# Team Competencies and Educational Threshold Concepts for Clinical Information Modelling

Philip J. SCOTT[a,1] and Kai U. HEITMANN[b]

[a] *Centre for Healthcare Modelling and Informatics, University of Portsmouth, Hampshire, United Kingdom*
[b] *Heitmann Consulting and Services / Gefyra GmbH, Germany, CEO HL7 Germany*

**Abstract.** Healthcare interoperability depends upon sound semantic models to support safe and reliable exchange of information. We argue that clinical information modelling requires a collaborative team of healthcare professionals, process and content analysts and terminologists and that 'separation of concerns' is unhelpful. We present six fundamental concepts that participants must understand to collaborate meaningfully in technology-agnostic information modelling.

**Keywords.** education, clinical information systems, interoperability, health information exchange, models, team dynamics

## 1. Introduction

Healthcare interoperability is about people far more than it is about technology [1]. The purpose of healthcare interoperability is to improve communication between clinicians and with patients, yet often there is inadequate clinician, and hardly ever patient, involvement in defining the requirements [2]. Information modelling comprises a substantial part of the requirements definition in healthcare interoperability design, regardless of the tools or technology adopted (e.g. FHIR, CDA, openEHR). Information modelling is related to model-driven software development [3] and the notation and concepts of the Unified Modeling Language (UML) [4] are commonly adopted or adapted. Clinical information modelling involves defining sets of data items for given care scenarios, their inter-relationships, constraints and vocabulary requirements [5]. This paper presents recommendations derived from practical experience of projects over many years and tested specifically in recent projects in the UK and Germany [6, 7].

We propose that there is a core set of team competencies: distinct roles that must work together in a team to address all the necessary aspects of an information model that truly represents the real world. We further propose that there is a minimum educational requirement: a set of fundamental concepts that participants need to understand to contribute productively to clinical information modelling projects. These concepts are about data structures and modelling methods and are not bound to any particular tooling or technology. These are "threshold concepts": 'portals to troublesome knowledge' [8].

---

[1] Corresponding Author, School of Computing, Buckingham Building, University of Portsmouth, Lion Terrace, Portsmouth PO1 3HE, United Kingdom; E-mail: Philip.scott@port.ac.uk.

## 2. Team Collaboration Is More Effective than 'Separation of Concerns'

We have found that a clinical information modelling team needs to contain three types of specialist. Firstly, and arguably most obviously, the project requires healthcare providers (doctors, nurses, therapists, pharmacists and other care professionals as necessary to the requirement under consideration) to explain clinical practice, operational priorities, implementation and safety issues and a user appraisal of different options for semantic representation. Secondly, the team requires process and content analysts to offer guidance on healthcare jargon and operational workflow in differing health systems. Thirdly, the project needs terminologists who bring expertise in the vocabularies used for formal representation of semantic concepts (e.g. SNOMED CT). Each of these specialists is necessary, but on their own insufficient.

Our experience supports findings from other projects [2] that an artificial 'separation of concerns' into clinical and technical stages or sub-teams is counter-productive – clinicians need guidance on usage of tools, technicians need guidance on the realities of healthcare practice. This work is interactive rather than linear and is enriched by positive recognition of what each specialism brings to the project [9]. Such team collaboration is a good example of situations where the whole is greater than the sum of its parts [10].

Recognizing from the outset that this is a multi-disciplinary endeavour is fundamental to achieving a successful approach and team spirit. We do not minimize the importance of clinical informatics experts, whether they are clinicians who have specialized in informatics or informatics practitioners who have specialized in healthcare. However, these specialists inevitably need to work with average clinicians who have relatively little interest in formal informatics concepts and with average technical advisors who have only superficial clinical knowledge.

There can be initial anxiety in these multidisciplinary teams. Healthcare providers obviously lead the content definition, but analysts and terminologists should be self-confident enough to believe that they are making substantial contributions and to behave accordingly. Healthcare providers need to welcome and encourage that.

Healthcare providers must learn how to communicate well with non-clinical experts. This requires overcoming a commonly observed condescending attitude towards non-clinicians, demonstrating behaviour that "educates" the rest of the team. This can of course work both ways: technical experts can be dismissive of clinician "ignorance". Neither of these is a constructive behaviour.

The team should perform team-building activities that go beyond the immediate task focus and help to overcome the "usual" way of doing things. Those "bonding" actions can be a meeting outside the usual setting, a dinner or other social event [11].

## 3. Threshold Concepts

We present six threshold concepts to illustrate our proposal of what to include in introductory training for clinical information modelling: data definitions, examples and evidence, grouping, data types, value sets and cardinality. Given space constraints, we only offer a basic outline of each concept. Other important concepts not addressed here are re-usability, inheritance and versioning. We also do not discuss related implementation issues about tooling, project management and curation.

## 3.1. Data Definitions

When defining the clinical content, consider the following aspects:

- Assign a **short name** with clear meaning to each data element. This needs to be simple text in natural language, not acronyms or abbreviations.
- Include a **description** for each data element that precisely explains the exact subject of the item and its scope. Thus a data set becomes a hierarchical glossary.
- Introduce an **identifier** (short number for example) for each data element, at least unique within the data set, eventually globally unique; this makes it much easier later to refer to the data element in further discussions.
- Agree upon standard **naming conventions** so that you are consistent across your project and, where necessary, wider scale developments.

## 3.2. Examples and Evidence

It is essential to give examples to keep the information model relevant for real world implementation. If a vaccination record data set defines a data element called "dose quantity", consider adding examples like "20 mL" or "200 I.U." to illustrate what is expected. When defining a data set, document a use case or rationale for the collection of the defined set of data overall and each data element individually. If the team cannot easily identify numerous use cases, real world justification is perhaps questionable.

## 3.3. Grouping

Proper data sets contain grouping. Initial draft data sets are often simple lists of data elements. We have observed that grouping and summarizing data items in hierarchies is for some reason not a natural approach for many healthcare providers, despite their familiarity with these concepts in chemistry and medicine. Using inappropriate tools like spreadsheets (that naturally produce flat lists) can reinforce this problem.

Grouping might seem superfluous, but a long flat list can easily become longer. Seeing only the top-level elements helps to understand and navigate the whole data set. Viewers can expand the "tree" to see its components. Good tools intuitively support this.

To illustrate the enhanced clarity from grouping data, consider the example of a person's "name" comprised of title, family name and a given name, along with other demographic properties. Rather than defining it as a flat list [Title, Family name, Given name, Date of birth, Gender], compose the "name" as a group:

- Name
    - o   Title
    - o   Family name
    - o   Given name

- Date of birth
- Gender

This grouping clarifies that "name" has several components. We can expand the "tree" to see the elements inside the group (denoted here by ➤):

➢   Name

- Date of birth
- Gender

## 3.4. Data Types

What is the expected *nature* of the data? In some tools, the lists of data types tend to use programming terms that are not self-explanatory to healthcare providers. The most common data types (for healthcare provider driven and defined data elements) are:

- **Count** – countable (non-monetary) quantities.
- **Code** – a code for a real-world thing; a code is a system of valid symbols that substitute for specified concepts, usually defined by a formal reference to a terminology or ontology, but may also be defined by the provision of text.
- **Dates and Times** – represents an absolute point in time specified to the day or additionally the time. The partial form is used for approximate dates.
- **Text/String** – may contain any amount of legal characters arranged as words or sentences. Formatting may be included (text) or not (string).
- **Quantity** – a measurement comprising a decimal number of arbitrary precision and a unit. There are some "special" quantities used in healthcare: for time durations *Duration*, for monetary amounts *Currency*.
- **Duration** – a quantity representing a period of time with respect to a notional point in time, which is not specified. A sign may be used to indicate the duration is "backwards" in time rather than forwards.
- **Identifier** – identifiers of real-world entities. "Pure" identifiers are meaningless, but that is not usually the intent.

Some less common data types are **ratio** (beware apparent "ratios" where the ratio is in the unit not the value: beats/minute, µmol/L), **Boolean** ("true" or "false"; often an apparent Boolean is not in fact binary, so is typed to allow for "Don't know" or similar), **ordinal** (rankings and scores with implied ordering) and **ranges** (such as a "normal range" for a vital sign). Quantities often need an associated allowed **unit** or set of units. Note that a terminologist can help here too: they know systems of predefined units to tell software builders what standardized units look like. While standardized units are very important for the communication between systems (often using a coding called UCUM [12]), the *display* of the units should always fit the usual way clinicians see them.

## 3.5. Value Sets

A data element typed as code may require a choice list of concepts you want to actually allow when it comes to real data. This is another good example of where a terminologist might help. Some terminologies such as SNOMED-CT already have a set of terms (along with codes) for specific purposes. Choice lists may be found as **value sets**, containing the choices along with their coded representation for proper implementation in software and reliable communication. Value sets can be "extensional" (pre-defined static lists) or "intensional" (dynamic lists generated by a rule, such as "descendants of *Allergy to substance"* in the SNOMED CT code system). Other simple examples of existing value

sets might be the ABO&Rhesus Blood Groups, Administrative Gender of a Person, Professional Specialties or Laboratory Tests.

### 3.6. Cardinality

An essential step of definition is to determine the "cardinality" of a data element: is it optional or mandatory, repeatable or not? This is typically expressed as "n..m", with n being 0 for optional data elements or 1 for mandatory ones, and m being 1 for non-repeatable items or 2, 3... for the number of allowed repetitions. The symbol * is used to express an unlimited number of repetitions of the item, so "n..*" means "n to many".

## 4. Conclusions

We argue that clinical information modelling requires a collaborative team of healthcare professionals, process and content analysts and terminologists and that 'separation of concerns' is unhelpful. We propose a core set of foundation knowledge for introductory training. This argument is our hypothesis derived from real world projects and we invite feedback on the validity and utility of these concepts and recommendations. Clinical information modelling is difficult, but need not be an arcane mystery.

## References

[1] P. J. Scott, M. Berger, I. Carpenter, L. Foley, A new approach to interoperable information standards for health and social care: normalizing culture, contracts and co-design, *Eur J Biomed Inform* **12** (2016), en28-en35.
[2] B. Christensen, G. Ellingsen, Evaluating model-driven development for large-scale EHRs through the openEHR approach, *Int J Med Inform* **89** (2016), 43-54.
[3] B. Selic, The pragmatics of model-driven development, *IEEE Software* **20** (2003), 19-25.
[4] Object Management Group, *Unified Modeling Language*. Available from: http://www.uml.org.
[5] B. Blobel, W. Goossen, M. Brochhausen, Clinical modeling - a critical analysis, *Int J Med Inform* **83** (2014), 57-69.
[6] P. J. Scott, S. Bentley, I. Carpenter, D. Harvey, J. Hoogewerf, M. Jokhani, D. Kalra, S. Kay, I. McNicoll, M. B. Ram, R. Q. Siddiqui, R. Wallhouse, R. Worden, Developing a conformance methodology for clinically-defined medical record headings: A preliminary report, *Eur J Biomed Inform* **11** (2015), 23-30.
[7] D. Brammen, H. Dewenter, K. Heitmann, V. Thiemann, R. Majeed, F. Walcher, R. Röhrig, S. Thun, Mapping Equivalence of German Emergency Department Medical Record Concepts with SNOMED CT After Implementation with HL7 CDA, *Stud Health Technol Inform* **243** (2017), 175-179.
[8] J. Meyer, R. Land, Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning, *Higher Education* **49** (2005), 373-388.
[9] P. M. Leonardi, Crossing the implementation line: The mutual constitution of technology and organizing across development and use activities, *Communication Theory* **19** (2009), 278-310.
[10] E. Ennen, A. Richter, The whole is more than the sum of its parts—or is it? A review of the empirical literature on complementarities in organizations, *Journal of Management* **36** (2010), 207-233.
[11] S. Newell, C. Tansley, J. Huang, Social capital and knowledge integration in an ERP project team: the importance of bridging and bonding, *British Journal of Management* **15** (2004).
[12] G. Schadow, C. McDonald, *The Unified Code for Units of Measure*, Available from: http://unitsofmeasure.org/ucum.html.