

**Validating Clinical Terminology Structures:
Integration and Cross-Validation of Read Thesaurus and GALEN**

JE Rogers¹ C Price² AL Rector¹ WD Solomon¹ N Smejko²

¹Medical Informatics Group, Department of Computer Science, University of Manchester, UK

²NHS CCC, Woodgate, Loughborough, Leicestershire, UK

<http://www.cs.man.ac.uk/mig/giu/> or e-mail GALEN@cs.man.ac.uk

A European pre-standard and an intermediate representation facilitated exchange of two independently authored compositional knowledge bases: one formal and automatically classified, the other manually classified. The exchange highlights different strengths and weaknesses in each approach, and offers a mechanism for partial, mutual quality assurance.

INTRODUCTION

Many new terminologies currently in development (e.g. DICOM SNOMED Microglossary, SNOMED RT¹, LOINC, ICNP², Read Thesaurus³), have adopted some form of compositional technique in preference to the enumerative approach embodied by traditional schemes such as ICD or the Read Codes version 1 or 2. The clinical expressivity of enumerative schemes is limited by whether appropriate, specific terms already exist in the scheme. Compositional schemes offer greatly increased expressivity by allowing complex descriptions to be fashioned from structured collections of more basic terms.

Standardisation work in Europe reflects this move to compositional techniques. The European Committee for Standardisation (CEN) has produced several standards and pre-standards following ENV 12264, itself a pre-standard for representing terminologies as a semantic network⁴. Of particular relevance to the work described in this paper is CEN ENV 1828, a pre-standard proposing a compositional structure for classifications of surgical procedures⁵.

The increased expressive power of a compositional approach comes at a price: the task of maintaining integrity and coherence in an informal scheme quickly becomes too complex to rely on human processing power alone. Conversely, fully formal representations of a scheme are difficult for authors to master.

Within Europe, two large organisations have independently been developing separate compositional schemes for surgical procedures: the GALEN Organisation, and the NHS Centre for Coding and Classification (NHS CCC). Both centres have based their schemes roughly on ENV 1828, but otherwise have used different approaches.

In the GALEN approach^{6,7,8}, the semantics of individual surgical procedure rubrics are first 'dissected' into an intermediate representation. This aims to protect the authors from the complexity of the final scheme. Dissections are then transformed automatically into a Description Logic, GRAIL⁹. The final classification of rubrics is derived automatically by the GRAIL classification engine. Correct classification

depends primarily on the quality of the semantic dissection of each rubric and, more significantly, the quality and richness of the reference model of medicine and anatomy to which those dissections are then mapped and within which they become integrated.

The GALEN Organisation is the commercial outgrowth of the GALEN programme. This European Union funded research has developed compositional terminologies and the tools and methodologies to maintain and deliver them. Results from earlier projects - the GRAIL formalism, GALEN Common Reference Model (CRM), High Level Ontology and Terminology Servers - are providing support for this task.^{10, 11, 12, 13, 14}

In the NHS CCC approach, the classification of rubrics is manually determined. The semantic meaning of the rubrics is described later in a set of qualifier-value pairs. Such 'atomisation' of the meaning is largely independent of the classification process, although a series of *post hoc* procedures are used to identify some inconsistencies between the atomisation and the classification.

The NHS CCC, funded by the United Kingdom Department of Health, is responsible for the maintenance and further development of the Read Codes. Additional functions include management of the UK Surgical Procedure Classification (OPCS4) and acting as the UK co-ordinating centre for ICD10 implementation. From 1992-95, the NHS CCC managed a series of multi-professional Terms Projects¹⁵ designed to capture natural clinical language to be incorporated into the new Read Thesaurus.

In the Thesaurus³, concepts are labelled with terms and meaningless coded identifiers and accommodated in a directed acyclic graph based on sub-type relationships. A template table uses object-attribute-value triples to assign semantic definitions and to provide partially compositional functionality by specifying optional qualifiers.

This paper reports on an experiment undertaken jointly with the NHS CCC and GALEN. It aimed to determine:

Figure 1 : Transformation of READ templates into dissections in the GALEN intermediate representation

Extract of 'atomic qualifier' data file from NHS CCC in template file	71000 Transethmoidal hypophysectomy X9002 Approach X812M Transethmoidal 9 A F 71000 Transethmoidal hypophysectomy X900S Method X793K Excision - action 9 A F 71000 Transethmoidal hypophysectomy X9019 Site Xa06A Pituitary structure 9 A F
GALEN 'dissection' resulting from transformation of same data	RUBRIC 'Transethmoidal hypophysectomy' SOURCE 'READ' CODE '71000' READ_MAIN excision action HAS_APPROACH transethmoidal SITE pituitary structure

- Whether the GALEN intermediate representation, designed initially for primary authoring, could be used as an exchange format between GALEN and Read.
- Whether such exchange would enable the content or classification of both systems to be compared and, thus, detect inconsistencies or errors of fact in either resource.

METHODOLOGY

Read Template Files

At the conclusion of the Clinical Terms Projects, lists of surgical procedures from fifteen specialty panels were integrated into a single, anatomically based branch of the hierarchy. Non-clinical concepts from earlier versions of the Read Codes were also incorporated, but with an *optional* rather than *current* status. The surgical procedures were manually classified using the following principles:

- Multiple classification was attempted for all *current* concepts, by assigning all appropriate *current* subordinates
- *Optional* concepts were assigned to a single parent only (to reflect the analytical functionality of earlier Read Code versions) and no attempt was made to assign full sets of subordinates to *optional* concepts.

An anatomical site was initially assigned to each procedure using the template table¹⁶. Subsequently, additional intrinsic characteristics were assigned based broadly on the ENV 1828 model¹⁷. In this development work, it was considered important to progress one prototype sub-domain to completion at an early stage. A hierarchy of all operations on the endocrine system was selected as it was of interest to a number of different specialties; included a representative range of surgical deeds and other intrinsic characteristics; involved a number of different body regions (intracranial, intra-abdominal etc) and yet was of manageable size.

Data for endocrine procedures were exported from the main Read Code database, including:

- An extract from the hierarchy table (holding parent/child subtype relationships) for all endocrine surgical procedures.
- An extract from the template table containing semantic

definitions (atomic dissections) for endocrine procedures.

Transformation to GALEN Dissections

The GALEN methodology⁶ initially paraphrases expressions manually to remove obvious ambiguity - e.g. 'amputation of hand and foot' usually meaning 'amputation of the hand OR foot' - then captures the semantics of the paraphrases as compositional 'dissections' using an intermediate representation. This format can later be transformed to the underlying formalism, GRAIL⁹.

For this experiment, the atomic qualifier data provided by NHS CCC was automatically transformed directly into the style of GALEN dissections (Figure 1). The transforming algorithm exploited the fact that both Read and GALEN had independently complied with the CEN ENV 1828 pre-standard: in Read, the atomic qualifier linked to the rubric via the 'Method' attribute was reliably equivalent to the 'MAIN deed' of a GALEN dissection.

The resulting 'Read dissections' were imported into a pre-existing software environment for processing dissections. The individual links (e.g. SITE, EXTENT) and descriptors (e.g. excision action, transethmoidal) of the Read dissections were extracted for mapping to their corresponding concepts within the GALEN Common Reference Model (CRM). Some *de novo* modelling of endocrine anatomy in the CRM was required to complete this task.

With the link and descriptor mappings in place, Read dissections were expanded into GRAIL and presented to a GRAIL engine. This allowed a new, automatic classification of the original Read rubrics to be derived, by integrating CCC's semantic dissections of each rubric into the existing semantic content of the GALEN CRM.

RESULTS

Intermediate representation as exchange format

The original Read template file contained 462 atomic qualifier triples from which 162 GALEN-style dissections were extracted and imported into the GALEN toolset. 71 unique descriptors from 398 occurrences and 7 unique links from 235

occurrences were identified in these dissections.

Further processing of the dissections - mapping them to and classifying them within the CRM - required all unique descriptors and links to be separately mapped to corresponding concepts in the CRM. Four principal differences in representational style were identified during this task, reflecting both arbitrary choices and necessary consequences of the respective underlying formalisms. These differences divide into those that prevented further processing of affected dissections, and those that did not:

Representational Differences Preventing Processing

Degree of atomisation. The authors in each scheme did not always make the same choices regarding whether, or how, to decompose a given concept into constituent atoms. The treatments of the modifiers unilateral/bilateral illustrate this: in Read, they modify the structure operated upon, whilst in GALEN they modify the deed itself. Further, the Read authors do not express unilateral/bilateral as separate atoms: they instead remain embedded within primitive entities. For example, Read uses the primitive descriptor 'bilateral adrenal glands' in dissecting 'Bilateral adrenalectomy':

READ_MAIN excision action
SITE bilateral adrenal glands

By contrast, the equivalent GALEN dissection would be:

GALEN_MAIN excision action
HAS_LATERALITY bilateral
ACTS_ON adrenal gland

Automatic transformation between such dissections - to recognise them as semantically equivalent - would have required linguistic and graph manipulation tools to manage their semantic and structural differences. These tools were outside the scope of this limited experiment.

Disjunction & Conjunction. Developers of the Read Thesaurus are constrained by the requirement to preserve forward compatibility from earlier versions of the scheme. For this reason, rubrics such as 'thyroid gland and parathyroid gland operations' are preserved in the hierarchy, although with an *optional* status. In common with other enumerative scheme rubrics, the word 'and' is misleading here: the rubric is a disjunctive term, meaning operations on either the thyroid or the parathyroid gland, rather than operations on both structures concurrently.

The Read developers express such rubrics by creating a disjunctive descriptor term within the thesaurus, corresponding to the notion of a disjunctive structure:

READ_MAIN clinical action
SITE thyroid and parathyroid structures

In GALEN, by contrast, conjunctive and disjunctive operators must be made explicit:

GALEN_MAIN clinical action
ACTS_ON thyroid structure / parathyroid structure

Such dissections, therefore, also could not be automatically transformed between schemes without new tools.

Representational Differences *not* Preventing Processing

Nesting. Read dissections typically have a flatter structure than those produced by GALEN-native authors:

READ_MAIN excision action
SITE thyroid structure
SURGICAL_PATHOLOGY disorder of thyroid gland

GALEN_MAIN excising
ACTS_ON lesion
HAS_LOCATION thyroid gland

This occurs because recursive nesting of atomic qualifiers is not supported by the Read file format. This difference did not affect classification of the processed dissections within this experiment.

Reflexivity. In the Read Thesaurus, part-whole relations are reflexive. For example, in the Read thesaurus, the term 'thyroid structure' subsumes both the organ as a whole as well as all of its recognised anatomical subparts¹⁸. GRAIL, on the other hand, is irreflexive and so the GALEN CRM does not provide a precisely equivalent term in such cases.

This difference in the formalisms was worked around within this experiment by mapping Read's 'thyroid structure' to [ThyroidGland] in the CRM. The GRAIL transitivity mechanism ensured that 'thyroid procedure' still subsumed 'procedure on part of thyroid', even though the descendants of the two elementary thyroid concepts in their respective anatomical hierarchies are not the same.

Comparison of Classification

149 of the original 162 Read dissections were successfully converted to GRAIL and classified. Thirteen dissections were not processed: three used descriptors whose meaning required clarification with an expert; ten could not be processed further without manual editing to transform a representational difference - four employed descriptors with embedded bilateral/unilateral and six contained disjunctive descriptors. Manual editing of these ten was not performed to allow an assessment of the suitability of the intermediate representation as a target for direct, automatic exchange. The machine-

derived classification of the 149 processed dissections was compared with the manual classification of the same 149 rubrics as provided by the NHS CCC.

Agreed (Common) Classifications. The two classifications were agreed on 95 parent-to-child-rubric relationships.

Missed Classifications. Eight relationships were present in the Read hierarchy but were not confirmed by the GRAIL classifier. 52 relationships found by the GALEN classifier were not present in the manually asserted Read hierarchy. Five main causes of missed classifications were identified:

Read dissection errors: Correct classification can not occur if the semantic information in the Read dissections is incorrect. Two 'Read only' relationships were missed in the derived hierarchy because, in both cases, a property ascribed to a parent had not been carried through to its child.

Twenty-four 'GALEN only' relationships were inferred because five Read dissections each contained a single semantic omission. The GRAIL concept generated from such erroneous semantic information is necessarily more general than the rubric's true meaning. The concept's derived classification is correspondingly higher in the hierarchy: spuriously with respect to the rubric's meaning as understood by humans, but correct with respect to the meaning as it was presented to the computer. In this higher position, it acquires children it did not previously have - often the terms that were originally its siblings.

For example, both dissections for partial- and total substernal thyroidectomy omitted to mention their approach element, so that the generated GRAIL concepts actually corresponded to the more general notions of partial- and total thyroidectomy, regardless of approach. The two terms between them acquired eleven false children, including 'Hemithyroidectomy' and 'Lobectomy of thyroid gland'. The remaining three semantic omissions were: 'excision of thyroglossal fistula' did not include fistula as the pathology atom (three false children); 'removal of thyroid nodule' did not include the nodule (six false children); 'Surgical biopsy of endocrine system NOS' did not include 'NOS - Operation' as a classification atom (four false children).

Incorrect descriptor-to-CRM mappings: a single 'Read-only' relation and two 'GALEN only' relations were attributed to incorrect mappings of two descriptors to the CRM. For example, the Read descriptor 'modification' had been mapped to the very general CRM notion of 'any form of surgery', when a more correct and specific interpretation would have

been 'any form of morphological change'.

Differences of opinion between anatomical models: Two 'Read only' relations arose because of differences of opinion regarding anatomy. One concerned Read's classification of the thymus as an endocrine gland; although clinically customary, some anatomical thinking does not support this and it is not present in the GALEN model. Eight 'GALEN only' parent-child relations also reflected anatomical dispute: six of these concerned the thyroglossal tract and associated structures and lesions, which GALEN modelled as part of the thyroid gland whilst Read did not.

Errors or inconsistencies in underlying knowledge models: The classifier missed two 'Read only' relations because of errors in either the Read or GALEN model. In one, persistent patent thyroglossal duct was not flagged as pathological and (therefore) not known to be a subtype of thyroglossal duct pathology in the CRM. In the other, modelling in the Read thesaurus was inconsistent regarding the notions 'excising' and 'removing'; as primitives, excising was given as a more specific form of removing. However, one rubric reading 'removing of...' had been classified manually as more specific than one reading 'excising of...'. Nine 'GALEN only' relations appeared justifiable on formal and semantic grounds. Study of the rubrics themselves confirmed they might be genuine omissions from the Read hierarchy. For example, 'Endocrine surgical biopsy' might reasonably subsume biopsies of the adrenal, parathyroid and thyroid glands.

GALEN transitivity and partonomy: the GRAIL transitivity mechanism has the effect that 'excision of gland' will subsume 'excision of part of gland'. One 'Read only' classification was attributed to this whilst nine 'GALEN only' parent-child relations arose: 'parathyroidectomy', 'thyroidectomy', 'hypophysectomy' and 'adrenalectomy' each erroneously subsumed excision of lesions located in the corresponding gland, whilst 'thyroid incision' also in error subsumed 'incision of thyroid lesion'.

The GRAIL classifier also suggested that Read's endocrine surgery chapter might expand to include rubrics from other chapters in the Read scheme. For example, it suggested that 'Endocrine system procedure' might subsume all testicular and all ovarian surgery, as the CRM considered these to be endocrine organs, and the Read dissections had defined endocrine surgery as any surgery on any endocrine organ. Few clinicians, however, would accept this classification. This example illustrates how unexpectedly subtle semantics in

apparently familiar rubrics may be missed in a formal approach, if taken out of context. A mechanism exists to capture such distinctions, but they must first be detected.

Misclassifications. Eight misclassifications arose as a result of semantic omission in the Read dissections: Two ‘inverted’ classifications were encountered, in which the parent-child relationship between two rubrics in the Read hierarchy was precisely inverted in the derived GALEN hierarchy. In one a ‘total excision of ...’ rubric was not supplied with the atomic qualifier for *extent:total*. In the other a rubric reading ‘removal of thyroid nodule’ had been given the atomic qualifier *method:excising*.

Six pairs of two different procedure rubrics were identified where both had an identical Read dissection. These had been projected to the same concept in the derived hierarchy.

DISCUSSION

This preliminary experiment demonstrates that, at least in this case, a knowledge base authored in the context of one compositional model may be re-expressed within another similar model. It also shows how difficulties may derive from technical differences in the underlying formalisms such as conjunction and disjunction, or in authoring style such as degree of atomisation.

The experiment showed that a common standard and an intermediate representation can facilitate interchange. However, devils remain in the detailed semantics such as those for ‘bilateral’ or ‘excision’. Whether all such semantic issues will ever be standardised remains in doubt.

The knowledge authoring process in compositional approaches is a very expensive undertaking. Our experience suggests that formal, automatic classification and manual classification have different strengths and weaknesses. Exchanging information between schemes provides a new and powerful mechanism by which each can systematically contribute to quality assurance and verification of the other.

Formal classification detects many missed classifications, analytical errors and omissions of the manual approach. However, the formal model on its own without qualification may suggest clinically spurious classifications (e.g. Testicular Surgery under Endocrine Surgery). Similarly, classifications may be missed that remain clinically customary even though they may no longer be formally justifiable (e.g. Thymic surgery under Endocrine Surgery). The manual classification provides an effective and necessary check on the formal model

both for correctness and for clinical appropriateness¹⁹.

In summary, our experience suggests that transformation between independently developed compositional systems is possible but not trivial, and that the resolution of discrepancies between the resulting classification structure can provide a useful quality assurance mechanism for both. The experiment is being repeated with 2500 imported Read dissections.

Acknowledgements

With thanks to all in the GALEN-IN-USE consortium. GALEN-IN-USE is funded as part of Framework IV of the EC Healthcare Telematics research program.

References

- 1 Spackman KA, Campbell KE, Cote RA. SNOMED RT: A Reference Terminology for Health Care. In Masys DR (Ed). *Proceedings of the 1997 AMIA Annual Fall Symposium*. Philadelphia: Hanley & Belfus, 1997: 640-644.
- 2 Mortensen RA (ed) *The International Classification for Nursing Practice ICNP with TELENURSE introduction*. Copenhagen: The Danish Institute for Health and Nursing Research, 1996
- 3 O’Neil MJ, Payne C, Read JD. Read Codes Version 3: A User Led Terminology. *Meth Inform Med* 1995; 34: 187-92.
- 4 CEN ENV 12264:1995. Medical Informatics - Categorial structure of systems of concepts - Model for representation of semantics. Brussels: CEN, 1995
- 5 CEN ENV 1828:1995 Health care informatics - Structure for classification and coding of surgical procedures. Brussels: CEN, 1995
- 6 Rogers JE, Solomon WD et al. (1997) Rubrics to Dissections to GRAIL to Classifications. Fifteenth International Congress of the European Federation for Medical Informatics, MIE-97 Thessaloniki, Greece
- 7 Rossi Mori A, Consorti F, Galeazzi E, (1997) Standards to support development of terminological systems for healthcare telematics. (proceedings of IMIA Working Group 6 meeting, Jacksonville, Florida)
- 8 Rogers J, Rector A (1997). Terminological Systems: Bridging the Generation Gap. Annual Fall Symposium of American Medical Informatics Association. Nashville TN, Hanley & Belfus Inc. Philadelphia PA: 610-614.
- 9 Rector A, and Nowlan WA (1993). The GALEN Representation and Integration Language (GRAIL) Kernel, Version 1. The GALEN Consortium for the EC AIM Programme. (Available from Medical Informatics Group, University of Manchester).
- 10 Rector A (1994). Compositional models of medical concepts: towards re-usable application-independent medical terminologies. Knowledge and Decisions in Health Telematics P. Barahona and J. Christensen (ed). IOS Press. 133-142.
- 11 Rector A, Gangemi A, Galeazzi E, Glowinski A and Rossi-Mori A (1994). The GALEN CORE Model Schemata for Anatomy:

Towards a re-usable application-independent model of medical concepts. Twelfth International Congress of the European Federation for Medical Informatics, MIE-94, Lisbon, Portugal, 229-233

- 12 Rector A (1995). Coordinating taxonomies: Key to re-usable concept representations. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95), Pavia, Italy, Springer. 17-28.
- 13 Rector A, Rogers JE, Pole P (1996) The GALEN High Level Ontology. Fourteenth International Congress of the European Federation for Medical Informatics, MIE-96, Copenhagen, Denmark
- 14 Rector A, Solomon WD, Nowlan WA and Rush T (1995). A Terminology Server for Medical Language and Medical Information Systems. *Methods of Information in Medicine*, Vol. 34, 147-157
- 15 Severs MP. The Clinical Terms Project. *Bulletin of Royal College of Physicians (London)* 1993; 27(2): 9-10.
- 16 Price C, Bentley T, Brown P et al. Anatomical Characterisation of Surgical Procedures in the Read Thesaurus. . In Cimino JJ (Ed). *Proceedings of the 1996 AMIA Annual Fall Symposium*. Philadelphia: Hanley & Belfus, 1997: 110-114.
- 17 Price C, Brown PJB, Bentley TE, O'Neil MJ. Exploring the Ontology of Surgical Procedures in the Read Thesaurus. *Proceedings of Conference on Natural Language and Concept Representation*. Florida: IMIA WG6, 1997: 215-221.
- 18 Schulz EB, Price C, Brown PJB. Symbolic Anatomical Knowledge Representation in the Read Codes Version 3: Structure and Application. *JAMIA* 1997; 4: 38-48.
- 19 Schulz EB, Barrett JW and Price C. Semantic Quality through Semantic Definition: Refining the Read Codes through Internal Consistency. In Masys DR (Ed). *Proceedings of the 1997 AMIA Annual Fall Symposium*. Philadelphia: Hanley & Belfus, 1997: 615-619.