A web-based translation interface for the MeSH vocabulary

Lukas Geist

Semantic Technologies Team, ZB MED

ZB MED

Information Centre
for Life Sciences

www.zbmed.de

Table of Content

► Short introduction to Medical Subject Headings (MeSH)

► Introduction to MeSH translation process

► Technical aspects of the web-based TermCurator platform for MeSH translations

► Demo of TermCurator platform

► English-German bilingual MeSH XML

Medical Subject Headings (MeSH)

► Controlled vocabulary for life sciences

► Published by the National Library of Medicine (NLM) since 1960

► Used in NLM's biomedical database PubMed

► MeSH entries managed in a hierarchical tree structure ⟶ Thesaurus

Use cases of MeSH

▶ Indexing of publications in life sciences

▶ Refining searches of publications

▶ Named Entity Recognition (NER)

MeSH terms

> Animals
> Disease Outbreaks
> Measles Vaccine
> Measles virus
> Measles* / epidemiology
> Measles* / prevention & control
> Primates*

Fig. 1: MeSH terms indexed to the publication "Measles in non-human primates"
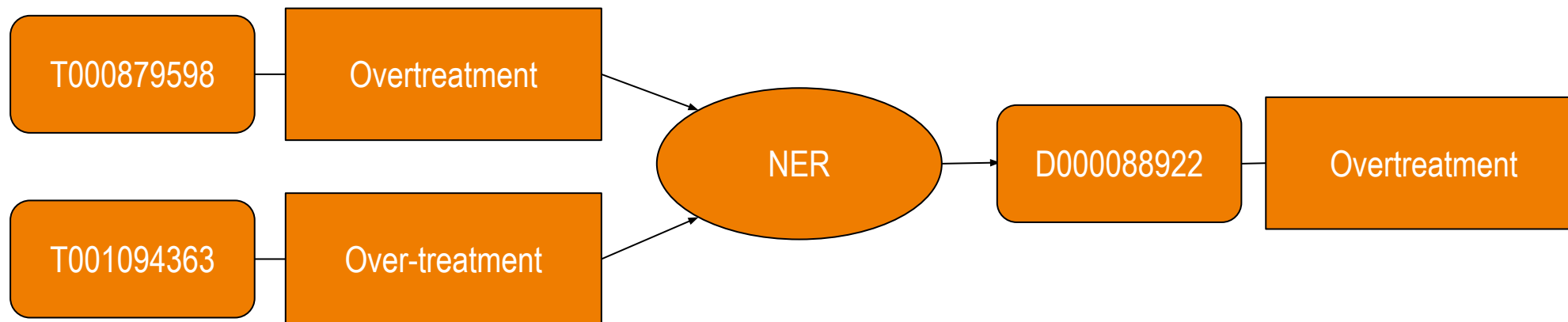
Use cases of MeSH - PubMed search
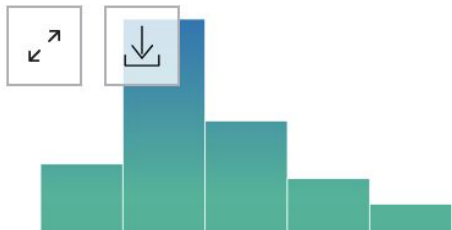
Fig. 2: Search query results on PubMed using the MeSH heading "Covid-19" with the subheading "diagnosis"

Use cases of MeSH - Named Entity Recognition (NER)

► Used by the PubMed search engine in which entry terms use their corresponding descriptors

► Used by Natural Language Processing (NLP) on life science documents

Using NER, neighbors of the entry term 'Overtreatment' will be mapped to the preferred descriptor term, which would treat a search using the alternative spelling 'Over-treatment' identically on the PubMed search engine

MeSH vocabulary types

▶ Four vocabulary types:
- – Headings (Descriptors)
  - • Concepts found in biomedical literature, translations are solely focused on headings
  - • "*Vaccine Efficacy, COVID-19 Testing, Dental Health Services*"
- – Subheadings (Qualifiers)
  - • Used in combination with headings to focus on aspects of an article
  - • "*Classification, Standards, Adverse Effects*"
- – Publication Characteristics
  - • Type of indexed publication
  - • "*Comment, Study, Review*"
- – Supplementary Concept Record (SCR)
  - • Created in a separate thesaurus to add new terms outside of the review process

Example MeSH heading entry 'Light Pollution'

Light Pollution

The inappropriate or excessive use of artificial lighting at night.

Year introduced: 2022

PubMed search builder options

Subheadings:

☐ adverse effects   ☐ ethics   ☐ prevention and control
☐ analysis   ☐ history   ☐ statistics and numerical data
☐ economics   ☐ legislation and jurisprudence

☐ Restrict to MeSH Major Topic.
☐ Do not include MeSH terms found below this term in the MeSH hierarchy.

Tree Number(s): N06.850.460.505

MeSH Unique ID: D000090465

Entry Terms:

- Pollution, Light
- Skyglow
- Artificial Night Sky Brightness
- Artificial Light at Night
- Artificial Night Lighting
- Artificial Night Lightings
- Lighting, Artificial Night
- Lightings, Artificial Night
- Night Lighting, Artificial
- Night Lightings, Artificial

Fig. 3: Snippet of MeSH header record for "Light Pollution"

MeSH heading

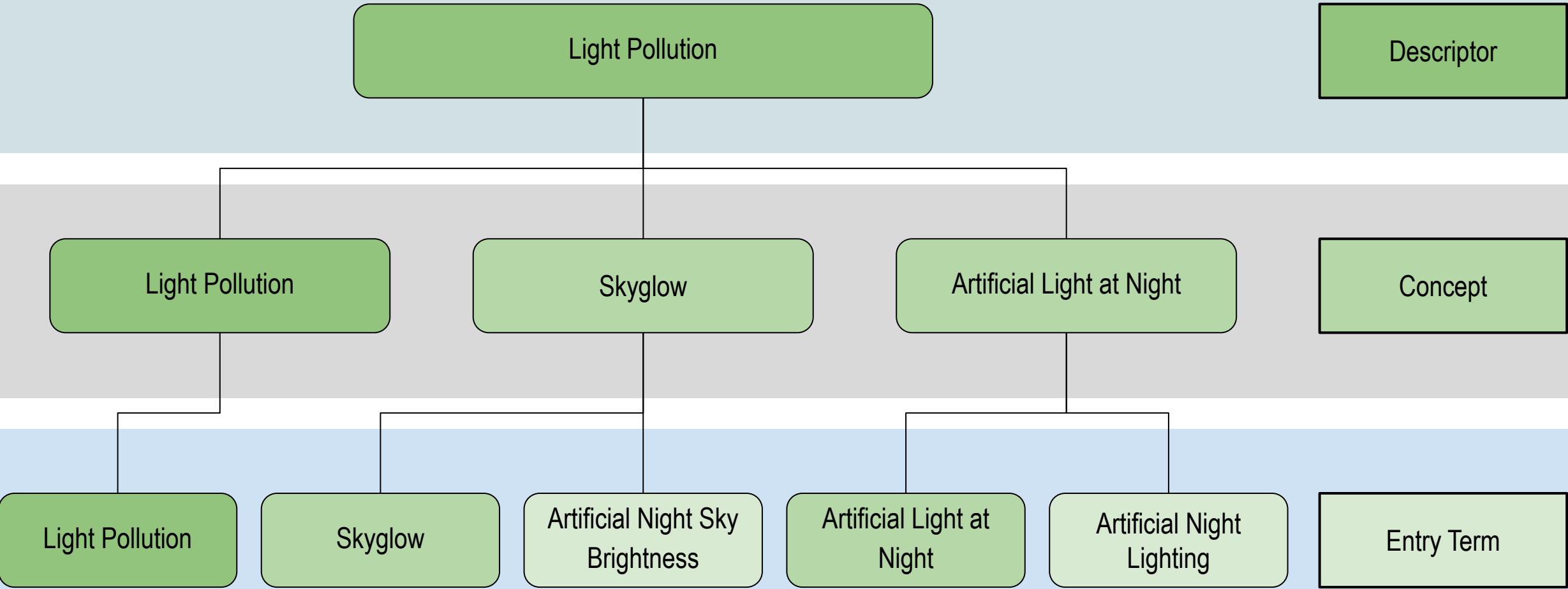ScopeNote - A description of the MeSH heading

Subheadings accompanying the MeSH heading

Descriptor ID

Entry terms listed under the descriptor 'light pollution', consisting of closely related concepts or synonyms

Descriptor hierarchy of heading 'Light Pollution'

Translation of MeSH entries

▶ MeSH translations requested by NLM

▶ Increases accessibility for researchers publishing works in other languages or searching for works in other languages

▶ Enables NER operations in other languages

▶ Currently bilingual versions with French, German, Norwegian, Spanish and Swedish are supported

German MeSH translation

► Translation of MeSH decentralized across multiple libraries and institutes

► Until 2018 translation into German performed by Federal Institute for Drugs and Medical Devices (BfArM)

► Since 2020 translation process taken over by ZB MED

► Development of a custom web-based MeSH translator: 'TermCurator'

► Publication of bilingual MeSH files in XML and RDF format

Overview of the translation pipeline

Awaiting annual release of new MeSH entries from NLM

Upload of bilingual MeSH entries in XML and RDF

Loading MeSH entries into curation pipeline with initial LLM supported translation by DeepL

Multi-level curation process and quality assurance

MeSH translation web application "TermCurator"

▶ User-guided development process

▶ German translations initialized using DeepL API to provide a starting point

▶ Multi-level curation process with quality assurance on user-submitted translations

▶ Easy and simple management of term assignment and splitting of translation tasks

▶ ScopeNotes and additional resources readily available to translators

Technical aspects of TermCurator platform

► Initially developed on Python Flask

► Later migrated to Vue3 framework Nuxt 3 using combination of JavaScript and TypeScript

► SQLite database built using Prisma ORM to store MeSH translation data

► Server deployment using Nginx

TermCurator Descriptor list

Search: [Enter Term or ID]   [FINALIZE SELECTED ROWS]

Select	ID	English Terms	German Terms	Interact
☐	D000087643	**Dorsolateral Prefrontal Cortex; DLPFC;** Prefrontal Cortex, Dorsolateral;	**Dorsolateraler präfrontaler Kortex; DLPFC;** Präfrontaler Kortex, dorsolateral;	Edit
☐	D000087662	**Centenarians;**	**Hundertjährige;**	Edit
☐	D000087742	**Octogenarians;**	**Oktogenarier;**	Edit
☐	D000087862	**Disinformation;** Fake News;	**Desinformation;** Fake News;	Edit
☐	D000087882	**Video-Assisted Techniques and Procedures;** Video Assisted Techniques;	**Videounterstützte Techniken und Verfahren;** Videounterstützte Techniken;	Edit
☐	D000087983	**Sound Recordings;**	**Tonaufzeichnungen;**	Edit
☐	D000088042	**Mentalization-Based Therapy;** Mentalization Based Therapy; Mentalization-Based Treatment;	**Auf Mentalisierung basierende Therapie;** Mentalisierungsbasierte Therapie; Auf Mentalisierung basierende Behandlung;	Edit
☐	D000088063	**Candida auris;** C auri;	**Candida auris;** C auri;	Edit

Fig. 4: TermCurator descriptor list interface for translators

TermCurator Translation Modal

Fig. 5: TermCurator translation modal interface for "Fermented Beverages"

ScopeNotes displaying the description of the descriptor heading for further information

The concept ID of the listed entry terms

Editable DeepL generated translation

Term types consisting of prefLabel (preferred concept/descriptor record) and synonyms

Additional synonyms may be added

Other helpful resources consisting of search queries to Google Scholar, ICD-10-GM, SWND and others

The German bilingual MeSH XML

▶ **Once all translations are approved,
a bilingual XML version is created**

▶ **German entry terms listed alongside English entries**

▶ **German terms receive their own unique identifier**

▶ **Currently 105,796 translated German MeSH entries**

▶ **Published by ZB MED on PUBLISSO**

```
▼ Concept [3]
  ▼ 0 {6}
      ConceptUI : M000746459
    ▶ ConceptName {1}
      ScopeNote : The inappropriate or excessive use of artificial lighting at night.
    ▶ ConceptRelationList {1}
    ▼ TermList {1}
      ▼ Term [4]
        ▼ 0 {8}
            TermUI : T001103613
            String : Light Pollution
          ▶ DateCreated {3}
          ▶ ThesaurusIDlist {1}
            _ConceptPreferredTermYN : Y
            _IsPermutedTermYN : N
            _LexicalTag : NON
            _RecordPreferredTermYN : Y
        ▶ 1 {6}
        ▼ 2 {7}
            TermUI : ger0095856
            String : Lichtverschmutzung
          ▶ DateCreated {3}
            _ConceptPreferredTermYN : N
            _IsPermutedTermYN : N
            _LexicalTag : NON
            _RecordPreferredTermYN : N
```

Fig. 6: XML tree snippet of the 2023 German MeSH XML publication

Publication of German bilingual MeSH XML

Deutscher MeSH - bilinguale Fassung, 2023 : XML (extensible markup language)

Fischer-Wagener, Brigitte ⬀ | Rebholz-Schuhmann, Dietrich ⬀ | Wollnik-Korn, Gabriele ⬀ | Albers, Miriam ⬀ | Baum, Roman ⬀ | Castro, Leyla Jael ⬀ | Geist, Lukas ⬀
Version 28.0 ; Stand: Mai 2023 .

Download ⬇
MeSH_Ger_2023_from_2023.03.21.xml.zip 18,93MB 🔓

|	
Titelzusatz weitere	XML (extensible markup language)
Verantwortlich	Mitwirkende beim Übersetzungsprozess: Brigitte Fischer-Wagener, Prof. Dr. Dietrich Rebholz-Schuhmann, Gabriele Wollnik-Korn ; Kuratierung: Brigitte Fischer-Wagener und Gabriele Wollnik-Korn ; Koordination: Dr. Miriam Albers ; Software-Entwicklung: Roman Baum, Dr. Leyla Jael Castro, Lukas Geist ; biochemische und molekularbiologische Begriffe wurden übersetzt mit freundlicher Unterstützung von Prof. Dr. Raimund Wagener, Zentrum für Biochemie, Medizinische Fakultät, Universität zu Köln ; ZB MED - Informationszentrum Lebenswissenschaften, U.S. National Library of Medicine, Deutsches Institut für Medizinische Dokumentation und Information

Fig. 7: Publication of the bilingual English-German MeSH XML of 2023 on PUBLISSO

Future work

► Translations phase of 2024 newly added MeSH terms

► Development of a new pipeline to handle changed terms according to their change types:
- Deleted entry terms
- Moved Concept ID or Descriptor ID
- Newly added synonyms to existing descriptors

► Processing of previous MeSH term changes

Thanks! Danke!
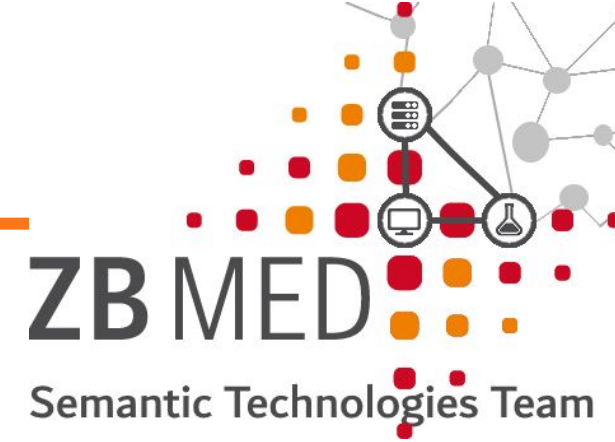
Lukas Geist
Semantic Technologies Team at ZB MED