Representing clinical information using SNOMED Clinical Terms®

with different structural information models

KR-MED 2008 - Phoenix

David Markwell *The Clinical Information Consultancy Ltd*

Laura Sato *NHS Connecting For Health*

Edward Cheetham *NHS Connecting For Health, IHTSDO and HL7 TermInfo Project Leader*

cliniClue®

Introduction:

Requirements for structure & terminology

- Effective representation of clinical information requires
  - A common structure to represent record entries in a consistent manner
    - To relate each record entry to a subject, author, time, place and other specific data items
  - A terminology to represent clinical concepts used in record entries
    - To relate the meanings of different concepts used in record entries, protocols, and queries in ways that facilitate consistent processing and reuse

Introduction:

Requirements for ‘terminology binding’

- Candidates for standard EHR structures include
  - HL7 Version 3 based models
    - Including – HL7 CDA and HL7 Clinical Statements
  - EN13606 based models
    - Including – *open*EHR
- The leading candidate EHR terminology is
  - SNOMED Clinical Terms®
- SNOMED CT can potentially be used in a variety of ways in either HL7 and *open*EHR structures
  - Different approaches can undermine the value of a standard terminology and structure
  - Consistent principles and methods need to be applied to ‘terminology binding’ if potential benefits are to be realised

Context: SNOMED CT & HL7 Version 3

- HL7 Version 3
  - Reference Information Model and method for development of communication specifications
- TermInfo Project
  - Looking at integration of terminologies with HL7 Version 3
  - Started late 2004 by NASA → adopted by HL7 Vocabulary TC
- Guide to Use of SNOMED in HL7 Version 3
  - Produced by TermInfo and refined by 3 ballot cycles
  - Adopted as HL7 Draft Standard for Trial Use (DSTU) in September 2007
- NHS Connecting for Health specified
  - SNOMED CT for coding clinical information
  - HL7 Clinical Statements & CDA based messages
  - Applied the early recommendations of TermInfo work

Context: SNOMED CT, EN13606 & *open*EHR

- EN13606
  - Five part European Standard for EHR Communication from CEN TC251
    - Also being submitted to ISO for wider adoption
  - A generalised EHR reference model
    - Uses archetypes for specific types of clinical information
- *open*EHR
  - A consortium based on open-source development of specification and tools around the EN13606 standard
    - Includes template to further refine archetypes
- NHS Connecting for Health
  - Used *open*EHR tools to capture content requirements expressed by clinicians
  - Found a variety of approaches had been applied to represent similar information in the resulting archetypes and templates

Managing overlaps

- Both TermInfo and the *open*EHR terminology binding work found overlaps
  - Situations that could be represented by combining different information model structures with different terminology components
    - E.g. Context such as ‘past history’ or ‘family history’ can be captured using specific data points in the structural model or using SNOMED CT context attributes to express a ‘situation with explicit context’
  - Different structural models driven by ‘user-interface’ considerations to collect similar information
    - E.g. Check-lists as compared to term or concept selection

Challenges presented by overlaps (1)

Same information represented in different ways

Retrieval and reuse may miss similar information represented in different structure/terminology combinations

For example, representing ‘Family history of asthma’

- A ‘family history’ *check-list* with ‘asthma’ marked ‘yes’
- A ‘family history’ *section* referring to the SNOMED CT concept ‘asthma’
- A *record entry* referring to the SNOMED CT concept ‘family history of asthma’
- A *record entry* containing a SNOMED CT expression such as ‘family history : associated finding = asthma’
- A *record entry* containing the SNOMED CT concept ‘asthma’ associated to a ‘family member’ by an information model structure

To avoid false negatives different representations must be transformed to a common model

Challenges presented by overlaps (2)

Same information represented in different ways

Risk of ambiguity from alternative interpretations

For example, representing an absent finding

- Information model attributes may indicate ‘absence’ or ‘negation’
- SNOMED CT finding context can represent ‘known absent’
- Does the combination of two representations of absence mean …
  - *double-negative*
  - *redundant restatement of the negative*
  - *additional emphasis of the negative*
- The logical interpretation is ***double-negative*** but may not be how it was intended
- It many not be clear which concepts are ‘negative’

  E.g. ‘conscious’ ‘**not** conscious’
  ‘unconscious’ ‘**no loss of** consciousness’

To avoid misinterpretation there need to be clear rules about the way information model and terminology semantics combine

Challenges presented by overlaps (3)

Different levels of semantic specificity

Structural and terminological representation of apparently similar information may have different levels of specificity

For example, representing the site of a procedure

- SNOMED CT distinguishes between
  - the '**direct site**' and '**indirect site**'
    e.g. for excision - what was excised & where it was excised from
  - '**site**' and '**morphology**'
    e.g. normal structures & abnormal structures
  - '**site**' and '**approach**'
  - procedures with a **defined site** and those applied to different sites
    e.g. 'appendectomy' → 'appendix'
- The information models considered
  - Do not make these distinctions in a systematic manner
  - Make distinctions that SNOMED CT does not support

Guidance on overlaps must take account of these differences in expressivity and the ways they affect particular use cases

General summary of alternative ways to manage overlaps

Processing requirements to manage each overlap depend on the permitted representation. In each case a single specified representation simplifies processing requirements.

#	Representation guidelines		Processing requirements	
	HL7 / *open*EHR representation	SNOMED CT representation	Validate & combine	Transform
1	Required	Required	Essential	-
2	Optional	Required	Essential	Dependent
3	Required	Optional	Essential	Dependent
4	Required	Prohibited	-	Dependent
5	Prohibited	Required	-	Dependent
6	Optional (either or both)		Essential	Essential
7	Optional (either one not both)		-	Essential

Managing gaps

- Both TermInfo and the *open*EHR terminology binding work found gaps
  - Situations that could not be fully represented due to lack of completeness in the structural model or the terminology
  - In some cases, a simply omission of a relevant concept in SNOMED CT
  - In other cases, an available SNOMED CT representation could not be accommodated in the model due to different assumptions about the concept domains or lack of support for post-coordination
- Gaps are usually easier to address than overlaps
  - Decide where a change is needed and submit a request
  - Some gaps highlight broader issues and it may not be appropriate to resolve these by piecemeal additions

Managing semantic granularity

- Structural information models
  - Use separate data points to represent different aspects of a record entry
  - Support instance information such as dates, times, people, places, numeric values and quantities
  - May support codes and post-coordinated expression in some data points and but not in others
- SNOMED CT expresses
  - Some aspects of a clinical situation in a single concept
  - Other aspects using post-coordinated expressions
  - Clinical concepts and situations but not instance specific data
- It is essential to match and combine these levels of semantic granularity in a consistent manner
  - This may require changes in the information model; and/or
  - Constraints on the use of SNOMED CT

Examples of binding to nodes in an *open*EHR template

- The next slides illustrate some of the different types of terminology binding that were found to be necessary to apply SNOMED CT binding to some of the *open*EHR templates developed for the NHS

Note

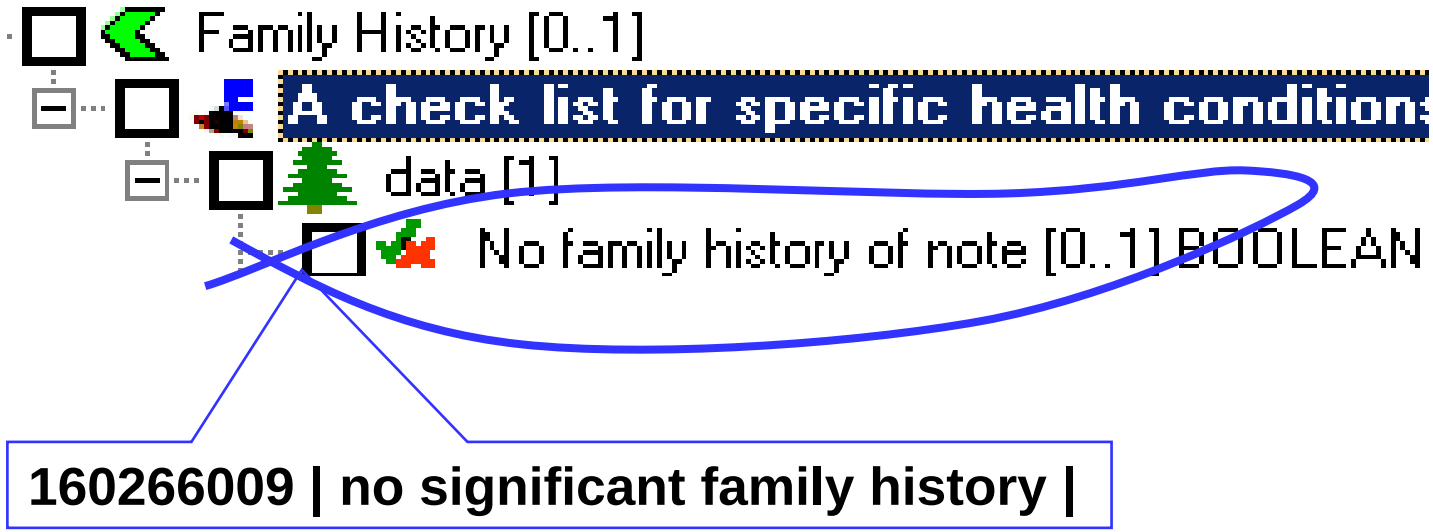
These templates were developed as a proof of concept in a short time and without reference to SNOMED CT.

The terminology bindings were applied retrospectively.

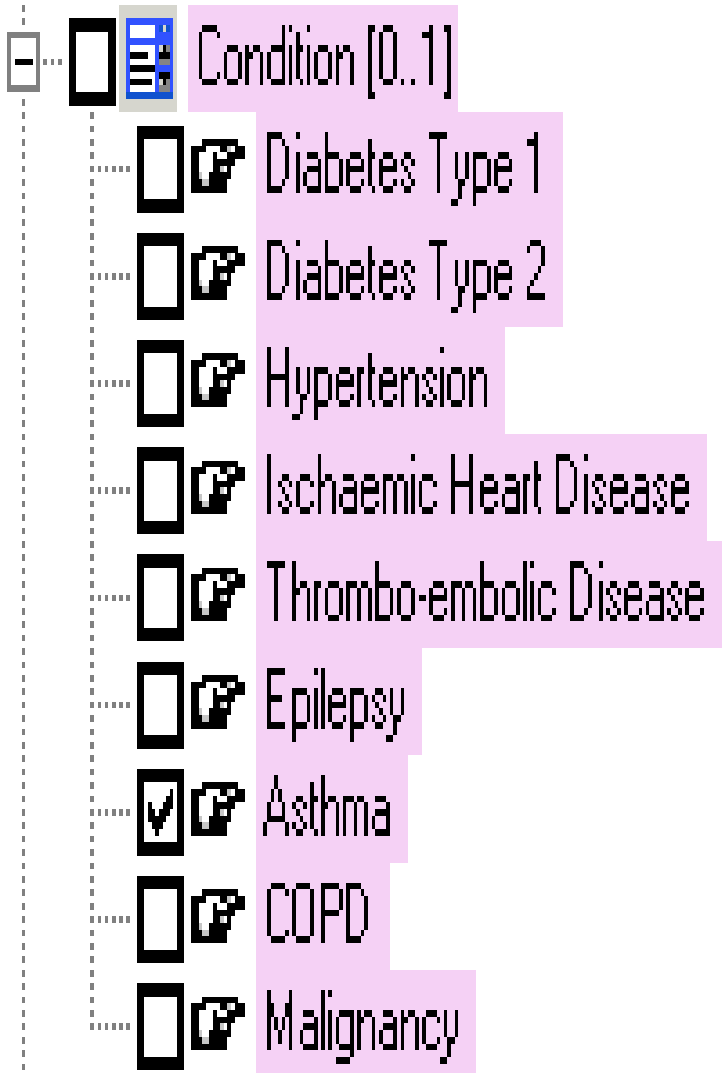
Ideally detailed models should be designed taking account of the terminology and following common patterns. However, it seems likely that, even if this is done, several different types of terminology binding will be necessary.

Node-fixed bindings

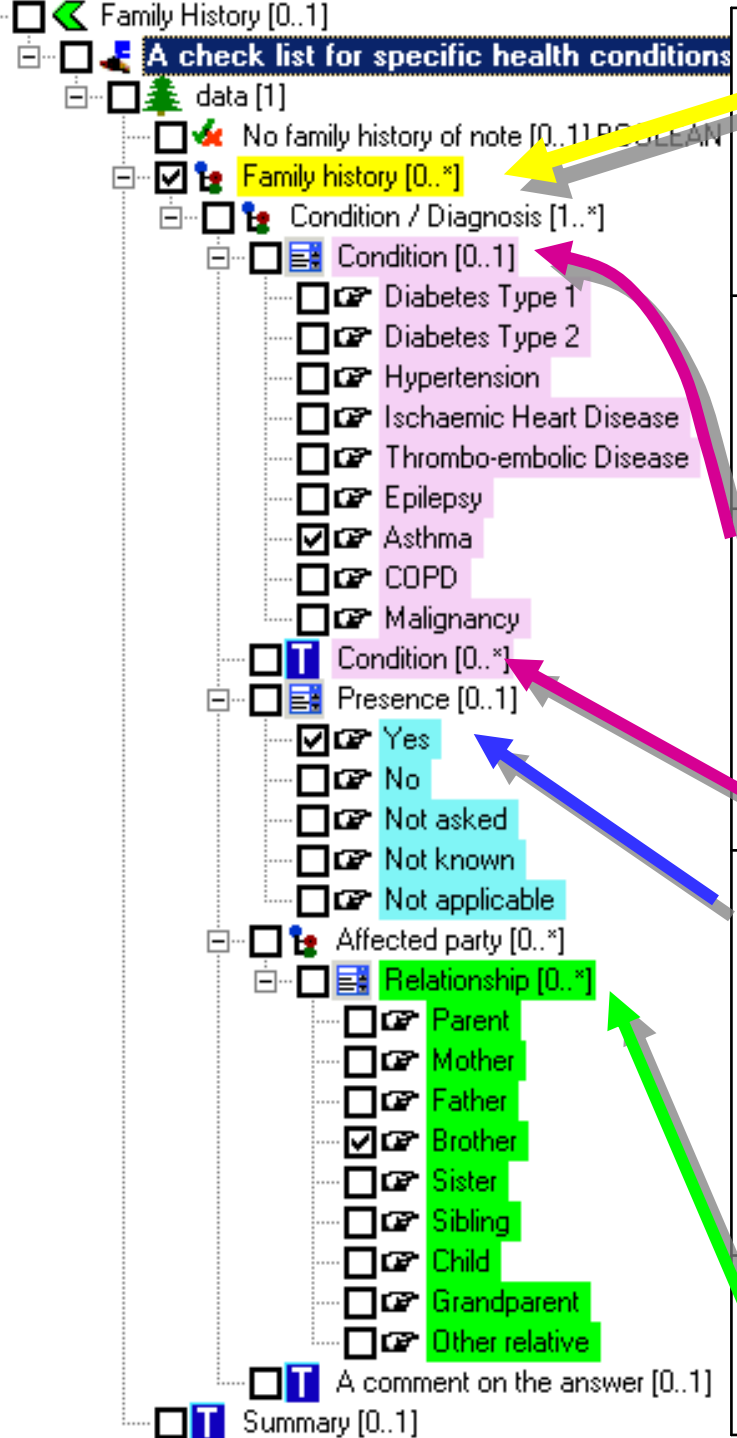
- **Node-fixed bindings assert that inclusion of a node has a specified fixed meaning**
- A node-fixed binding specifies the SNOMED CT expression that is applied if the bound node is included. In the case of a node with a Boolean value the expression is applied if the node has the value 'true'.



Semantic constraint bindings* and Choice-fixed bindings



<64572001 | disease| (*semantic - constraint
46635009 | diabetes mellitus type 1|
44054006 | diabetes mellitus type 2|
38341003 | hypertensive disorder|
414545008 | ischaemic heart disease|
371039008 | thromboembolic disorder|
84757009 | epilepsy|
195967001 | asthma|
13645005 | chronic obstructive lung disease|
86049000 | neoplasm, malignant (primary)|



Each instance of the **Family history** node represents an item of positive or negative family history which SNOMED CT expresses as a [243796009 | situation with explicit context].

The full expression that represents this depends on the values assigned to three subsidiary nodes

- Condition / Presence / Relationship

The **condition** node provides the 'associated finding' attribute for the SNOMED CT expression. This value may be either:

- Selected as a choice from a list; or
- Choosing any disease concept from SNOMED CT.

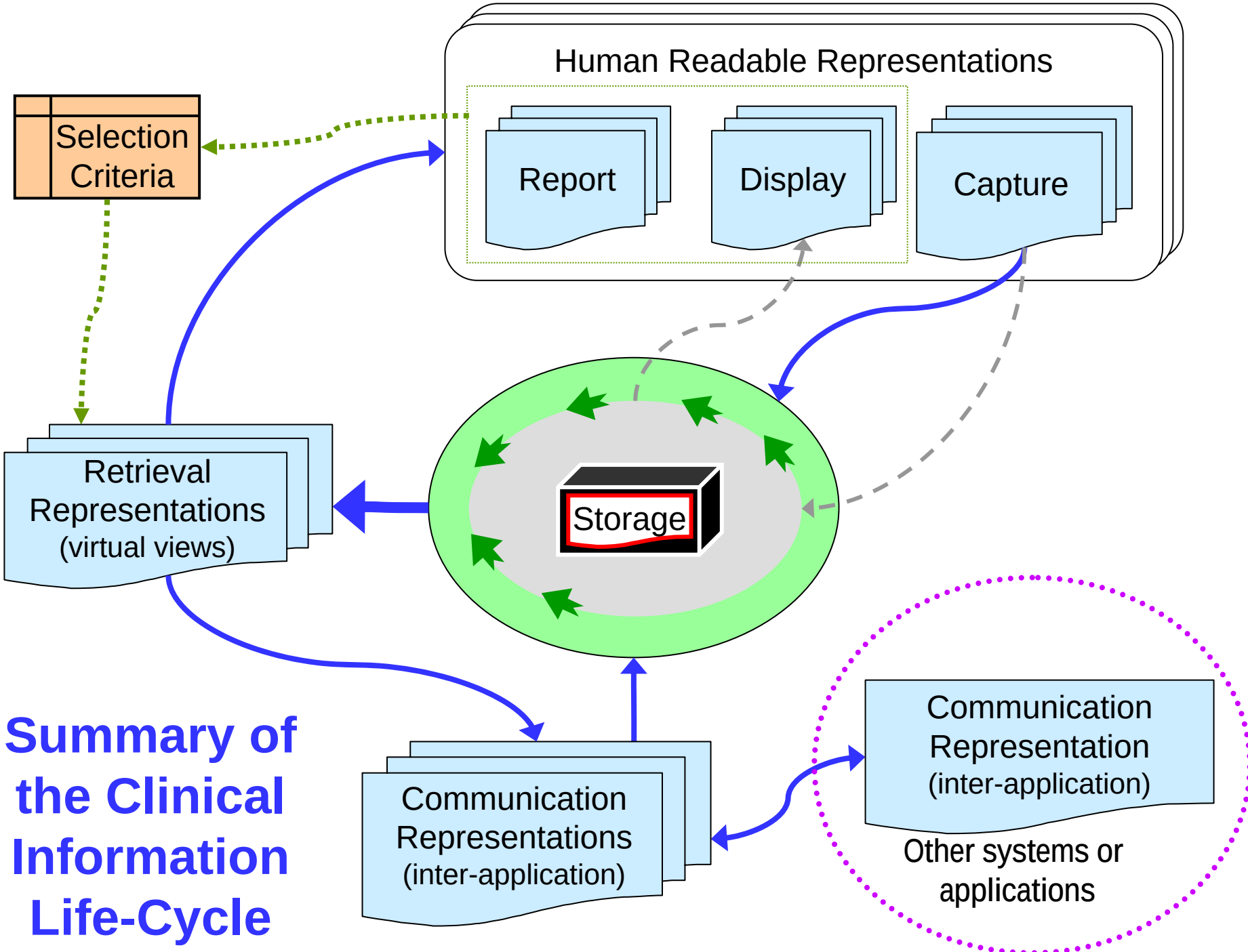
The **presence** node provides the 'finding context' value for the SNOMED CT expression.

- There are directly equivalent values for 'Yes', 'No' and 'Not known'
- Issues related to 'Not asked' and 'Not applicable' are discussed elsewhere in this document.

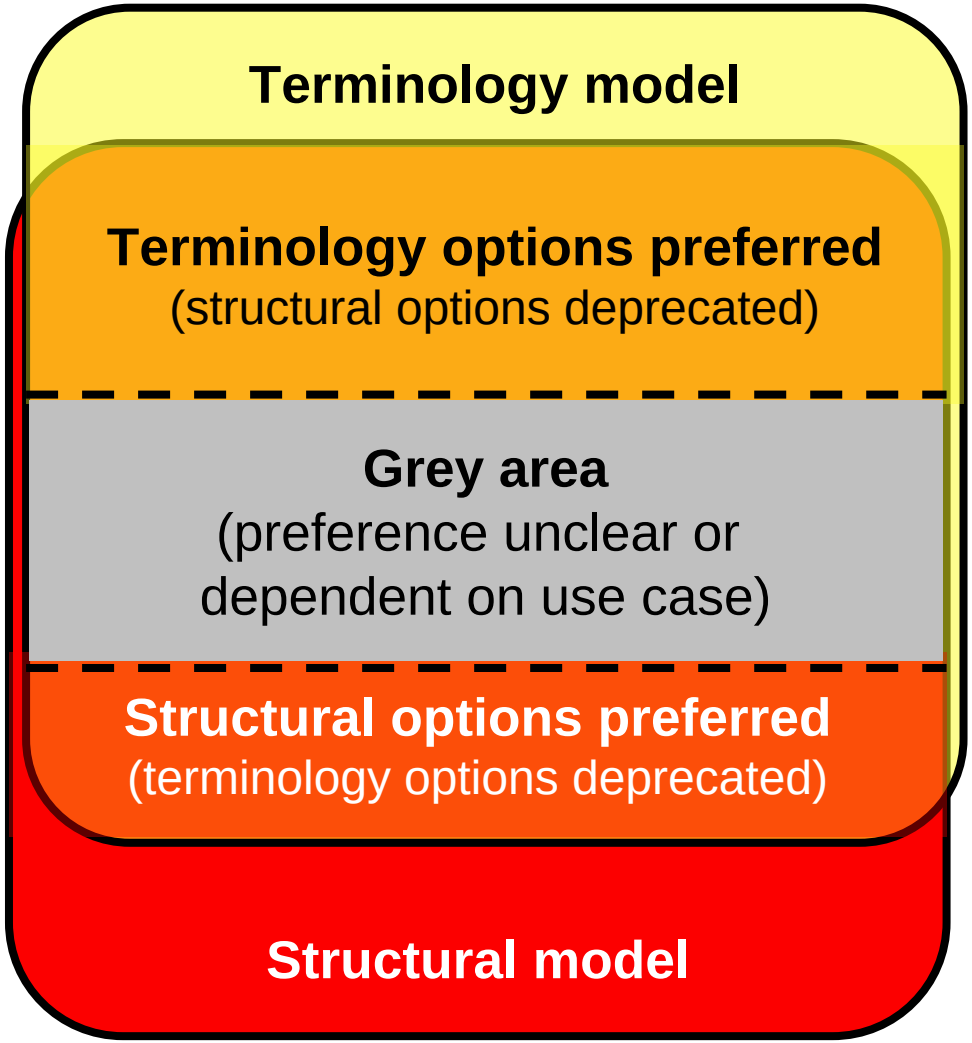
The affected party **relationship** node provides the 'subject relationship context' value for the SNOMED CT expression.

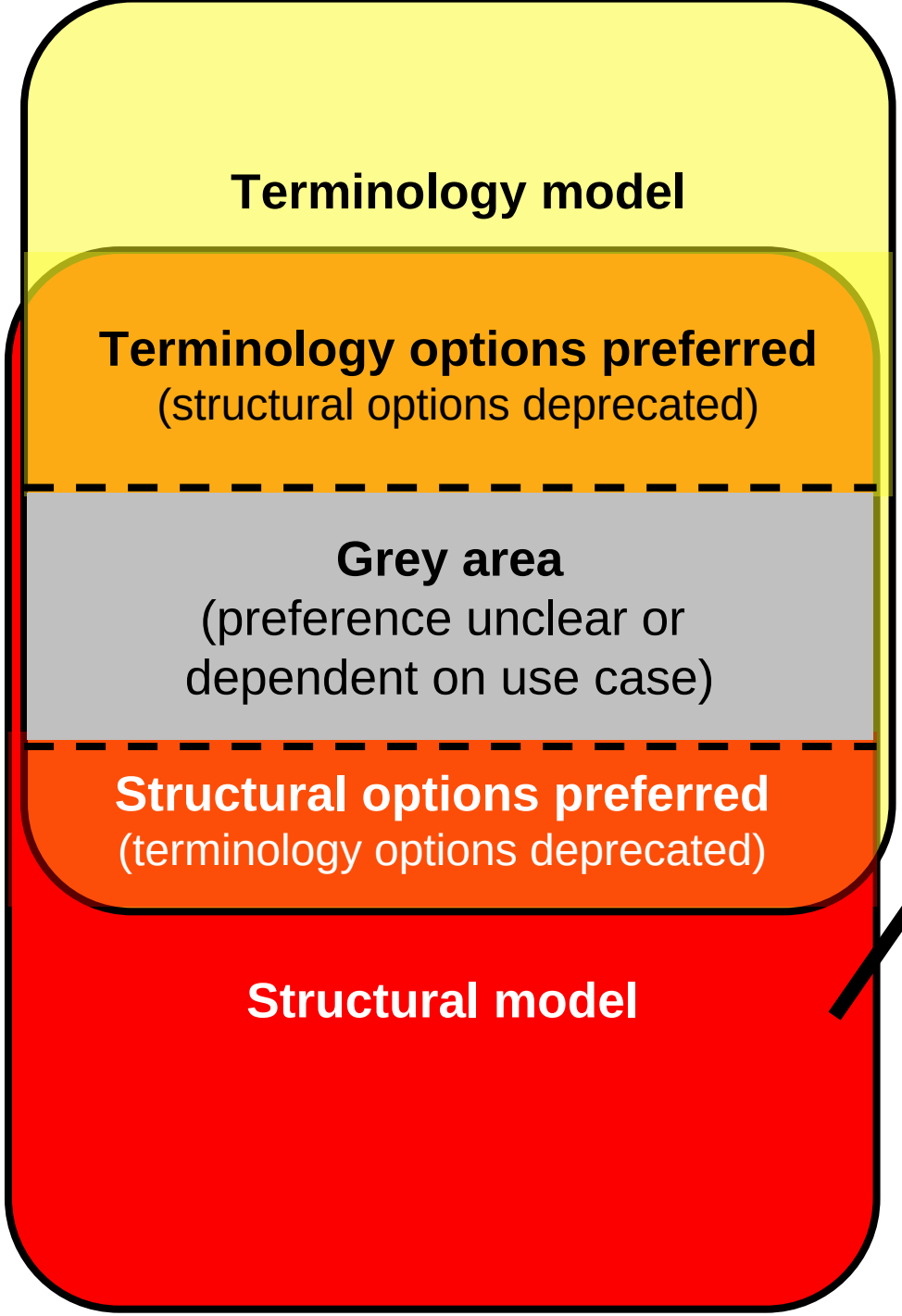
Detailed entries, summaries & check-lists

- Information model design is sometimes driven by particular use cases or data capture paradigms
  - For example, check-lists, summaries and detailed record entries
- There is a clear user requirement to support these different ‘models of use’ but the need for a common ‘model of meaning’ is required to enable effective retrieval and reuse
- The way terminology is applied to a information model needs to take account of the complete life-cycle of clinical information


Human Readable Representations
Report
Display
Capture
Selection Criteria
Retrieval Representations (virtual views)
Storage
Communication Representations (inter-application)
Communication Representation (inter-application)
Other systems or applications
Summary of the Clinical Information Life-Cycle

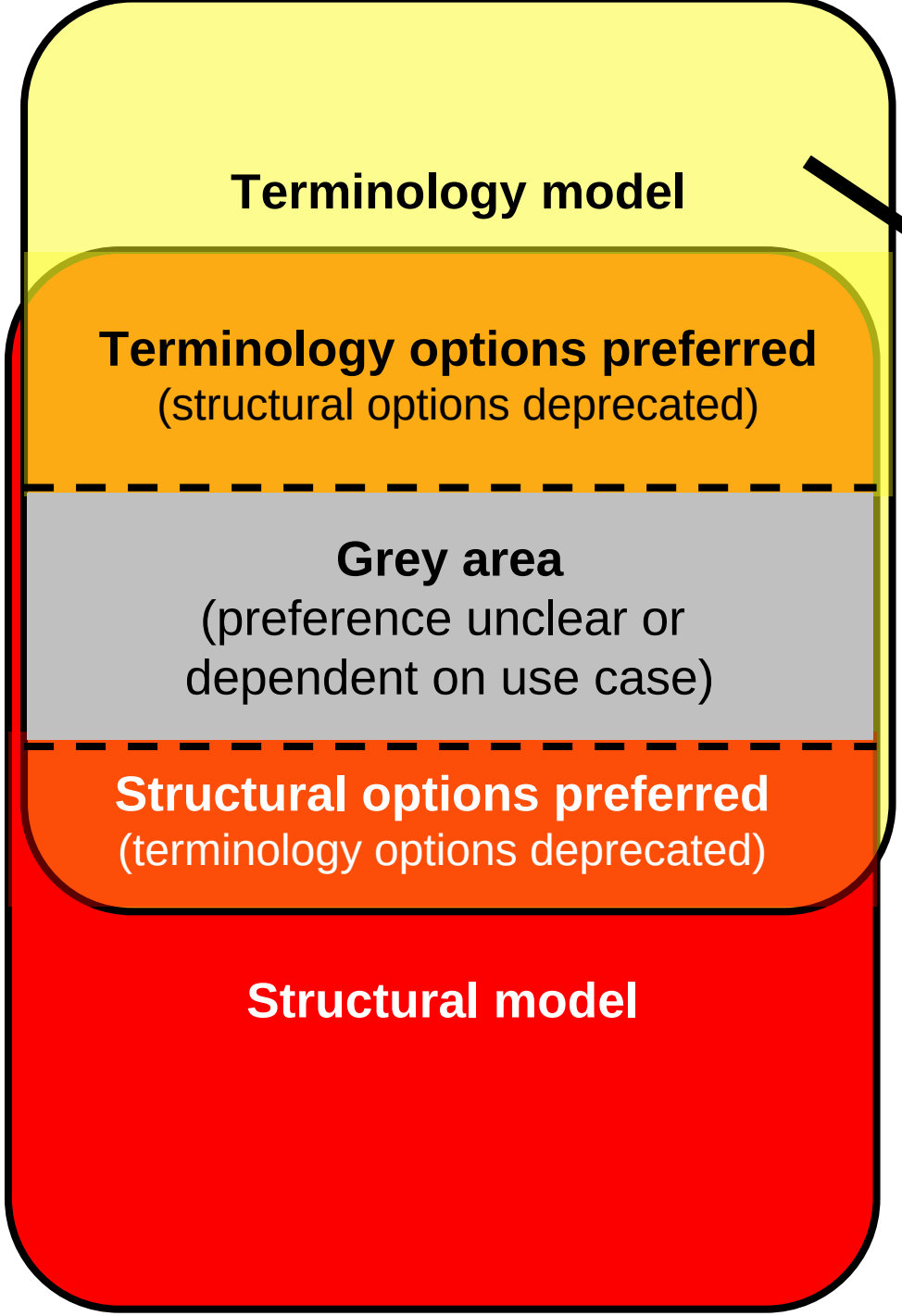
Balancing structure and terminology





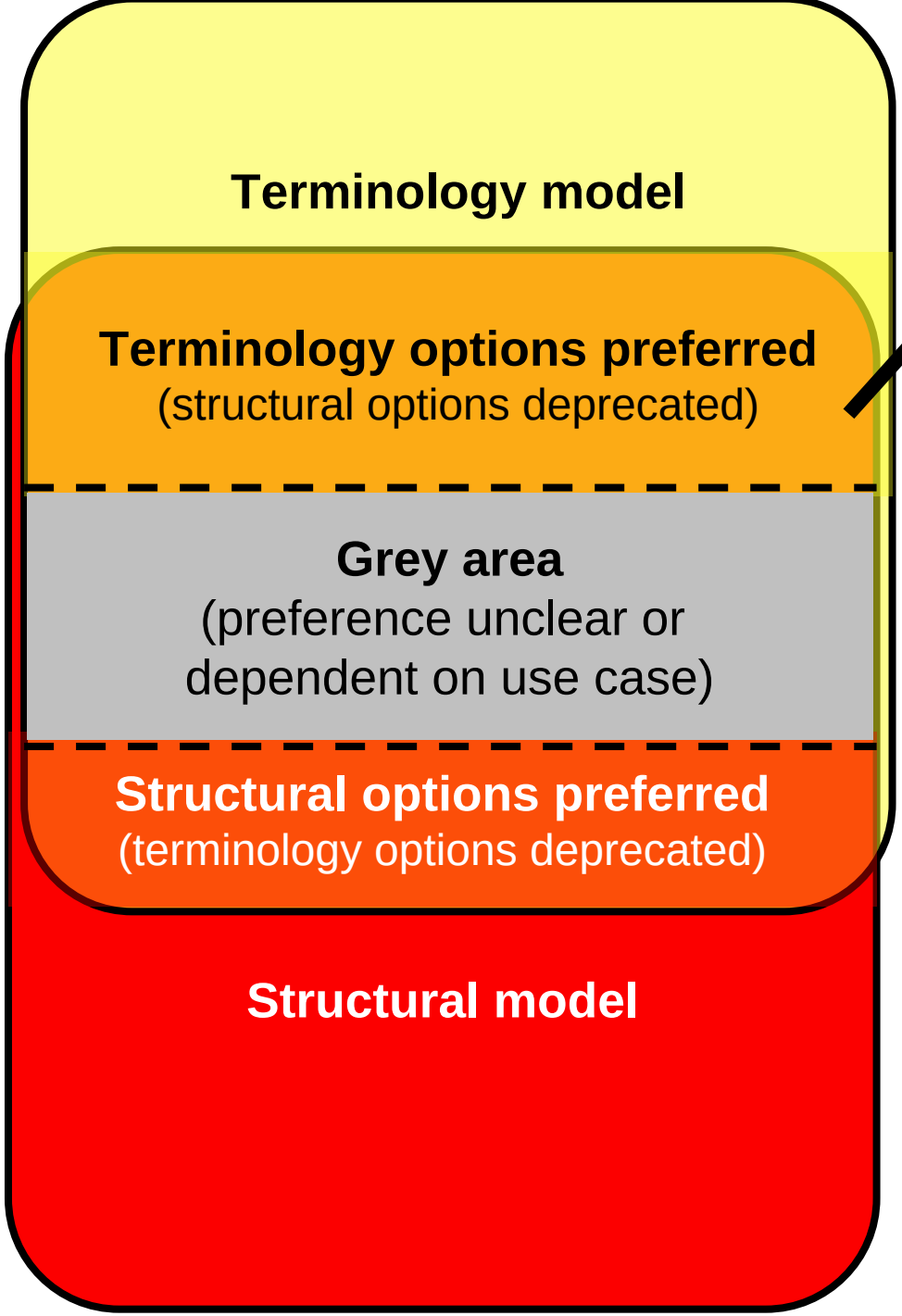
Structural model

- Attributes with specific data types
  - For example, dates, times, durations, quantities, text markup.
- Identifiable instances of real-world entities
  - For example, people, organisations, places.
- Overall record and/or communication architecture
  - For example, EHR extract, EHR composition, openEHR reference model, CDA documents, HL7 messages.
- Representation of constraints on use of particular classes or attributes in given use cases
  - For example, formalism for templates applied to constrain openEHR archetypes or HL7 CDA documents.



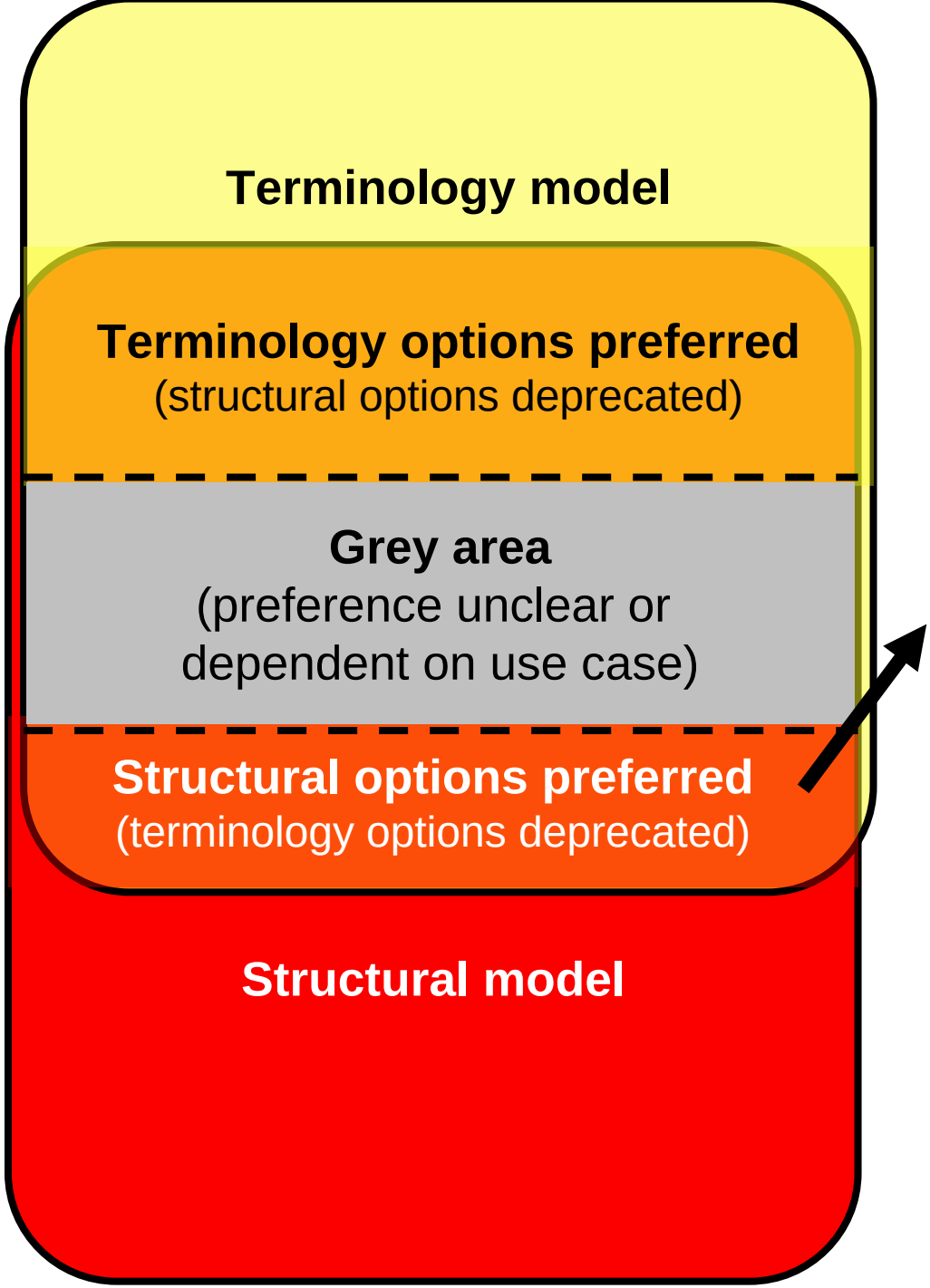
Terminology model

- Specific concepts:
  - For example, diseases, symptoms, signs, procedures, drugs, etc
- Semantic relationships between concepts
  - For example, relationship between "viral pneumonia", "lung", "virus", "infectious disease".
- Representation of constraints on use of terminology
  - For example, concept model and value-set definition formalism.



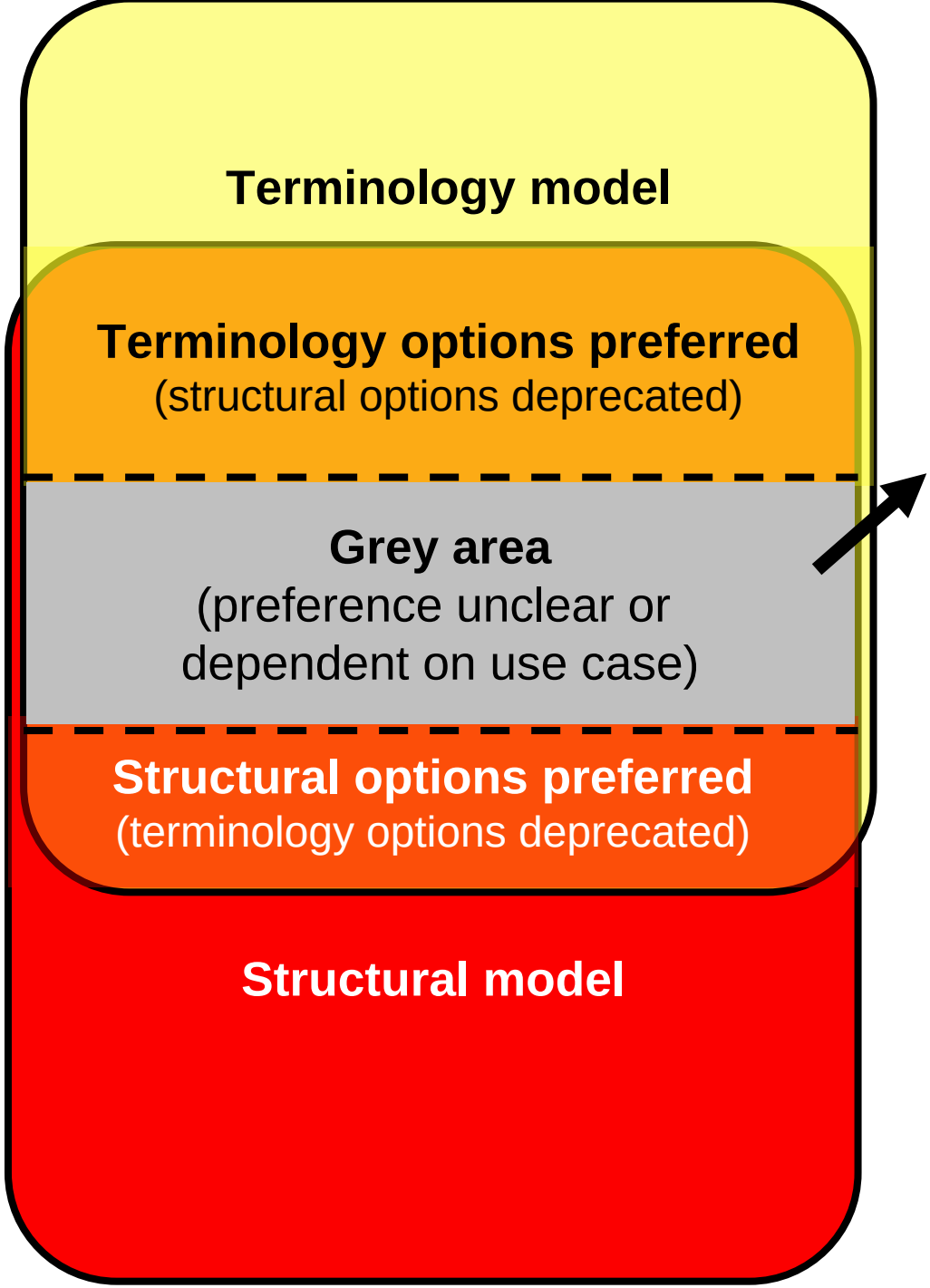
Terminology model preferred

- Constraints on combination of concepts in instances including abstract model of post-coordination and permissible attributes and ranges for refinement of concepts in specified domains:
  - For example, restrictions on "finding site" refinement of "appendicitis", conventions on representation of laparoscopic variants of procedures.



Structural model preferred

- Representation of relationships between distinct instances of record entries and other classes
  - For example, grouping of record entries related by timing, problem or other organising principles.



Grey area

- Representation of contextual information related to instances of clinical situations
  - For example, family history, presence/absence, certainty, goals, past/current, procedure done/not-done, etc.
- Representation of additional constraints on post-coordination of concepts for specific use cases
  - For example, constraints on terminology use specific to immunisation and related adverse reaction reporting.

Practical principles of terminology binding (1)

- ***1. Understandability***
  - Understandable to those familiar with the terminology and the structural models being integrated
- ***2. Reproducibility***
  - Tested on members of the intended target audience to ensure they are interpreted and applied consistently
- ***3. Usefulness***
  - Need not cover all possible use cases but should support the most common business objectives in the intended scope of use
- ***4. Reusability and common patterns***
  - Representations that can be reused consistently in many contexts should be preferred to those specific to one use case

Practical principles of terminology binding (2)

- ***5. Transformability and normal forms***
  - If alternative representations are permitted, unambiguous transform rules should be specified to a common representation
- ***6. Tractability***
  - Requirements for tooling to transform or validate instances should be computationally tractable
- ***7. Practicality***
  - Existing tools and applications (with reasonable enhancements) should be able to implement the recommendations
- ***8. Scalability***
  - e.g. Combinatorial explosion of pre-coordinated concepts should not be required

Practical principles of terminology binding (3)

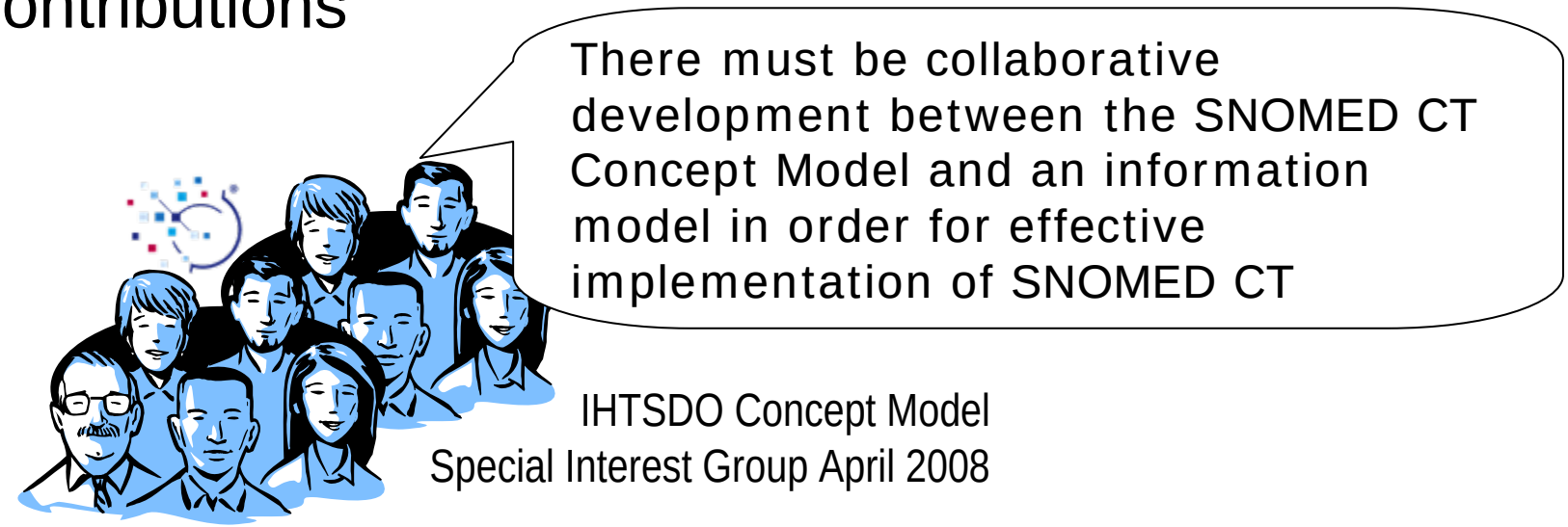
- ***9. Limiting arbitrary variation***
  - Where more than one approach appears to be equally valid based on other criteria, a single approach should be recommended to avoid unnecessary variation
- ***10. Responsive participating standards***
  - The participating structural and terminology standards should provide prompt mechanisms to enable notification and correction of gaps and inconsistencies
  - Implemented systems and participating standards should be sufficiently agile to allow rapid and reasoned development of effective compositional solutions

---

- **Requirements for specific guidelines**
  - The general principles of terminology binding need to be instantiated by specific guidelines that support practical integration between SNOMED CT and a selected set of information model

Conclusion

- The practical consequence of interdependencies between terminology and structural models are often underestimated
  - Information models cannot be terminology neutral
  - SNOMED CT implementation is dependent on tight integration with standard information models
- Developers of clinical terminologies and clinical information models should adopt policies that facilitate ‘dependency aware evolution’ of their contributions



Contact details and acknowledgements

- Author and presenter
  - Contact: David Markwell – david@clininfo.co.uk
- The HL7 TermInfo Project was facilitated by Health Level 7 of Ann Arbor, USA
  - Contact: Ed Cheetham – ed.cheetham@nhs.net
- The work described on binding terminology to *open*EHR archetypes and templates relates to models under development at NHS Connecting for Health
  - Contact: Laura Sato – laura.sato@nhs.net