

# TIB

GERMAN NATIONAL LIBRARY OF  
SCIENCE AND TECHNOLOGY



## Towards a Comprehensive Knowledge Organisation System for the Engineering Domain

E. Bernauer (WTI Frankfurt), M. Mehlberg (TIB Hanover),  
M. Runnwerth (TIB Hanover), G. Schmidt (WTI Frankfurt)  
*LIS'2015 Workshop, Colchester, 2015-09-02*



# Structure of the talk

- 1. Outline of the challenge to be tackled**
- 2. Objectives and time horizon**
- 3. Past and current activities (joint venture between TIB and WTI)**
- 4. Planned future activities**

# The German National Library of Science and Technology (TIB)

- **Areas of collection:** all areas of engineering, as well as architecture, chemistry, information technology, mathematics and physics
- **special collections:** research reports, norms and standards, patents, primary data, conference reports and literature from the regions of East Asia and Eastern Europe
- **special field of expertise:** hard-to-obtain, not commercially available grey literature in the library's subject areas
- **customers:** national and international research and industry
- **GetInfo:** portal for technical and scientific specialised and research information (currently 160 million data sets, including AV media, 3D-models and research data)

# Starting point: holdings related to Engineering

## Search in GetInfo:

- referenced for subject area 'Engineering' (all information resources, including dependent works like journal articles, articles in edited works): approx. **26 million data sets**
- referenced for subject area 'Engineering' (independent works like monographs, edited books, university papers): **approx. 2.8 million data sets**

*We would like to assist our customers interested in engineering-related topics better in finding the information they are looking for!*

Parental level

Artificial language

KOS-based semi-automatic query expansion as a tool to support information research

Search Term

Programming language

Functional programming language

Child Level

Object-oriented programming language

Show  
Stock

91. [Programming languages and systems : 8th Asian symposium, APLAS 2010, Shanghai, China, November 28 - December 1, 2010 ; proceedings](#)  
/ Kazunori Ueda. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2010
92. [Rewriting logic and its applications : 8th international workshop, WRLA 2010, held as a satellite event of ETAPS 2010, Paphos, Cyprus, March 20 - 21, 2010 ; revised selected papers](#)  
/ Peter Csaba Ólveczky. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2010
93. [Programming languages and systems : 8th Asian symposium, APLAS 2010, Shanghai, China, November 28 - December 1, 2010 ; proceedings](#)  
/ Kazunori Ueda. - Berlin [u.a.] : Springer, 2010
94. [Modellextraktion aus natürlichen Sprachen : eine Methode zur systematischen Erstellung von Domänenmodellen](#)  
/ Tom Gelhausen. - Karlsruhe : KIT Scientific Publishing, 2010
95. [Rewriting logic and its applications : 8th international workshop, WRLA 2010, held as a satellite event of ETAPS 2010, Paphos, Cyprus, March 20-21, 2010 ; revised selected papers](#)  
/ Peter Csaba Ólveczky. - Berlin [u.a.] : Springer, 2010
96. [JavaGI: a language with generalized interfaces](#)  
/ Stefan Wehr. - 2010
97. [Formal methods for quantitative aspects of programming languages : 10th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2010, Bertinoro, Italy, June 21 - 26, 2010 ; advanced lectures](#)  
/ Alessandro Aldini. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2010
98. [Languages and compilers for parallel computing : 22nd international workshop, LCPC 2009, Newark, DE, USA, October 8 - 10, 2009 ; revised selected papers](#)  
/ Guang R. Gao. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2010
99. [Formal methods for quantitative aspects of programming languages : 10th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2010, Bertinoro, Italy, June 21, 2010 ; advanced lectures](#)  
/ Alessandro Aldini. - Berlin [u.a.] : Springer, 2010
100. [Reversible computing : fundamentals, quantum computing, and applications](#)  
/ Alexis De Vos. - Weinheim : Wiley-VCH, 2010

Parental level

Object-oriented programming

Programming language

Search Term

Object-oriented programming language

Child Level

C++

Java

Ruby

Swift

- 11. [Formal verification of object-oriented software : international conference, FoVeOOS 2011, Turin, Italy, October 5 - 7, 2011 ; revised selected papers](#)  
/ Bernhard Beckert. - Heidelberg [u.a.] : Springer, 2012
- 12. [ECOOP 2012 - object-oriented programming : 26th European conference, Beijing, China, June 11 - 16, 2012 ; proceedings](#)  
/ James Noble. - Heidelberg [u.a.] : Springer, 2012
- 13. [Objective-C phrasebook : \[updated for ARC, OS X 10.7, and iOS 5\]](#)  
/ David Chisnall. - 2. ed., 1. printing. - Upper Saddle River, NJ [u.a.] : Addison-Wesley, c 2012
- 14. [ECOOP 2011 - object-oriented programming : 25th European conference, Lancaster, UK, July 25 - 29, 2011 ; proceedings](#)  
/ Mira Mezini. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2011
- 15. [Models in software engineering : workshops and symposia at MODELS 2010, Oslo, Norway, October 3 - 8, 2010 ; reports and revised selected papers](#)  
/ Juergen Dingel. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2011
- 16. [ECOOP 2011 - object-oriented programming : 25th European conference, Lancaster, UK, July 25-29, 2011 ; proceedings](#)  
/ Mira Mezini. - Heidelberg [u.a.] : Springer, 2011
- 17. [Models in software engineering : workshops and symposia at MODELS 2010, Oslo, Norway, October 3-8, 2010 ; reports and revised selected papers](#)  
/ Juergen Dingel. - Heidelberg [u.a.] : Springer, 2011
- 18. [Formal verification of object-oriented software : international conference, FoVeOOS 2010, Paris, France, June 28 - 30, 2010 ; revised selected papers](#)  
/ Bernhard Beckert. - [Online-Ausg.]. - Berlin [u.a.] : Springer, 2011
- 19. [Formal verification of object-oriented software : International Conference FoVeOOS 2010, Paris, France, June 28-30, 2010 ; revised selected papers](#)  
/ Bernhard Beckert. - Berlin [u.a.] : Springer, 2011
- 20. [JavaGI: a language with generalized interfaces](#)  
/ Stefan Wehr. - 2010

Show  
Stock



# WANTED!

- Our goal: support information needs of our customers with a special interest in engineering more effectively (e.g. with applications like the one shown above) and improve the quality of search results
- In order to achieve this goal, a **comprehensive and semantically rich Knowledge Organisation System for the engineering domain** is needed.
- Until now, there is no comprehensive KOS for this vast domain! (comparable with, for example, the Medical Subject Headings (MeSH) for the domain of medicine)

# A comprehensive KOS for engineering: desired features

- enables, for example, implementation of a KOS-based query expansion and search term recommendation as a new service in our search portals
- allows for a conceptual exploration of the knowledge structures (serendipity effects)
- readable by humans and machines (via user interfaces, APIs)
- modelled using web-based representation languages/ Semantic Web standards
- (possibly also allows for machine reasoning)
- [...]



# Approaching the goal as a joint venture

## Cooperation partners:

### **Information service provider WTI (Wissenschaftlich-technische Information) at Frankfurt**

- services offered to customers include e.g.
  - databases with bibliographic references of technical literature
  - tools for information search and analysis
  - abstracting service

and

### **German National Library of Science and Technology (TIB) at Hanover**

# Joint venture activities: objectives and timeframe

## **Short term (2015 / 2016) / Preparatory Phase:**

- further development of existing KOS for the engineering domain (particularly TEMA Thesaurus maintained at WTI)
- development of a light-weight ontology for a domain of manageable size (pilot project)

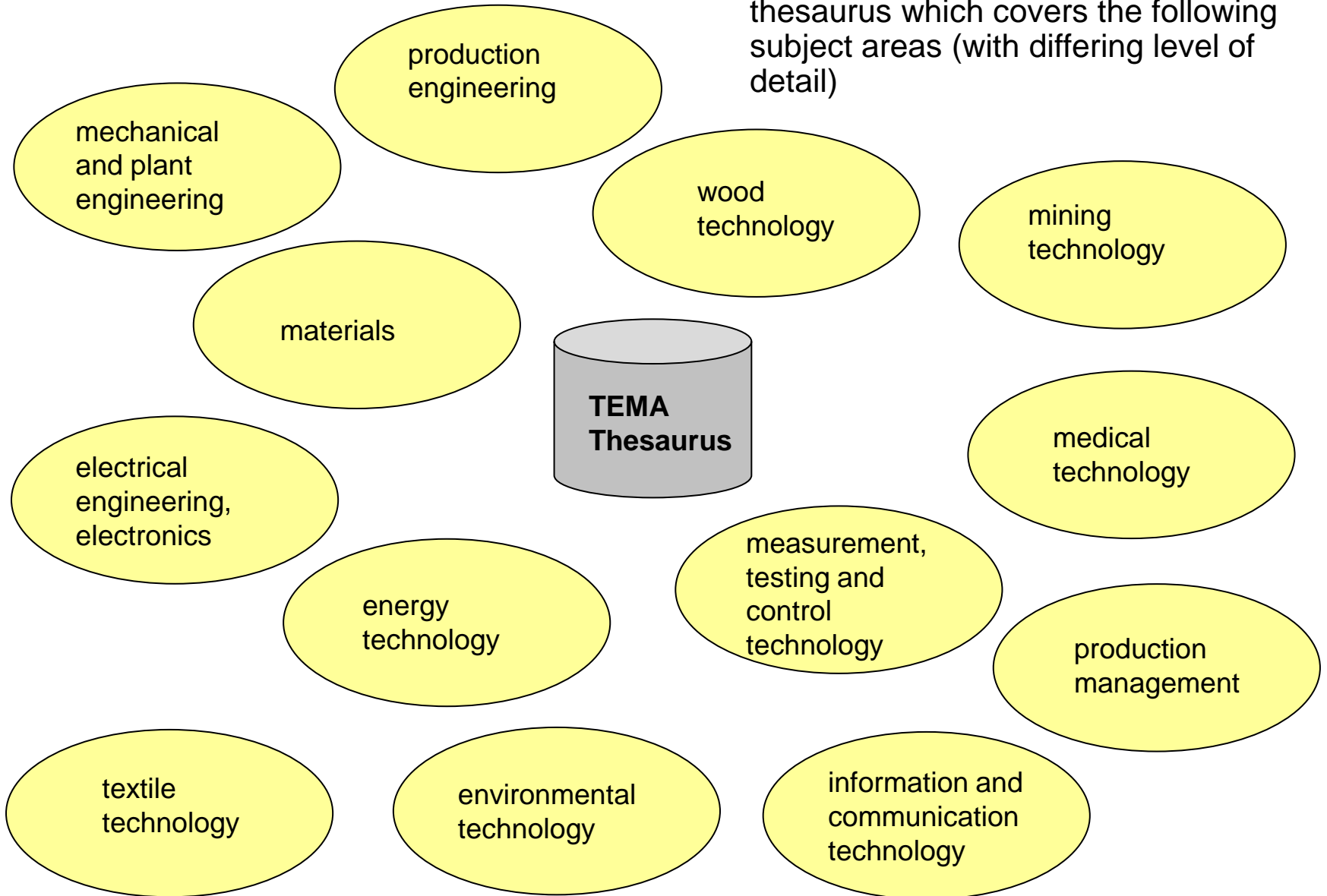
## **Medium to long term:**

- transform whole TEMA Thesaurus into a (light-weight) ontology (cooperation between ontology developers and subject specialists (and their communities) will be necessary)
- set up a prototype environment which allows for the alignment of several engineering-related KOSs and for a distributed and collaborative maintenance and actualisation of different KOSs

## **Thesaurus Engineering and Management (TEMA):**

- starting point and central element on our way towards a comprehensive KOS for the engineering domain
  - valuable and unique resource for engineering because of its size and comprehensiveness
  - comprises approx. 53.000 concepts and 167.000 terms (German and English)
  - was ported to a new management platform in 2014 (Averbis Terminology Platform)
  - part of WTI's product portfolio

The TEMA Thesaurus is a macro-thesaurus which covers the following subject areas (with differing level of detail)



# TEMA Thesaurus: Example entry with thesaural relations

008405 AD Brennelement  
 011720 AE fuel elements  
 707978 ND 3EE (Kraftwerke, Kernreaktoren)  
 707978 NE 3EE (Power plants, nuclear reactors)  
 227494 BD Brennstoffelement (Kerntechnik)  
 010991 BD Reaktorbrennelement  
 422137 BE nuclear fuel elements  
 009120 O Kernbrennstoff  
 131108 U abgebranntes Kernbrennelement  
 008413 U Brennstab (Brennelement)  
 129397 U Kernbrennstoff-Pellets  
 073942 V Kernbrennstoffaufbereitung  
 001710 V Kernreaktor  
 418974 V Magnox  
 010256 V Umhüllung von Brennelementen

AD	Preferred term (German)
AE	Preferred term (English)
ND	Notation of WTI 's subject classification (German)
NE	Notation of WTI 's subject classification (English)
BD	Synonym (German)
BE	Synonym (English)
O	Broader term
U	Narrower term
V	Related term

# Project activities in 2014 (Phase 1): overview

- transfer of the **TEMA Thesaurus** to a **new management platform** (Averbis Terminology Platform (ATP))
  - definition of technical and functional requirements for the new platform (Working Group “Usability”: thesaurus manager, subject specialists)
  - definition of workflows for thesaurus management
  - testing functionality and usability of new platform
- enrichment of TEMA Thesaurus with subject headings from the **subject section ‘engineering’ of the GND** (Integrated Authority File maintained by the German National Library and other libraries)
- establish maintenance of TEMA Thesaurus on the new platform as a part of everyday business

# Transfer of TEMA Thesaurus to a new platform (Averbis Terminology Platform)

**Hierarchy - TEMA 1.0 (xyz)**

Modern Classical

- ▣ abgebranntes Kernbrennelement
- ▣ keramischer Kernbrennstoff
- ▣ Spachtelmasse
- ▣ Kläranlage
- ▣ Knotenpunkt
- ▣ Koeffizient
- ▣ Zusammenstoß
- ▣ Konfiguration
- ▣ Konverter
- ▣ Kraftmaschine
- ▣ Krankenhauseinrichtung
- ▣ Kreisprozess
- ▣ Kühlanlage
- ▣ Kühlkörper
- ▣ Lagerstätte

Sibling limit  
25

**Search results (1)**

ConceptID	Label	Terminolo
008405	Brennelement	TEMA 1.0

1 - 1 von 1 Ergebnissen

---

**Concept details (TEMA 1.0 (xyz))**

008405 Brennelement

Terms/Synonyms (5) Relatives (5 / 7) Attributes (1) Xrefs (2) Linked Data

Pref	Label	Attributes	Type
	<input checked="" type="radio"/> Brennelement	011720 [translation (P)]	AD
	<input type="radio"/> Brennstoffelement (Kerntechnik)		BD
	<input checked="" type="radio"/> fuel elements		AE
	<input type="radio"/> nuclear fuel elements		BE
	<input type="radio"/> Reaktorbrennelement	422137 [translation (P)]	BD

## Some features of the new platform

- **xref** ('cross references'): allows mappings to other vocabularies and classifications (at the moment used for mapping groups of concepts to WTI's subject classification)
- **Export to SKOS** and other formats: publishing of thesaurus on the web using tools like SKOSMOS in principle possible
- (envisaged, but not yet implemented: LOD interface)



# Mapping GND – TEMA: some details

## **Objective:**

enrichment of TEMA Thesaurus with subject headings from the subject section “engineering” of the GND (Integrated Authority File maintained by the German National Library)

## **Method to reach the objective:**

- provision of the corresponding subject section (TIB)
- setting up of an automated mapping procedure between both vocabularies and generation of proposal lists (Averbis)
- examination and intellectual validation of proposals by subject specialists (WTI, TIB)
- Integration of new terms and concepts into TEMA
- establishment of a workflow ensuring a constant mapping from engineering-related GND subject headings onto TEMA (not yet implemented)

# Activities in 2014: Transfer to new platform and mapping GND – TEMA

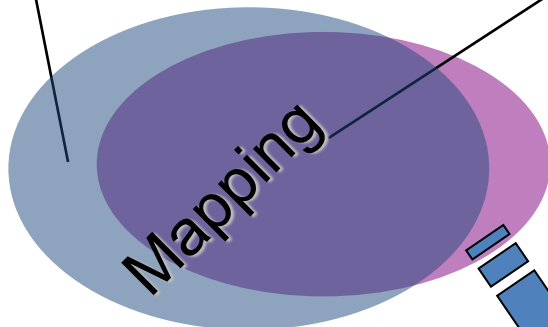
## Content

**Thesaurus  
Engineering and  
Management (TEMA)**

**166.014 Terms**  
(German and English)

**Subject Section  
„Engineering“ from the GND  
(Integrated Authority File)**

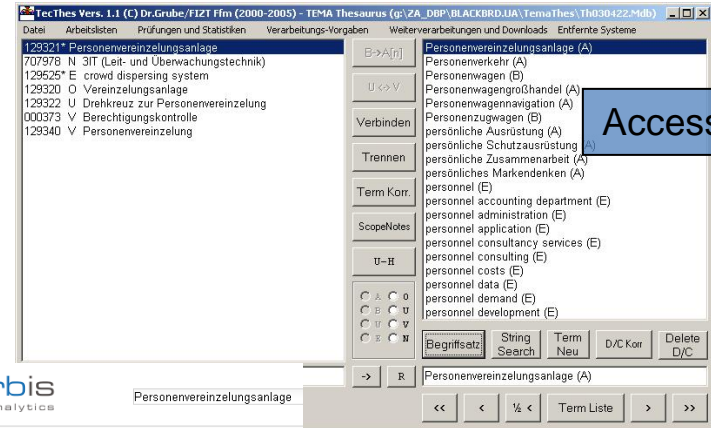
**133.596 Terms** (German  
and English)



**Result of mapping procedure: 11.419  
terms** from GND not contained in  
TEMA Thesaurus (8 % of the subject  
section)

**Intellectual validation and enrichment of  
TEMA Thesaurus: 5.550 new synonyms,  
1.769 new concepts**  
(4.100 proposal terms were rejected)

## Platform



**Porting of TEMA  
Thesaurus to new  
management  
platform**

**Averbis Terminology Platform (ATP)**  
(browser-based frontend)

# Project Activities in 2015 / 2016 (Phase 2).

## Further development of existing KOSs. **Term extraction.**

### **Objective:**

- testing and evaluating the benefits of using text-mining software for a continuous updating of existing KOSs

### **Method to reach the objective:**

- Provision of sample corpus comprising 200 machine-readable texts (doctoral dissertations, research reports) from 10 engineering subject areas (TIB, WTI)
- **Automated term extraction** (frequent nouns and noun phrases) and mapping onto GND and onto TEMA Thesaurus (tool: Averbis Extraction Platform)
- terms not contained in **GND** and/or **TEMA** are to be considered candidates for new subject headings (GND) or new terms (TEMA)
- examination and intellectual validation of proposals by subject specialists (WTI, TIB)
- evaluation by subject specialists: efficiency improvement? If yes: establishment of workflow

# Activities 2015 / 2016. Further development of existing KOSs. 'Internationalization' of TEMA Thesaurus.

## **Objective:**

- creation of an English-language version of TEMA Thesaurus (feasibility of mapping to controlled vocabularies like LCSH)

## **Method to reach the objective:**

- automated generation of English preferred terms from the set of already existing English synonyms by WTI (task completed)
- quality of translation and evaluation of frequency of usage for generated English preferred terms still have to be checked (analysis of a first random sample showed promising results)

- identification of thematically relevant thesauri and ontologies that could possibly lend themselves for a mapping / interlinking:
  - *Civil Engineering Database* (CEDB), maintained by ASCE (American Society of Civil Engineers)
  - *NASA Thesaurus*
  - *IEEE Thesaurus* (Institute of Electrical and Electronics Engineers, US)
  - *Inspec Thesaurus* (maintained by IET The Institution of Engineering and Technology, UK)
  - *Engineering Index Thesaurus* (Elsevier)
  - micro-thesauri developed at the Library of the University of Applied Sciences at Wildau (e.g. Thesaurus Informatik)
  - [...]

- crucial task: enriching and restructuring of TEMA Thesaurus heading for a semantically richer KOS / ontology
  - e.g. complete hierarchical structure for effective use of KOS in information retrieval; inventory of typed relations deemed useful for information seekers
- building-up of know-how and competence in the area of formal and conceptual ontology engineering

## **Pilot project:**

- development of a (light-weight) ontology for a knowledge domain of manageable size (as a testbed)
- chosen domain: **E-mobility**

# Planned future activities (and challenges)

## Objective:

(in view of the fact that Engineering is a vast domain with overlaps to thematically related domains, e.g. medical technology < > medicine, and that knowledge in the field is growing steadily and subject to constant change):

- expand information environment by aligning various complementary KOSs
- establish linkages/mappings to concepts of other KOSs ensuring cross-domain and interdisciplinary interoperability
- maintain link to vocabularies already used for indexing purposes

## Method to reach the objective:

- choose approach which allows for (semantic) interoperability
  - „non-equivalent pairs model“ or „backbone model“?
  - promising approach: ‘**Linked Open Ontology Cloud**’  
(Semantic Computing Research Group, Aalto University)

# „Linked Open Ontology Cloud“ (Frosterus et al. 2015) as a reference architecture for distributed ontology management?

- central idea: „provide a shared cross-domain ontology for data annotations based on a set of interlinked domain ontologies“
- emphasis on collaboration: **coherent aggregated ontology** comprises a set of **component ontologies** that are maintained by different domain communities
- integration and merging of component ontologies is achieved through aligning them to a **General Upper Ontology**
- ontologies considered both as individual entities but also as integral parts of the cloud (forming together a harmonized global ontology: all concepts part of a single hierarchy)
- LOOC differs from the usual approach of linking independently developed ontologies



# Use case: The KOKO ontology cloud

- developed during FinnONTO research project (2003-2012), now part of national ontology service Finto
- ontology development coordinated by the National Library of Finland
- center of **KOKO Cloud: General Finnish Ontology** (YSO) serves as the General Upper Ontology (provides top-level-concepts and concepts needed in many domains)
- **16 domain ontologies** extend YSO 's concepts into more detailed subconcept hierarchies (domain ontologies were developed from traditional thesauri and are maintained by the experts responsible for the original thesauri)
- KOSs constituting the LOOC classified as **light-weight RDF ontologies** and are represented using SKOS
  - „[...] a light-weight ontology is a hierarchy of concepts, with subsumption, partitive, and associative relations like a traditional thesaurus. [...] They can typically be represented using RDFS, simple OWL constructs, or SKOS.“

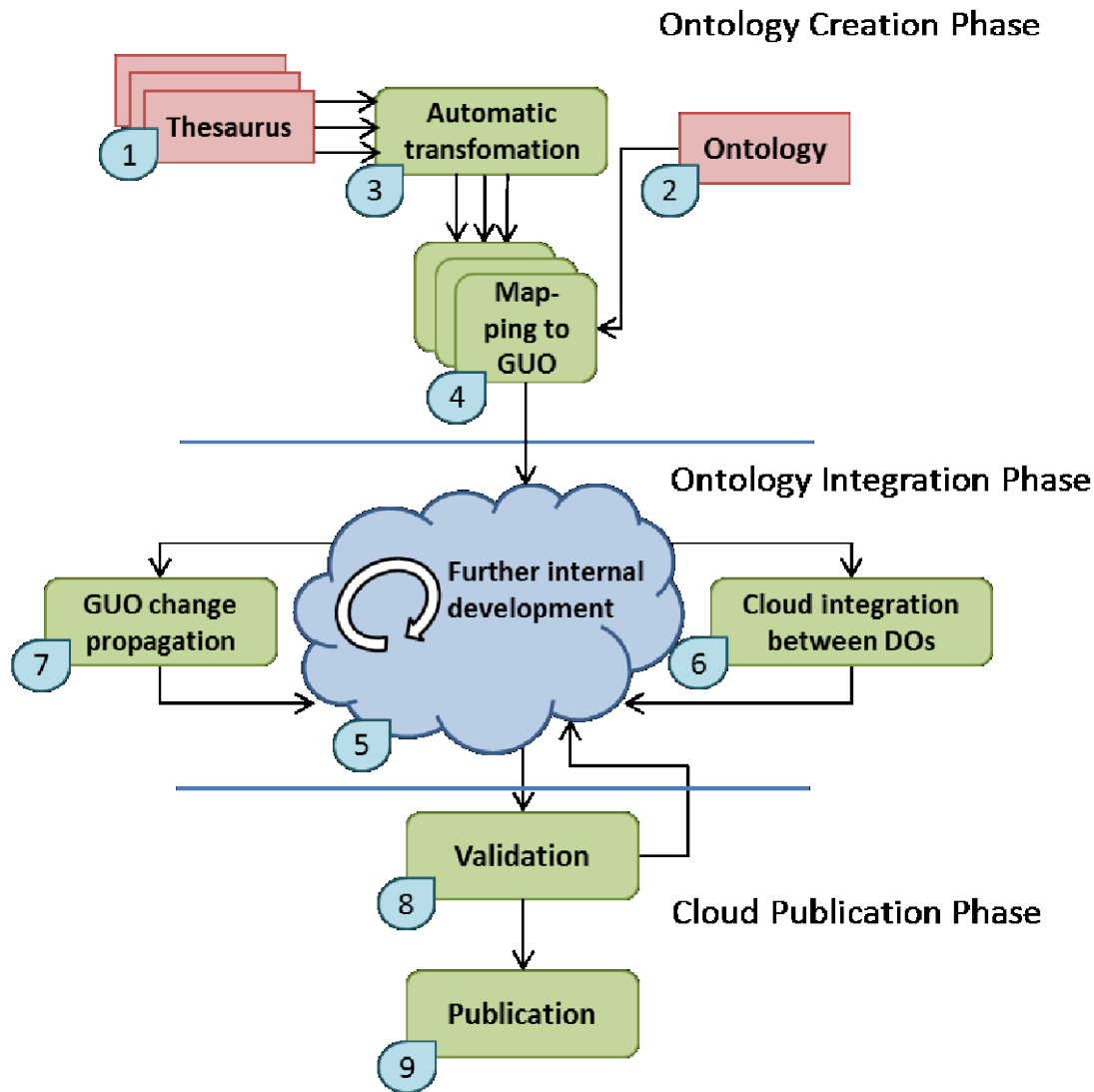
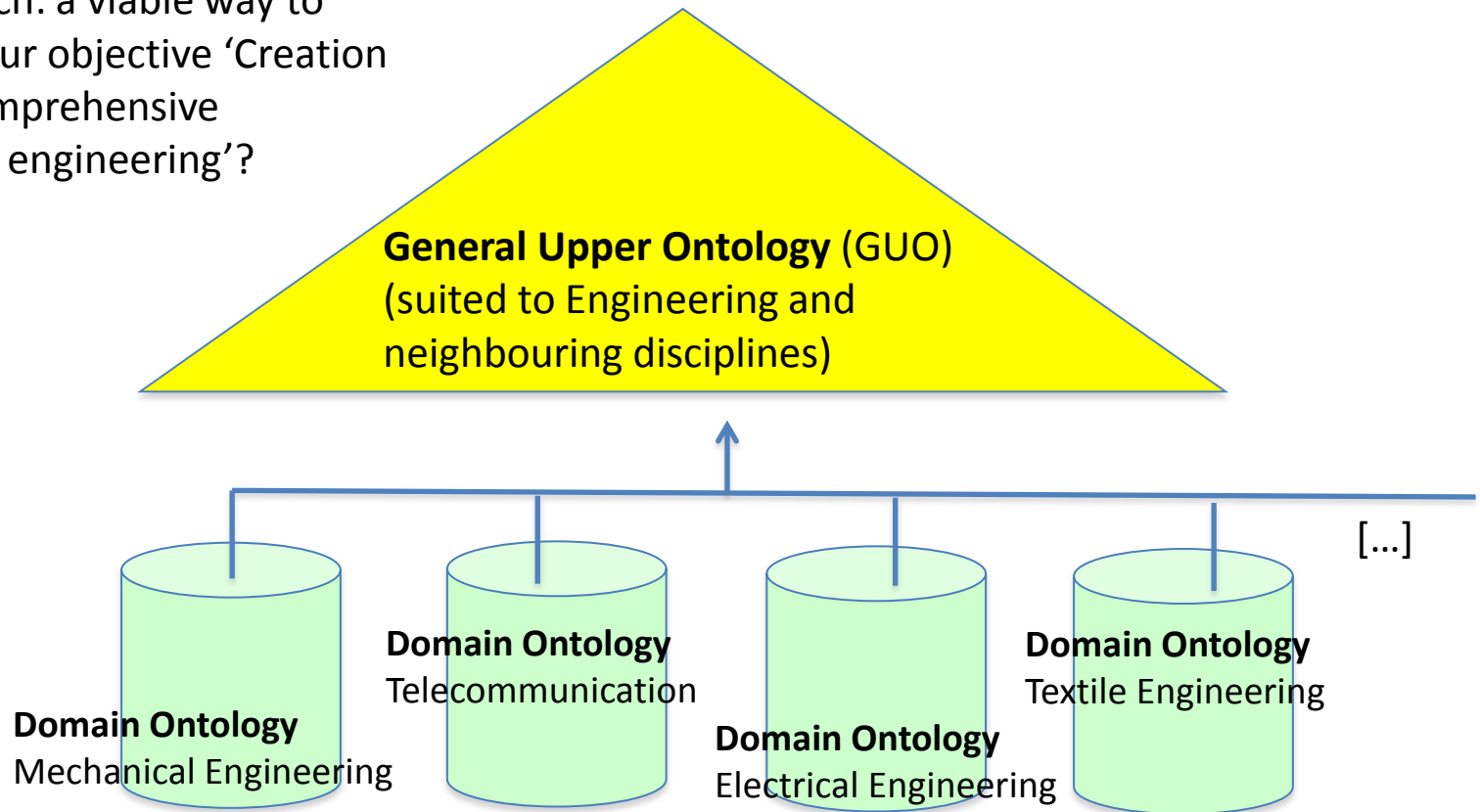


Fig. 1 The model of Linked Ontology Cloud formation and management

'Linked Open Ontology Cloud'-  
approach: a viable way to  
reach our objective 'Creation  
of a comprehensive  
KOS for engineering'?



# Summary

- We have taken first steps on our way towards a comprehensive KOS for the engineering domain (particularly further development of TEMA Thesaurus)
- Regarding the decision on how to proceed, we are at a crossroads.
- Because of the vastness of the domain, a distributed and collaborative approach to KOS / ontology development and managing appears attractive to us (like the „Linked Open Ontology Cloud“-approach).
- We are looking for institutions interested in adopting such an approach.
- Much work lies ahead.

# References

Matias Frosterus, Jouni Tuominen, Sini Pessala and Eero Hyvönen: *Linked Open Ontology Cloud - Managing a System of Interlinked Cross-domain Light-weight Ontologies*.

March, 2015. Submitted.

<http://www.seco.tkk.fi/publications/submitted/frosterus-et-al-loo-cloud-model.pdf>

# TIB

GERMAN NATIONAL LIBRARY OF  
SCIENCE AND TECHNOLOGY



Thank you for your attention!

