



Contents lists available at ScienceDirect

## Journal of Biomedical Informatics

journal homepage: [www.elsevier.com/locate/yjbin](http://www.elsevier.com/locate/yjbin)

## The caBIG terminology review process

James J. Cimino<sup>a,\*,1</sup>, Terry F. Hayamizu<sup>b,1</sup>, Olivier Bodenreider<sup>c</sup>, Brian Davis<sup>d</sup>, Grace A. Stafford<sup>b</sup>, Martin Ringwald<sup>b</sup><sup>a</sup> National Institutes of Health, Laboratory for Informatics Development, Clinical Center, Room 6-2551, 10 Center Drive, Bethesda, MD 20892, USA<sup>b</sup> The Jackson Laboratory, Bar Harbor, ME, USA<sup>c</sup> Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, USA<sup>d</sup> 3rd Millenium Inc., Waltham, MA, USA

## ARTICLE INFO

## Article history:

Received 13 June 2008

Available online 25 December 2008

## Keywords:

Terminology

Ontology

Auditing

Evaluation

## ABSTRACT

The National Cancer Institute (NCI) is developing an integrated biomedical informatics infrastructure, the cancer Biomedical Informatics Grid (caBIG<sup>®</sup>), to support collaboration within the cancer research community. A key part of the caBIG architecture is the establishment of terminology standards for representing data. In order to evaluate the suitability of existing controlled terminologies, the caBIG Vocabulary and Data Elements Workspace (VCDE WS) working group has developed a set of criteria that serve to assess a terminology's structure, content, documentation, and editorial process. This paper describes the evolution of these criteria and the results of their use in evaluating four standard terminologies: the Gene Ontology (GO), the NCI Thesaurus (NCIt), the Common Terminology for Adverse Events (known as CTCAE), and the laboratory portion of the Logical Objects, Identifiers, Names and Codes (LOINC). The resulting caBIG criteria are presented as a matrix that may be applicable to any terminology standardization effort.

Published by Elsevier Inc.

## 1. Introduction

The National Cancer Institute (NCI) is developing an integrated biomedical informatics infrastructure, the cancer Biomedical Informatics Grid (caBIG<sup>®</sup>), to expedite the cancer research community's access to key biomedical informatics platforms. caBIG's common, extensible framework will integrate diverse data types and support interoperable analytic tools, allowing research groups to make use of the rich collection of emerging cancer research data while supporting their individual investigations [1,2].

caBIG takes an object-oriented approach to ensure interoperability among software development projects. Developers and users are encouraged to share data that have been constructed in an object-oriented manner. The object data are described with publicly available, commonly agreed upon concepts from controlled terminologies and the description of the data (the metadata) are stored in a publicly available, ISO11179-compliant metadata repository called the cancer Data Standards Repository (caDSR) [1]. By this method, data and systems are described by commonly understood concepts and definitions, and the descriptions are publicly available for users and other developers to re-use.

Agreed upon, controlled terminologies and common data elements are essential tools for implementing terminological and semantic consistency in caBIG. The Vocabularies<sup>2</sup> and Common Data Elements Workspace (VCDE WS) was established to evaluate and disseminate the use of data standards within caBIG, including developing standards for the representation of ontologies and terminologies used throughout the caBIG system.

The construction and maintenance of vocabularies, ontologies and terminologies is a non-trivial task [3]. However, there are many "good" terminologies available, and many "best practices" as to what a "good terminology" is and what it is not have been described [4]. The VCDE WS has taken on the task of evaluating terminologies in order to suggest to caBIG participants which terminologies to use in describing their software systems and data to ensure clear and unambiguous understanding. To this end, the VCDE WS has proposed a number of criteria that every caBIG standard terminology should satisfy. The VCDE WS has performed a series of "terminology reviews" to exercise and evaluate the criteria. This paper describes the evolution of the caBIG terminology review process, a description of the criteria in their present form, the

\* Corresponding author.

E-mail address: [ciminoj@mail.nih.gov](mailto:ciminoj@mail.nih.gov) (J.J. Cimino).<sup>1</sup> These authors contributed equally to lead the authorship of this paper and the work reported herein.<sup>2</sup> While some authors use the terms "terminology" and "vocabulary" interchangeably, we adhere to commonly accepted standardized definitions for these terms and will generally use the former term throughout this paper, except when referring to organizations and standards that, for historical reasons, continue to use the latter term.

process by which the criteria are applied, and a summary of the findings from the four terminology reviews carried out to date.

## 2. Source of criteria

### 2.1. Initial VCDE WS efforts

In January of 2005, the VCDE WS began to survey the software development that was occurring in caBIG to understand the needs and uses of controlled terminology in caBIG. In May 2005, the VCDE WS began a series of face-to-face meetings and teleconferences to answer the question: What are the criteria for a “validated” or “standard” terminology? In June, Frank Hartel and Jim Oberthaler from NCI suggested that the VCDE WS examine past efforts at evaluation metrics for terminologies, including those described in “Desiderata for Controlled Medical Vocabularies in the Twenty-first Century” [5], the College of American Pathologists “Understandability, Reproducibility, Usability” (URU) criteria [6], and the International Standards Organization’s “Health Informatics—controlled Health Terminology—structure and High Level Indicators” [7].

These documents and principles, and the needs of software applications, were considered by the VCDE WS participants, who then formulated 10 categories for evaluating biomedical terminologies. Over the course of the next year, terminology review criteria were conceived for each of the ten categories. In June 2006, the criteria were judged to be sufficiently well-developed to provide to terminology experts. The next tasks were to establish a process for a caBIG terminology review, and then to test the criteria on an actual terminology. A team (TFH and MR) at the Jackson Laboratory, being participants in the caBIG VCDE WS and knowledgeable in the field of biomedical terminologies, took on these tasks.

### 2.2. Development of a defined terminology review criteria matrix

The resources identified by the VCDE WS (see above) presented a number of useful criteria to be weighed in considering whether to propose or accept a specific terminology as a caBIG standard. In the next phase of this effort, each of the various terminology review criteria was evaluated in depth. The documents provided, as well as the references cited therein, were extensively reviewed. The “Desiderata” publication [5] was used as an initial guide with regards to appropriate interpretation of the URU criteria. Overall, this detailed assessment was critical in order to gain an appropriate understanding of the criteria, particularly as they would pertain to terminologies being considered as caBIG standards. Notably, some overlap in content between these resources was discovered, as well as ambiguities, in some cases related to details specific to the field being discussed in the publication. During the evaluation process, questions regarding the proposed criteria were directed to the VCDE WS and/or the NCI, and uncertainties regarding interpretations were discussed, including feedback on which criteria should be required versus desired. Additional resources with information pertinent to biomedical terminologies and terminological systems were also reviewed. Issues such as the appropriateness of the criteria with regards to the scope and focus of the evaluation were discussed between project participants. As a result of this work, a well-defined set of approximately 100 “baseline” terminology review criteria were determined to be of particular relevance to the caBIG community for evaluating biomedical terminologies.

In order to facilitate evaluation of the criteria themselves, as well as use of the criteria for terminology assessment, some organizational modifications to the criteria list were made. Within the proposed criteria, 10 high level categories had been identified:

- Understandability, reproducibility, usability (URU)
- Quality of documentation
- Maintenance and extensions (change management)
- Accessibility and distribution
- Intellectual property considerations
- Considerations regarding mapped technologies
- Quality assurance and quality control (QA/QC)
- Concept definitions
- Community acceptance
- Reporting requirements

These categories provided a framework within which the criteria could be organized. Some criteria could justifiably be placed in more than one category. In these cases, some redundancy (so that a given criterion could fall into more than category) was allowed. In addition, to simplify use of the criteria, each criterion was phrased as a question that, in most cases, could be answered by a standard binary (i.e. “meets” vs. “does not meet” the criteria) response for the specific terminology being assessed. Finally, in order to further expedite use of the proposed terminology review criteria, they were organized as a “checklist” in a Microsoft Excel<sup>®</sup> worksheet.

A draft version of the proposed terminology review criteria “checklist” was made available for critique, both by NCI terminology experts and members of the Gene Ontology Consortium (see Section 3.1, below). Comments on the specific terminology evaluation criteria were solicited, including answers to questions such as:

- Do you think this is a valid criterion?
- Do you think this criterion should be required or merely desired?
- Do you think it is a reasonable criterion given its significance and the effort it would take to reach compliance?
- Are there criteria for which compliance could be accomplished without a major effort?
- Are there criteria for which you plan to achieve compliance?

In addition, more general questions with regard to evaluation of biomedical terminologies were put forth, such as:

- Do you think the proposed set of criteria is appropriate for terminology review in general?
- Are there criteria that you would consider unrealistic?
- Are there other criteria that should be taken into account?

Responses to these questions were valuable in the final assessment of the terminology review criteria and, as a result of this feedback, individual criteria and the criteria list organization were revised.

At the conclusion of this phase of the project, a “version 1.0” set of review criteria, formatted as a “checklist” within a spreadsheet and henceforth referred to as the Terminology Review Criteria Matrix (TRC Matrix), was submitted to caBIG. In order to further evaluate these criteria, as well as the utility of the matrix (selected portions of which are shown in Fig. 1), the criteria were subsequently used in evaluation of the Gene Ontology and other terminologies.

## 3. Application and evolution of criteria

### 3.1. The Gene Ontology (GO)

The next task was to evaluate a large biomedical terminology using the proposed criteria. The Gene Ontology (GO) [8,9] is used by all of the major model organism databases, as well as other large bioinformatics resources, as a structured, controlled terminology with which to classify gene products with regard to their

PROPOSED REVIEW CRITERIA (version 1.0)			USING THE CRITERIA TO ASSESS THE GENE ONTOLOGY (GO)					RECOMMENDATIONS FOR USING THE CRITERIA				
			meets criterion	partially meets criterion	does not meet criterion	criterion not applicable	criterion not assessed	Suggested rewording for criteria:	fulfillment should be required	fulfillment is recommended	may not apply to all vocabularies	fulfillment may be difficult to assess
Understandability, Reproducibility and Usability (URU)	purpose and scope	Is the purpose and scope of the vocabulary clearly stated in operational terms so that its fitness for particular purpose can be assessed and evaluated?	✓					✓				
	vocabulary content coverage	Does the vocabulary provide comprehensive or explicit in-depth coverage of the domain of interest it claims to address as stated in purpose and scope of the vocabulary segment?							✓		✓	
		Are there formal methods in place for expanding and refining the vocabulary?	✓						✓			
		Are there explicit, reproducible methods for recognizing and filling gaps in content?	✓						✓			
	formal definitions	Are there formal rules for inferring subsumption from the definitions?		✓					✓			
Accessibility and Distribution	Is the vocabulary freely available for download in a format(s) that can be readily used by the community:	RRF			✓							
		OWL	✓					✓				
XML		✓						✓				
		If the vocabulary is not available in an appropriate format, is there a clear and reliable path by which the format can be transformed into one of the above?				✓		✓	✓			
Community Acceptance	Has a scientific community accepted the vocabulary as a de facto standard?	✓						✓				
Reporting Requirements	Has a health regulatory body required this vocabulary for reporting?			✓				✓				
OVERALL ASSESSMENT OF GO USING THE REVIEW CRITERIA			80	5	4	12	1					
			were met	were partially met	were not met	were not applicable	was not assessed	SUMMARY OF CRITERIA RECOMMENDATIONS	40	59	21	1
								should be required	are recommended	may not be applicable	may be difficult to assess	

Fig. 1. Selected portions of a terminology review criteria matrix showing (A) examples of the proposed criteria and their use in assessing the Gene Ontology (GO), as well as (B) some of the recommendations for revising and for using the criteria that were derived from this effort.

molecular function, the biological processes in which they are involved, and the cellular components with which they are associated. Consequently, GO was selected as one of the terminologies for which the proposed review criteria would be evaluated and two of us (TFH and MR) were assigned the task of carrying out the evaluation.

The initial assessment of GO relied on the terminology itself and on information that was readily available on-line or in the form of publications. Additional information believed to be required for the evaluation was elicited directly from members of the GO Consortium. Documentation for this evaluation included the manner in which each criterion was used to evaluate GO, the steps and efforts needed to obtain pertinent information, as well as difficulties encountered during this process. Data pertaining to each of the criteria, including detailed source information, were methodically collected and stored in separate columns in the TRC Matrix. Data from this file were used for the subsequent assessment of GO.

Based on this work, a draft of the terminology evaluation document was developed. The draft was then made available to selected members of the GO Consortium for review and comments. Specifically, the reviewers were asked for the following types of feedback:

- Is our evaluation correct and complete?
- Are there misinterpretations on our part?
- Are there inaccuracies because we relied upon documentation that is not up-to-date?

- Is there additional information that should be included in the evaluation?

Answers to these questions were helpful in augmenting and finalizing the terminology evaluation results.

At the completion of this phase of the project, a report containing detailed results of the analysis of GO for each criterion was presented to members of the VCDE WS. This report included an overall assessment as to how GO met or did not meet each criterion, along with excerpts or statements from the sources of information that were used in developing the assessment. In summary, we determined that GO fulfilled 80 of the 102 proposed terminology review criteria, five criteria were determined to be partially met, and four criteria were not met by GO. In addition, twelve criteria were considered to be “not applicable” with regards to GO. Only one of the criteria was not assessed. See Figs. 2–4 and Appendix I for details of the results of the GO evaluation.

As shown in Fig. 1A, the criterion that was designated “not assessed” focused on whether the terminology did, in fact, “provide comprehensive or explicit in-depth coverage of the domain of interest it claims to address.” We concluded that an adequate assessment of GO with regards to this criterion was beyond the scope of the work possible given the resources at hand and the time frame available for this task. Furthermore, we determined that focusing on methodologies for recognizing gaps in content, and for expanding and refining the terminology (both covered in subsequent criteria: “Are there formal methods in place for

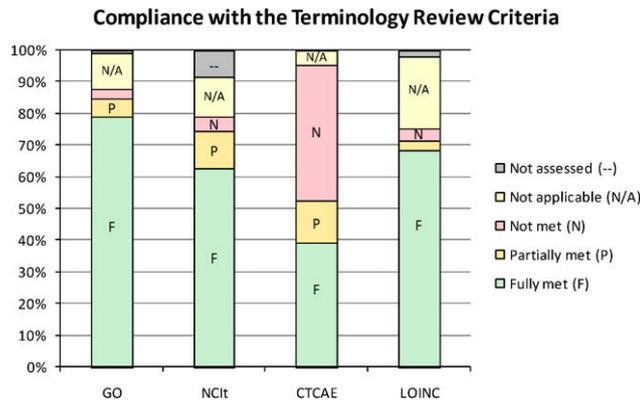


Fig. 2. Compliance with the terminology review criteria by terminology.

expanding and refining the terminology?” and “Are there explicit, reproducible methods for recognizing and filling gaps in content?”), was equivalently relevant.

The evaluation of GO was relatively straightforward for most of the proposed criteria. For some, however, we determined that the criterion was either not clearly defined as stated, was very difficult to adequately assess, and/or would clearly not be applicable to all terminologies. Other issues encountered were specific to the version of the terminology review criteria initially proposed. These included redundancy between separate criteria, as well as semantic problems that could be overcome by re-wording the criteria. As a result of these efforts, suggestions for revising the terminology review criteria were presented. For example, as shown in Fig. 1A (un-

der the category “Accessibility and Distribution”), there were originally three separate criteria pertaining to the formats in which a terminology should be available. The reviewers determined that GO did not meet all of these; however, as availability in any of these formats was judged to be sufficient, these criteria were combined and the resulting single criterion was reworded (see Fig. 1B).

As pointed out above, we recognized that no terminology would be likely to fulfill every one of the criteria. Thus, with regards to the caBIG terminology review process overall, we concluded that defining specifically which of the criteria should be absolutely required versus recommended was a critical issue. Terminologic research to date largely ignores the trade-off between what is desired and what is practical to achieve. We therefore turned to the consensus of the collective expertise among the caBIG participants in general and the VCDE WS members in particular, since these would ultimately be the users of whatever terminologies were found to be acceptable.

A draft version of specific recommendations regarding the individual terminology review criteria was generated (see Fig. 1B). This list of recommendations also indicated those criteria for which fulfillment might be difficult to assess (e.g. the “content coverage” criterion discussed above), as well as those that would not be applicable to every terminology being assessed. In the example presented in Fig. 1B, the second criterion relating to “Accessibility and Distribution” (“Is there a clear and reliable path by which the format can be transformed into one of the [acceptable formats]?”) would only be pertinent if the preceding one (“Is the terminology freely available for download in a format that can be readily used by the community?”) was not fulfilled. The project report further specified that “fulfillment of either of these criteria is recom-

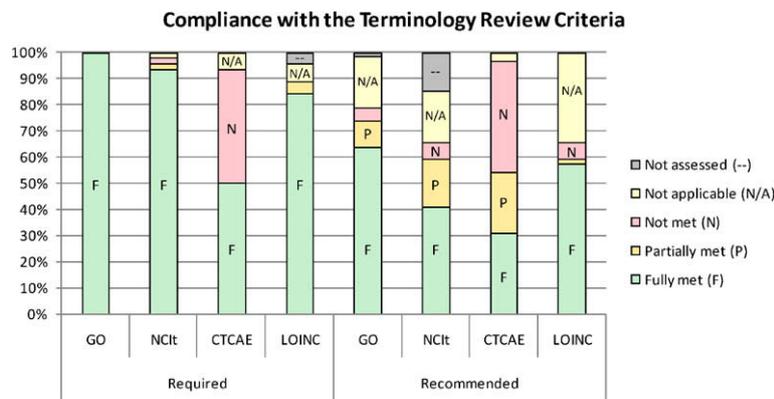


Fig. 3. Compliance with the terminology review criteria by type (required vs. recommended).

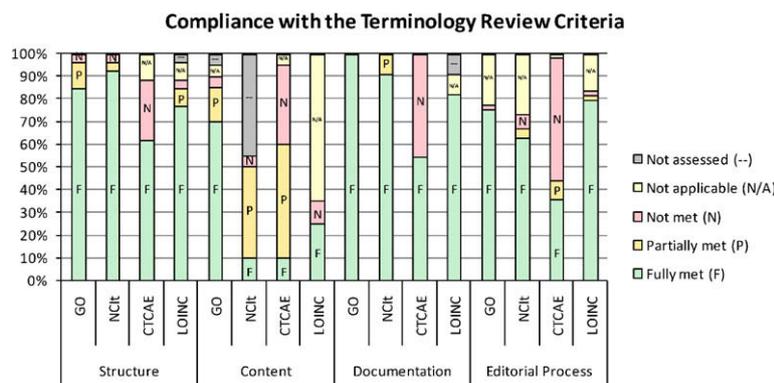


Fig. 4. Compliance with the vocabulary review criteria by section.

mended, but fulfillment of at least one of them should be required.”

Finally, suggestions for how the criteria, in the form of a TRC Matrix, might be used in the terminology review process were also provided. Among these suggestions was that the matrix be used both by the terminology submitter (and/or developers) and by terminology reviewers to manage information collected and to record their assessment of the individual criteria.

### 3.2. NCI Thesaurus

The NCI Thesaurus (NCIt) is a public domain reference terminology developed by the NCI Center for Bioinformatics (NCICB) as part of the Enterprise Vocabulary Services (EVS) Project [10]. NCIt provides broad coverage of the cancer domain and related topics, including findings, drugs, therapies, anatomy, genes, pathways, cellular and subcellular processes, proteins, and experimental organisms. NCIt was designed to be used in systems supporting basic, translational, and clinical research and is part of the Cancer Common Ontologic Representation Environment (caCORE) of the caBIG project. Like other modern terminologies, NCIt uses description logics for its development [11]. The formalism used for distributing NCIt is the Web Ontology Language (OWL) [12].

The evaluation of NCIt presented here was based on version 2.0 of the TRC Matrix, with the primary objective of assessing the degree to which NCIt complied with these criteria. A secondary objective was to evaluate the criteria themselves, and especially their applicability to terminologies other than the Gene Ontology on which they had been originally tested. Additionally, we attempted to design operational mechanisms for automatically assessing some of the criteria. For example, the presence of multiple hierarchies could be assessed by looking for concepts with more than one parent (or is-a) relationship.

Version 06.09d of NCIt, which contained over 54,000 concepts and 150,000 concept names, organized into 20 subsumption hierarchies, was investigated by one of us (OB). The evaluation primarily relied on the examination of the original OWL file, manually (once loaded into Protégé) or programmatically (through programs developed to operationalize the assessment of some of the criteria). Part of the evaluation was also based on the documentation (caCORE documentation [13], technical documents [14–16] and articles published about NCIt in the scientific literature by authors affiliated with the NCI [10,11,17–19] and other authors [20–21]). We did not rely on personal communication with the developers.

Overall, we found NCIt to be fully compliant with 64 of the 102 original review criteria (e.g., polyhierarchy, concept permanence, absence of restrictions to free dissemination) and partially compliant with eleven criteria (e.g., rejection of “Not Elsewhere Classified” (NEC) terms, textual definitions). NCIt did not meet five criteria: context representation, multiple views, description of the validation process in the documentation, review by independent experts, and use for mandatory reporting. Finally, thirteen criteria were not applicable to NCIt and nine criteria were not evaluated, including many of the criteria related to textual definitions, because they were difficult to evaluate systematically. See Figs. 2–4 and Appendix I for details of the results of the NCIt evaluation.

Compliance with some criteria was assessed programmatically. For example, the existence of textual definitions (similar to those found in a dictionary) was assessed by exploring the following properties of the OWL class that contain such text: DEFINITION, LONG\_DEFINITION, ALT\_DEFINITION and ALT\_LONG\_DEFINITION. Sixty-two percent of all classes had at least one such definition. Analogously, the existence of necessary and sufficient conditions for OWL classes (`owl:equivalentClass`) was used to assess the presence of formal definitions. Eighteen percent of the classes were de-

finied. Using string matching on the concept names, we verified that only eight terms contained “NEC” or “not elsewhere classified”. Overall, we were able to create operational assessments for twelve criteria.

Most of the criteria in the original set were applicable to NCIt. Some redundant criteria were later removed during subsequent revisions of the review criteria. Some criteria were difficult to assess thoroughly, including qualitative aspects of textual definitions (e.g., absence of circular definitions). The creation of operational assessments was facilitated by the availability of the easily parseable OWL file. Most criteria definitions created for NCIt could be applied to other terminologies represented in OWL (e.g., GO). The existence of operational assessments is an important factor in the scalability and reproducibility of the review process, because such definitions support the consistent evaluation of large terminologies such as NCIt.

### 3.3. The Common Terminology for Adverse Events (CTCAE)

The Common Terminology for Adverse Events (CTCAE) Version 3.0 is a coding system for reporting adverse events that occur in the course of cancer therapy. It was derived from the Common Toxicity Criteria (CTC) v2.0 and is maintained by the Cancer Therapy Evaluation Program (CTEP) at the NCI [22]. The VCDE WS wished to evaluate CTCAE in order to determine its suitability for use in recording adverse events in cancer therapy protocols; this evaluation was carried out by one of us (J.J.C.).

The application of the TRC criteria to CTCAE was initially problematic, because CTCAE was less of a terminology and more of a coding system for recording data through postcoordination of Adverse Event (AE) terms with Grade terms. In order to fully evaluate CTCAE, it was necessary to consider not only the enumerable terms, but the ways they could be combined. Although most of the 1058 AE terms can be combined with any of the five Grade terms (ranging from 1-Mild to 5-Death), this is not always the case. For example, the AE term “Glaucoma” can never be paired with “Death”, since the former could never be severe enough to directly cause the latter. Furthermore, the actual meanings of the Grade terms vary, depending on the AE to which they are applied. For example, Grade terms for the AE term “Lymphopenia” are quantitative (with values like “>800/mm<sup>3</sup>”, “500–800/mm<sup>3</sup>”, etc.), while those applied to the AE term “Nausea” are qualitative (“Loss of Appetite without Alteration in Eating Habits”, “Oral Intake Decreased without Significant Weight Loss, Dehydration or Malnutrition”, etc.). The TRC criteria do not cover this sort of postcoordination process.

Fortunately, the NCI addressed this issue when including CTCAE in the NCI Metathesaurus (NCI Meta). In essence, NCI Meta includes all of the legal permutations of AE-Grade combinations as precoordinated terms, each with its own unique identifier. The NCI also organized these reified terms into a hierarchy that reflected the CTCAE’s organization into Categories (e.g., “Cardiac Arrhythmias”) and superordinate terms (e.g., “Ventricular Arrhythmia”), with the unmodified AE terms (e.g., “Bigeminy”) placed at the level above the precoordinated terms (e.g., “Grade 1 Bigeminy”). It was this set of terms that was evaluated, according to the criteria, as they existed at that time.

For the most part, while CTCAE in its native form did not meet the “Understandability, Reproducibility and Usability” (URU) criteria, the NCI Meta version did—largely because NCI Meta itself conforms to good terminology practices. For example, NCI Meta imposes concept orientation, concept permanence and meaningful identifiers on the terminologies it subsumes. The main deficiency was related to the lack of formal, structured definitions. In other general areas, such as quality of documentation, accessibility, intellectual property considerations, QA/QC, textual definitions, community acceptance, and reporting requirements, CTCAE fared

well. However, it failed all of the maintenance criteria because, at that time, CTEP had no plans or mechanisms for updates. See Figs. 2–4 and Appendix I for details of the results of the CTCAE evaluation.

The experience with the evaluation of CTCAE resulted in a number of process-related recommendations, such as the need to convert coding systems into formal terminologies, improving access to the terminology and its documentation, and methods for reporting results. More importantly, this experience suggested a number of changes to the TRC criteria that influenced their evolution into a form that was subsequently applied to the evaluation of LOINC (see below). The principal recommendation was to reduce overlap among, and clarify distinctions between, the criteria by arranging them into four major categories: Structure (criteria that relate to the *model* of the terminology), Content (criteria that relate to the *terms contained* in the terminology), Documentation (material external to the terminology that helps with comprehension and use, including user manuals and published evaluations), and Editorial Process (all aspects of the mechanisms by which the terminology is created, maintained and distributed). As a result, the criteria were reorganized from nine major and 46 minor groupings of 99 specific items, into four major and 34 minor groupings of 105 specific items.

#### 4. Current state of the criteria

The criteria are presented to both the terminology developers and the review team in a worksheet matrix, encompassing four primary categories, Structure, Content, Documentation, and Editorial Process (see below). As described above, these categories evolved from the original criteria set forth by the VCDE WS. These include the principles of Understandability, Reproducibility and Usability, with further recognition of the need for quality documentation, change management, QA/QC and community acceptance. As of this writing, the current TRC Matrix is version 3.3.

Each category is divided into one or more subsections, which themselves may have subdivisions to provide additional granularity for the criteria. Overall, there are 105 criteria provided, each in the form of a question to facilitate evaluation. Each criterion is individually assessed and designated with one of five endpoints:

- Meets criterion
- Partially meets criterion
- Does not meet criterion
- Criterion not applicable
- Criterion not assessed.

Additionally, the worksheet provides the user with space to note supporting documentation.

The organization of the TRC Matrix into categories and subcategories leads the reviewer through a logical progression to enable a scalable review process of terminologies proposed for use within the caBIG. As mentioned above, terminologies are not expected to meet every criterion, and a second worksheet matrix provides recommendations on which criteria must be fulfilled (44/105) and which are strongly recommended (61/105) to be fulfilled by a chosen terminology. The worksheet further notes where criteria may not apply or may be difficult to assess. See Appendix I for the complete spreadsheet with recommendations.

##### 4.1. Structural Criteria

The Structural Criteria are related to the data model of the terminology. The structure of the terminology is evaluated separately from the actual content. From both Cimino [5] and the ISO/TS

17117 Technical Specification [7], the desired structure of a terminology should be based on the notion of the concept as the basic unit of terminology (concept orientation). Hence this section of the matrix begins by asking the reviewer, “Is terminological information organized around meaning of terms?” Each concept is expected to have a single meaning that is non-vague, non-ambiguous and non-redundant.

The matrix continues, presenting the reviewer questions with which to evaluate the terminology. The questions are posed so that an affirmative response indicates that the terminology satisfies the following requirements:

- The data model should allow for concept permanence, accommodating name changes and retirement of concepts.
- Each concept should have a unique identifier, which does not contain semantic information (i.e. is free of hierarchical or other implicit meaning).
- The terminology is organized hierarchically. The basic principle underlying the hierarchy should be explicitly stated. Frequently, a strict hierarchy of terms would limit the usability of a terminology [5].
- If a polyhierarchical organization is appropriate, concepts should be allowed multiple parental terms. The meaning of the concept, however, should remain the same regardless of the parent from which it is reached.

A terminology used to annotate an evolving field, be it molecular biology or clinical research, must of necessity likewise be able to evolve. These changes should be described in detail and referable to consistent versions of the terminology.

The matrix continues, leading the reviewer through additional required attributes for structure, asking for explicitly defined relationships between terms; appropriate granularity of terms; and suitability and consistency of multiple views (if multiple views are provided).

The final criteria in this section deal with formal definitions, redundancy and extensibility. The recommendations for caBIG usage strongly suggest formal definitions, which are helpful for identifying redundancy. Extensibility is related to evolution, but here the emphasis is that the underlying structure will not limit the terminology (as, for example, the decimal hierarchical codes of ICD-9-CM unfortunately will do).

##### 4.2. Content Criteria

Terminology content is evaluated separately from structure. Although some of the criteria seem redundant with those described above, the point is to ensure the content actually contained in the terminology coincides with what it is purported to contain.

The content should provide comprehensive coverage of the domain it proposes to address. Again, since the domain will likely be a moving target, there should be methods implemented to expand the terminology, recognizing and filling in the gaps.

In the Content section, the reviewer considers whether polyhierarchy is appropriately used, with every term being in all appropriate classes. The reviewer also determines how well the terminology is defined and ensures that it avoids problematic terms like “not elsewhere classified” (NEC) and “other.” Corresponding to the levels of compatibility designated for caBIG tools, the matrix currently designates required definitional elements at the Bronze, Silver and Gold levels [23].

##### 4.3. Documentation Criteria

The documentation provided with a terminology aids the reviewer in determining the acceptability of a terminology and the

end user in appropriately using the terminology. One expects the purpose and scope of the terminology to be clearly stated in operational terms, with the intended use and intended users delineated.

The Documentation section of the matrix goes on to evaluate the fitness of the documentation, asking the reviewer to determine if the descriptions of seven criteria are adequate:

- Terminology structure and organizing principles
- Use of concept codes/identifiers
- Use of semantic relationships
- Output format(s)
- Applications, contexts or domains where the terminology would not be appropriate
- Any relationships/links to other resources
- Methods for extending the terminology

As noted above, versioning is required to keep track of the evolution of the terminology. The documentation should be adequate to describe how a version differs from the one it replaces. It should also describe methods or tools available to utilize the terminology.

#### 4.4. Editorial Process Criteria

The Editorial Process Criteria relate to the activities involved in designing, creating, maintaining, and distributing the terminology. Again, some of the criteria appear redundant with other sections, but the point is to determine whether the processes described are adhered to.

Cimino described the graceful evolution of content and structure [5,24]. The editorial process should allow for changes for good reasons (simple additions, refinement, pre-coordination, disambiguation, obsolescence, discovered redundancy, and minor name changes), while discouraging changes for bad reasons (redundancy, name changes that alter the meaning of the concept, code reuse, code changes). Updates and modifications need to be referable by precise version identifiers, with the resulting implication that earlier versions should be available (permanent storage) [7].

This section seeks to evaluate the processes for quality assurance and quality control (QA/QC) and to determine who contributes to those processes. Given the recognition that the terminology needs to evolve gracefully, there should ideally be an organization with a commitment to maintain the terminology. Recommended QA/QC processes include internal checks and validations (including documentation of these processes), independent expert review, and processes to improve in response to feedback. If the terminology is an extension or overlay of other terminologies, it is advantageous to provide explicit representation and have processes to keep up to date with the other terminologies. There should be evidence of a thoughtful editorial process, carried out by experts in the domain of interest, with mechanisms for accepting and incorporating external contributions, such as error reporting and user requests for additional content.

The terminology must be available to be useful. Ultimately, the caBIG community requires access to the terminology via an enterprise terminology service such as the NCI LexBIG-powered EVS [25], but many formats are readily usable (e.g., RRF, OWL, XML, OBO). Availability also includes issues of intellectual property considerations, as caBIG promotes open-access. Community acceptance is another consideration for caBIG, as well as the requirements of any health regulatory body to utilize a terminology for reporting. However, the fact that a terminology is used does not necessarily mean that its use is desired; caBIG may choose to promote use of a new terminology if an old one is found to be defective with respect to the TRC criteria.

#### 4.5. Current status

The format of the TRC Matrix enables individual reviewers to provide evaluations that can be collated and used as a starting platform for reaching consensus. Having high granularity of both the individual criteria and the level to which each criterion is graded provides an underlying objectivity for the evaluation of the terminology, while providing the review team a detailed enough analysis to ensure that “perfect does not become the enemy of the good.”

Enhanced documentation of the matrix is currently being developed to further clarify the criteria and provide expanded examples. Nonetheless, experience within caBIG has already demonstrated the utility of this matrix. The matrix is also intended to inform the terminology developers of the criteria necessary to become a caBIG-approved terminology. Enhanced documentation will also naturally aid the developers; moreover, the developers and review teams are encouraged to communicate with each other during the review process.

### 5. Application of the criteria to Lab-LOINC

#### 5.1. Evaluation process

The TRC matrix (version 3.3) has subsequently been applied in an evaluation of the Logical Objects, Identifiers, Names and Codes (LOINC) [26] with specific focus on the laboratory portion (Lab-LOINC). The LOINC development team provided the LOINC User's Guide, access to their browser tool (RELMA) [27], several published articles related to evaluation and use of LOINC, and some narrative discussion of some of the criteria. The Lab-LOINC content was loaded into the Stanford University BioPortal [28] and made available to the evaluation team (T.F.H., G.A.S., M.R., and J.J.C.). Team members were instructed to analyze the data collected and determine whether or not the individual criteria are fulfilled by the terminology. Based on the experience gained through previous terminology review work, team members were advised that:

- For some criteria, such as those regarding the documentation itself, reliance on information provided by the terminology submitter is probably sufficient to demonstrate whether the terminology meets the criterion.
- For other criteria, information provided by the submitter is not necessarily sufficient to prove that a given criterion is fulfilled by the terminology, and this may need to be evaluated further.
- For some criteria, assessment of a terminology might be amenable to quantitative and/or automated evaluation. This approach should be pursued if possible, and if it augments and speeds up the evaluation process.
- In most cases, methodical spot-checking of the terminology, in conjunction with thorough evaluation of the documentation, will probably be sufficient to determine whether or not the terminology fulfills a given criterion.

After several initial conference calls to discuss the evaluation process, the members of the review team each completed the matrix independently, using the available materials. The matrices were merged and discussed via conference call in order to reach consensus.

#### 5.2. Results of Lab-LOINC evaluation

For the most part, Lab-LOINC met the structural criteria. The main areas of deficiency related to the formal representation of knowledge, including explicitness of relations and formal rules

for inferring subsumption based on definitions. In fact, Lab-LOINC does appear to meet these criteria, but the description of the Lab-LOINC structure does not confirm this.

A formal analysis of Lab-LOINC content coverage was beyond the scope of the evaluation; however, spot-checking showed LOINC to generally have good coverage of the laboratory test domain. In addition, the Lab-LOINC structure, which makes explicit such aspects as analyte, specimen and method, helped reassure the review team that ambiguity and redundancy were unlikely.

LOINC documentation fell short in many areas. Only personal knowledge of some members of the review team helped interpret what material was available. Subsequent discussions with the LOINC developers helped supplement the documentation, but the TRC matrix was not updated to reflect this, in order to reinforce the need for adequate, written terminology documentation.

In general, Lab-LOINC fared well with the editorial process criteria. However, the deficiencies in the LOINC documentation hindered appropriate assessment of many of the specific points, and only personal experience of some team members, supplemented with discussions with the LOINC developers, helped resolve them favorably. See Figs. 2–4 and Appendix I for details of the results of the Lab-LOINC evaluation.

### 5.3. Lessons learned from Lab-LOINC evaluation process

The Lab-LOINC evaluation process was more than just an assessment of a fourth terminology for caBIG. It also served as an evaluation of the evaluation process itself, including the ability of new reviewers to understand and apply the TRC criteria. Several lessons were learned from this process that are being used to inform future VCDE WS terminology evaluations:

- In general, documentation about terminologies is usually lacking. The most common form is publications describing the application of the terminology to some problem. Often, there is little public record about how the terminology is designed, constructed, and maintained. Less is written about evaluation of terminologies. As a result, the evaluation team was often unaware of key aspects of the terminology structure and content. Terminology developers should therefore be encouraged to provide documentation that supports each of the criteria in the evaluation matrix.
- The terminology developer is in the best position to provide authoritative (although potentially biased) assessments of the whether the terminology meets each of the evaluation criteria. Unfortunately, the developer may have little experience with the kinds of evaluation criteria that are actually in the matrix. Due to the fact that the developer's assessments will be crucial to the team's understanding of the terminology, an experienced member of the assessment team should engage the developer and thoroughly explain each of the criteria, resulting in an authoritative (although potentially biased) response (with explanation and supporting documentation) on each criterion.
- The question arose as to whether it was fair to use RELMA in the evaluation process, since it was a tool that could not be used for evaluation of other, possibly competing, terminologies. In general, the team felt that it was acceptable to use the terminology's native browsing tools (when available) in addition to any formal representation available in a standard browsing tool.
- In this review, important hierarchical information appeared to be missing from the BioPortal version of Lab-LOINC. The LOINC developer did not identify the specific hierarchical information that was available in Lab-LOINC but not available in BioPortal. This lack of information seriously hampered the team's ability to assess Lab-LOINC's suitability with respect to several criteria. Hence, we recommend that, prior to evaluation, the terminology

developer be given the opportunity to use the browsing tools to be made available to the reviewers, in order to determine if the tool is providing an accurate, complete representation of the content. In cases where a discrepancy is noted, the developer should work with the review team to try to make additional content available through the reviewers' browsing environment and, failing that, to provide the missing content to the review team through some other mechanism.

- Application of the TRC criteria depends upon a certain amount of experience with terminologic research principles. It is therefore crucial that some team members have experience with formal terminology evaluation, preferably including first-hand experience with the caBIG terminology criteria matrix.
- There is often much about a terminology that can only be learned with experience. This is especially true in cases where the terminology developer has provided less complete documentation about the creation, model, maintenance, and methods for data reuse. Therefore, it is extremely helpful to have one or more team members who are intimately familiar with the terminology under study. These members can provide insight from their personal experience with the terminology.
- The evaluation team can benefit from having one or more members who have experience with coding data in the domain of interest, preferably in the context of relevant caBIG data elements. This will be especially important for assessing the actual breadth and depth of the content of the terminology that, on close inspection, may be missing key terms or entire topic domains.
- Evaluation materials were readily distributed to team members. However, since these materials may be incomplete, it would be helpful to have feedback from the terminology developer, during the evaluation process.
- As noted above, one or more team members should have previous experience with terminology evaluation processes in general and the VCDE Workspace criteria in particular. The will allow the more experienced team members to train the other reviewers on the evaluation matrix and review process.
- It is clear from the Lab-LOINC evaluation process that many of the criteria in the matrix were somewhat ambiguous, especially to those who have not spent much time in terminology evaluation. While the more experienced team members can explain the criteria, the documentation of the matrix should be enhanced to include line-by-line instructions, complete with examples.
- The consensus process was a fruitful one, with consensus generally achieved when team members with evaluation experience explained more fully the meanings of various criteria, while those with specific experience with the terminology could supplement the documentation by identifying additional information buried in the terminology. However, it may be unrealistic to rely on such experience. Fortunately, if the above recommendations are followed, team members should have an easier time working independently and should produce more consistent results. In any event, the consensus building should be continued as a crucial step in resolving possible misunderstandings in the evaluation.

The conclusion of the review team was that, over all, the evaluation process appears to be based on sound criteria. Improving communication with the terminology developer and adding more explanatory notes to the evaluation matrix should improve the process and make it more streamlined.

## 6. Discussion and conclusions

Although some cancer research data are currently coded with controlled terminologies, it is the intent of caBIG to extend coding to new data as much as possible and to promote the use of high-quality terminologies for this purpose. Although terminologic re-

search has yielded several useful lists of criteria for assessing the quality of terminologies, little has been published about the systematic application of these criteria for use in auditing existing terminologies. The VCDE WS attempt to do so has identified many challenges to the practical application of criteria, resulting in the evolutionary process described in this paper. We believe that this process has resulted in a set of criteria that captures the intent of published terminology criteria and in a reproducible methodology for using them to identify strengths and weakness in standard terminologies.

The recounting of the evaluations carried out so far (of GO, NCI, CTCAE and LOINC) are included here to illustrate the development and adaptation of the TCR criteria (although the inclusion of the results of our evaluations may also benefit those readers interested in adopting any of these four terminologies). The VCDE WS is continuing to audit terminologies for use in caBIG (see [Appendix II](#) for information on how terminologies may be submitted to the caBIG review process). We anticipate that the terminology review process and the individual criteria will continue to evolve to meet the needs of the caBIG community. Nonetheless, the current process has proven to be well received by the reviewers, is readily implementable, promises scalability, and we believe can be adopted by other groups and individuals interested in conducting formal audits of terminologies.

### Acknowledgments

The authors thank the caBIG community, in particular participants of the VCDE WS for their valuable insight into the community needs for controlled terminologies, and Drs. George Komatsoulis and Frank Hartel for their leadership and scientific insights in developing community standards.

We especially thank Sherri DeCoronado and others at the NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT) and Enterprise Vocabulary Service (EVS) for input while refining the review criteria and during the GO evaluation, and members of the GO Consortium—specifically Judith Blake, Harold Drabkin, David P. Hill, Michael Ashburner, Suzanna Lewis and Chris Mungall—for comments and valuable discussion.

In addition to the coauthors (J.J.C., T.F.H., G.A.S., and M.R.), the Lab-LOINC evaluation team included Salvatore Mungal (Duke University), Stuart Turner (University of California at Davis), and Kristel Dobratz (Gulfstream Bioinformatics). The Lab-LOINC reviewers thank Clem McDonald for providing LOINC materials, the National Center for Biomedical Ontology at Stanford University for making BioPortal available for the review process, and Thomas Johnson (Mayo Clinic) for help loading Lab-LOINC into BioPortal.

Drs. Bodenreider, Cimino, Davis, Hayamizu, Ringwald, and Stafford were supported in whole or in part by federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. GS-35F-0306J. Drs. Bodenreider and Cimino were supported in part by funds from the intramural research program at the NIH Clinical Center and the National Library of Medicine (NLM).

### Appendix I. Current caBIG process

In order to achieve a common understanding of what data are being passed by computers among caBIG participating institutions, there needs to be agreement—and enforcement—of common semantics. caBIG has decided—based on community-developed and agreed-upon compatibility guidelines—that the best way to achieve this shared semantic goal was to rely upon publicly available, controlled terminology standards. The VCDE WS within caBIG was given the responsibility to develop a terminology review process by which terminologies would be evaluated. Those terminol-

ogies conforming to recognized best practices would be designated as “caBIG terminology standards,” which software development projects within caBIG are then strongly encouraged to use for description of data and services within caBIG. The terminology review process is based upon other, familiar review practices within caBIG VCDE WS, but also covers needs that are unique to terminology review.

The process begins by first engaging the submitter (the owner of the terminology or some other party seeking to use the terminology in caBIG) of a putative caBIG standard terminology. The submitter of the terminology is encouraged to work with the VCDE WS of caBIG to develop and submit a formal terminology standard submission package to a publicly accessible Gforge site [29]. The package should include:

- Justification of why the terminology should be a caBIG standard
- An electronic version of the terminology (or pointer to a download site)
- Tools (or pointers to tools) available to use the terminology
- Documentation describing the terminology
- Relevant publications
- A version of the terminology review criteria matrix completed by the submitter to indicate the submitter's conception of the terminology's compliance with the criteria

Wherever possible, coding systems should be converted to terminologies before submitting for evaluation. There are several electronic formats that are easily imported into the caBIG LexBIG tool for browsing and querying:

- Web Ontology Language (OWL); primary focus on DL
- Open BioMedical Ontologies (OBO); version 1.0
- Open BioMedical Ontologies (OBO); version 1.2
- Unified Medical Language System (UMLS) Metathesaurus Rich Release Format (RRF)
- Unified Medical Language System (UMLS) Semantic Net
- LexGrid XML
- Protégé frames (Importation of Protégé requires customization to map each frame-based model during the import process; currently supports the Foundational Model of Anatomy (FMA) and RadLex models.)

As noted earlier, documentation is frequently lacking for terminologies, and this will impede the review. However, close engagement by the terminology submitter can help mitigate the lack of documentation.

The submitter is encouraged to make a formal presentation at a biweekly VCDE WS teleconference (an open, public forum) to familiarize the caBIG and VCDE WS participants with the terminology and submit a case for its importance as a caBIG standard. The formal review starts by assemblage of a team of voluntary reviewers, with one member designated as lead, from the VCDE WS (and other interested parties). The team will ideally be composed of some members with experience in formal terminology evaluation, some with experience with the terminology being evaluated, and some with experience in the domain(s) covered by the terminology. Obviously, these roles may overlap. The VCDE WS highly encourages the terminology submitter to be closely involved with the review team, although he/she is not formally part of the review. The possibility of conflict of interest is present, but thus far has not arisen.

The lead directs the other reviewers to the material and instructs the group in evaluating the terminology and filling out the TCR Matrix. The reviewers work independently to evaluate the material. In most cases, methodical spot-checking of the terminology, in conjunction with thorough evaluation of the documentation, will be sufficient to determine whether or not the

terminology fulfills a given criterion. If the terminology lends itself to quantitative and/or automated evaluation (as described in the NCI evaluation, above), this approach should be pursued if possible, to speed up the evaluation process.

The group meets midway through the evaluation (about two to four weeks) to assess progress. If there are no serious issues raised, the reviewers continue individual evaluations and submit their final evaluations to the lead, who collates the results. The group then meets to discuss results and reach a consensus on their assessment of the criteria—a crucial step in resolving possible misunderstandings in the evaluation. The inclusion of the terminology submitter at this step can help resolve any issues that might arise. Finally, the terminology review group resolves the issues and presents a consensus recommendation to the VCDE WS as to whether the terminology should be promoted as a caBIG terminology standard based on the TRC results.

As of April 2008, the caBIG community has evaluated the four terminologies as described above. Three of these (GO, the NCI Thesaurus, and Lab-LOINC) have been accepted for designation as caBIG Terminology Standards. Only CTCAE was not accepted, as it was found to be lacking in one key area (an acceptable maintenance process). The CTCAE developers have decided to address this issue and CTCAE will likely be reviewed again at a later date. Additional terminologies slated for review as of this writing, include the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) from The International Healthcare Terminology Standards Development Organization [30] and RxNorm, a standardized nomenclature for clinical drugs produced by the National Library of Medicine [31]. The terminologies chosen for review thus far were selected from a list created by surveying the caBIG development community. The VCDE WS has approached the terminology developers in each case to initiate the review. Certainly, the caBIG community will welcome developers to submit their terminologies for consideration for caBIG standards.

## Appendix II. Supplementary data

Summary of results of the caBIG terminology evaluations. *Note:* this appendix contains a spreadsheet, sent separately, which we have rendered as four pages of images, also sent separately. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.jbi.2008.12.003.

## References

- [1] Komatsoulis GA, Warzel DB, Hartel FW, Shanbhag K, Chilukuri R, Fragoso G, et al. caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. *J Biomed Inform* 2008;41(1):106–23.
- [2] Von Eschenbach AC, Buetow K. Cancer Informatics Vision: caBIG<sup>SM</sup>. *Cancer Inform* 2006;2:22–4.
- [3] Raiez F, Arts D, Cornet R. Terminological system maintenance: a procedures framework and an exploration of current practice. *Stud Health Technol Inform* 2005;116:701–6.
- [4] Elkin PL, Brown SH, Chute CG. Guideline for health informatics: controlled health vocabularies—vocabulary structure and high-level indicators. *Medinfo* 2001;10(Pt 1):191–5.
- [5] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998;37(4–5):394–403.
- [6] Walker D. GP Vocabulary Project—Stage 2 Report: SNOMED Clinical Terms (SNOMED CT); November, 2004. Available from: [http://www.adelaide.edu.au/health/gp/research/current/vocab/2\\_02\\_2.pdf](http://www.adelaide.edu.au/health/gp/research/current/vocab/2_02_2.pdf).
- [7] International Standards Committee (ISO) Technical Committee TC215. TS 17117:2002(E): Health Informatics—controlled Health Terminology—Structure and High Level Indicators. Geneva, 2002.
- [8] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25–9.
- [9] Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res*. 2001;11(8):1425–33.
- [10] Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40(1):30–43.
- [11] Hartel FW, De Coronado S, Dionne R, Fragoso G, Golbeck J. Modeling a description logic vocabulary for cancer research. *J Biomed Inform* 2005;38(2):114–29.
- [12] Golbeck J, Fragoso G, Hartel F, Hendler J, Oberthaler J, Parsia B. The National Cancer Institute's Thesaurus and ontology. *Web Semantics: Science, Services and Agents on the World Wide Web* 2003;1(1):75–80.
- [13] National Cancer Institute. Enterprise Vocabulary Services. In: caCORE 3.1 Technical Guide; 2006; p. 49–75.
- [14] Roles declared in the NCI Thesaurus. Available from: <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusSemantics/Roles.pdf>.
- [15] NCI Thesaurus Terms of Use. Available from: [ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusTermsOfUse\\_files/ThesaurusTermsOfUse.pdf](ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusTermsOfUse_files/ThesaurusTermsOfUse.pdf).
- [16] NCI Thesaurus property definitions. Available from: <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusSemantics/Properties.pdf>.
- [17] de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus: using science-based terminology to integrate cancer research results. *Medinfo* 2004;11(Pt 1):33–7.
- [18] Fragoso G, de Coronado S, Haber M, Hartel F, Wright L. Overview and utilization of the NCI Thesaurus. *Comp Funct Genomics* 2004;5(8):648–54.
- [19] Hartel FW, Fragoso G, Ong KL, Dionne R. Enhancing quality of retrieval through concept edit history. *AMIA Annu Symp Proc* 2003:279–83.
- [20] Ceusters W, Smith B, Goldberg L. A terminological and ontological analysis of the NCI Thesaurus. *Meth Inform Med* 2005;44(4):498–507.
- [21] Kumar A, Smith B. Oncology ontology in the NCI thesaurus. In: *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. 2005; 2005: p. 213–20.
- [22] Cancer Therapy Evaluation Program. Common Terminology Criteria for Adverse Events v3.0 (CTCAE). August 9, 2006. Available from: <http://ctep.cancer.gov/forms/CTCAEv3.pdf>.
- [23] caBIG Compatibility & Certification Guidelines. Available from: [http://cabig.nci.nih.gov/guidelines\\_documentation](http://cabig.nci.nih.gov/guidelines_documentation).
- [24] Cimino JJ. An approach to coping with the annual changes in ICD9-CM. *Methods Inf Med* 1996;35(3):220.
- [25] The NCI LexBIG Enterprise Vocabulary Server. Available from: [http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/vocabulary](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary).
- [26] Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem* 1996;42(1):81–90.
- [27] McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem* 2003;49(4):624–33.
- [28] The National Center for Biomedical Ontology—BioPortal. Available from: <http://www.bioontology.org/bioportal.html>.
- [29] Cancer Biomedical Informatics Grid (caBIG) Vocabulary Standards Project Information. Available from: <http://gforge.nci.nih.gov/projects/vocabstandard>.
- [30] Spackman KA. SNOMED CT milestones: endorsements are added to already-impressive standards credentials. *Health Inform* 2004;21(9):54–6.
- [31] Bouhaddou O, Warnekar P, Parrish F, Do N, Mandel J, Kilbourne J, et al. Exchange of computable patient data between the Department of Veterans Affairs (VA) and the Department of Defense (DoD): terminology mediation strategy. *J Am Med Inform Assoc* 2008;15(2):174–83.