

NCI Thesaurus Terminology Guidelines for Content Creation and Enrichment

There are three categories of data that cannot be coded in NCI Thesaurus (NCIt): 1- invalid data, 2-unconfirmable data, 3-incompatible data.

NEW CONCEPTS

1. Concepts that are ambiguous or unclear will not be accepted as part of the NCIt. NCIt aims for clarity and precision.
2. NCIt will not create a new concept for which there is an existing semantically equivalent concept.
3. Terms for concepts will follow U.S.-style English spelling rules and will not include non-conventional and/or excessive punctuation.
4. Terms that are excessively complex (e.g., containing conjunctions or disjunctions, such as "and," "or," or "and/or") generally will not be included as conceptual entities. This includes exclusion due to extensive pre-coordination of the term.
5. Terms accepted should be widely used in the relevant domain, as evidenced by related authoritative references, except in the cases where the concept is a newly recognized or an emerging entity. NCIt will not add concepts where requested terms are without widely accepted usage and meaning. Exceptions to this rule include new or emerging concepts for which there is only study documentation, manufacturer publication, emerging research, etc.
6. NCIt Preferred Terms will generally not contain syntactic variation or unusual punctuation, and they will not follow foreign spelling rules (i.e. British, French, etc.) or contain special characters.

NEW SYNONYMS

1. For any term, synonyms must have the same meaning as the NCIt Preferred Term.
2. Synonyms cannot be broader or narrower than the existing concept. The general rule for synonym additions should be followed: If A equals B, then B must always equal A.
3. Synonyms should only be added that are scientifically sound and in general use across the knowledge domain or industry. Acronyms following this rule may be added as synonyms.
4. Lexical variants of existing synonyms, including syntactic variation (word order) and unusual punctuation, synonyms following foreign spelling rules (i.e. British, French, etc.) or terms containing special characters, generally will not be added to NCIt. In general, NCIt avoids duplication and redundancy of terms within a concept.
5. The addition of syntactic variation and punctuation in synonyms of chemicals, genes and proteins, and molecular abnormalities is allowed if recognized authoritative sources include those terms as synonyms.
6. Terms with typographic errors will not be included in NCIt.