

## Onto-Builder

### A Software Tool for the Development of Data Dictionaries

#### Motivation

The management of clinical trials involves various facilities and medical experts (e.g. pathologists, radiotherapists). Presently, the large amount of data generated in these trials is not documented in a uniform manner, and underlying concepts often are not uniformly or precisely defined. To overcome the resulting inconsistencies in the understanding of medical terms and interpretation of data it is necessary to construct a semantically founded Data Dictionary.

#### Objectives

By using a common Data Dictionary the following goals can be achieved:

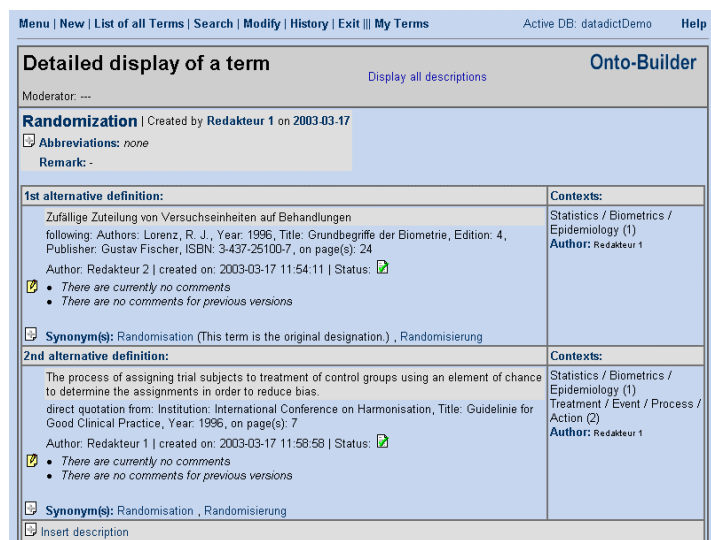
- reusability of terms and their meanings, e.g. in the definition of new trials
- uniform reference of terms for all trial-specific tools, input windows
- uniform reference of terms for the definition of Standard Operating Procedures (SOPs) within the Lymphoma Alliance
- quality assurance in carrying out clinical trials
- comparison of differing definitions of a term in various sources, thereby supporting comparative analysis and consensus building
- harmonization of the study-center databases within the Lymphoma Alliance
- harmonization of trial protocols and case report forms

#### Methods and Realization

The Onto-Builder is a tool which makes the construction and administration of Data Dictionaries possible.

In the first stage the Onto-Builder was implemented as a web-based application to allow for the decentralized compilation of contents. Data management in the Onto-Builder takes place in a relational database, and allows for several independent Data Dictionaries to be constructed simultaneously. Synonymous designations of a concept can be captured, as well as multiple alternative definitions with corresponding source and context information relating to the validity of a definition.

The standardized input of terms and their natural-language definitions takes place decentrally by trained professionals and experts in the relevant fields. To facilitate subsequent analysis, their work is performed according to a developed Guideline containing lexical and semantic rules for the definition of terms and relations. In addition, the Onto-Builder provides a multi-level quality assurance cycle to support the consensus-building process for uniform definitions in natural language.



The screenshot shows the 'Detailed display of a term' for 'Randomization' in the Onto-Builder. The interface includes a navigation menu at the top, a moderator field, and a 'Display all descriptions' link. The term is created by 'Redakteur 1' on 2003-03-17. It features two alternative definitions:

1st alternative definition:	Contexts:
Zufällige Zuteilung von Versuchseinheiten auf Behandlungen following: Authors: Lorenz, R. J., Year: 1996, Title: Grundbegriffe der Biometrie, Edition: 4, Publisher: Gustav Fischer, ISBN: 3-437-26100-7, on page(s): 24 Author: Redakteur 2   created on: 2003-03-17 11:54:11   Status: <input checked="" type="checkbox"/>	Statistics / Biometrics / Epidemiology (1) Author: Redakteur 1
<ul style="list-style-type: none"> <li>There are currently no comments</li> <li>There are no comments for previous versions</li> </ul>	
<input checked="" type="checkbox"/> <b>Synonym(s):</b> Randomisation (This term is the original designation.) , Randomisierung	
2nd alternative definition:	Contexts:
The process of assigning trial subjects to treatment of control groups using an element of chance to determine the assignments in order to reduce bias. direct quotation from: Institution: International Conference on Harmonisation, Title: Guideline for Good Clinical Practice, Year: 1996, on page(s): 7 Author: Redakteur 1   created on: 2003-03-17 11:58:58   Status: <input checked="" type="checkbox"/>	Statistics / Biometrics / Epidemiology (1) Treatment / Event / Process / Action (2) Author: Redakteur 1
<ul style="list-style-type: none"> <li>There are currently no comments</li> <li>There are no comments for previous versions</li> </ul>	
<input checked="" type="checkbox"/> <b>Synonym(s):</b> Randomisation , Randomisierung	
<input type="checkbox"/> Insert description	

Figure 1: A term with two alternative definitions in the Onto-Builder

In the second stage the data model was expanded in order to enhance its expressive power. The focus here was on the realization of a concept-centered approach which provides for the improved representation of synonymy and multi-language contents. Other important improvements included the representation of semantic relations as well as different views and contexts (corresponding to, among other things, the divergent professional backgrounds of trial management personnel and institutions).

The Data Dictionary model and the Onto-Builder are designed to be domain-independent, allowing for implementation in other fields.

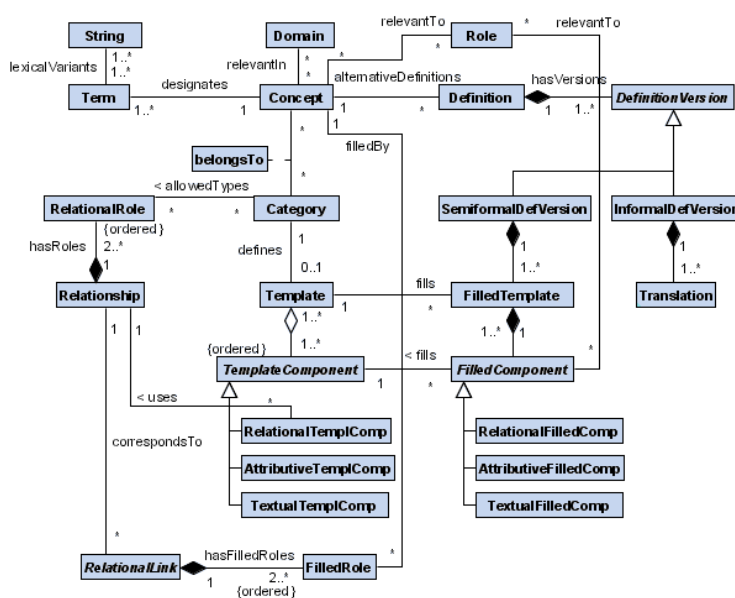
## Evaluation

The Onto-Builder is used by a wide range of user groups, including the Research Networks on Malignant Lymphoma and Heart Failure as well as the Coordination Center for Clinical Trials Leipzig (KKSLL). The experience and analysis of term definitions performed in connection with these groups are exploited in the implementation of the extended data model.

## Further Developments

Within the scope of the research group Onto-Med ([www.onto-med.de](http://www.onto-med.de)), the extended context-sensitive Data Dictionary model is being implemented, which will result in the introduction of the Onto-Builder Version 2. Moreover, existing projects with the Research Networks on Malignant Lymphoma, Heart Failure and Pediatric Oncology as well as the KKSLL are being expanded and new ones with other interested research networks are being actively developed.

The scientific goals of the Data Dictionary project include the ontological analysis of natural-language term definitions with the aim of compiling a domain-specific ontology for clinical trials. The goal is the formal, ontologically-founded representation of the semantically analyzed term definitions for reuse in other software applications in the clinical trial context.



### Project team:

Roland Muecke, [Ronald Speer](mailto:ronald.speer@imise.uni-leipzig.de)

### Project management:

Prof. Heinrich Herre, Prof. Markus Loeffler

### Contact address:

IMISE, University Leipzig, Haertelstr. 16-18,  
04107 Leipzig

Tel.: +49 (0) 341 97 16 105, Fax: +49 (0) 341 97 16130

E-mail: [ronald.speer@imise.uni-leipzig.de](mailto:ronald.speer@imise.uni-leipzig.de)