



caBIG[™] cancer Biomedical
Informatics Grid [™]

An Initiative of the National Cancer Institute

VCDE Small Group: CTCAE as a Controlled Vocabulary Gap Analysis

caBIG - Vocabulary and Common Data Elements Small Group

Lewis Frey, Sal Mungal, Michael Riben, Larry Wright, Stuart Turner, Ann Setser,
Ranjana Srivastava, Michael Keller

Monday, 29 September 2008

Note: The evaluation of the CTCAE (Common Terminology for Adverse Events) version 3.0 was performed late 2006 and presented to the caBIG Workspace in January and April 2007. This is a summary of that review performed by **Jim Cimino** and the CTCAE review team. Much of the following information is derived from that report and the source documents can be referenced from Vocabulary Standardization section of GForge.

- The CTEP version of CTCAE (ver 3.0) does not inherently meet the requirements of a controlled terminology (i.e. a finite, coded, enumerated and explicit set of terms)
- CTCAE's content *does* cover the domain of interest well and is structured sufficiently enough to begin the task of *reifying* the content toward a structured terminology.
- While the CTEP version does not meet vocabulary criteria, the reification into the NCI Thesaurus bridged many of the deficits required.
- This presentation is an overview of that analysis and status of a Vocabulary and Common Data Element Small Group to provide a gap analysis based on a prior review by Jim Cimino, the current NCI/caBIG vocabulary standardization process and the development of recommendations to guide the CTCAE revision to ensure its instantiation to a sound terminological structure and especially satisfaction of criteria to ensure certification by the caBIG community.

The Vocabulary and Common Data Elements Workspace's formal vocabulary (terminology or ontology) evaluation process.

- A process capable of investigating, describing and providing recommendations for:
 - Taxonomies, terminologies
 - Simple to complex hierarchies
 - Controlled vocabularies
 - Concepts as nodes associated with a unique **identifier**, a unique **term**, well crafted **definition**, **synonyms**, annotations, etc.
 - Ontologies
 - Nodes connected by edges to form a directed graph with improved *semantic understanding through associations or relations* (e.g. “is-a”, “part-of”, “isDefinedBy”)

**Cardiac arrhythmia:
Supraventricular
tachycardia**

hasGrade

**Grade 3 (severe): "Symptomatic and
incompletely controlled medically, or
controlled with device (e.g., pacemaker)"**



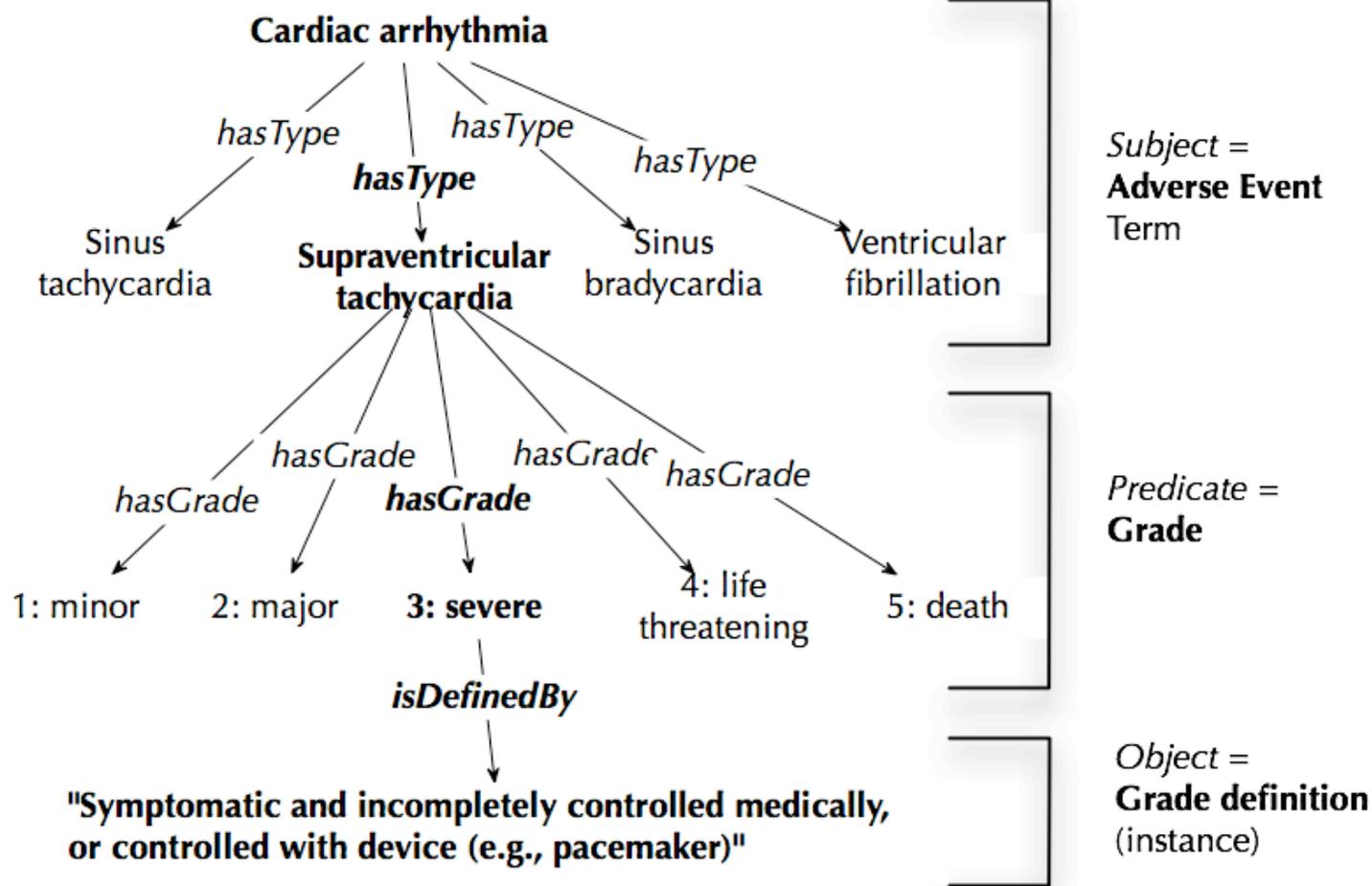
Subject =
**Adverse Event
Term**



Predicate =
Grade



Object =
**Grade definition
(instance)**



- Set of ~ 105 criteria
- Based on core set of rigorous principles (e.g. Cimino's Desiderata)
 - Evaluation of *understandability, reproducibility, and usability* (the "URU")
 - Quality of Documentation
 - Maintenance and Extensions
 - Accessibility and Distribution
 - Intellectual Property Considerations
 - Quality Assurance and Quality Control
 - Textual Definitions
 - Community Acceptance
 - Reporting Requirements
 - Harmonization with other standards

Great opportunity now to build an even better resource

- Current content revision
- Harmonization with MedDRA
- Formal vocabulary analysis, evaluation and recommendations
- Lessons learned from the reification into NCIt
- Existing and emerging NCI vocabulary practices and support
- Novel method of curation using the Semantic Mediawiki
- Ability to leverage existing tools and technology to build a well structured vocabulary or ontology
- Collectively provide guidance for a formal governance environment that is efficient, open and rigorous

Who: VCDE Small Group

What: Gap Analysis

Why: Provide guidance in parallel to CTCAE 3.0 to 4.0 content revision to ensure that CTCAE 4.0 is a product that meets or exceeds the requirements of a sound controlled vocabulary or ontology.

How: Identify deficiencies from exhaustive review of CTCAE 3.0 by Cimino as foundation. Identify existing tools (NCI and others) and technology (e.g. ontology development) to make pragmatic yet innovative recommendations to improve longevity, extensibility and invulnerability.

Where: All VCDE related artifacts (e.g. working documents, normative references, literature, schedule, etc.) will be open for review and discussion and posted (e.g. Gforge). Various methods for notification of updates (email, workgroup meetings, email subscription, RSS)

- The reification of CTCAE 3.0 into the NCI Thesaurus mitigated many of problems related to vocabulary structure (e.g. non-semantic identifiers, mapping to concepts). And, much of this because of excellent vocabulary practices, post-reification quality control, etc.
- However, no process to allow continued development (one time)
- Some information lost (e.g. Also Consider, Navigation Notes and Remarks)
- No change in content to ensure consistency of term representation (e.g. semantic types, plurality, composite terms, use of remarks as formal annotations, etc.
- Approach (precoordination) a pragmatic one. Still require more rigorous evaluation to determine if this is still the best or if additional time/investment to model in more detail (e.g. ontology) is warranted

What is known: Some examples

- Instantiation into a controlled vocabulary (or ontology) alone will enable a formal governance and editorial process and facilitate documentation.
- Will better enable independent evaluation of content quality as well as critical review of revisions (difficult to compare changes if terms, rather than identifiers, are used)
- Use of “Other” terms violate core vocabulary principles, however rules to mandate use of “Specify” will at least assist with composition with parent (or super ordinate) terms. However, both content and model can be improved to discourage this approach
- Full definitions for terms (AE and supra-ordinate) necessary and can leverage existing concepts in the NCI Thesaurus, especially if structured well
- Clearly define delineation between MedDRA and CTCAE, both in sharing of terms/concepts as well as usage.
- Vocabulary team work in parallel with content developers to ensure consistency in terms (e.g. clinical vs. laboratory findings, plurality, polyhierarchy, etc.)

What is known: Some examples

- Better curate and model AE-Grade terms so they alone are clear and unambiguous (do not require examination of grade definitions to determine context)
- Provide structure to cross-reference to other NCI concepts
- Ensure intellectual property issues (e.g. copyright, licensure) are well crafted to ensure openness yet prevent infringement on exclusive rights of the terminology and to ensure usage is appropriate.
- Governance process will be enabled and guided in part by structure and output formats alone. This will allow publishing of revisions, tools to allow broad community participation and contribution, identify gaps, ensure that special rules are captured and computable (e.g. description logic, other). Group to decide on output format (tool agnostic), recommended tools for curation, browsing, search and retrieval, etc.
- Editorial process will require excellent documentation, explicit rules for how content is updated, how the community can participate, what and how decisions about changes were made, provision or permanent archival storage of versions and version management, etc.