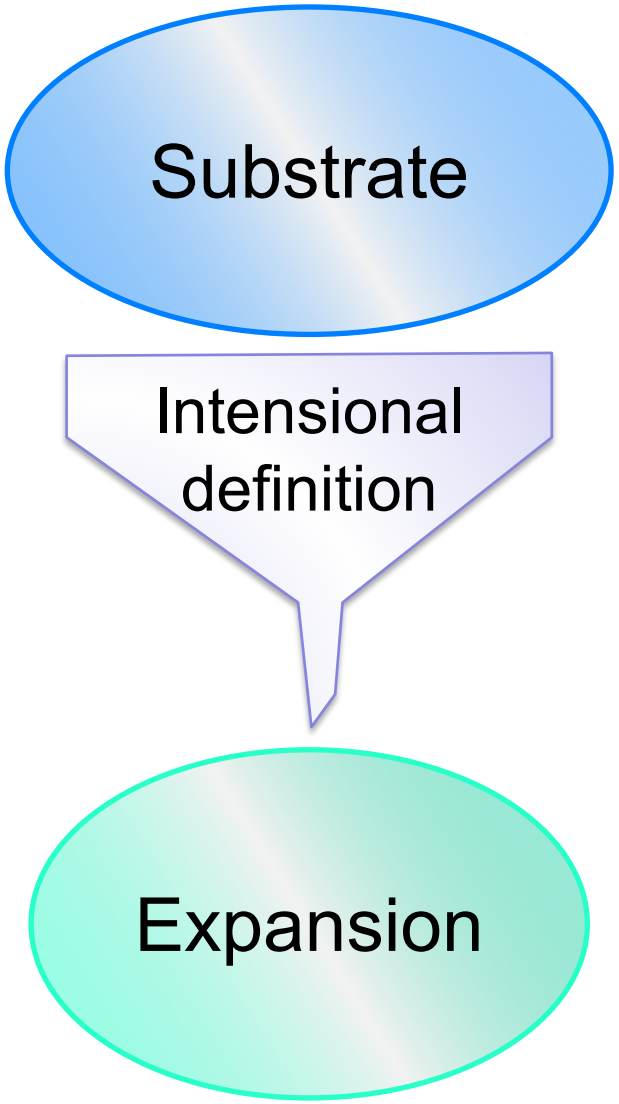
- The superset of members to which an intensional subset definition is applied

Intensional definition

- The set of rules (query) that defines whether a member of the substrate is included in the subset

Expansion

- The result of applying an intensional definition to a given substrate



Intensional Subset Definitions: Substrate and Expansion (General Example)

Substrate

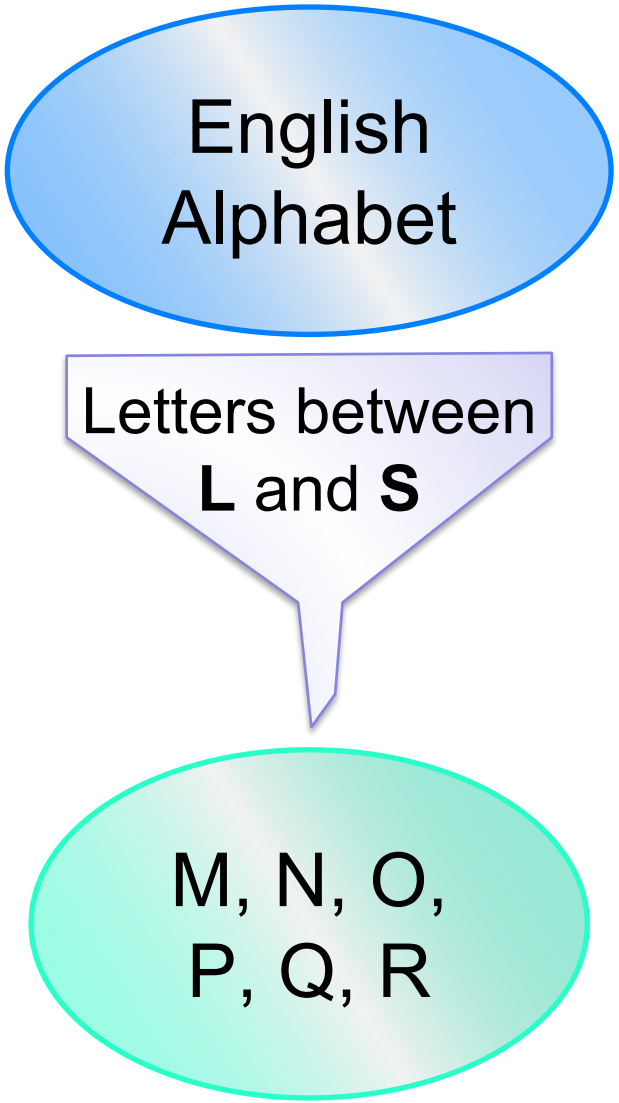
- The superset of members to which an intensional subset definition is applied

Intensional definition

- The set of rules (query) that defines whether a member of the substrate is included in the subset

Expansion

- The result of applying an intensional definition to a given substrate



Intensional Subset Definitions: Substrate and Expansion (SNOMED CT Example)

Substrate

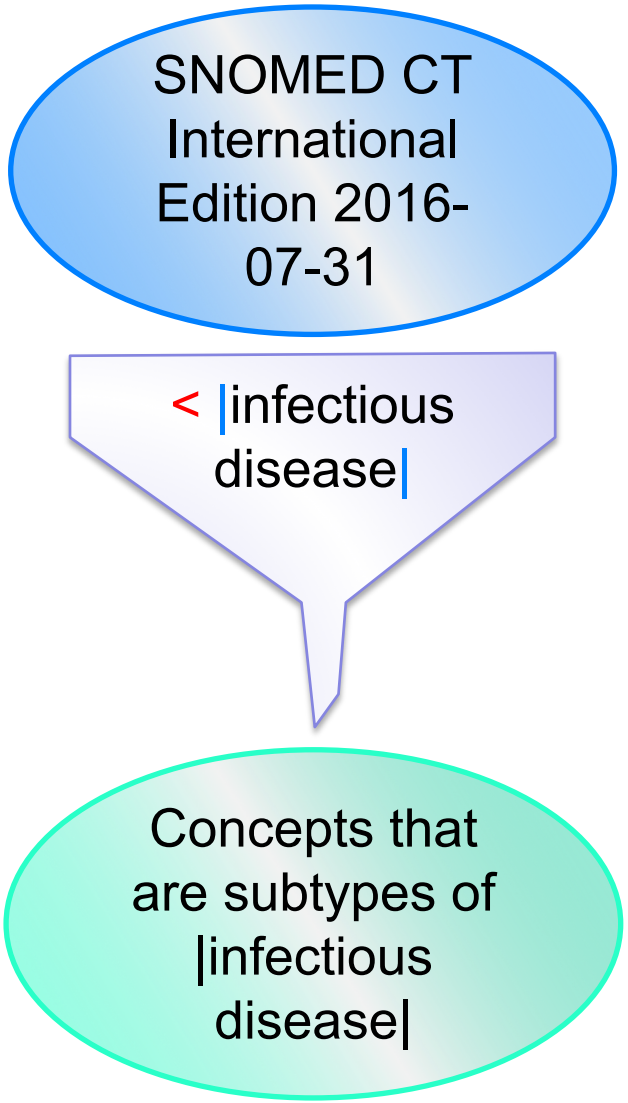
- The superset of members to which an intensional subset definition is applied
  - A specified version of a SNOMED CT edition

Intensional definition

- The set of rules (query) that defines whether a member of the substrate is included in the subset
  - A SNOMED CT expression constraint

Expansion

- The result of applying an intensional definition to a given substrate
  - Concepts in specified version of the edition that comply with the constraint



Query Specification Type Reference Set Representing Intensional Subset Definitions

To save space, **id** values are not shown for this example

The **query** column contains a text string using SNOMED CT expression constraint language to represent rules applied to the substrate to define the required subset

id	effective Time	active	moduleId	refsetId	referenced ComponentId	query
…	20160731	1	19999999103	59999999104	69999999101	< 40733004 |infectious disease|
…	20160731	1	19999999103	59999999104	79999999109	<40733004 |infectious disease| :363698007 |finding site| = <<20139000 |structure of respiratory system|

The default substrate for the query is the module identified by **moduleId** and all modules on which that module depends

A single identified **query reference set** may contain several members (rows) each with a separate query

The **referencedComponentId** identifies an *actual* or *virtual* reference set representing the expansion created by applying the query to the default substrate

Reference Set Requirements Beyond Subsets

Requirements for Ordered Lists

- Requirements for ordered lists of descriptions
  - Presenting terms in an order that is rational or helpful for a particular purpose in user interface controls including:
    - Simple lists
    - Drop down lists
    - Popup menus
- Requirements for ordered lists of concepts
  - Presenting concepts in an order that is rational or helpful for a particular purpose irrespective of the term displayed

    Examples
    - Body parts that have a natural order
      - Fingers, cranial nerves and vertebrae
    - Enumerated values
      - Mild, moderate, severe

Requirements for Prioritization

- Prioritization is similar to order but multiple components may have the same rank
- Requirements for prioritization of concepts
  - Making it easier to find concepts that are most commonly used in a particular specialty, department or data entry scenario
  - Highlighting concepts that are preferred options for a particular purpose without preventing access to a wider selection of concepts
- Priorities may be implemented in various ways
  - Highlight high priority items
  - Only show prioritized items initially with a display more option

Ordered Component Type Reference Set Representing Display Order or Priority

Fingers sorted A-Z	
21356012	**Fifth finger**
136021011	**Fourth finger**
138873019	**Second finger**
108884010	**Third finger**
127053016	**Thumb**

The **order** column is an integer that specifies:
- Display order
- Priority rank: 1 (first) highest rank

id	effectiveTime	active	moduleId	refsetId	referencedComponentId	order
…	20160731	1	19999999103	809999999103	127053016	1
…	20160731	1	19999999103	809999999103	138873019	2
…	20160731	1	19999999103	809999999103	108884010	3
…	20160731	1	19999999103	809999999103	136021011	4
…	20160731	1	19999999103	809999999103	21356012	5

- **Thumb**
- **Second finger**
- **Third finger**
- **Fourth finger**
- **Fifth finger**

Requirements for Adding Information to a Referenced Component

- Displaying a textual note or comment about a listed or selected concept
  - For example, a guidance note on requesting a particular procedure or service
- Marking descriptions with indications of whether the terms they contain are acceptable or preferred in a specified language or dialects
  - For example, distinguishing term usage between different languages, dialects, local or specialty groups
- Marking particular concepts with specific values to provide processable and/or displayable information
  - For example, marking inactive concepts with indicators of the reasons for inactivating them

Annotation Type Reference Set
Adding Unstructured Information to a Concept

Id	Term
82272006	Common cold
6142004	Influenza
55604004	Avian influenza
3928002	Zika fever

The **annotation** column is an string that contains a free text note about the referenced component.

id	effective Time	active	moduleId	refsetId	referenced ComponentId	annotation
…	20160731	1	19999999103	119999999106	82272006	Trivial just treat symptoms.
…	20160731	1	19999999103	119999999106	6142004	Risk in elderly. Advise immunization in future years.
…	20160731	1	19999999103	119999999106	55604004	Report to public health is suspected. Admit if any respiratory distress.
…	20160731	1	19999999103	119999999106	3928002	Indicate if pregnant or potentially pregnant. Note travel in affected area.

Language Type Reference Set Indicating Acceptability of Descriptions

The **refsetId** specifies the refset for a language

referencedComponentId refers to a description to which an acceptability applies

acceptabilityId refers to a concept that represents |acceptable| or |preferred|

id	effective Time	active	moduleId	refsetId	referenced ComponentId	acceptabilityId
…	20160731	1	900000000000207008	**900000000000509007**	132967011	900000000000548007
…	20160731	1	900000000000207008	**900000000000509007**	132972019	900000000000549004
…	20160731	1	900000000000207008	**900000000000508004**	132972019	900000000000549004
…	20160731	1	900000000000207008	**900000000000508004**	132973012	900000000000548007

132967011|Appendectomy| |*Preferred*|
132972019|Excision of appendix| |*Acceptable*|

132973012|Appendicectomy| |*Preferred*|
132972019|Excision of appendix| |*Acceptable*|

Attribute Value Type Reference Set
Adding Structured Data to a Concept

The **valueId** column refers to a concept that represent the additional data applicable to the concept

id	effective Time	active	moduleId	refsetId	referenced ComponentId	valueId
…	20160731	1	900000000000207008	900000000000489007	1720006	900000000000482003
…	20160731	1	900000000000207008	900000000000489007	4515009	900000000000484002
…	20160731	1	900000000000207008	900000000000489007	4961001	900000000000485001
…	20160731	1	900000000000207008	900000000000489007	5381002	900000000000483008

1720006 |Dextroposition of aorta| |*Duplicate*|
4515009 |Keratoderma punctata| |*Ambiguous*|
4961001 |Psychiatric function| |*Erroneous*|
5381002 |Enterobacter gergoviae| |*Outdated*|

Requirements for Associations Between Components

- Historical associations between components
  - For example, associating an concept that has been inactivated as a duplicate with the active concept that it duplicates
- Grouping concepts together
  - For example, representing categories containing concepts that are used for reporting
- Presenting alternative hierarchical views of concepts of descriptions at the user interface data entry

Association Type Reference Set Representing Unordered Directed Associations

referencedComponentId refers to a component that is the source of the association

The **targetComponentId** refers to a component that is associated with the referenced component

id	effective Time	active	moduleId	refsetId	referenced ComponentId	targetComponentId
…	20160731	1	19999999103	149999999107	82272006	85562004
…	20160731	1	19999999103	149999999107	6142004	70327001
…	20160731	1	19999999103	149999999107	55604004	70327001
…	20160731	1	19999999103	149999999107	3928002	70327001

Ordered Association Type Reference Set Representing Ordered Directed Associations

referencedComponentId refers to a component that is the source of the association

targetComponentId refers to a component that is associated with the referenced component

id	effective Time	active	moduleId	refsetId	referenced ComponentId	target ComponentId	order
…	20160731	1	19999999103	819999999101	127053016	85562004	1
…	20160731	1	19999999103	819999999101	138873019	70327001	1
…	20160731	1	19999999103	819999999101	108884010	70327001	2
…	20160731	1	19999999103	819999999101	136021011	70327001	3
…	20160731	1	19999999103	819999999101	21356012	70327001	4

order indicates the order of the association

Associations and Maps to Other Code System

referencedComponentId refers to a component that is the concept that is mapped

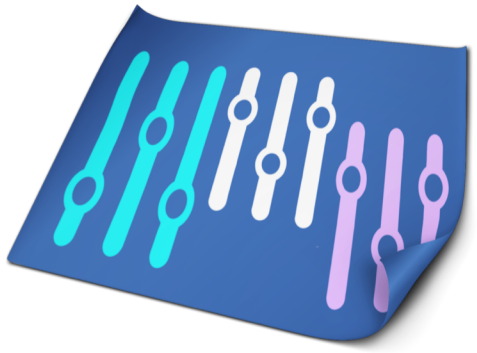
mapTarget is a code in the other code system

id	effective Time	active	moduleId	refsetId	referenced ComponentId	mapTarget	*… other attributes …*
…	20160731	1	19999999103	169999999108	9389005	B00.5	
…	20160731	1	19999999103	169999999108	13608004	B00.5	
…	20160731	1	19999999103	169999999108	13710008	B00.7	
…	20160731	1	19999999103	169999999108	3401001	B00.8	
…	20160731	1	19999999103	169999999108	16592009	B00.8	

Other attributes depend on the complexity of maps between SNOMED CT and the other code system. The attributes may represent: correlation, options, groups, rules, advice, etc.

SNOMED International

Reference Set Design and Defined Types



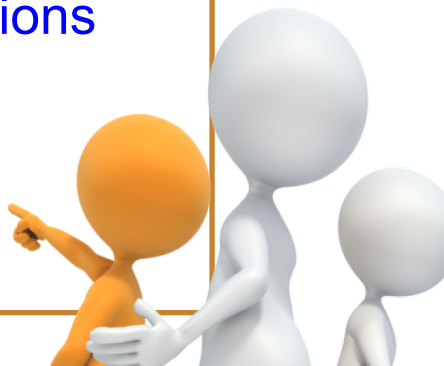
Reference Set Types

General versioning information	id	
	effectiveTime	
	active	
	moduleId	
Reference set identifier	refsetId	
Component reference	referencedComponentId	
Additional information depending on **reference set type**	*<attribute-1 … attribute-n>*	• *component (id)* • *string* • *integer*

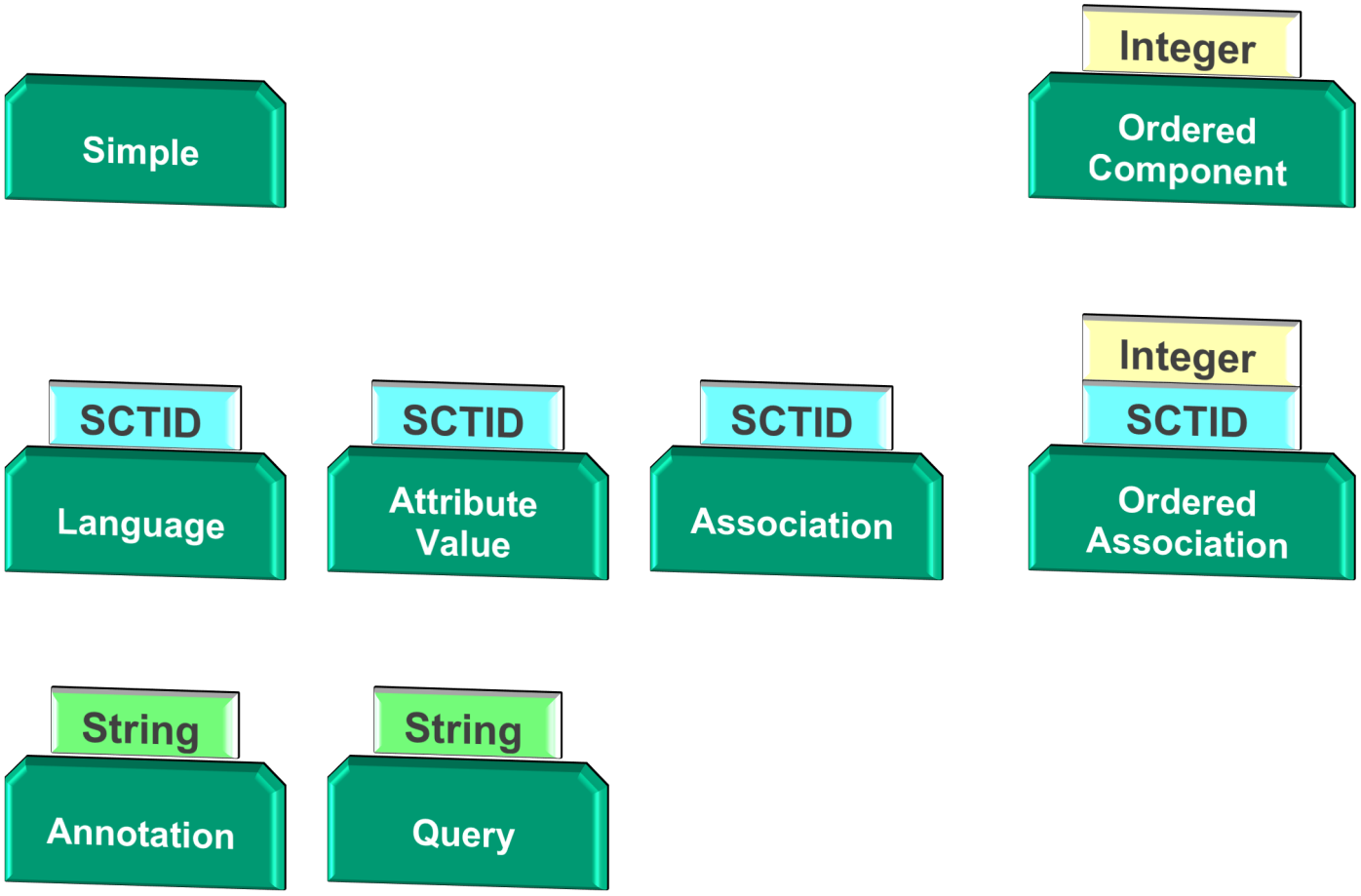
Each additional attribute has one of a set of defined basic data types

Reference set types are defined with particular patterns of additional attributes to meet specific requirements

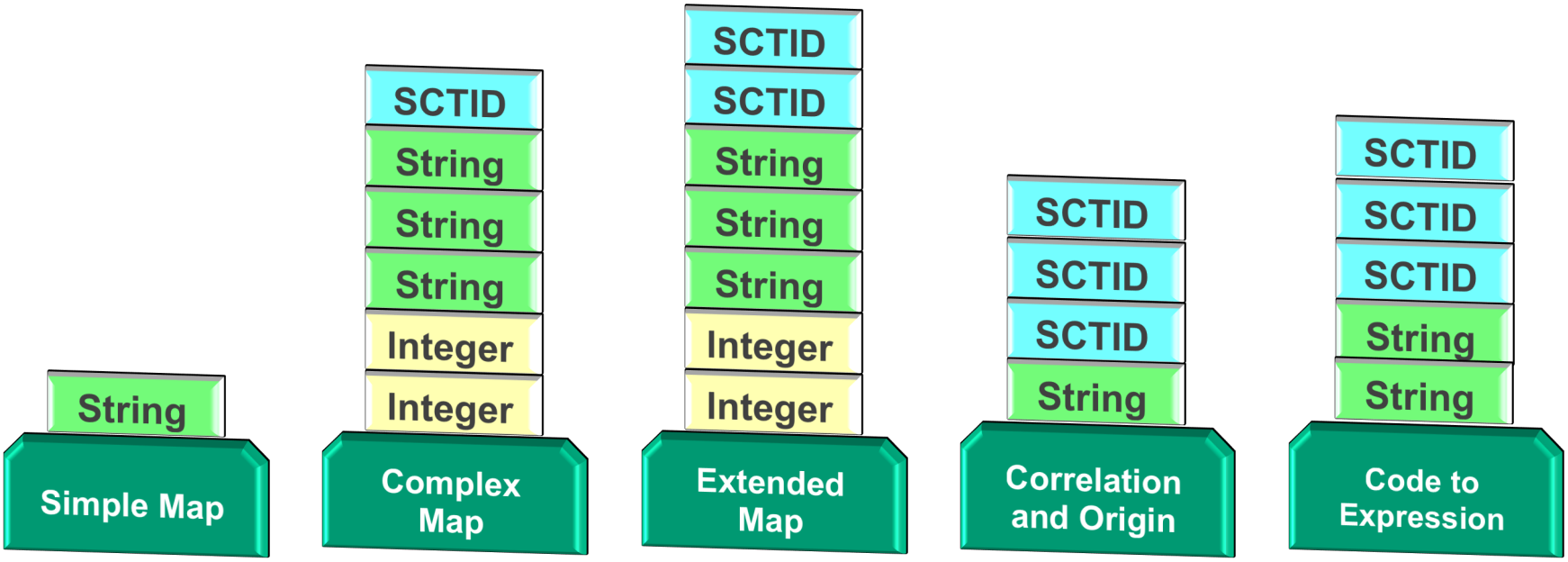
- Reference set types that meet many common requirements are predefined in SNOMED CT release file specifications
- New reference set types can be defined to meet emerging requirements
- Customized reference set types can also be defined as part of an extension



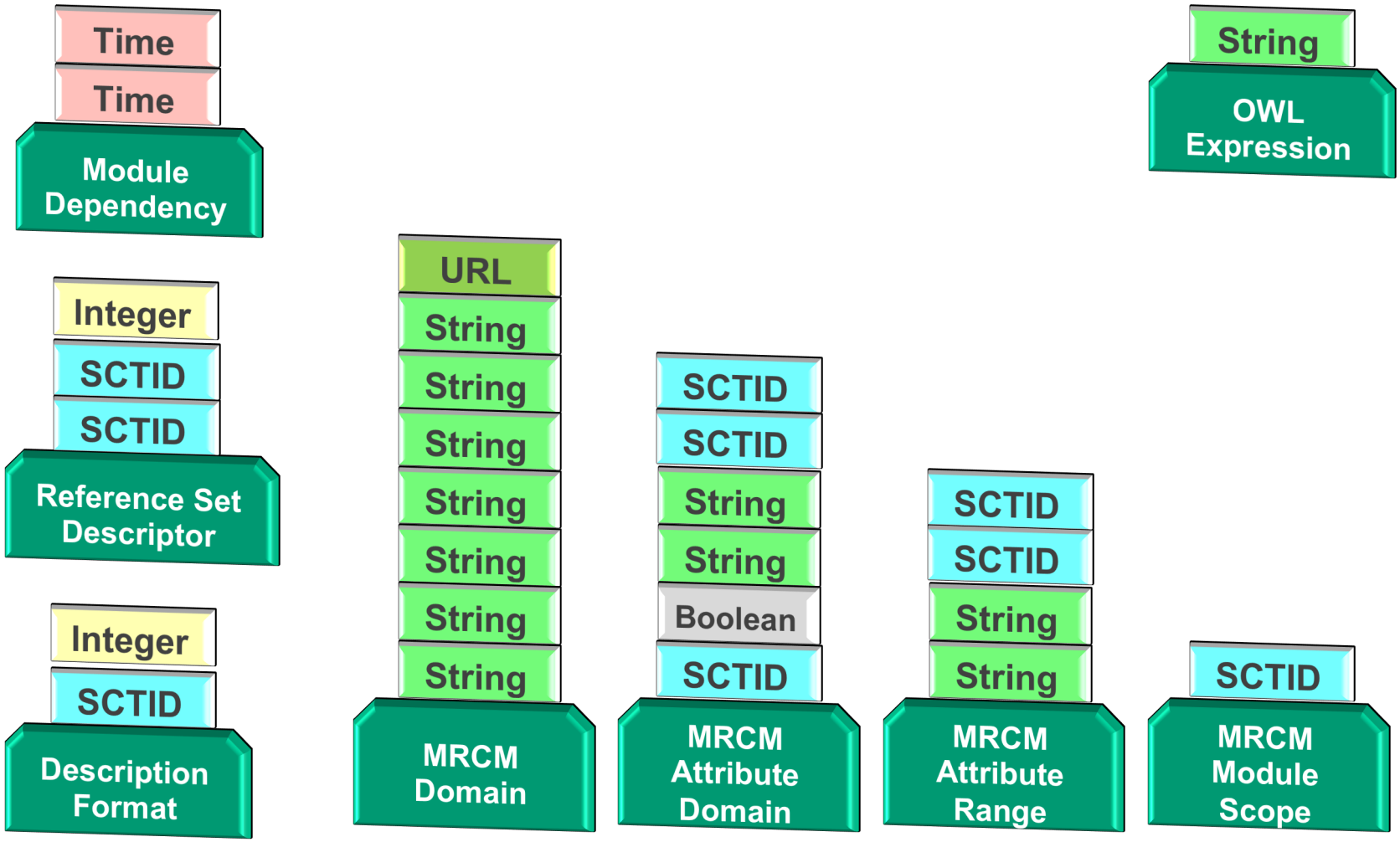
Widely Used Reference Set Types



Mapping Reference Set Types



Reference Set Types with Technical Uses



Reference Set Customization to Meet Additional Requirements

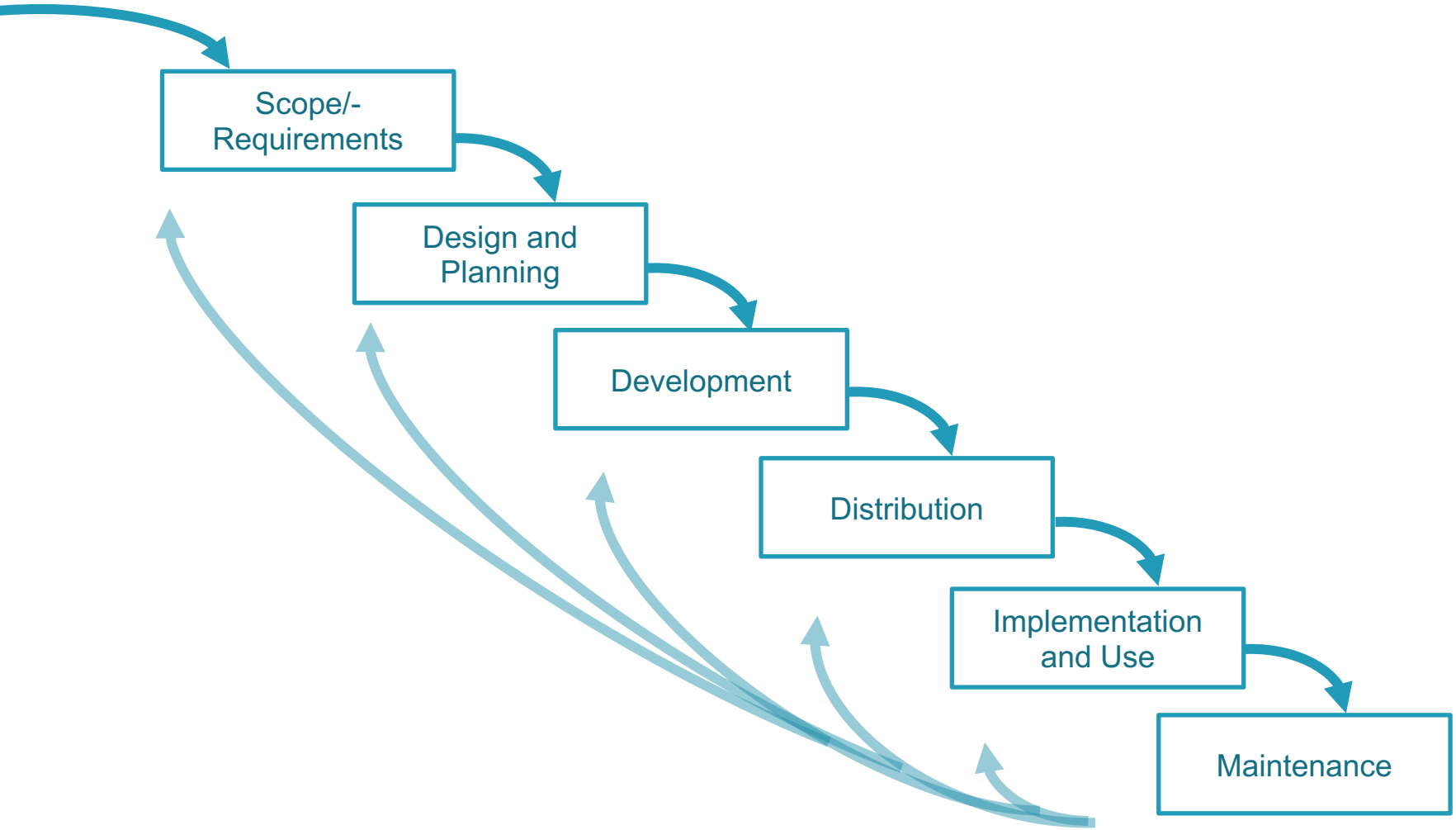
- The reference set specification defines a special reference set that can be used to specify the structure of other reference sets
  - The **reference set descriptor reference set**
- Therefore …
  - Reference sets offer a standard SNOMED CT format for representing and versioning collections of data linked to SNOMED CT components
  - Reference sets can meet a wide variety of requirements. There may be other ways to meet those requirements but the added value of reference sets is the standard systematic approach to versioning and change tracking

Overview

- Part 1. Introduction to Reference Sets
  - Purpose of refsets
  - Subsets, value set and reference sets
  - Requirements design
  - Requirements, types and examples
- Part 2. Reference Set Life Cycle
  - Overview of development process
  - Practical examples
  - Questions

Refset Lifecycle

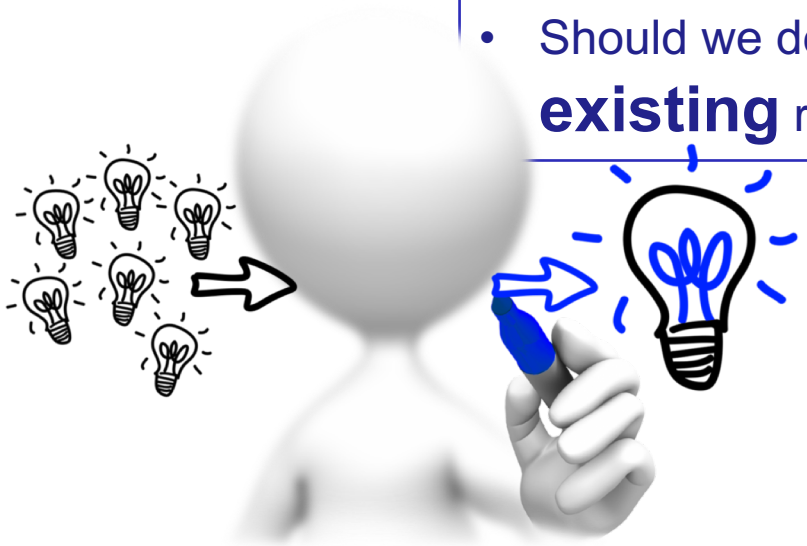
Reference Set Lifecycle



Scope and Requirements

Scope/-
Requirements

- What is the main **purpose** of the reference set?
- Who are the **users** of the reference set?
- What is the **scope** of content of the reference set?
- Who is **responsible** for developing and maintaining the reference set?
- Should we develop a **new** reference set **or** does any **existing** reference set meet our requirements?



Design and Planning

Scope/-
Requirements

Design and
Planning

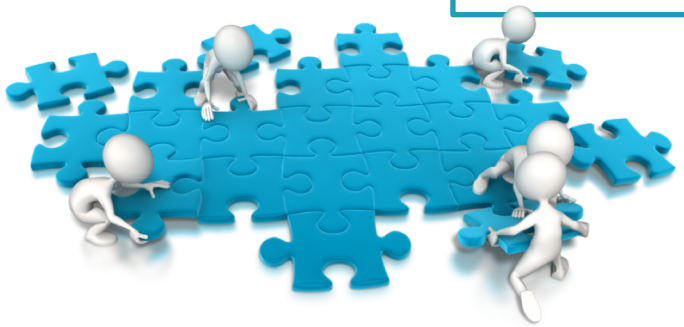
- Resources
- Dependencies
- Reference set type
- People
- Development approach and tools
- Timeline
- Quality assurance
- Distribution, integration and maintenance

Development

Scope/-
Requirements

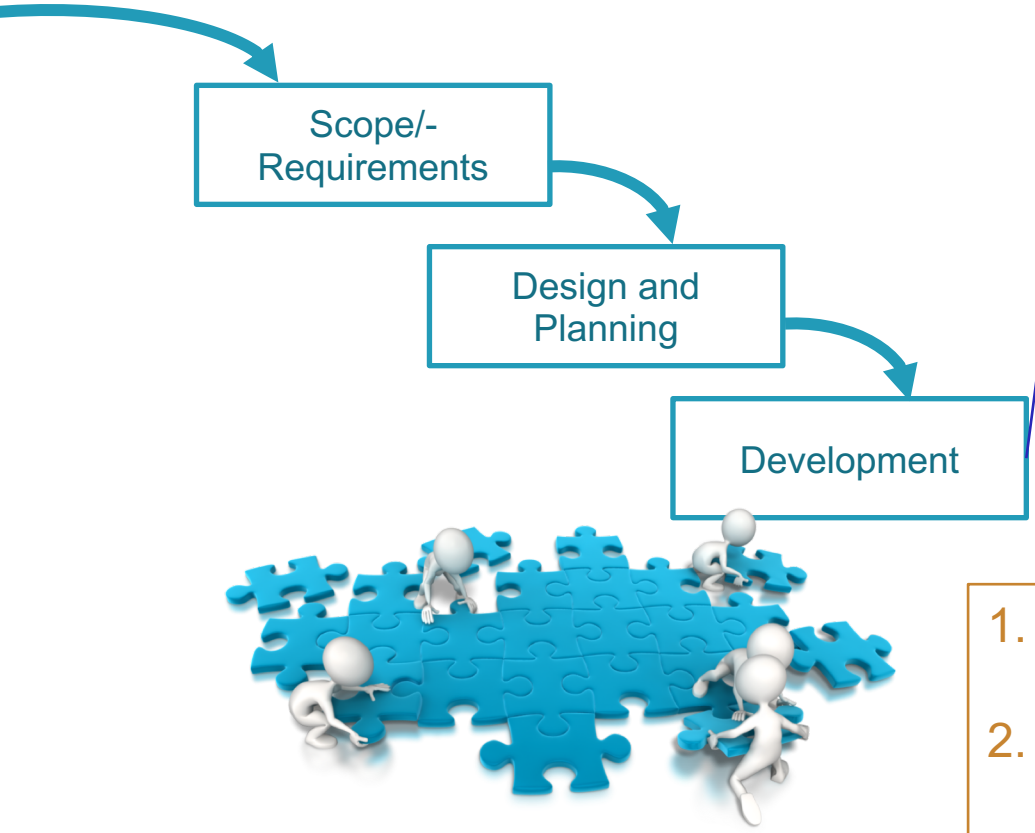
Design and
Planning

Development



- Reference Set Creation
- Development Approaches
- Development Methods
- Review and Quality Assurance

Development - Reference Set Creation



- **Reference Set Creation**
- Development Approaches
- Development Methods
- Review and Quality Assurance

1. Define the reference set in the metadata hierarchy
2. Define the reference set attributes within the metadata hierarchy
3. Create the Descriptor for the reference set
4. Add members to the reference set

Development Approaches

Scope/-
Requirements

Design and
Planning

Development

- Reference Set Creation
- **Development Approaches**
- Development Methods
- Review and Quality Assurance

Create new
reference set

Adopt existing
reference set

Adapt existing
reference set

Copy existing
reference set

- No existing reference set meets your requirements
- The scope of the reference set is limited and clearly specified

Create new reference set

- The reference set meets your requirements
- You are confident that the existing reference set will be well-maintained
- The reference set is part of a module that is included in the SNOMED CT Edition that you use

Adopt existing reference set

Adapt existing reference set

Copy existing reference set

- The reference set almost meets your requirements, but you wish to modify it to fully fulfill your requirements

- The reference set meets your requirements
- You are NOT confident that the existing reference set will be well-maintained
- The existing reference set is NOT part of a module that is included in the SNOMED CT Edition that you use

Development Methods

Scope/-
Requirements

Design and
Planning

Development

- Reference Set Creation
- Development Approaches
- **Development Methods**
- Review and Quality Assurance

EXTENSIONAL
DEFINITION

INTENSIONAL
DEFINITION

Development Methods:

Extensional Definition

- **Membership is defined by enumeration**
  - Identifying each of the members individually

- General finding of observation of patient (finding)
  - Finding of general observation of appearance (finding)
    - Beautiful appearance (finding)
    - Bizarre personal appearance (finding)
    - Blue bloater (finding)
    - Comfortable appearance (finding)
    - Finding of appearance relating to age (finding)
    - Finding of general form of body (finding)
    - General appearance of patient (finding)
    - Generally clean appearance (finding)
      - On examination - clean and tidy (finding)
    - Neglected appearance (finding)
      - On examination - care poor (finding)
    - No injuries apparent (finding)
    - Normal appearance (finding)
    - Normal extension (finding)
    - Normal flexion (finding)
    - On examination - appearance (finding)
    - Parkinsonian features (finding)
    - Personal appearance about reported age (finding)
    - Personal appearance dirty (finding)
      - Clothing dirty (finding)
      - Hands dirty (finding)
      - On examination - unclean (finding)
    - Pink puffer (finding)
    - State of dress finding (finding)
      - Clothing dirty (finding)
      - Dressed - appearance (finding)
      - Undressed (finding)
    - Well cared for appearance (finding)
      - Adequately dressed (finding)
      - On examination - care good (finding)
      - Smartly dressed (finding)

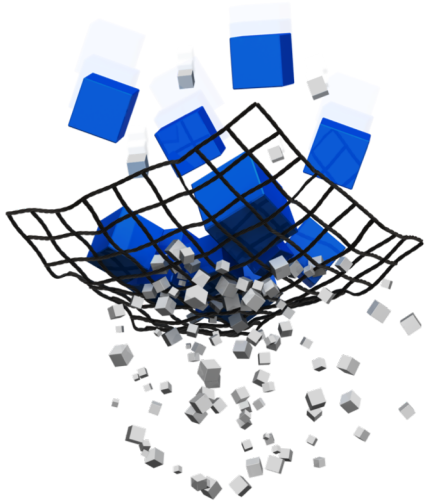
Example:

- 80670004 | Bizarre personal appearance (finding) |
- 398168006 | Generally clean appearance (finding) |
- 102892007 | Neglected appearance (finding) |
- 386549008 | Normal appearance (finding) |
- 78158007 | Personal appearance dirty (finding) |
- 397749000 | Well cared for appearance (finding) |

Development Methods:

Intensional Definition

- **Membership is defined by a set of rules**
  - The rules are expressed as a query that computes the reference set members
- **SNOMED CT Expression Constraint**
  - A computable rule that can be used to define a bounded set of clinical meanings
  - http://snomed.org/ecl



Examples:

<< 387714009 |Operation on lung (procedure)|

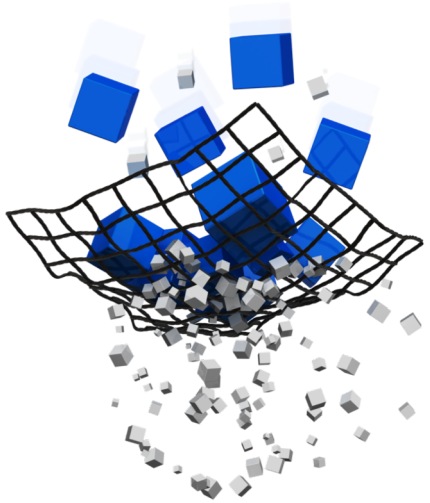
< 19829001 |Disorder of lung (disorder)| :
116676008 |associated morphology| = 79654002 |edema|

Development Methods:

Intensional Definition

SUBSTRATE

SNOMED CT
International Edition
2016-07-31



INTENSIONAL DEFINITION

< 19829001 | Disorder of lung (disorder) | :
116676008 | Associated morphology | =
79654002 | Edema |

EXPANSION

Neurogenic pulmonary edema (disorder)	233705000
Adult respiratory distress syndrome (disorder)	67782005
Hemorrhagic pulmonary edema (disorder)	276637009
Postimmersion-submersion syndrome (disorder)	89687005
Fluid overload pulmonary edema (disorder)	233712009
Pulmonary edema (disorder)	19242006

Development Methods:

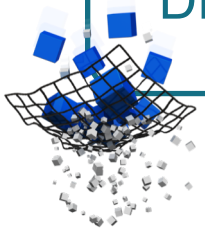
Refset Refinement

OVERALL APPROACH

REFINEMENT APPROACH

EXTENSIONAL DEFINITION

INTENSIONAL DEFINITION



INCLUSION

EXCLUSION

EXTENSIONAL

INTENSIONAL

EXTENSIONAL

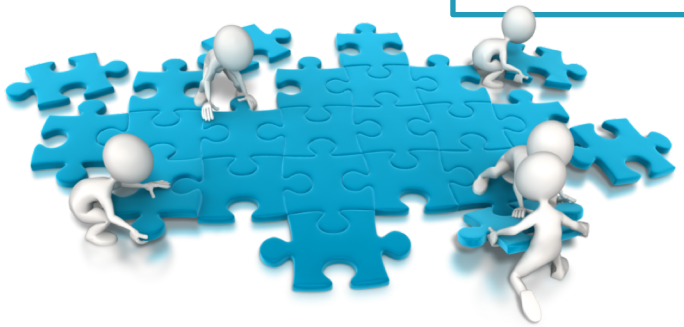
INTENSIONAL

Development – Quality Assurance

Scope/-
Requirements

Design and
Planning

Development



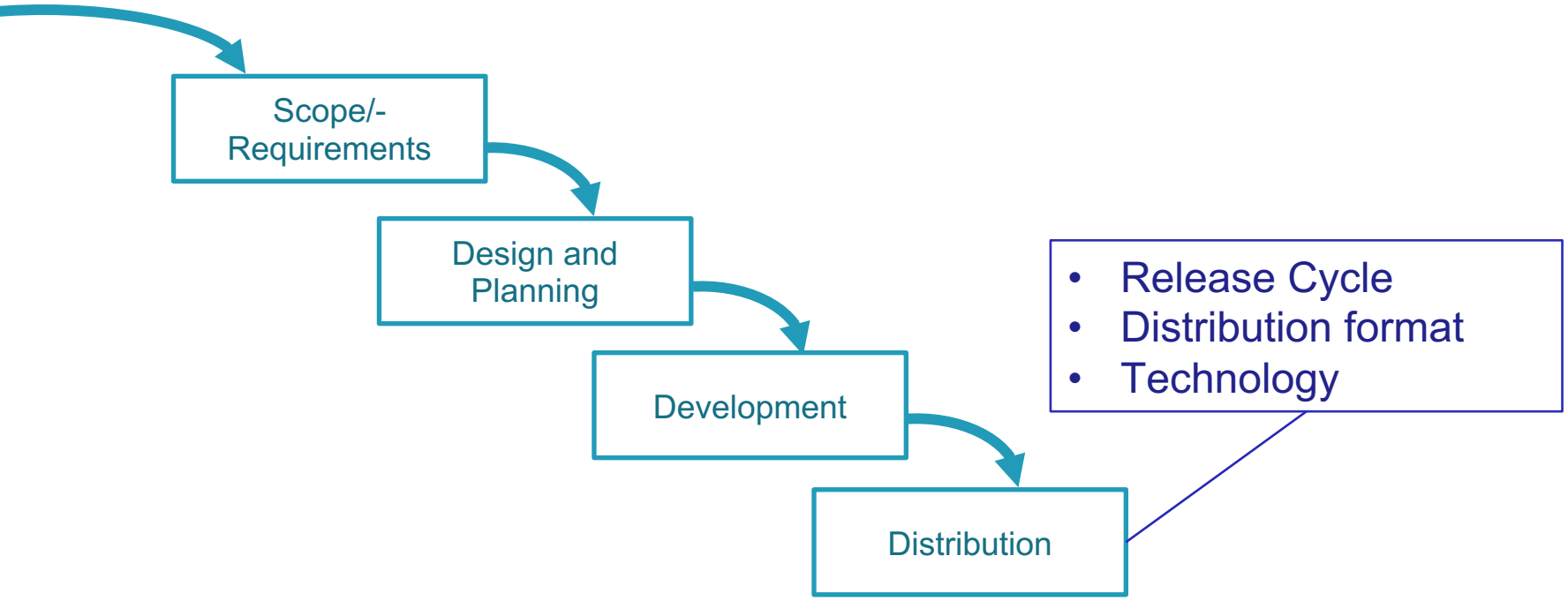
- Reference Set Creation
- Development Approaches
- Development Methods
- **Review and Quality Assurance**

1. Single author, single reviewer
2. Cross-review
3. Dual blind review with an adjudicator
4. Workshop

Important that people involved with review are aware of the requirements, scope and intended uses of the reference set

!

Distribution



Implementation and Use - Guidance

Scope/- Requirements

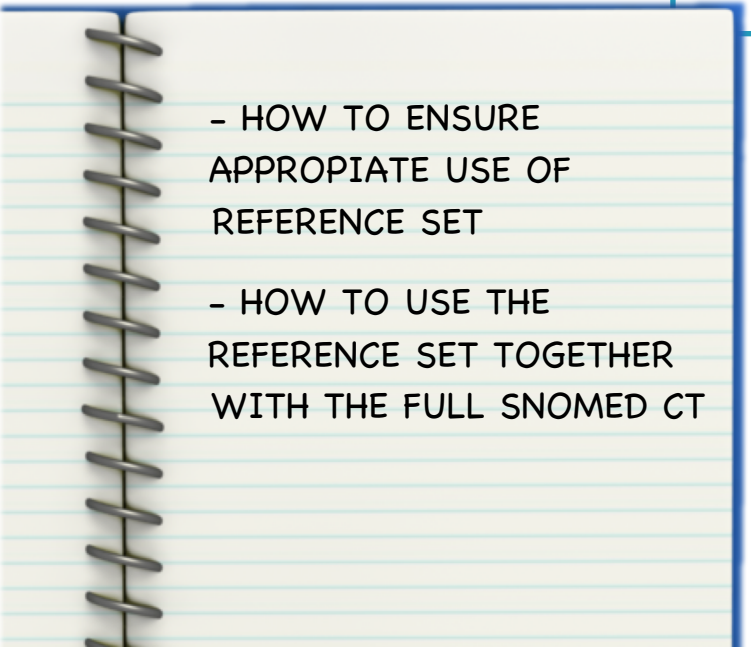
Design and Planning

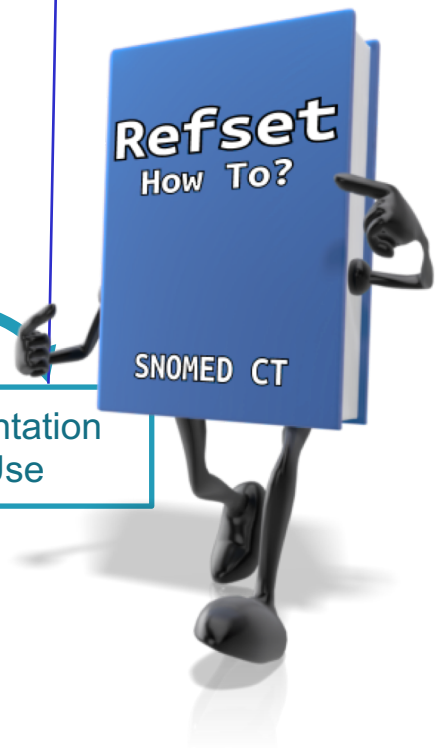
Development

Distribution

Implementation and Use

- Implementation Guidance
- Optimization for Implementation

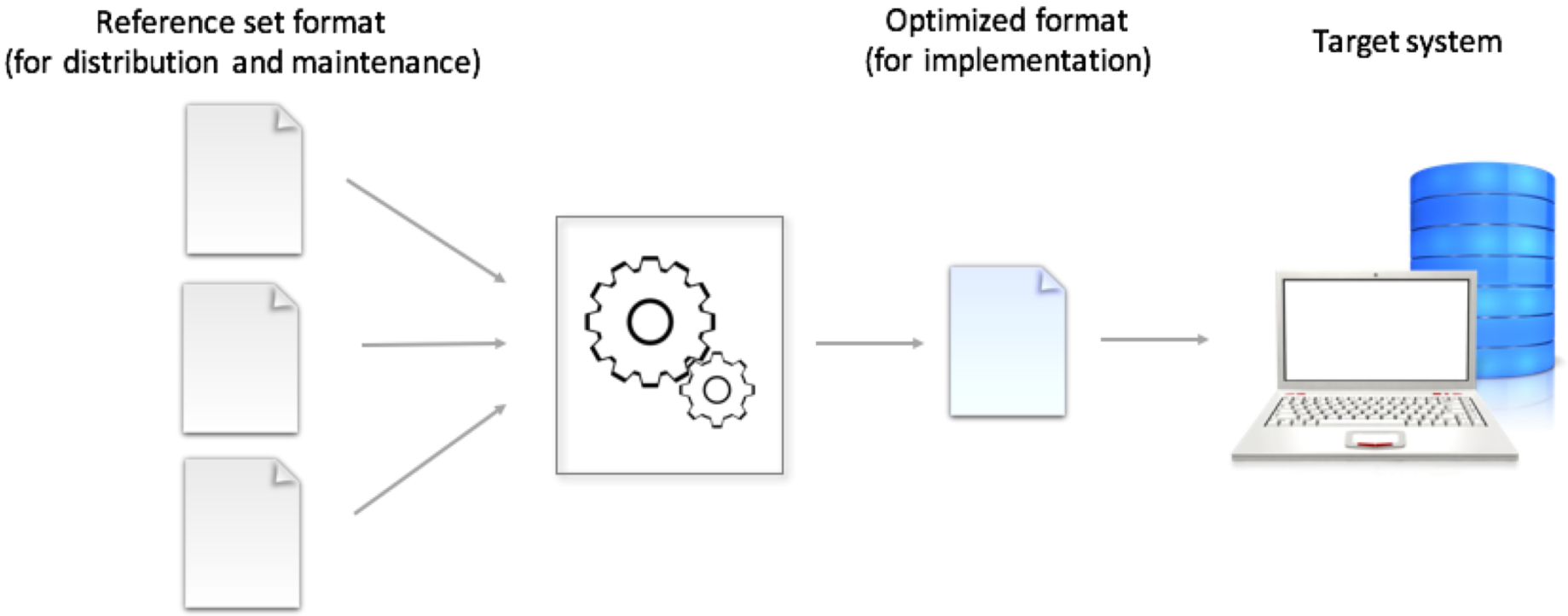




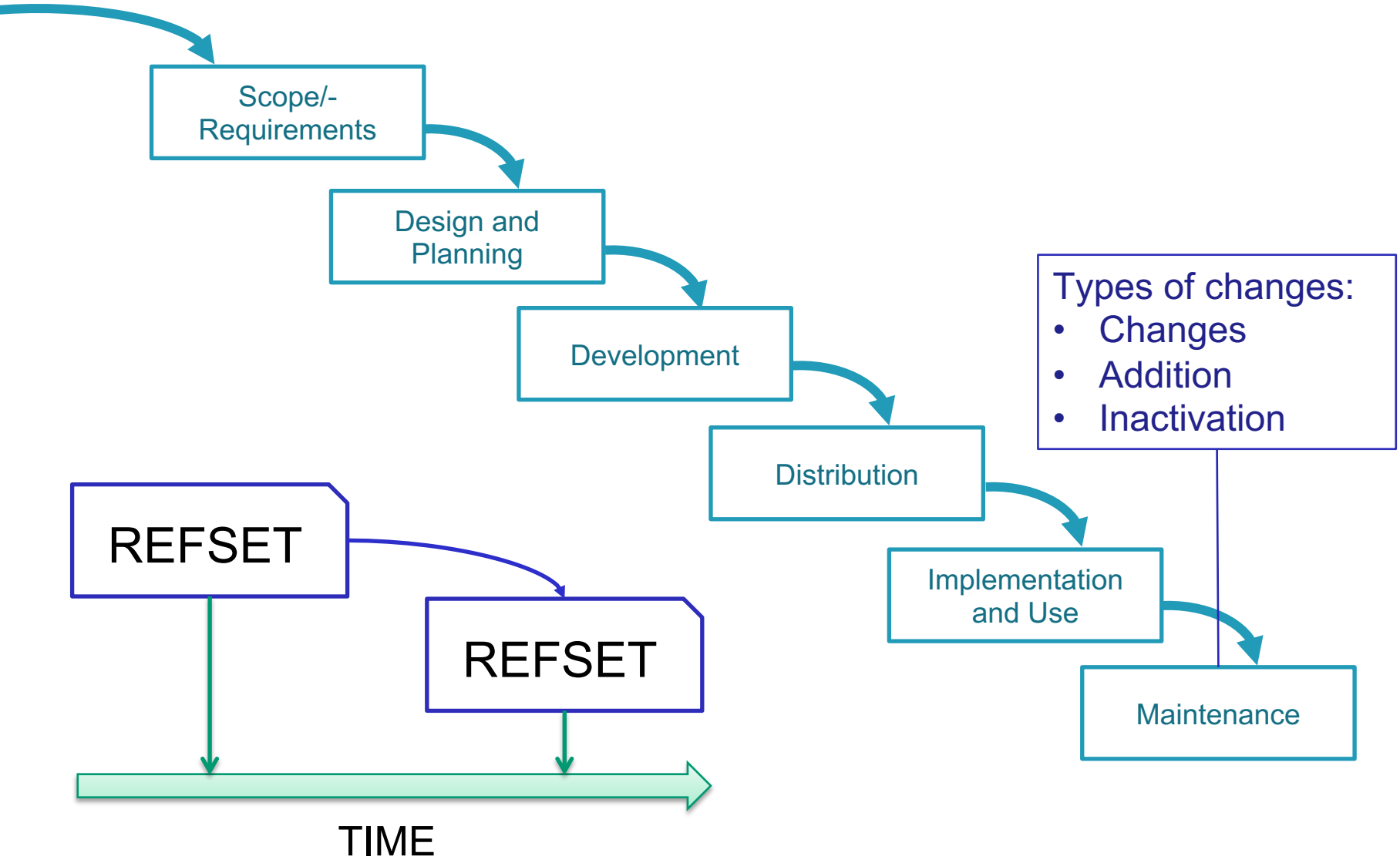
Implementation and Use – Optimization

- Implementation Guidance
- **Optimization for Implementation**

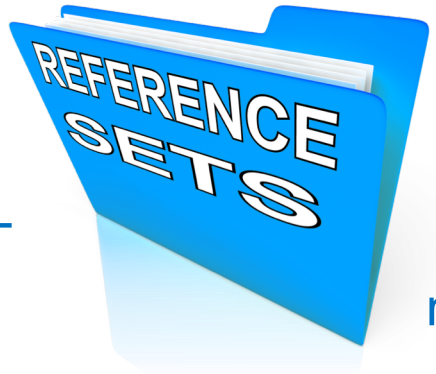
Implementation and Use



Maintenance



Maintenance

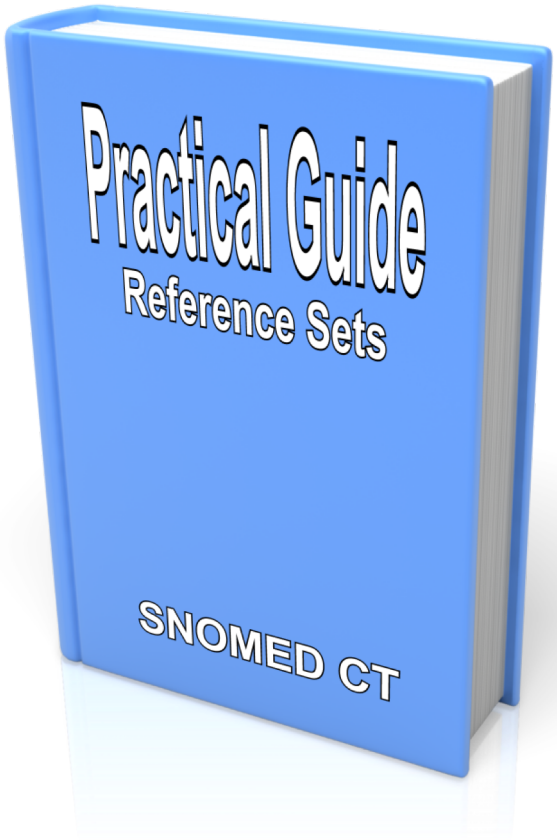


Changes to SNOMED CT may require changes to reference sets

Requests from users may require changes to reference sets

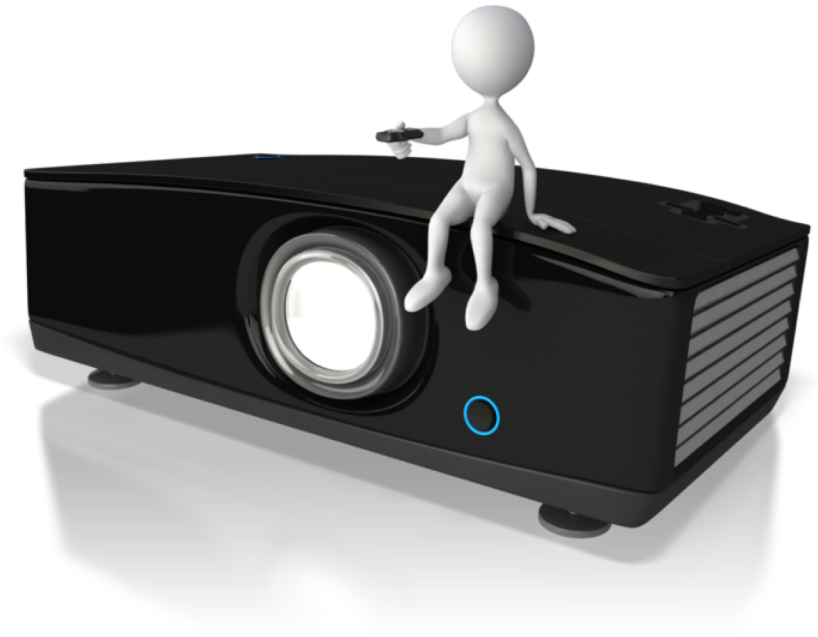
Overview

- Part 1. Introduction to Reference Sets
  - Purpose of refsets
  - Subsets, value set and reference sets
  - Requirements design
  - Requirements, typ
- Part 2. Reference
  - Overview of develop nt pr ess
  - Practical example
  - Questions



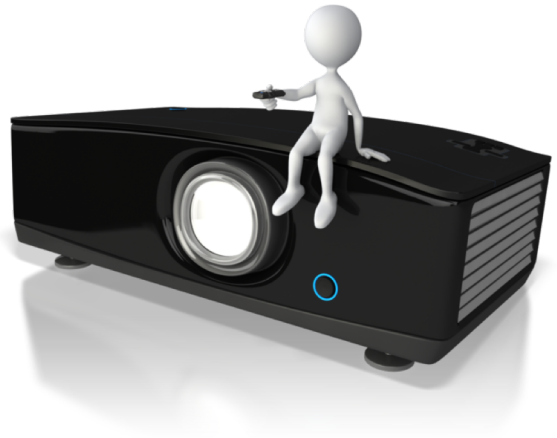
SNOMED International

Practical Example



Demonstration

- SNOMED International Refset Management Tool
  - http://snomed.org/tools
  - https://refset.ihtsdotools.org
- Refset Management Tool
  - Enables the management & creation of reference sets against the International Edition of SNOMED CT and Member extensions.
  - Provide a directory of existing reference sets that can be searched and downloaded to be used by others



Demonstration

- Extensional Refset Example:
  - Rare Diseases concepts

    7199000, 9014002, 13213009, 16631009, 22053006, 23238000, 24700007, 30188007, 31323000, 44785005, 51615001, 58606001, 62067003, 63702009, 65389002, 65880007, 74911008, 75053002, 76670001, 77128003, 80651009, 82275008, 86044005, 128241005, 190794006, 195353004, 230418006, 230791000, 234542004, 236403004, 239928004, 252246005, 387732009, 387759001, 396338004, 410795001, 417357006, 423590009, 699310000, 716997004
  - Concept to add

    190905008 | Cystic fibrosis (disorder) |

- Intensional Refset Example
  - Lung Procedure Refset

    << 387714009 | Operation on lung (procedure) |
  - Refset of lung disorders with an associated morphology equal to descendantsOrSelf of inflammation.

    < 19829001 : 116676008 = << 23583003

Overview

- Part 1. Introduction to Reference Sets
  - Purpose of refsets
  - Subsets, value set and reference sets
  - Requirements design
  - Requirements, types and examples
- Part 2. Reference Set Life Cycle
  - Overview of development process
  - Practical example
  - Questions

SNOMED International

Questions?