NCI Thesaurus
Editorial Policy Statement
March 27, 2007

NCI Thesaurus concepts

An NCI Thesaurus concept is the basic unit that makes up the NCI Thesaurus. It is intended that each concept have its own distinct and stable meaning. Every NCI concept receives a unique, permanent NCI concept code, a preferred term, a semantic type, and one or more parent concepts; each also has a unique concept name and id, but this may change soon. Most concepts already have a formal English-language definition and relevant synonyms, and it is our goal to provide these for all concepts. Most concepts also contain various annotations and concept to concept relationship information.

Creation of Concepts

Concepts in the NCI Thesaurus are created under the following conditions:

1. If required by groups that we support, such as NCI divisions, offices, and centers, wider efforts such as caCORE and caBIG, and outside partners such as FDA and CDISC;
2. If published in the scientific literature or other key sources of biomedical information; or
3. If useful to better characterize and complete coverage of existing content areas, or to expand content coverage to areas not previously addressed by the vocabulary.

Once a candidate for a new concept has been identified, the database is searched to ensure that the concept, or an equivalent variant, does not already exist.

When creating a new concept, it is permissible for more than one concept to have the same preferred name, which is generally in singular form. However, each concept has a unique code and a unique meaning. Every concept must have a parent concept (super concept) and is placed in the parent-child is-a hierarchy ("treed") in the most specific possible place(s). For example, a helicase is a kind of enzyme but more specifically, it is a type of hydrolase. Therefore, it should be treed as is a hydrolase. Once treed, a concept inherits all of the attributes of its parents.

Each top level division of the NCI Thesaurus is a disjoint class; therefore, a concept can only exist in one class and can only be double treed within a single class. If a term has two distinct meanings, requiring treeing in multiple classes, then two separate concepts are created.

Splits and Merges

In the course of curating the NCI Thesaurus, there are times when EVS editors may need to merge or split existing concepts. An individual editor with appropriate subject matter expertise will split a concept if it includes strings with two different meanings, or will merge two concepts that have the same meaning. This usually occurs when there is a direct request from a user, new terms appear in the literature, or outdated terms for concepts are overtaken by the scientific process. Occasionally, splits and merges are performed to correct errors made in earlier editing.

When a split occurs, one or more new concepts will be created and the appropriate terms and attributes will be moved to the new concept. The new concept automatically becomes a sibling of the original concept and automatically inherits all of the children of the original concept. The validity of these inherited relationships is reviewed and modified as required by the editor.

Concepts are merged when it is determined that they are synonymous (i.e., they don't have a meaning distinct from each other). Unless otherwise specified, by default, the older concept survives and gains all of the annotations and concept to concept relationships of the newer concept, and the newer concept is retired. Exceptions occur if:

1) The newer concept has been referenced in more caCORE metadata, data models, or other systems; or
2) The newer concept is better formed or more fully modeled.

Merging is a two-step process, requiring review by at least two editors. First, the editor "flags" a concept for merging. Next, during baseline update – the periodic synchronization of all editors' changes – the Workflow Manager (a senior editor) will review the candidate merges. The manager may accept or reject the suggestion. If accepted, the concept properties and relationships will be transferred to the merge target concept and the old concept will be treed as a retired concept.

Retirements

Once a concept has been created and published it must remain available to users. Therefore, published concepts never "disappear" from the NCI Thesaurus. Rather than being deleted they are "retired". A concept is retired when it is determined that it is not needed, e.g.:

1. Another concept exists with the same meaning;
2. The concept itself is poorly formed or otherwise not useful or is ambiguous;
3. The concept is an unneeded header;
4. The concept has been superseded by a more current scientific conceptualization.

Retirement is a two-step process, requiring review by at least two editors. First, the editor flags a concept for retirement. Next, during baseline update the Workflow Manager reviews all of the pre-retirements. The manager accepts or rejects the suggestion. If accepted, the concept will be treed as a retired concept. Once retired, a concept is not "unretired," nor is it further edited. If it is later found that the retirement was in error or the concept has regained usefulness, a new concept with a new concept code but substantially the same meaning will be created.

Feedback & Support

The EVS team invites collaboration and welcomes input from our users. For general questions or comments about NCI Thesaurus or the EVS, contact a member of the EVS staff:

EVS Federal Staff

NCI Office of Communications

Margaret Haber*	mhaber@mail.nih.gov
Larry Wright*	lwright@mail.nih.gov

NCI Center for Bioinformatics

Frank Hartel*	hartel@mail.nih.gov
Sherri de Coronado	decorons@mail.nih.gov
Gilberto Fragoso	fragosog@mail.nih.gov

*Co-Directors

EVS Technical Support is available through the NCICB Application Support Group, and will answer and/or pass on any questions or comments:

- Email: ncicb@pop.nci.nih.gov
- Phone: 301-451-4384 or toll free: 888-478-4423.