Proposal for Terminology Binding Syntax

Dr. Rahil Qamar Siddiqui
Charlie McCay
UK

Outline

- Definitions
- Requirements
- Terminology Binding Approaches
- Proposed Syntax
- Design Decisions
- Benefits
- Issues

What are Information Models and Terminology Binding

- *Information model :* A formal description of how information may be structured, interrelated and accessed.
  - ISO13606 archetypes, *open*EHR archetypes, HL7-based models

- *Terminology binding :* Linking of information model components to one or more concepts in a terminology.
  - **SNOMED CT** used for binding.

What are Expression Constraints?

- *Expression:* Collection of references to one or more concepts used to express a clinical idea.

- *Expression Constraint:* Computable rule that can be applied to a SNOMED CT *expression* to test it's compliance with rules that may relate to its meaning and/or compositional structure.

Requirements of Terminology Binding

- The *meaning* of information model components should be unambiguously defined.
- The *values* of information model components should be unambiguously defined using SNOMED CT concepts.
- The *semantic relationships* in information models should be unambiguously represented.
- The *semantic equivalence* between different information models should be unambiguously determined by enabling.

Out-of-scope for Terminology Binding

- Mapping SNOMED CT to other terminologies and coding and classification schemes such as ICD-10, Read Codes, and others.
- Providing a specification for the creation of reference sets.
- Validating the consistency within models and between models using an underlying ontology.
- Validating semantic interoperability between two systems using different information and recording structures

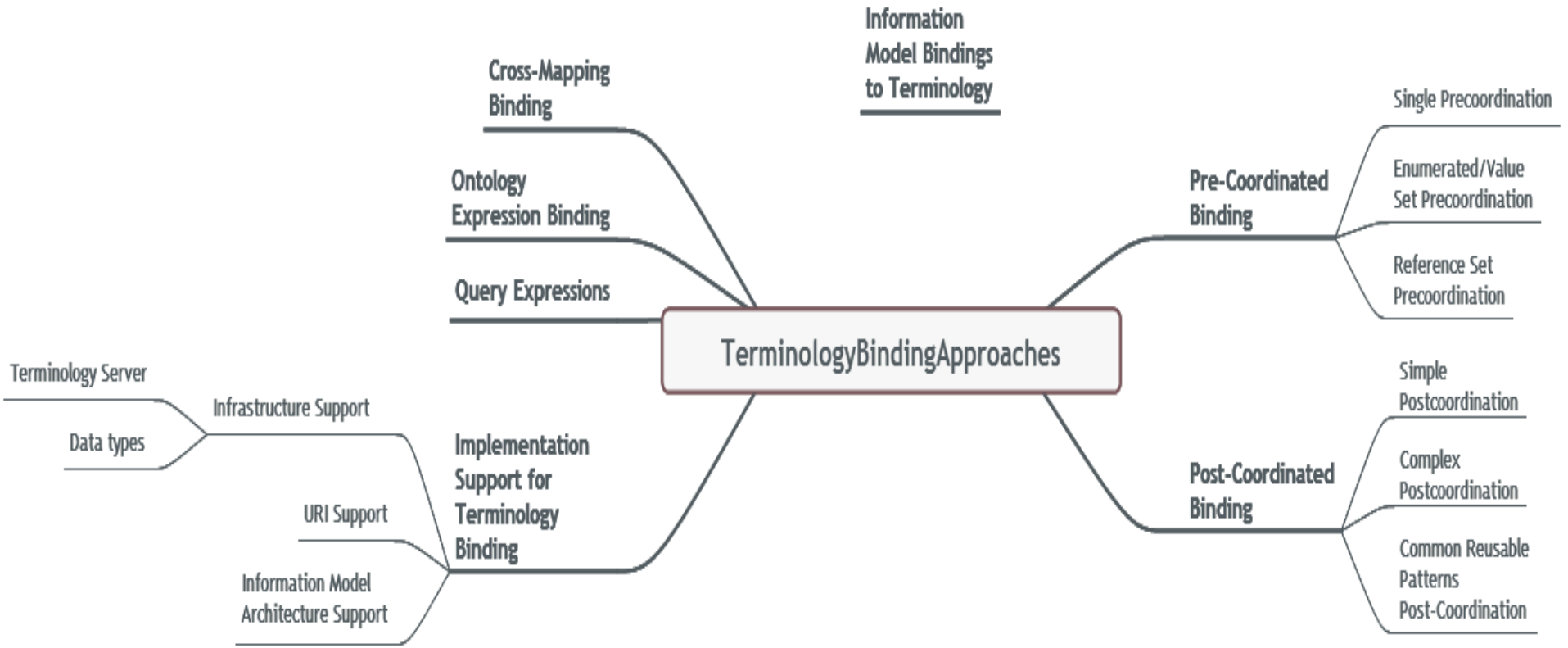
Requirements for Expression Constraints

- The structure of an expression constraint must follow the SNOMED CT *concept model*.
- Constraints should be represented either *extensionally* or *intensionally* or by reference to a *simple reference set*.
- Constraints may be *open* or *closed*.
- Constraints may be interpreted as *semantic*, *literal*, or *concrete* interpretations.
- Each expression constraint should have a *unique identifier*

Requirements for Syntax Serialisation

- Multiple format serialisation should be supported to facilitate *ease of downstream implementation and integration*.
- A human-readable view of the syntax should be provided to enable *review of binding expressions* by non-technical stakeholders.
- Programmatic views of the syntax may be provided to enable *education and further processing* by technical stakeholders.
- Data interchange formats of the syntax should be provided to enable *integration* into existing systems.

Terminology Binding Approaches



Information Model Bindings

- Information models bind to terminology concepts to help standardise clinical data.

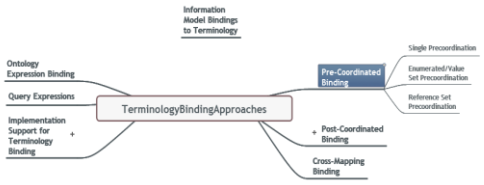
- Projects based on information model binding to SNOMED CT include (but are not limited to):

  - HL7 Terminfo
  - The NHS Logical Record Architecture (LRA), UK
  - openEHR Archetypes
  - ISO13606 Archetypes
  - HL7 CDA Templates
  - MOHH Logical Information Model (LIM) Archetypes, Singapore
  - Clinical Information Modeling Initiative (CIMI) Archetypes
  - CDISC/HL7/ISO BRIDG Model

Precoordinated Terminolgy Binding

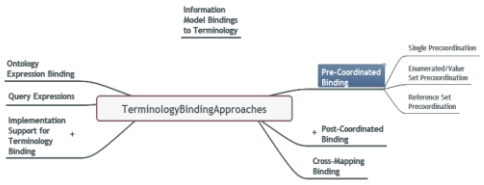
- Use of single or an enumerated list of concept.

- **Single Precoordination:** The simplest form is he use of a single Identifier.
  - NHS HL7 CDA Template : *Risk To Patient* defined using 716661000000109 | risk to patient |
  - *Blood Pressure* archetype : *Systolic* archetype node mapped to 163030003 | on examination - Systolic BP reading (finding)|



Precoordinated Terminolgy Binding

- **<u>Enumerated/Value Set Precoordination:</u>** Enumerated list of SNOMED CT identifiers specified either extensionally or intensionally.
  - HL7 Terminfo project : *Document* represented using the expression = < 419891008 | record artifact | (intensional).
  - NHS HL7 CDA Blood Pressure Template: *Blood Pressure observation* represented using an enumerated list { 75367002 |Blood pressure|, 163035008 |Sitting blood pressure|, 163034007 |Standing blood pressure|} (extensional).

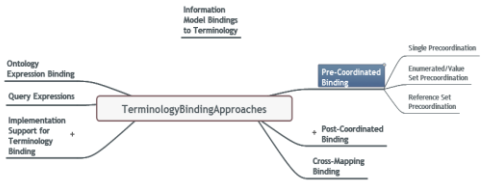


Precoordinated Terminolgy Binding

- **<u>Reference Set Precoordination:</u>** Associating a node to a reference set.
  - NHS LRA: *Allergies and Adverse Reaction Event* model permitted the use of two simple reference sets

    = (243796009 | situation with explicit context | :
    { 246090004 | associated finding | =
    (((^ 1111000000132 | Allergy Event |)
    OR (^1021000000139 | Adverse Reaction Event |))

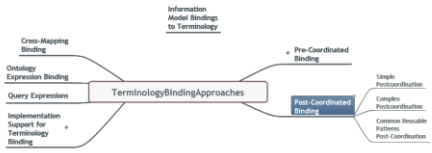
  - NHS HL7 CDA: *Investigation* template is represented using the **1031000000137** | Investigations| subset



Postcoordinated Terminology Binding

- Two or more concepts in combination with each other to jointly define the meaning of a clinical phenomenon.
- Avoids proliferation of precoordinated concepts for local use
  - Reduces cost and effort of creating and maintaining several hundreds of local concepts.

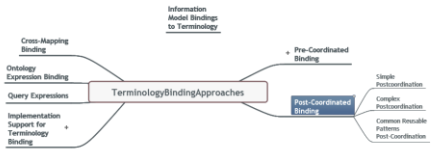
- **Simple Postcoordination**: Two or more concepts combined with each other using one or more defining relationships to provide a common contextual meaning
  - Snow OWL Extended SNOMED CT Composition Grammar (ESCG) Expression:
    - terminology:2.16.840.1.113883.6.96?escg=<<38341003
  - openEHR Archetype: *Procedure* node representing the method and procedure site :
    - 71388002 :{260686004=129304002,363704007=66754008}



Postcoordinated Terminology Binding

- **Complex Postcoordination:** Includes refinements, qualifications, and a combination of two or more concepts.
  - NHS LRA model: *Allergies and Adverse Reaction Event* expression constraint.

```
= ( 243796009 | situation with explicit context | :
      { 246090004 | associated finding | = (
        ( ( ^ 1111000000132 | Allergy Event | )
          OR ( ^ 1021000000139 | Adverse Reaction Event | )
        ) :
            { 246075003 | causative agent | =
              ( ( < 410607006 | organism | )
                 OR ( < 78621006 | physical force | )
                 OR ( < 105590001 | substance | )
                 OR ( < 373873005 | pharmaceutical / biologic product | )
                 OR ( < 260787004 | physical object | )
              )
              , 246112005 | severity | =
                 ( ( 255604002 | mild | )
                    OR ( 6736007 | moderate | )
                    OR ( 24484000 | severe | )
                 )
              , 363698007 | finding site | = (
                 ( ( < 280115004 | acquired body structure | )
                    OR ( < 91723000 | anatomical structure | )
                    OR ( 91722005 | physical anatomical entity | )
                 ) :
                       { 272741003 | laterality | =
                          ( ( < 182353008 | side | )
                             AND ( ! << 51440002 | right and left | )
                          )
                       }
                    )
              }
            )
            , 408729009 | finding context |
              = ( ( 410592001 | probably present | )
                 OR ( 415684004 | suspected | )
                 OR ( 410591008 | definitely present | )
                 OR ( 410605003 | confirmed present | )
```

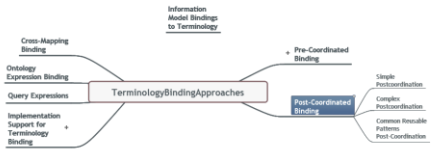


Postcoordinated Terminology Binding

- **<u>Common Reusable Patterns Postcoordination:</u>** Defines reusable patterns for creating postcoordinated expressions
  - Ensures syntax correctness and helps avoid excessive precoordination.

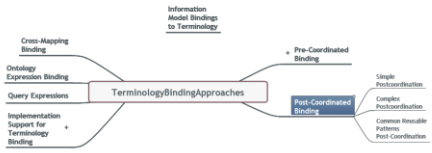
  - Family History expressions project: F*amily history of malignant adenomatous cancer* expressed using long normal form pattern.

57177007 | family history with explicit context | :
{ 246090004 | associated finding | = 443961001 | malignant adenomatous neoplasm |
, 408729009 | finding context | = 410515003 | known present |
, 408731000 | temporal context | = 410511007 | current or past |
, 408732007 | subject relationship context | = 444148008 | person in family of subject | }



Postcoordinated Terminology Binding

- **IHTSDO Common Reusable Patterns :**
  - Clinical finding present: Clinical-finding-present (<finding>)
  - Clinical finding absent: Clinical-finding-absent (<finding>)
  - Clinical finding unknown: Clinical-finding-unknown (<finding>)
  - History of: History-of (<finding>)
  - No history of: No-history-of (<finding>)
  - Family history of: Family-history-of (<finding>)
  - No family history of: No-family-history-of (<finding>)
  - Observable + value: Finding-present-observable-value (<observable>, <value>), observation-result-site-value(<observable>,<site>,<value>)
  - Procedure done: Procedure-done(<procedure>)
  - Procedure not done: Procedure-not-done(<procedure>)
  - (Drug or procedure) contraindicated: Drug-contraindicated(<substance>)
  - Pain-pattern (<site>,<side>,<laterality>)



Binding through cross-mapping

- Associations to chosen terminology concept(s) achieved by mapping internal vocabularies to standard terminologies.
  - Rapid model/content development without dependency on a single terminology
  - Safeguards terminology changes-related revisions

  - The BMJ Best Practice Integration: *Condition* mapped to 59621000 | essential hypertension (disorder) | in the URI format

    http://ec.api.bmj.com/api/v1.1/en-gb/{identity id}/definition-for-condition/SNOMEDCT/59621000.xml

Ontology Expression Binding

- Semantic structure of concepts represented using ontologies
  - Ensures that new clinical phenomenon created or expressed does not conflict with existing phenomena in the semantic hierarchy.

  - SemanticHealthNet: Expression below states that if a *heart failure diagnosis* is about some clinical situation then this clinical situation is of the type heart failure situation.

shn:*HeartFailureDiagnosis* subClassOf **isAboutSituation**_only sct:*HeartFailure*
Sct:*HeartFailure* subClassOf btl:*Situation*
Shn:*HeartFailureDiagnosis* subClassOf btl:*InformationObject*

where
shn: SemanticHealthNet Ontology. It represents the parts of clinical information models,
sct:SNOMED CT, as the chosen clinical ontology, and
btl: BioTopLite, providing general classes, relations, and constraints for the SemanticHealthNet ontology.

Query Expressions

- Syntax to use simple or complex query expressions on SNOMED CT coded data.
  - Powerful mechanism to help with data analysis and decision support.

  - Draft IHTSDO reference set query language specification: Query *"all concepts in the 'Immune hypersensitivity reaction' hierarchy that have an explicit ungrouped 'Causative agent' relationship defined to any target concept"*

```
Intersection (
 DescendantsAndSelf (418925002|Immune hypersensitivity reaction|),
 HasDirectRel (246075003|Causitive agent|, All)
)
```

Implementation Support

- **Infrastructure Support:** These include (but are not limited to) support through datatypes, terminology servers, and so on.
- **URI Support:** Supports the identification, exchange and persistence of data including those bound to terminologies.
- **Information Model Architecture Support:** Underlying architectural framework of information models influence the degree and extent of terminology binding available.

Syntax

- Examples
- BNF

Example: All fully defined

This query expression returns all fully defined concepts in the Clinical finding sub-hierarchy

<<404684003|Clinical finding| **AND fullydefined**

Example: Three levels of findings

This query expression returns the first three levels of the Clinical findings hierarchy.

<<*3 404684003|Clinical finding|

Example: Combine concepts with reference set

This query expression returns all the members of the viral disease sub-hierarchy, together with members of a pre-defined reference set called “My Virus Refset”.

<<34014006|Viral disease| **OR**
^60140068|My Virus Reset|

Example: Select using relationships

All concepts that contain a group with a ‘Finding site’ of ‘Inguinal canal structure’ and an ‘Associated morphology’ of ‘Hermial opening’. Both these relationships must be in the same group.

all:{ 363698007|Finding site| = 90785001|Inguinal canal structure|,116676008|Associated morphology| = 414402003|Hermial opening|}

Unary Operators

- “^” reference set members
- “!” not
- “>>” supertype or self
- “>” supertype
- “<” subtype
- “<<” subtype or self

Binary Operators

- “>>* n” supertype or self within n levels
- “>* n” supertype within n levels
- “<* n” subtype within n levels
- “<<* n” subtypes or self within n levels
- “top n” first n from the set of expressions
- “tail n” last n from the set of expressions

Other additions

- Named referencesets
  - All, fullydefined, primative, active
- String operators
  - filterOnMatch, filterOnNoMatch
- AND, OR
- Permissive use of brackets

Design Decisions

- Compositional Grammar
- Consistency across SNOMED CT languages
- Identifiers for Packages and Expressions
- Packaging approach…
- Parameterised Expressions…
- Co-occurrence Constraints
- Type System…
- Sets of Sets

Decision: Packaging approach

- Package expressions used in specific message / message set
- Controlled scope (public / private expressions)
- Configuration Management

- Alignment with SNOMED CT Modules

Parameterised Expressions

- Expression Libraries
- Reuse of expression fragments
  - Avoid cut/paste errors
  - Simplify maintenance (fewer change between releases)
  - Logical packaging of concerns

Type System

- Improves validation and early discovery of errors
- Improves feedback on errors
- Explicit difference between a “Constraint Expression” and a “Compositional Grammar” expression

Binding Metadata and Context

- Provenance
  - Who, Why, When for the binding / expression
- Governance
  - How will the expressions be maintained?
- Expression/Binding Type
  - Concrete, Literal, Semantic
- Defaults and inherited constraints
  - The meaning of “all”

Issues when developing bindings

- Finding similar bindings to reuse
- Missing concepts
- Dealing with new releases of SNOMED CT
- Unfamiliar Syntax / Grammar
- Mappings to local/other terminologies
- Version Control
- …. Please tell us…

Benefits of IHTSDO approved formalism

- Pool of examples and experience to draw on
- Actively maintained specification
- Evolve with SNOMED CT Concept Model and best practices
- Testing and certification
  - People, tools and bindings
- Larger market for:
  - Tools, training and learning resources
  - People and skills

Thank you
Your comments!

Dr. Rahil Qamar Siddiqui
rahil.siddiqui@hscic.gov.uk
Charlie McCay
charlie@ramseysystems.co.uk