

The Practice of Informatics

JAMIA

Application of Technology ■

Read Code Quality Assurance: From Simple Syntax to Semantic Stability

ERICH B. SCHULZ, MB, BS, JAMES W. BARRETT, MB, BCHIR, MSC,
COLIN PRICE, MPhil, FRCS

Abstract As controlled clinical vocabularies assume an increasing role in modern clinical information systems, so the issue of their quality demands greater attention. In order to meet the resulting stringent criteria for completeness and correctness, a quality assurance system comprising a database of more than 500 rules is being developed and applied to the Read Thesaurus. The authors discuss the requirement to apply quality assurance processes to their dynamic editing database in order to ensure the quality of exported products. Sources of errors include human, hardware, and software factors as well as new rules and transactions. The overall quality strategy includes prevention, detection, and correction of errors. The quality assurance process encompasses simple data specification, internal consistency, inspection procedures and, eventually, field testing. The quality assurance system is driven by a small number of tables and UNIX scripts, with "business rules" declared explicitly as Structured Query Language (SQL) statements. Concurrent authorship, client-server technology, and an initial failure to implement robust transaction control have all provided valuable lessons. The feedback loop for error management needs to be short.

■ JAMIA. 1998;5:337-346.

Recent years have seen rapid developments in clinical vocabularies,¹ with increases in both breadth and depth of coverage and also data structure complexity. Modern controlled medical vocabularies support many functions, including order communication, recording of clinical summaries and more detailed health care records, decision support, clinical research,

and data aggregation for resource management and epidemiologic purposes.

The Read Thesaurus² is an extensive controlled medical vocabulary comprising more than 220,000 concepts in a directed acyclic graph subtype hierarchy. It aims to be a comprehensive collection of concepts uniquely identified by codes and clearly labeled by clinical terms. Several reports have described the features of its version 3 Read Code file structure,^{3,4} which provides flexibility and includes the facility to post-coordinate and semantically define concepts. There is, ideally, a one-to-one relationship between each *code* and the pre-coordinated *concepts* in the thesaurus, and the provision of synonyms allows a many-to-many relationship between *codes* and *terms*. Post-coordination using predetermined combinations specified as object-attribute-value triples in a *template table* allows the construction of additional, more detailed concepts in a coherent, controlled manner.

Affiliation of the authors: National Health Service (NHS) Centre for Coding and Classification, Loughborough, England.

Correspondence: Dr. Colin Price, NHS Centre for Coding and Classification, Woodgate, Loughborough, Leicestershire LE11 2TG, England. e-mail: (Colin.Price@nhsccc.exec.nhs.uk).

Reprints: Dr. Erich B. Schulz, National Centre for Classification in Health (Brisbane), Queensland University of Technology, Kelvin Grove 4059, Australia.

Received for publication: 11/24/97; accepted for publication: 3/23/98.

Concepts within the thesaurus are semantically defined, or decomposed into their constituent base concepts. Recent reports have described our experience in the domains of surgical procedures^{5,6} and disorders.⁷ These semantic definitions allow the vocabulary to be introspective,⁸ or capable of self-validation.

The thesaurus, in common with the UMLS⁹ and SNOMED,¹⁰⁻¹² has undergone iterative, evolutionary development. Its content is now derived from a number of sources, including previous versions of the Read Codes (four-byte and version 2),^{13,14} statistical classifications,¹⁵⁻¹⁷ and substantial specialist clinical input during the United Kingdom Terms Projects.¹⁸⁻²⁰ Large-scale manual integration of concepts from these different sources²¹ was performed during the initial assembly of the thesaurus, but the subsequent development of automated methods for completing the task and refining the product was always anticipated.

We have recently reported the organizational aspects of user-driven maintenance of the Read Codes.²² In this paper, we discuss the development and application of the Read Thesaurus quality assurance processes at the U.K. National Health Service Centre for Coding and Classification (NHS CCC).

Quality Matters!

Quality assurance of clinical coding systems (including controlled vocabularies and clinical classifications) has not previously attracted significant attention. The traditional use of coded clinical data only for aggregation to support population-based analysis for epidemiology, resource allocation, and audit is arguably reasonably fault-tolerant. With scarce resources available for data collection, a certain miscoding rate is accepted, and the effects of such coding errors are dampened somewhat by data aggregation. The very inflexibility of early coding systems rendered them less error-prone, and when errors did occur, correction proved difficult.

Modern coded vocabularies, however, are likely to be a central component of clinical patient-centered information systems,²³ driving report generation, user interfaces, and decision support. Errors or omissions may lead to malfunctioning of the application and ultimately to inappropriate clinical care decisions. In general, the conclusions drawn from using a clinical terminology, whether for an individual patient or a population, will be only as good as the terminology itself. Quality assurance must, therefore, become a major concern for today's developers of controlled vocabularies. This task becomes increasingly challenging as vocabularies become more complex and more flex-

ible to meet clinical needs, and their dynamic nature means that quality assurance must become an ongoing activity, with continuous examination of the impact of changes between formal reviews.

We believe that a high-quality controlled vocabulary is one whose quality can be largely ensured by forcing it to obey its schema. Although we believe that this type of quality assurance is necessary, whether it is adequate to guarantee useful functionality awaits formal evaluation.

Read Thesaurus Database

We have stratified the data that are subject to quality assurance into three types: dynamic editing data, developmental data, and release data.

Dynamic Editing Data

The master databases are edited concurrently by a team of clinical authors and are the sources of all NHS CCC Read Code products. The laws of entropy dictate that these data are at constant risk of degradation, so the focus of the quality assurance process is to minimize the number of errors.

Scheduling of error correction activity involves a degree of tradeoff. Often, the time required to correct a large batch of errors is only marginally greater than that required to correct a single error. However, such corrections often result in the unforeseen generation of further errors, potentially delaying release if errors are allowed to accumulate until just before planned data export. Therefore, a balance must be struck between the overhead of fixing errors as they arise and the need for repeated cycles of error correction at times of data release. The goal is always to ensure that live data are completely correct before export for release so the quality assurance of the release data is simply a check of the export and transformation program.

Developmental Data for Discussion and Feedback

New sections of the thesaurus are often added provisionally, and demonstration software is issued to allow review and comment by external agencies including specialist clinicians and systems developers.²² These provisional data do not need to be entirely correct, providing that the terms and the relationships between them are correctly represented. For browsing software to function, however, these data still need to conform to a basic specification.

Release Data for Clinical Information Systems

Data for incorporation into live clinical information systems must be the most reliable and are, therefore,

subject to the strictest quality assurance procedures. Each release of working data must vary only in a controlled way; in particular, identifiers should not be removed and should retain their meaning.

Sources of Errors

A team of technical and authoring personnel at the NHS CCC use custom-written software in a client-server environment (Figure 1) to maintain three versions of the Read Codes and the associated tables holding cross-mappings to formal classifications.²² A number of factors contribute to the occurrence of errors within the master database, including human errors (author and technical), evolving rules, novel transactions, hardware and software problems, incomplete transactions and, finally, concurrency and version control problems.

Human Errors

Inexperienced new authors and technical staff are a readily identifiable source of errors. Rather than being inhibiting, a robust error detection system enables new personnel to be more rapidly integrated into the working team as both new staff members and their supervisors can be assured that errors will be rapidly identified. The spectrum of human error ranges from simple typographic inaccuracies in terms, such as misspellings and the inclusion of double, leading, or trail-

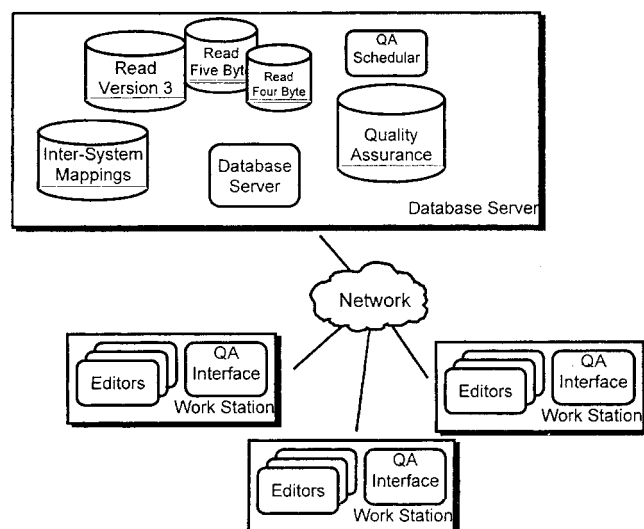


Figure 1 Overview of coding system architecture. Windows PC workstations operated by authoring and technical personnel communicate with the UNIX Oracle 7 database server across a TCP/IP network using ODBC (Open Database Connectivity). Workstations run several editing and viewing tools developed in a mixture of Visual Basic, Delphi, C++, and Microsoft Access. The quality assurance scheduler is a shell script cron job resident on the server.

ing spaces, to more complex structural errors. These include failing to apply subclass links when creating hierarchies, a common problem with new authors who may, for example, misclassify the symptom *abdominal pain* as a subordinate of *appendicitis*.

Evolving Rules

The thesaurus is dynamic, with new sections added, existing sections restructured and consolidated, and minor enhancements to the file structure introduced to support user-driven requests for additional functionality. These changes usually require introduction of new quality assurance rules, and the first global application of a new rule to the database sometimes identifies a significant number of violations.

Novel Transactions

Large-scale changes to the thesaurus, such as reorganized branches of the hierarchy, are usually undertaken as batch transactions on the database. Performing such a new process for the first time occasionally generates unanticipated errors. For example, in bypassing the checks usually imposed by our editing software, current concepts have sometimes been inadvertently placed as subordinates of optional (“retired”) concepts, effectively excluding them from the current classification hierarchy.

Hardware and Software Problems

Software. Newly developed in-house editing software, often produced to allow a new phase of specific thesaurus enhancement, sometimes results in data corruption. Nevertheless, it is generally more expeditious to roll out new versions of the software and correct any errors that arise during the later debugging phases than it is to undertake exhaustive testing prior to use on the master database.

Incomplete Transactions. A transaction is a set of database updates that changes the database from one valid state to another. The addition of a new concept to the thesaurus, for example, requires insertion of records into a minimum of four tables, and unless the server completes all four inserts, the database will be left in an invalid state. For instance, if a new concept insertion is aborted by a network failure, the new concept may be created but not placed anywhere in the hierarchy. Many earlier problems of this type could have been avoided had robust handling of database transactions been implemented with our first generation client-server editor. Indeed, the unavailability of database connectivity layers that reliably supported transaction commit and rollback was a major source of errors and decreased overall productivity.

Concurrency and Version Control

In our experience, version control has resulted in greater difficulties than concurrency. We have found it easier for the in-house database server to handle concurrent access and editing than to cope with issues of offline editing and re-integration. For example, offline assignment of semantic definitions may assume inheritance from higher nodes in the hierarchy; if the hierarchy is altered in the database before the work is imported, these assumptions may be incorrect.

Quality Development Strategy

There are three key facets in this process: prevention, detection, and correction.

Prevention

Adequate author training combined with good authoring tools goes a long way to preventing many errors. For example, our editing software now prevents authors from adding double or trailing spaces to terms by stripping them as they are added to the database. Release of developmental data enables specialist review of work prior to final release of codes for service use.

Detection

Errors should be reported as soon as possible, ideally immediately and prior to being committed to the database. Database server constraints help trap a number of potential violations. Routine overnight running of quality assurance rules allows detection of new errors, such as placement outside the current hierarchy during a batch transaction, within 24 hours. Review of work by senior authors prior to release provides an additional safeguard; this is particularly useful for detecting inappropriate subtype links. Finally, feedback from users and specialist advisers is facilitated by distribution of browsing software and electronic subsets of the thesaurus.

Correction

Certain conditions in the database may be detected and automatically corrected, and it is sometimes expeditious to write modular scripts to perform these operations rather than embed the logic in the editing tool. The remaining errors usually require experienced author intervention for appropriate resolution. Ideally, this should be facilitated by tools that both display relevant information to the author and allow required edits to be performed.

In a team environment, the allocation of errors to the appropriate author is a task in itself. Errors for correction may be allocated on the basis of the hierarchic

position of the concept, the table in which the error occurred, or the specific rule.

QA Processes and Tools

QA Database Schema

Quality assurance data for all NHS CCC-maintained code sets and mapping tables are held in a single server schema. The principal tables are RULES, RULESCOPE, ERRORS, and EXCEPTIONS. Additional INSPECTION tables are described later.

RULES

This is the master table of more than 500 rules, each consisting of an English text description, a Structured Query Language (SQL) statement, and an identifying number. The majority of rules are expressed as SQL statements in the generic form:

```
INSERT INTO ERRORS (RuleID, Details)
SELECT (RuleID), (Details)
FROM (Relevant tables)
WHERE (Error condition)
```

The *rule identifier* (RuleID) serves to join the error condition with the rule attributes in the RULES table. The polymorphic details field must be interpreted in conjunction with the rule identifier and contains a variable-length string of information enabling location of the error. For tests applied to single entities, this field may contain a single identifier. Rules applied to relationships concatenate the relevant identifiers into a single string.

The ability to embed user-defined functions in Oracle SQL greatly increases its functionality, and we could not have expressed many of our rules in ANSI SQL. Rules vary in their complexity and some are very simple; for example:

Read Code must be 5 characters long and not null:

```
INSERT INTO ERRORS (RuleID, Details)
SELECT 278, readcode
FROM concept
WHERE readcode IS NULL OR
LENGTH (RTRIM (readcode)) != 5
```

The RULES table also contains fields for the following rule attributes:

- On what nights should the rule run?
- Does the rule apply to release, developmental, or live data?
- Is this an error (mandatory to correct) or a warning?
- In which database schema should the rule run?

ERROR_NO	RULE	NUM	LASTRUN
V3_1012	All current concepts must have at least one current parent	1	13/03/98 04:52:55
V3_2015	All preferred terms must be Current or Developmental	1	13/03/98 05:46:27
V3_E030	All three character target codes must have a refineflag of M.	1	13/03/98 08:19:41
V3_6012	Valid characters in terms are: abcdefghijklmnopqrstuvwxyz0123456789^# ?[]!"<>+ =, & * - : () % \ /	15	13/03/98 06:23:36
590	(Sins QA : Procs - sites) Values of all atoms should be null, equal to, or sub-ordinate of at least one value of a super-ordinate atom	128	12/03/98 00:50:59
975	New releasable terms must have passed inspection by 2 authors unless in FB or V2 or extinct preferred term.	189	13/03/98 02:42:24
997	All new Persistent / Redundant entries that fail inspection must be corrected	2	13/03/98 04:38:42

Figure 2 ERRORS table for March 13, 1998. The four "V3_" errors relate to simple data specification. Error 590 reports on a semantic coherence error. Error 975 identifies entities requiring inspection, and error 997 reports failed inspections.

Overnight execution of rule queries enables the processing to be performed while the database server is quiet. All changes to rule SQL are logged in an audit trail.

RULESCOPE

The table RULESCOPE brings order to the implementation of more than 500 rules, each of which may affect multiple fields in multiple tables or data in multiple database server schemas. RULESCOPE is maintained manually each time a new rule is introduced and serves as an index for RULES, providing information such as:

- Which rules test the Four Byte Read Code set?
- Which rules test both the hierarchy table and the concept table?

ERRORS

This table of only four fields is populated by running the UNIX QA scripts each night. These examine RULES to identify all rules scheduled to run, delete the previous day's errors from the ERRORS table, and then rerun the rule SQL. Therefore, this table contains only a snapshot of the database status in the early morning of each working day rather than a dynamic view (Figure 2). After all rules have run, the script performs two more important actions. First, it puts a hierarchy range number²⁴ against each error, so that quick domain-based extraction based on the version 3 hierarchy is possible. Then it marks all exceptions to rules on the basis of the exception table.

EXCEPTIONS

This table is used when there are known exceptions to particular rules—perhaps for historical reasons or because of inadequate resources to correct all errors. It contains just four fields, of which the first two match the errors table and the others maintain a limited audit trail.

Types of Quality Assurance

Our quality assurance mechanisms can be grouped into four broad categories: simple data specification, automatic checks of coherence, human inspection, and field-testing.

Data Specification

The simplest set of rules includes data definition and referential integrity, easily expressible by database constraints. The database engine simply and adequately enforces allowable sets of values within fields (e.g., a term must be a preferred "P" or synonymous "S" label for a concept), the uniqueness of primary keys (e.g., Read codes must be unique), and the existence of foreign keys (e.g., a Read code in the hierarchy table must exist in the concept table). Wherever possible this functionality has been exploited.

Many of the rules that check the data model are, however, more complex and require the use of SQL (or occasionally PL/SQL stored procedures). An example of a rule in this category is "Preferred terms must not be attached to other concepts, except as synonyms of extinct concepts."

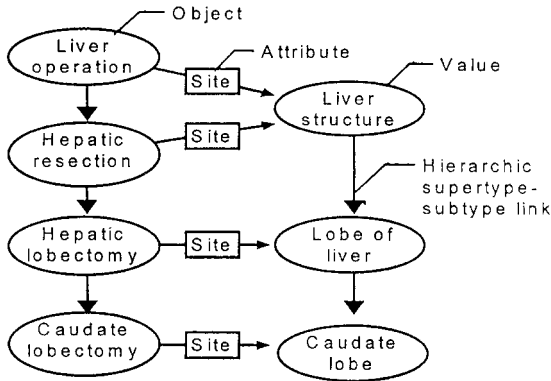


Figure 3 Object-value parallelism. Semantic links specify the intrinsic anatomic characteristic of each liver operation, with increasing detail at lower hierarchy levels.

Simple data specification errors, although the focus of much of our earliest QA activity, have been significantly reduced by refinement of our error reporting processes and debugging of editing software.

Automatic Checks of Coherence

A number of rules are applied to test the semantic coherence of the thesaurus. This covers notions such as redundant hierarchy links, relationships between semantic definitions and hierarchies, self-contained hierarchies, and duplicate semantic definitions.

Traditional knowledge bases often adopt a truth-maintenance strategy, where each new fact is tested prior to integration. Should the new fact be inconsistent, it is rejected. This approach is clearly incompatible with the rapid development and magnitude of the Read Thesaurus, in which tests for internal consistency now play a valuable role in both new development and refinement of the existing content.

While in a large Thesaurus of more than 200,000 concepts, internal inconsistencies are inevitable, their impact in real-world clinical applications remains uncertain and their often complex nature means that correction is potentially expensive.

Internal Consistency Between Semantic Definitions and Subtype Hierarchy. An important tenet in the construction of the thesaurus is to maintain parallelism between the classification of the objects and that of their intrinsic values.²⁵ Two complementary rules are employed to ensure completeness and correctness of both the subtype hierarchy and the semantic definitions. The first rule confirms parallelism between object and value hierarchies (Figure 3). It states that each intrinsic characteristic of a concept must be the same as, or more detailed than, the corresponding characteristic of a superordinate.

The second rule aims to auto-classify concepts based on their semantic definitions (Figure 4 and Appendix). The rule ensures that a concept with characteristics more detailed than those of another concept is a subordinate. Before this rule is applied, the potential superordinate must have been fully defined with respect to a common ancestor.

Although these two rules introduce order and assist in refinement of the underlying concept model, widespread semantic definition is labor-intensive and carries large resource implications. Limitations of the object-attribute-value syntax used for semantically defining concepts⁵ and inability to detect internally consistent errors (Figure 5) further decrease the utility of these rules.

Duplicate Semantic Definitions. The disparate origins of concepts in the thesaurus led to a number of duplicates at the time of initial integration. These have been gradually detected either by visual means, often because they have proximity in the hierarchy or, as semantic definition work has progressed, by discovery of identical definitions. An example of such a duplicate is shown in Figure 6. A separate table within the version 3 data structure enables documentation of the redundant code and its persistent equivalent.

Detecting duplicate semantic definitions allows identification of either incorrect or under-specific semantic definitions, or true duplication of concepts.

Enforcement of Pure Semantic Types. Even in a taxonomy with multiple classifications, there are intrinsic limitations as to the allowable multiple placements of a single concept. The same entity cannot, for example, be both a clinical finding and a clinical procedure or both a microorganism and an anatomic concept. A rule enables restriction of multiple parents for these semantic types, leaving them in self-contained hierarchy branches.

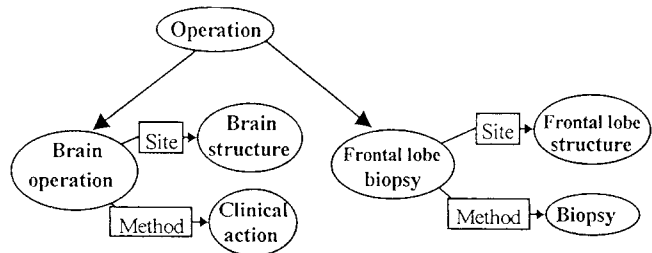


Figure 4 Incomplete classification. The rule will detect that *Frontal lobe biopsy* should be classified as a type of *Brain operation* because the semantic definition values of *Frontal lobe biopsy* (Anatomical site: *Frontal lobe structure*, Method: *Biopsy*) are classed as types of the respective *Brain operation* values. A formal representation of this rule is detailed in the Appendix.

Inspections

While tests for internal consistency provide interesting challenges for our programmers, true assurance of clarity, completeness, and correctness is only achievable through visual inspection of the contents by appropriate domain experts. The costs of such inspection can vary enormously. Depending on the complexity of the task, an author may inspect and assess between 10 and 120 records an hour. Recourse to detailed references (such as textbooks and journals) can significantly increase the time required. In controversial cases, and if the process is not controlled, many person-hours could be spent by committees of experts resolving some of the finer points.

Extension of QA beyond basic technical specification conformance and internal consistency requires rules such as:

- Any change to released terms must be inspected by two clinical authors, or
- For the same Read Code, version 2 terms that are not identical to the version 3 terms must be inspected by two clinical authors.

These inspection tasks are laborious for all involved and require careful management (Figure 7) to prevent needless repetition of work. Given the continuous

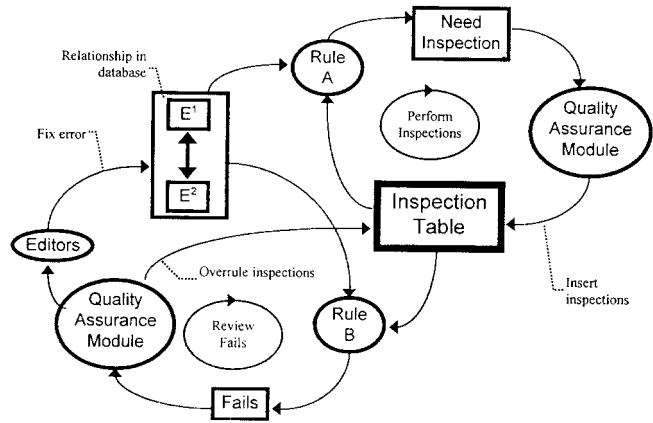


Figure 7 Overview of the inspection process.

multifaceted evolution of the Read Thesaurus, it is inadequate simply to record, say, that a certain branch of the hierarchy was inspected by a particular individual at a given time. Rather, the exact terms inspected must be recorded along with the author's interpretation.

Inspection Logging. The inspection tables now hold thousands of records, representing many hundreds of hours of author work and providing a valuable resource and audit log.

The introduction of a new type of inspection into this system requires, first, an analysis of what precisely is to be inspected. Then a new inspection table is designed, supported by two new QA rules and appropriate extensions to the meta-database within the user interface program.

Inspection Tables. The inspection tables all follow a similar pattern and are the key to the entire system. Each inspection record consists of:

- The information—including terms—actually inspected by the author;
- The result: pass, fail, or equivocal;
- The inspecting author, date and time, and comments.

In most cases the inspected information consists of a relationship between two entities and is contained in multiple fields. For example:

- Inter-Read version
V2 term and V3 term
All V2 terms and V3 preferred term
- Inter-Read release
V3 term in previous release and this release

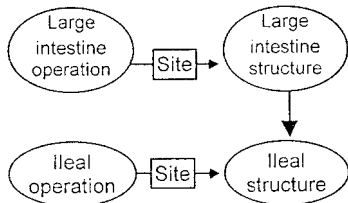


Figure 5 Internally consistent error. The ileum is part of the small intestine but has been incorrectly classified in both the object and value (anatomy) hierarchies. This error is internally consistent and passes automated quality assurance checks.

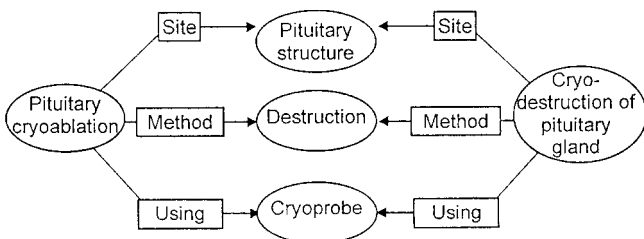


Figure 6 Detected redundancy. Two Read Thesaurus terms originating from different sources, Pituitary cryoablation and Cryodestruction of the pituitary gland, have identical semantic definitions and, in fact, represent the same concept.

```

INSERT INTO errors (RuleId, Details)
SELECT DISTINCT
  572, m.termid || '|'
      || m.readcode || '|'
      || m.termcode
FROM
  V3llive.terms t3,
  v2llive.v2_keys v2,
  v2termv3 m
WHERE t3.termid = m.termid
AND v2.readcode = m.readcode
AND v2.termcode = m.termcode
AND v2.term != t3.term
AND 2 >
  (SELECT COUNT(*)
   FROM quality.v2v3terminspection i
   WHERE i.v2rc = m.readcode
        AND i.v2tc = m.termcode
        AND i.v3tid = m.termid
        AND i.v2t1 = v2.term
        AND i.v3t1 = t3.term
        AND i.result = 'P'
        AND i.overruled = 'N');

```

Figure 8 SQL for rule 572.

- Intra-Read concept

V3 synonym and V3 preferred term

The recording of fails, as well as passes, allows staggering of the inspection and correction processes. It is important to enable inspectors to flag records as being uncertain. This limits the assignment of false positives and false negatives and prevents repeated presentation of the same problem to an author who has already indicated an inability to offer an opinion. If a decision is revised, the record is not deleted but is marked as being overruled and further details are added.

Rules. Two rules are required to support each inspection table, one to identify pending inspections and another to check that failed records have been corrected. For example, rule 572 states, “*All non-identical term maps in the version 2 to version 3 term mapping table must have at least two non-overruled passes in v2v3terminspection.*” Therefore, for rule 572 the set of entities requiring inspection is *non-identical term maps in the version 2 to version 3 term mapping table*, and the required set of inspections is *two non-overruled passes*. The resulting SQL is illustrated in Figure 8.

Both rules must ignore records in the inspection table that are marked as overruled and also those in which the entries have changed since the time of inspection. These rules are computationally expensive, but overnight execution prevents the database server from becoming overloaded during the working day.

Modification of the rule criteria allows application of inspections to limited domains, such as:

- New concepts pending release
- Current released concepts
- Particular branches of the hierarchy (e.g., all diseases)
- Particular clinical speciality subsets (e.g., all ophthalmology-related concepts)

These filtering clauses allow the NHS CCC to schedule and prioritize its work.

Quality Assurance Interface Module. Generating reports for authors, filing away their interpretations, and then taking appropriate action is time-consuming and potentially error-prone. A request for “*... a report of all term mappings in the diseases chapter that have not yet been passed twice, have not failed at all but have been rated as equivocal by at least two authors, including all the inspections that have been performed this morning,*” becomes a page of SQL.

The quality assurance interface module (Figure 9) generates and allows processing of these reports, including initial inspections, reviewing failures, and counting outstanding inspections. Authors are able to limit the set to be inspected in a number of ways, by excluding, for example, those already viewed by the inspecting author, those already failed by any author, those that fall outside a particular hierarchy branch, those marked equivocal by two or more authors, or those not marked equivocal by two or more authors.

Technical staff set up new types of inspection by inserting new meta-data about the inspection process. These meta-data enable generation of appropriate SQL statements for execution on the database server and also dynamic configuration of graphical user interface forms.

Field-testing

It is generally easier for our human domain experts to identify an error than to detect omissions, and we have found that external review by domain experts generally ensures correctness rather than completeness. While experts are often interested in new and topical concepts, field testing quickly reveals the absence of frequently occurring entities that experts have overlooked. Recent *in vitro* studies have identified incompleteness as a significant shortcoming of available controlled clinical vocabularies.^{26,27} The Read Thesaurus is now undergoing *in vivo* operational testing in a number of sites, and a recent analysis of 1,022 individual feedback items reported between June 1, 1995, and August 31, 1996, revealed just two spelling errors and 129 queries relating to administrative

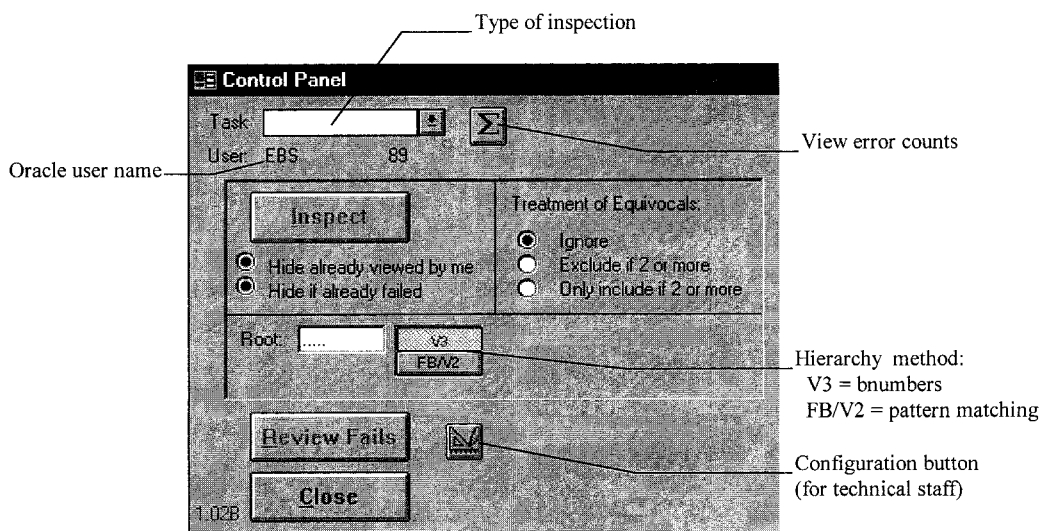


Figure 9 Inspection database control panel form.

cross-mappings; the remainder identified omissions.²² Responding to such feedback is a vital component of our quality improvement strategy.

Lessons Learned

Managing the consolidation and enhancement of the Read Thesaurus since its introduction in 1994 has provided valuable lessons for the NHS CCC. In particular, the establishment and refinement of additional QA processes have provided challenges for both authoring and technical staff.

The maintenance of a dynamic comprehensive clinical vocabulary requires a team of authors using sophisticated tools. The overheads of migrating to a client-server environment capable of supporting multiple concurrent authorship should not be underestimated. The implementation and management of reliable networks and database servers require experienced, dedicated personnel.

A further lesson is to provide as short a feedback loop as possible. This is important because errors are best rectified while the original authoring thought processes are still fresh and before secondary errors result.

As might be expected, it has proved far easier to define rules than to enforce them within existing manpower constraints. It is also far easier to maintain the ever-changing business rules within a rule base than to incorporate them in the logic of the editing software. Each rule is expressed as a single procedure that tests all relevant records in the involved tables, enabling efficient processing by the SQL engine. This does, unfortunately, preclude these rules from being applied during each transaction.

Conclusion

As clinical vocabularies become larger, more complex, and more flexible, quality assurance becomes increasingly challenging and must also be an integral component of both development and maintenance. The ultimate goal of the quality assurance process is to ensure that the products produced by the NHS CCC are fit for purpose. This is dependent on completeness, correctness, consistency, and conformity to technical specification. Automatic tests of internal consistency are an essential adjunct to this process. They are not a panacea, however, and detected errors can rarely be corrected automatically.

Prevention of errors is the ultimate goal of our quality strategy for the Read Thesaurus. However, vocabulary designers must recognize that the interaction between human authors and computers in the production and maintenance of increasingly sophisticated schemes will always carry a risk of errors and that an important aspect of vocabulary design is to allow for their correction. After all, "the man who makes no mistakes does not usually make anything" (E. J. Phelps, 1822–1900).

Simon Warrick and Nick Smejko made significant technical contributions to the development of the Read Thesaurus Quality Assurance rules database. The authors thank their reviewers for their helpful comments.

References ■

1. Cimino JJ. Coding systems in health care. In: Van Bommel JH, McCray AT (eds). *Yearbook of Medical Informatics 1995*. Stuttgart, Germany: Schattauer, 1995:71–85.

2. Calman KC. New national thesaurus. CMO's Update. 1994; 4:1.
3. O'Neil MJ, Payne C, Read JD. Read Codes Version 3: A User Led Terminology. *Methods Inf Med.* 1995;34:187-92.
4. Schulz EB, Price C, Brown PJB. Symbolic anatomic knowledge representation in the Read Codes version 3: structure and application. *J Am Med Inform Assoc.* 1997;4:38-48.
5. Price C, Bentley TE, Brown PJB, Schulz EB, O'Neil MJ. Anatomical characterisation of surgical procedures in the Read Thesaurus. In: Cimino JJ (ed). *Proc AMIA Annu Fall Symp.* 1996:110-4.
6. Price C, Brown PJB, Bentley TE, O'Neil MJ. Exploring the ontology of surgical procedures in the Read Thesaurus. In: Chute CG (ed). *Proceedings of the Conference on Natural Language and Medical Concept Representation.* Jacksonville, Fla: IMIA Working Group 6, 1997:215-21.
7. Brown PJB, O'Neil MJ, Price C. Semantic representation of disorders in version 3 of the Read Codes. In: Chute CG (ed). *Proceedings of the Conference on Natural Language and Medical Concept Representation.* Jacksonville, Fla: IMIA Working Group 6, 1997:209-14.
8. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an introspective, multipurpose, controlled medical vocabulary. *Proc 13th Annu Symp Comput Appl Med Care.* 1989: 513-8.
9. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inform Med.* 1993;32: 281-91.
10. Côté RA, Rothwell DJ, Palotay JL, Becket RS, Brochu L. The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. 4 vols. Northfield, Ill: College of American Pathologists, 1993.
11. Lipow SS, Fuller LF, Keck KD, Olson NE, Erlbaum MS, Tuttle MS, et al. Suggesting structural enhancements to SNOMED International. *Proc AMIA Annu Fall Symp.* 1996: 901.
12. Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Gálapagos: computer-based support for evolution of a convergent medical terminology. *Proc AMIA Annu Fall Symp.* 1996:269-73.
13. Bentley TE, Price C, Brown PJB. Structural and lexical features of successive versions of the Read Codes. In: Teasdale S (ed). *Proceedings of the Annual Conference of the Primary Health Care Specialist Group.* Worcester, England: PHCSG, 1996:91-103.
14. Schulz EB, Barrett JW, Brown PJB, Price C. The Read Codes: evolving a clinical vocabulary to support the electronic patient record. In: *Conference Proceedings: Toward an Electronic Health Record Europe.* Newton: CAEHR, 1996:131-40.
15. World Health Organization. *International Classification of Diseases, 9th Revision, Clinical Modifications.* Geneva, Switzerland: WHO, 1975.
16. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision.* Geneva, Switzerland: WHO, 1992.
17. Office of Population Censuses and Surveys. *Classification of Surgical Operations and Procedures, 4th Revision.* London, England: Her Majesty's Stationery Office, 1990.
18. Severs MP. The Clinical Terms Project. *Bull R Coll Physicians Lond.* 1993;27(2):9-10.
19. Stannard CF. Clinical Terms Project: a coding system for clinicians. *Br J Hosp Med.* 1994;52:46-8.
20. Buckland R. The Language of Health. *BMJ.* 1993;306:287-8.
21. Stuart-Buttle CDG, Brown PJB, Price C, O'Neil M, Read JD. The Read Thesaurus: Creation and Beyond. In: Pappas C (ed). *Medical Informatics Europe '97.* Amsterdam, The Netherlands: IOS Press, 1997:416-20.
22. Robinson DB, Schulz EB, Brown PJB, Price C. Updating the Read Thesaurus: User-interactive maintenance of a dynamic clinical vocabulary. *J Am Med Inform Assoc.* 1997;4:465-72.
23. Cimino JJ. Coding systems in health care. *Methods Inf Med.* 1996;35:273-84.
24. Schulz EB, Smejko N, Price C, et al. Indexing the directed acyclic graph hierarchy of the Read Thesaurus. *Proc AMIA Annu Fall Symp.* 1996:853.
25. Schulz EB, Barrett JW, Price C. Semantic quality through semantic definition: refining the Read Codes through internal consistency. *Proc AMIA Annu Fall Symp.* 1997:615-19.
26. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. *J Am Med Inform Assoc.* 1996;3:224-33.
27. Campbell JR, Carpenter P, Sniderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. *J Am Med Inform Assoc.* 1997;4:238-53.

APPENDIX

Notation for Auto-classification Rule

Where *O* represents *object*, *a* represents *attribute*, and *V* represents *value*, *O^b* must be a subclass of *O^p* if:

1. For every $O^p - a^x - V^{xp}$ there exists an $O^b - a^x - V^{xb}$ such that V^{xb} is equal or subordinate to V^{xp} ;
2. O^b and O^p are subordinates of O^r ; and
3. The set $(a^1 - V^{1p}$ to $a^n - V^{np})$ captures all differences between O^r and O^p .

