Neuroinformatics
Copyright © Humana Press Inc.
All rights of any nature whatsoever are reserved.
ISSN 1539-2791/03/01:61–64/$20.00

Commentary

Adding NeuroNames to the UMLS Metathesaurus

William T. Hole[*] *and Suresh Srinivasan*

National Library of Medicine, Bethesda, MD

Introduction

NeuroNames is a valuable component of the Metathesaurus®, one Knowledge Source in the National Library of Medicine's Unified Medical Language System (UMLS®). The Metathesaurus organizes biomedical names from many component vocabularies by meaning, uniting those with the same meaning in *concepts*. The Spring 2002 release contains 2.1 million names for 871,584 Concepts from 102 sources. NeuroNames provides an authoritative reference vocabulary for neuroscience, with rich synonymy and valuable links to further information such as BrainInfo. Its very high standards permit quality assurance including the identification of unrecognized synonyms from other Metathesaurus vocabularies. The addition of NeuroNames enables navigation from many other vocabularies to NeuroNames and its linked resources, or from NeuroNames to other resources that use different vocabularies.

The UMLS Metathesaurus

The Metathesaurus represents the names, relationships, and attributes from its source vocabularies, unites synonyms in Concepts, and adds other interconnecting relationships. It also adds disambiguating names, relationships, and *semantic types* for all concepts; it is a useful repository of many standard vocabularies in a common format. The Metathesaurus is not an NLM-authored encyclopedia of biomedicine; its nature, qualities, and scope are at heart the sum of its component vocabularies.

Criteria for including vocabularies in the Metathesaurus range from the practical to the theoretical, including thesaurus principles that have emerged in our work as important for biomedical thesauri. Starting with the practical, an entire source needs to be available in clean and well-documented file formats. Formatting errors, data value problems, and undefined character sets are issues we

[*] Author to whom all correspondence and reprint requests should be sent. E-mail: wth@nlm.nih.gov

encounter all too frequently. The representation of all information (the *schema*) needs to be as complete and detailed as possible, with different data elements distinguished and linked by unique identifiers. Data integrity issues often appear where there is not rigorous automated and expert human quality control.

We prefer sources which are *concept oriented*—i.e., organized by meaning rather than as vaguely aggregated related meanings. Names should be *face valid* to biomedical professionals—neither idiosyncratic nor with implied context. There should be definitional information such as text definitions, definitional relationships or attributes, or links to other defining or supplementary information. Good candidates also supplement other Metathesaurus sources by adding breadth or depth, by adding names including explicit synonyms, or by contributing new hierarchies, relationships, and attributes.

Vocabularies should be authoritative and maintained by a stable, recognized entity. There should be a clear model of releases and updates and responsive technical and content support contacts are needed. The UMLS License Agreement system model for protection of intellectual property must be acceptable to the sponsors. Finally, vocabularies should be useful in biomedical informatics and should be in actual use; they should help to create or link to currently available electronic resources such as patient records, journals, texts, and research databases.

Merging NeuroNames into the Metathesaurus

It is surprisingly difficult to find vocabularies that meet most of the criteria listed above. NeuroNames met these criteria unusually well. In particular, the clean, concept-oriented information; the authoritative content; and the clear schema with links to BrainInfo and other digital resources for further and definitional information are outstanding.

A few minor problem areas related to nontransparent naming were noted. First, the NeuroNames abbreviations are clearly very useful as labels or shorthand entry forms but may be opaque or ambiguous in general biomedicine; examples are “6n” for “abducens nerve,” or “ZI” for “zona incerta.” These abbreviations do not have clear face valid meanings and so were added as attributes rather than as formal names in the Metathesaurus. There are also a few cases of two or three character synonyms (e.g., “AV” for “anteroventral nucleus” or “Pt” for “paratenial nucleus”) which may be cryptic or ambiguous. Finally, the important species restriction for Human-only and Macaque-only meanings is indicated in naming by an added “(H)” or “(M)”; this may not clearly indicate the meaning to generalists, so we added *fully specified names* for the Metathesaurus, e.g., “Occipital gyrus (Macaque only)” and also added an attribute with this information.

After a new vocabulary is inserted, all new content is reviewed by our editors who must resolve all conflicts. This process is remarkably easy with NeuroNames, since the content is authoritative and concept-oriented; we in fact always found that NeuroNames is correct when other vocabularies disagree.

Overall, 59.2% of NeuroNames meanings merged with other vocabularies. Most frequent merges were with SNOMED International, MeSH, and the NHS Clinical Terms (Read codes). Volumetric concepts matched more frequently (62.3%) than Superficial concepts (47.9%).

As always, we hope that authoritative additions will provide increased detail (granularity) and this did indeed occur with NeuroNames as is shown by the monotonic decrease from 70.6% to 46.1% in the percentage of matches at each level of the Brain Hierarchy, when starting below the top “BRAIN” and the second level Volumetric Substructures and Superficial Features.

Another remarkable contribution to the Metathesaurus is shown by the merging of previously unrecognized synonyms from other vocabularies which occurred when NeuroNames synonyms were added for the 2000 Metathesaurus. The explicit synonymy and the editing of these concepts resulted in 607 merges of previously distinct concepts.

We are eager to include the enhanced NeuroNames 2002 in the Metathesaurus, which is being updated and is released quarterly. We especially anticipate the recently added content, including the ancillary terms and foreign language names.