

The terminology requirements for caBIG compatibility are laid out in the Compatibility Guide [https://cabig.nci.nih.gov/guidelines\\_documentation/caBIGCompatGuideRev2\\_final.pdf](https://cabig.nci.nih.gov/guidelines_documentation/caBIGCompatGuideRev2_final.pdf)

### **For Bronze compliance:**

Terminologies must be used that meet two broad criteria:

1. Use of publicly accessible controlled vocabularies as well as local terminologies.
2. Terminologies must include definitions of terms that caBIG VCDE workspace guidelines

On page 9 the Guide goes on to say that:

At the Bronze level of maturity, the information resource utilizes public controlled vocabularies in parts of the data collection and reporting process, but may supplement them with local vocabularies. All terminologies, including those developed locally, should include definitions of terms that are sufficient to distinguish the meaning of that concept from other concepts in the terminology (a 'working definition'). At a practical level these definitions must meet the following criteria:

1. They are stated in the singular
2. They describe what the concept is, not just what it is not
3. They are stated as a descriptive phrase or sentence(s)
4. They contain only commonly understood abbreviations that are themselves defined in the terminology
5. They can be expressed without embedding definitions of other concepts (i.e. any other concepts that are required must also exist in the terminology)
6. They must not involve circular reasoning (i.e. they do not use the term in its definition)

### **For Silver compliance:**

Terminologies must be used that meet two more stringent criteria:

1. Terminologies reviewed and validated by the caBIG VCDE Workspace used for all appropriate data collection fields and attributes of data objects.
2. Term definitions must meet VCDE Workspace guidelines.

Silver-level maturity introduces the requirement for review and approval of terminologies by the caBIG VCDE workspace. Local or private terminologies that are not available to the caBIG community may not be implemented. The NCI Enterprise Vocabulary Services (EVS) provides a management system for approved terminologies, but note that not all EVS-hosted terminologies have necessarily been reviewed and approved for caBIG.

The VCDE workspace will use the criteria described below (Understandability, Reproducibility, Usability, documentation, accessibility, maintenance and Intellectual Property) to determine if a vocabulary should be approved. Definitions in silver level vocabularies should:

1. Describe the essential nature of the concept

2. Be concise, precise and unambiguous
3. Be expressed without embedding rationale, functional usage, or procedural information
4. Use the same terminology and consistent logical structure to describe similar concepts.
5. The presence of description logic relationships to other concepts in the vocabulary may be leveraged to produce the English language definition. In any case the English language definition of the concept must not conflict with description logic relationships asserted about the concept.

As indicated above, the VCDE workspace may (at its discretion) accept a vocabulary that meets these requirements for most of its terms, or that has a clear plan for meeting these requirements.

**For Gold compliance:**

Terminologies must be used that meet two additional criteria:

1. Full adoption of caBIG terminology standards as approved by the VCDE workspace.
2. Terminologies must be available through a caGrid service.

Gold compatibility is similar to Silver, but with the added requirements that registered standards approved for caBIG-wide usage are implemented wherever they are available and that the terminology be accessible through a caGrid vocabulary service.

## vCDE compliance criteria:

### 1. URU (Understandability, Reproducibility, Usability) after Keith Campbell

The first criterion, understandability, makes reference to whether a concept (or other design feature of the terminology) can be fully and unambiguously comprehended by users of the terminology. Understandability is tested by checking to see whether users believe they can tell whether the concept is relevant or not relevant to a given experiment, patient or situation. Reproducibility indicates whether multiple users apply the concept to the same situations. Finally, Usability refers to the level of helpfulness and appropriateness conveyed in a concept or feature.

### 2. Desiderata after Jim Cimino and ISO/TS 17117:2002

The ideal terminology:

- *states its purpose and scope clearly in operational terms*
- *provides comprehensive or explicit in-depth coverage of the domain of interest*
- *is concept oriented*
- *provides concept permanence*
- *uses non-semantic concept identifiers*
- *may be uni- or poly-hierarchy*
- *uses formal definitions (consisting of a definitive set of relationships to other concepts)*
- *defines the logical definition of subsumption*
- *rejects of Not Elsewhere Classified (NEC) terms*
- *supports multiple granularities (multipurpose terminologies only)*
- *supports multiple consistent views (multipurpose terminologies only)*
- *provides context representation ( formal, explicit information about how concepts are used.*
- *provides clear detailed descriptions of what changes occur and why*
- *provides a mechanism by which redundant coding of external data can be recognized*

### 3. Quality of Documentation

The terminology ought to be well described from a variety of viewpoints:

- *Organization publishing the terminology provide a statement of intended use (including use cases?), intended users, and scope of coverage*
- *The list of concepts/terms and definitions is available*
- *The documentation describes:*
  - *the terminology's structure and organizing principles;*
  - *its use of concept codes/identifiers;*
  - *its use of semantic relationships (if any);*
  - *the format(s) in which it may be obtained;*
  - *any restrictions on its use;*
  - *its relationships/links/mapping (if any) to other terminology resources, etc.?*
- *New versions are accompanied by documentation that describes how the new version differs from the one it replaces.*
- *Methods for extending the terminology are documented*
- *A description of methods or tools for acquisition and application of the vocabulary is provided*

#### **4. Maintenance and Extensions**

The terminology publisher ought to:

- *Follow defined change management for: Additions, Refinements, Pre-coordination, Disambiguation, Obsolescence, Redundancy discovery, Name changes*
- *Be produced by an organization that:*
  - *is committed to maintain it over time.*
  - *sees creation and maintenance of terminology a core function or, if the developer/publisher is not in the terminology business, has it made creditable arrangements with another organization for maintenance and enhancement*
  - *lack of maintenance commitment should be disqualifying*
- *Take special steps if their terminology extends or overlays other terminologies. Those producing such extensions must:*
  - *must have a formal methodology for expanding content*
  - *take on the burden of keeping their extensions up-to-date and consistent with future releases of the underlying terminology*

#### **5. Accessibility and Distribution**

Open formats facilitate distribution. The lack of a clear migration/transformation pathway to a servable format is a sufficient reason to reject the terminology being considered.

*At a minimum, the terminology should be freely available for download in a format that can be readily used by the community.*

*Examples are*

- *RRF (Rich Release (MR+) Format*
- *OWL (The format used by the World-Wide Web Consortium.)*
- *OBO (The format used by the Open Biology Ontology group).*
- *XML and accompanying DTD.*

*Other formats are possible, but there should be a clear and reliable path by which the format can be transformed into one of the above. Availability in one of these formats ensures that the terminology will be able to be served interactively across the caBIG grid by suitable server software and accompanying API.*

#### **6. Intellectual Property Considerations**

Any caBIG standard Terminology should be:

- *freely disseminated without restriction (distribution applies only to the dissemination of the terminology itself. It does not apply to the authoring environment)*
- *there should be no fee associated with dissemination and use*

If a terminology's license is for a limited term, and the rules may change when the term expires, this represents a significant risk. The implications of that risk should be carefully considered before nominating or accepting the terminology as a caBIG standard.

## **7. Quality Assurance and Control**

QA/QC documentation for a vocabulary should describe:

- *internal checks performed to detect and eliminate errors in terminology*
- *review and validation process used to achieve good quality.*

Terminologies should be reviewed by independent experts from the field in which the terminology will be used.

## **8. Reporting Requirements**

Specification of a terminology as a regulatory standard of legal requirement is an important determinant of its adoption as a caBIG standard. Several questions may be used to assess if a terminology is needed to satisfy a reporting requirement.

- *has a health regulatory body required this vocabulary for reporting?*
- *can the regulatory body be local? (eg IRB)*
- *what is the regulatory body?*
- *what are the requirements?*
- *is this vocabulary used in this application meant to be used for reporting to a regulatory body? If so, which one(s)?*
- *if a regulatory body accepts a “rival” vocabulary, is that grounds to reject a putative vocabulary?*

## **9. Community Acceptance**

Community acceptance of a terminology is an important determinant of its suitability for caBIG. Several questions may be used to assess the centrality of a terminology in a community:

- *How many “community standards” are there in the community in question?*
- *Does the community have a formal statement about its standard terminology?*
- *Does the terminology in question have the highest use in the community?*
- *Is there a large body of relevant data tagged with the terminology?*
- *Alternatively, is there something controversial about the terminology that would never allow a terminology to be accepted in some places?*