

Determining the Quality of a Terminology

James J. Cimino, M.D.

Department of Biomedical Informatics

Columbia University College of
Physicians and Surgeons

Requirements for High-Quality Terminology

- Synonymy (not redundancy)
- Multiple levels of granularity
- Data model has terms too
- Multiple hierarchies
- Include definitional knowledge
- Support automated translation
- Avoid “Not Elsewhere Classified” (NEC)

But how do we measure these?

The Desiderata

Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine*. 1998 Nov;37(4-5):394-403.

- I. Content
- II. Concept Orientation
- III. Concept Permanence
- IV. Nonsemantic Concept Identifiers
- V. Polyhierarchy
- VI. Formal Definitions
- VII. Reject "Not Elsewhere Classified"
- VIII. Multiple Granularities
- IX. Multiple Consistent Views
- X. Representing Context
- XI. Graceful Evolution
- XII. Recognize Redundancy

Formal Terminology Evaluations

- Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. *JAMIA*. 1996;3:224-233.
- Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions and clarity. *JAMIA*. 1997;5:238-251.
- Sujansky W. NCVHS Patient Medical Record Information (PMRI) Terminology Analysis Reports. National Committee for Vital and Health Statistics, December, 2002 (<http://www.ncvhs.hhs.gov/031105rpt.pdf>).
- Arts DG, Cornet R, de Jonge E, de Keizer NF. Methods for evaluation of medical terminological systems -- a literature review and a case study. *Methods Inf Med*. 2005;44(5):616-25. Review.
- Cornet R, Abu-Hanna A. Two DL-based methods for auditing medical terminological systems. *AMIA Annu Symp Proc*. 2005;:166-70.
- Cornet R, de Keizer NF, Abu-Hanna A. A framework for characterizing terminological systems. *Methods Inf Med*. 2006;45(3):253-66.

Content Coverage of Clinical Classifications (Chute, et al., 1996)

- Do terminologies contain codes for concepts?
- How would one evaluate this question?
- Parsed arbitrary text into arbitrary concepts
- Diagnoses, Findings, Modifiers, Procedures, Other
- 0, 1, 2 scale

Phase II Evaluation (Campbell, et al., 1997)

- Completeness - coding done by experienced coders, reviewed by vocabulary creator
- Taxonomy - presence of appropriate super and subclasses
- Mapping - connection between clinical and financial
- Definitions
- Clarity - ambiguity

Missing from these Evaluations

- Measures of reproducibility
 - Due to redundant terms
 - Due to redundant coding
- Structural desiderata
- Documentation
- Maintenance

NCVHS PMRI Evaluation (Sujansky, 2002)

- Attempt to determine candidate “core” terminologies
- Administrative and legacy terminologies considered
- Domains: diagnoses, symptoms, observations, tests, results, specimens, methods, organisms, anatomy, medications, chemicals, devices, supplies, social and care-management, standard assessments
- Criteria:
 - Coverage
 - Desiderata
 - Organizational criteria
 - Process (Responsiveness) criteria
- Questionnaire sent to terminology developers
- Two step evaluation: essential and detailed study

NCVHS PMRI Desiderata

- Concept orientation *
- Concept permanence *
- Non-Ambiguity *
- Explicit version IDs *
- Meaningless identifiers
- Multi-Hierarchies
- Non-Redundancy
- Formal Concept Definitions
- Infrastructure/tools for collaborative development
- Change sets
- Mappings to other terminologies

* = Essential criteria

NCVHS PMRI - Essential Criteria

- Medcin
- SNOMED-CT
- NCI Thesaurus
- LOINC
- Multum Lexicon
- NDDF
- NDF-RT
- RxNorm
- SNODENT
- HL7 v.3 Codes

(31 failed, including MedDRA, Medi-Span, NDC, 9 nursing terminologies, DICOM, NCPDP, CPT, DSM, ICD-10-CM, ICD-10-PCS, ICPC)

(See Appendix I of these slides for detailed study)

Review and Selection of 3 Methods (Arts, et. al, 2005)

- 24 studies
- Coverage and correctness of concepts, terms and relations
- Chose three methods:
 - Concept Matching
 - Formal Algorithmic Evaluation
 - Expert Review
- All three methods were complementary

Description-Logic Evaluation (Cornet and Abu-Hannah, 2005)

- Frame-based terminology converted to DL
- Examination of DL definitions for
 - Duplicate descriptions (implying synonymy)
 - Underspecified descriptions
- Strict interpretation of definitional information (e.g., mutually disjoint siblings , “anding” slot values, etc.)
- Modeled anatomic terms as structure-entity-part triplets
- Applied methods to Foundational Model of Anatomy
- 494 concepts with non-unique definitions
- 307 inconsistent definitions

Framework for Classifying Terminologies (Cornet, et al., 2006)

- Distinguished terminology, thesaurus, classification, vocabulary, nomenclature, and coding system
- Distinguishes formalism, content, and functionality
- For content, distinguishes concept coverage, concept token coverage, and postcoordination coverage

Cancer Biomedical Informatics Grid (caBIG)

- US National Cancer Institute initiative to speed discoveries and improve outcomes
- Links researchers, physicians, and patients
- Network of infrastructure, tools, and ideas
- Collection, analysis, and sharing of data and knowledge from laboratory to bedside
- Vocabulary and Common Data Elements Workspace (VCDE-WS)
 - Sets standards for common data elements
 - Developers encouraged to use standards

VCDE-WS Evaluation Efforts

- Understandability, Reproducibility, and Usability (URU)
- Documentation
- Maintenance and Extensions (Change management)
- Accessibility and Distribution
- Intellectual Property Considerations
- Quality Control and Quality Assurance
- Concept Definitions
- Community Acceptance
- Reporting Requirements

Vocabulary Checklist (draft)

Revised Terminology Review Criteria -- version 3.2			Revised Recommendations			
			fulfillment should be required	fulfillment is recommended	may not apply to all terminologies	fulfillment may be difficult to assess
B. Content – criteria related to the information contained in the terminology	B.1. Content coverage	B.1.a. Does the terminology provide comprehensive or explicit in-depth coverage of the domain of interest it claims to address as stated in purpose and scope of the terminology segment?		<input type="checkbox"/>		<input type="checkbox"/>
		B.1.b. Are there formal methods in place for expanding and refining the terminology?	<input type="checkbox"/>			
		B.1.c. Are there explicit, reproducible methods for recognizing and filling gaps in content?		<input type="checkbox"/>		
	B.2. Polyhierarchy	If it is allowed and appropriate,, is it used? – That is, is every term in all the classes to which it should belong?		<input type="checkbox"/>		
	B.3. Rejection of NEC terms	Are "not elsewhere classified" (NEC) and "other" terms avoided? Does the terminology provide a way to represent information not explicitly covered in the terminology?	<input type="checkbox"/>			
B.4. Context representation	Does the terminology provide formal, explicit information about how concepts are used?		<input type="checkbox"/>	<input type="checkbox"/>		

Vocabulary Checklist (draft)

- A. Structure – criteria related to the data model of the terminology
- B. Content – criteria related to the information contained in the terminology
- C. Documentation — criteria related to information available about the terminology
- D. Editorial Process - criteria related to the activities involved in designing, creating, distributing and maintaining the terminology

(See Appendix II of these slides)

Next Steps

- Develop Standard Operating Procedure (SOP) for group review of terminologies
- Review terminology with small group of volunteers to test the SOP and “train the trainers”

Conclusions

- Determining the content coverage of a terminology is a complex task
 - Inclusivity
 - Consistent coding
- Terminology evaluation is more than just about coverage:
 - Structure
 - Documentation
 - Maintenance

Thanks to Team Members

- Brian Davis, PhD (3rd Millennium)
- Martin Ringwald, PhD (Jackson Labs)
- Terry Hayamizu, MD, Ph D (Jackson Labs)
- Grace Stafford, PhD (Jackson Labs)

Appendices

I. Details of NCVHS Evaluation

II. Details of caBIG Criteria (draft)

**Appendix I:
Details of NCVHS Evaluation**

NCVHS PMRI - Coverage

<u>Terminology</u>	<u>Concepts</u>	<u>Terms</u>
Medcin	216,00	N/A
SNOMED-CT	345,000	914,000
NCI Thesaurus	27,000	84,000
LOINC	33,000	?
Multum Lexicon	121,000	N/A
NDDF	500,000	N/A
NDF-RT	100,000	N/A
RxNorm	41,000	138,000
SNODENT	3,900	6,500
HL7 v.3 Codes	6,500	6,000

NCVHS PMRI - Desiderata

<u>Terminology</u>	<u>MI</u>	<u>MH</u>	<u>NR</u>	<u>FD</u>	<u>I/T</u>	<u>CS</u>	<u>MT</u>	<u>Total</u>
Medcin	2	0	2	0	2	2	2	10
SNOMED	2	2	2	2	2	2	2	14
NCI	2	2	2	2	2	2	2	14
LOINC	2	2	2	1	1	2	1	11
Multum	2	2	2	1	2	2	0	11
NDDF	2	2	2	1	1	2	1	11
NDF-RT	2	2	2	1	2	2	1	13
RxNorm	2	0	2	1	2	2	2	11
SNODENT	2	2	2	2	0	2	2	12
HL7 v.3	2	2	2	0	0	0	0	16

0 - bad or none, 1 - Some, 2 - Good

MI=Meaningless identifier, MH=Multihierarchy, NR=Nonredundancy, CD=Formal definitions, I/T=Infrastructure/Tools, CS=Change sets, MT=Mappings to terminologies

NCVHS PMRI - Organizational

<u>Terminology</u>	<u>LC</u>	<u>IP</u>	<u>3P</u>	<u>Total</u>
Medcin	0	1	2	3
SNOMED	2	1	2	5
NCI	2	2	2	6
LOINC	2	2	2	6
Multum	1	1	1	3
NDDF	0	1	2	3
NDF-RT	2	0	2	4
RxNorm	2	2	2	6
SNODENT	?	0	2	2-4?
HL7 v.3	2	2	2	6

0 - Has/requires, 1 - Some, 2 - Not Requires

LC=High licensing costs, IP=Intellectual property restrictions,
3P=Third party platform/tools

NCVHS PMRI - Responsiveness

<u>Terminology</u>	<u>UF</u>	<u>VS</u>	<u>AT</u>	<u>Total</u>
Medcin	2	2	1	5
SNOMED	2	2	2	6
NCI	2	2	1	5
LOINC	2	2	2	6
Multum	2	2	0	4
NDDF	2	2	2	6
NDF-RT	2	2	1	5
RxNorm	2	2	0	4
SNODENT	0	2	0	2
HL7 v.3	2	2	1	5

0 - <Yearly/one source/No training, 1 - two sources/modest training,
2 - Yearly/three sources/Extensive training

UF=Update frequency, VS=Varied sources, AT=Availability of training

Appendix II:
Details of caBIG Criteria (draft)

Structure Criteria (1)

- A.1. Concept orientation –Is terminologic information organized around meaning of terms?
- A.2. Concept permanence - Is the meaning of a concept, once created, inviolate and does the data model accommodate name changes and retirement?
- A.3. Nonsemantic concept identifiers - Does each concept have a unique identifier that is free of hierarchical or other implicit meaning and are not re-used?
- A.4. Polyhierarchical organization - Is it allowed? Is it appropriate?
- A.5. Graceful evolution - How are updates applied to the content?
- A.6 Explicitness of relations – Are the meanings of inter-term relations explicit?

White = required

Blue = recommended

Structure Criteria (2)

- A.7. Multiple granularities - If the terminology is intended to serve multiple purposes, does it provide different levels of granularity appropriate for the different purposes?
- A.8. Multiple consistent views - If the terminology is intended to serve multiple purposes, does it provide multiple views suitable for the different purposes?
- A.9. Formal definitions - Does term representation provide a definitive set of relationships to other concepts that, taken together, are both individually necessary and collectively sufficient to distinguish the concept from all other concepts?
- A.10. Recognition of redundancy - Is the structure sufficiently rich to support detection of redundant meaning?
- A.11 Extensibility - Does the structure avoid imposing limits on the ability of the terminology to cover the domain? (e.g the decimal hierarchical codes of ICD9-CM)

Content Criteria (1)

- B.1. Content coverage - Does the terminology provide comprehensive or explicit in-depth coverage of the domain of interest it claims to address as stated in purpose and scope of the terminology segment?
- B.2. Polyhierarchy - If it is allowed and appropriate, is it used? – That is, is every term in all the classes to which it should belong?
- B.3. Rejection of NEC terms - Are "not elsewhere classified" (NEC) and "other" terms avoided? Does the terminology provide a way to represent information not explicitly covered in the terminology?
- B.4. Context representation - Does the terminology provide formal, explicit information about how concepts are used?

Content Criteria (2)

B.5. Textual Definitions - Does the terminology provide a clear textual definition of each term in the terminology and are the textual definitions sufficient to distinguish the meaning of each concept from other concepts in the terminology?

B.6 Formal Definitions - Does each term in fact have a definitive set of relationships to other concepts that, taken together, are both individually necessary and collectively sufficient to distinguish the concept from all other concepts?

Documentation Criteria

- C.1. Purpose and scope - Is the purpose and scope of the terminology clearly stated in operational terms so that its fitness for particular purpose can be assessed and evaluated?
- C.2. Statement of intended use - Is there a statement of the terminology's intended use, intended users and scope?
- C.3. Documentation descriptions - Does the available documentation describe terminology structure and organizing principles, use of concept codes/identifiers, use of semantic relationships, output format(s)
- C4. Version documentation - Are new versions accompanied by adequate documentation that describes how the new version differs from the one it replaces?
- C.5 Tool documentation - Is there a description of methods or tools for acquisition and application of the terminology?

Editorial Criteria (1)

- D.1. Process for maintenance and extensions - Does the editorial process enable changes for 'good' reasons and discourage change for 'bad' reasons, and does it maintain concept permanence while correcting recognized redundancy?
- D.2. Quality Assurance and Quality Control - Are there internal checks to detect and eliminate errors in modeling and/or editing, is there a process for review by independent experts from the field in which the terminology will be used, and is there a process in which the terminology developer can improve the terminology in response to the findings and recommendations of the review?
- D.3. Methods for extending the terminology - Is the terminology evolving to maintain domain coverage?

Editorial Criteria (2)

- D.4. Organization criteria - Is maintenance of the terminology a core part of the organization's business?
- D.5. Extensions to other terminologies - If the terminology extends or overlays other terminologies, do they have a formal methodology for expanding content?
- D.6. Availability of lists of concepts, terms and definitions - Is the terminology included in an EVS-type terminology server? If this is not possible, then flat files (such as used by the UMLS) should be available.
- D.7. Methods and tools for acquisition and application - Is the terminology freely available for download in a format(s) (e.g. RRF, OWL, XML, OBO) that can be readily used by the community? Has an effective user interface been built? Is there support for computer interface and system implementers?

Editorial Criteria (3)

- D.8. Intellectual Property Considerations - Is the terminology available to all classifications of users (e.g. government agencies, for-profit and not-for-profit institutions, academia, private citizens, etc.), without fee, permission requirement, or restrictions?
- D.9. Community Acceptance - Has a scientific community accepted the terminology as a de facto standard?
- D.10. Reporting Requirements - Has a health regulatory body required this terminology for reporting? If so, which one(s)?
- D.11. Editorial Process - Is there evidence of a thoughtful editorial process, carried out by experts in the domain of interest and terminology representation, ideally with input from potential users of the terminology?
- D.12. Mechanisms for accepting and incorporating external contributions - these include error reporting and requests for additional content