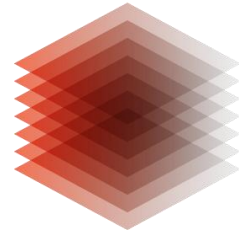


LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

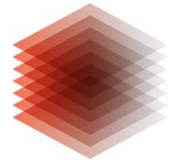


TIB

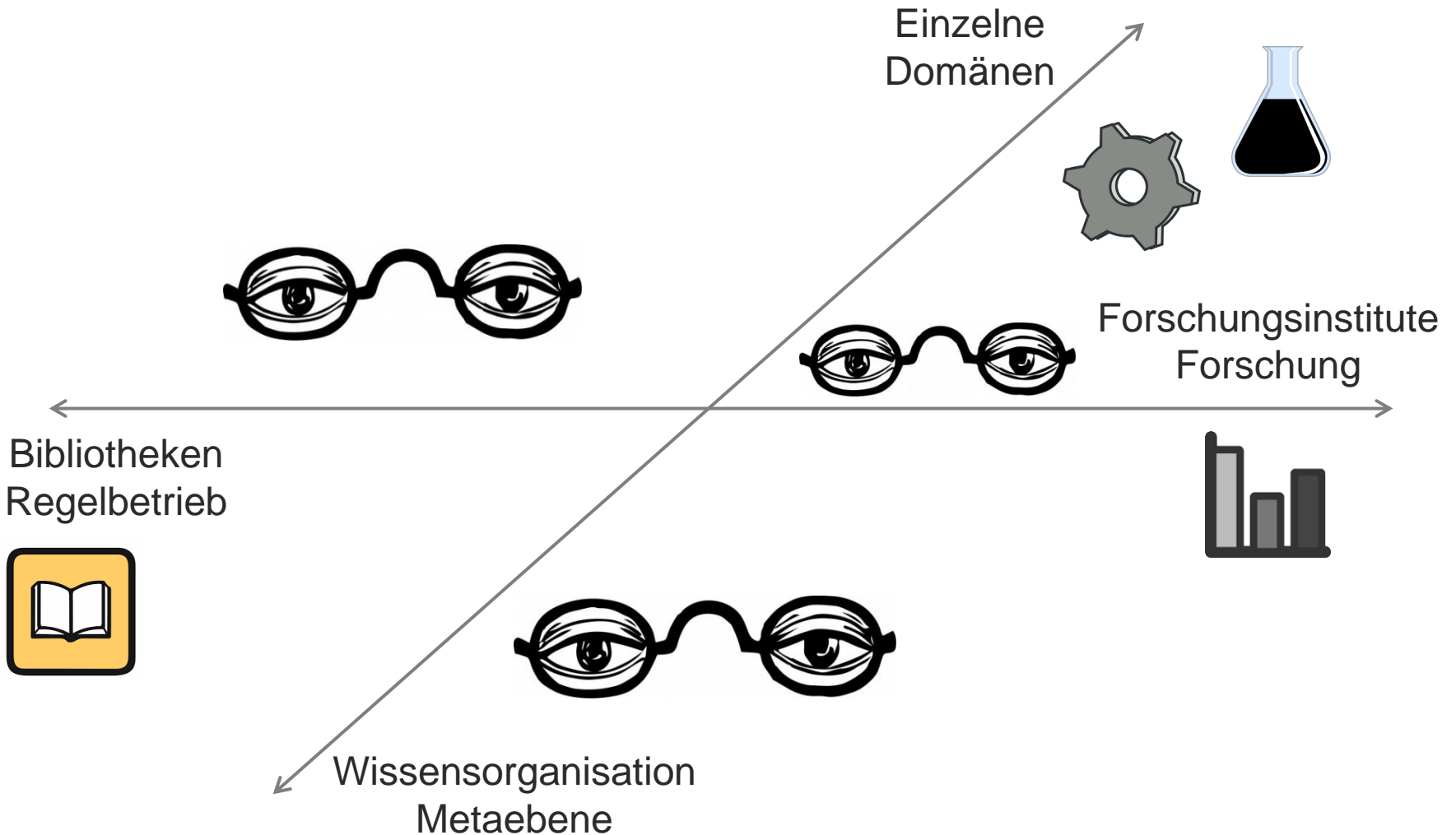
Community-getriebene Entwicklung von Sacherschließungswerkzeugen rund um die GND

Von der Normdatei zum Thesaurus, und vom Thesaurus zur Ontologie – strukturelle Stolpersteine

Dr. Anna Kasprzik
Berlin, 12. Juni 2018
107. Bibliothekartag – „offen&vernetzt“



In welchem Bereich bewegen wir uns? Verschiedene Perspektiven



Community-getriebene Entwicklung von Sacherschließungswerkzeugen rund um die GND

Anna Kasprzik // TIB

Jürgen Kett // DNB

Mila Runnwerth // TIB

Das GND-Entwicklungsprogramm



Hintergrund: "Mungstener Brücke" von Niels Wörheide (CC BY-SA 3.0)

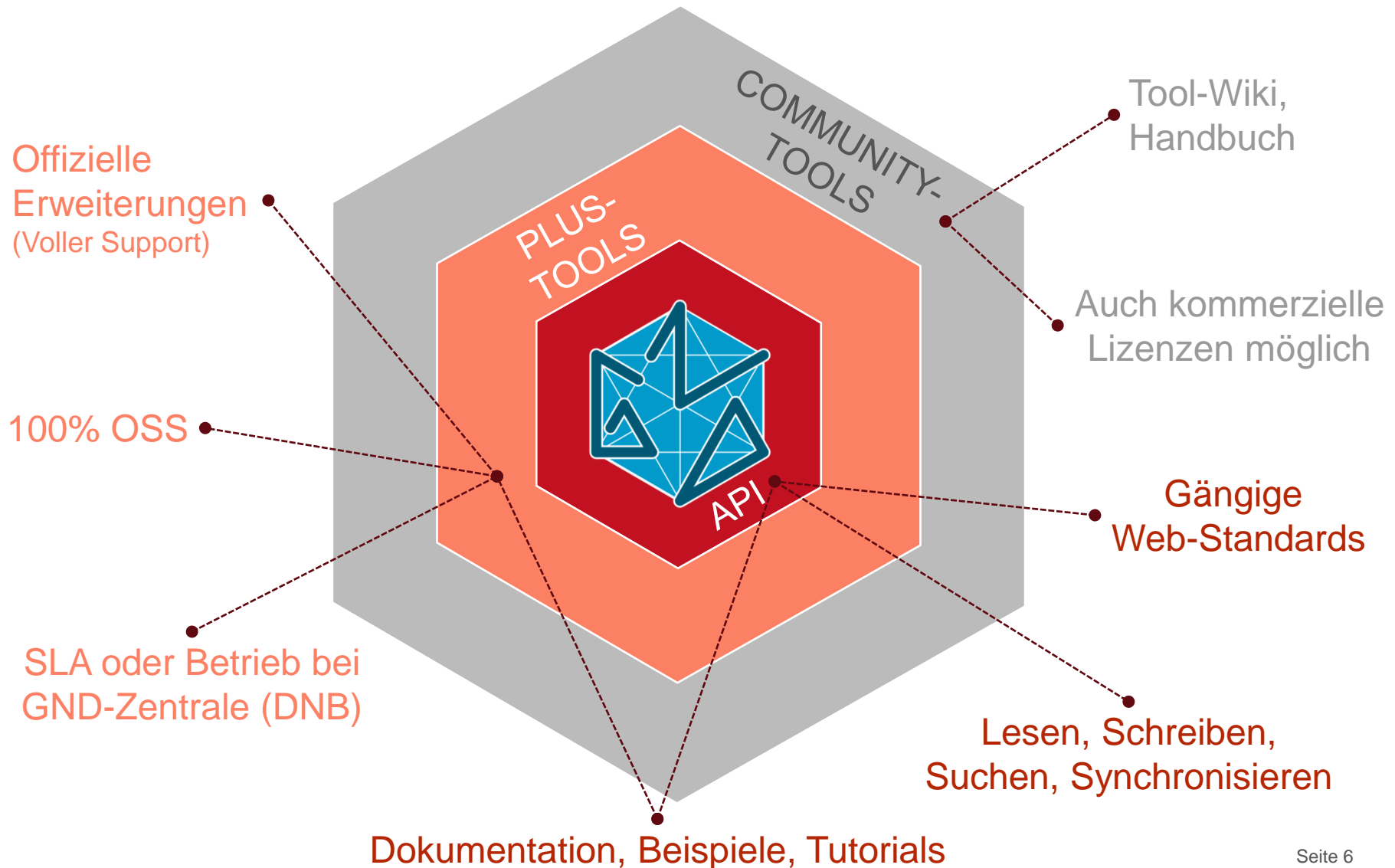
Das GND-Entwicklungsprogramm

Organisatorische, fachliche und technische Modernisierung

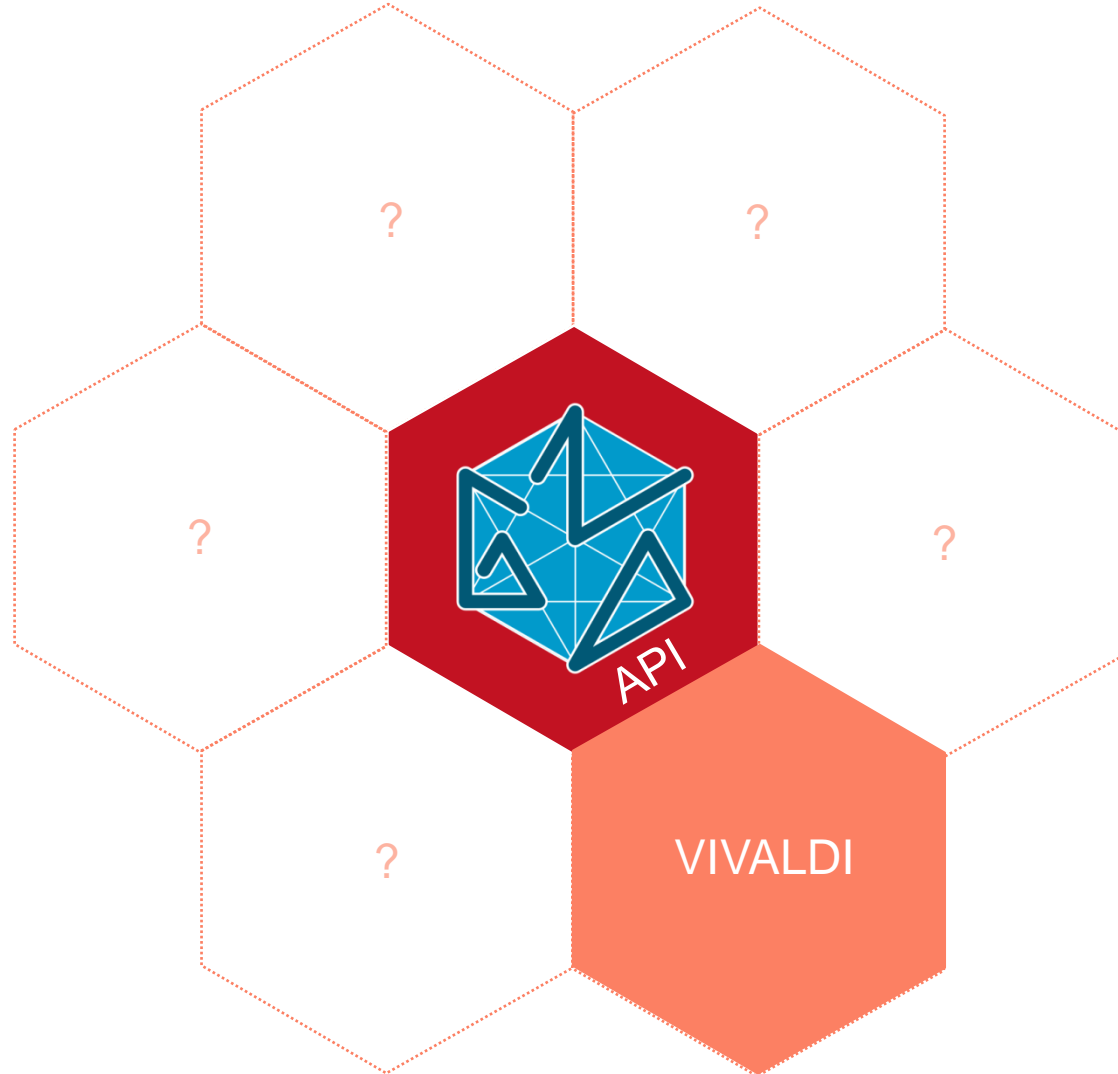


Öffnung der GND für andere Sparten und Communities
... und für neue Kooperationsformen ...

GND als Entwicklungsplattform



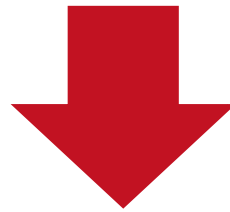
Entwicklung von Sacherschließungswerkzeugen als ideales Referenzprojekt



Ausgangssituation

Zwei bestimmende Faktoren für die Sacherschließung von heute:

- Die Qualität der Sacherschließung in Titeldatensätzen ist heterogen; aufgrund der Publikationsflut verschärft sich das Ungleichgewicht
- Die GND weist in ihrer Struktur Unschärfen und Lücken auf



Die Pflege von Titeldaten und Normdaten
muss erleichtert werden, sowohl für
Fachreferenten als auch für Maschinen!



Normdatenpflege: Schwerer Stand der GND

Der Mehrwert der GND als Wissensorganisationssystem ist (außerhalb der Bibliothekswelt) schwer zu vermitteln:

- Sie kommt naturgemäß **im Hintergrund** zum Einsatz, für Nutzer nicht sichtbar / selten explizit genannt
- Sie bildet teilweise Konzepte (insbesondere gebräuchliche englische Terme) aus der **aktuellen Forschung nicht ab**
- Nutzer ziehen **kaum einen Mehrwert aus den GND-Relationen**, diese sind weder niederschwellig zugänglich noch intuitiv dokumentiert
- Die **Relationen in der GND sind nicht der gleichen strengen Kontrolle unterworfen** wie die redaktionelle Betreuung der Schlagworte selbst



Was wird im Fachreferatsalltag bei der Sacherschließung gebraucht?

Titeldatenebene	Normdatenebene
<ul style="list-style-type: none">• Unterstützung bei der Schlagwortrecherche• intelligente Vorschläge aus verschiedenen Quellen• direkte Übernahme der gefundenen GND-Schlagwörter in das eigene System	<ul style="list-style-type: none">• einfache Schlagwortansetzung• einfache Korrektur von GND-Datensätzen• generell: einfache Pflege der GND für den eigenen Fachbereich / die eigenen Zwecke



Beide Bereiche erfordern eine gute ÜBERSICHT über das bereits bestehende Terminologiegut!

Einschub: Was heißt „einfach“?

Einfach heißt auch und vor allem:

- Möglichst **wenig Client-Wechsel**, Minimierung des Arbeitsaufwands („Anzahl Clicks“)
- Möglichst **Live-Anbindung** aller (Verbund- und Lokal-!) Systeme an die jeweiligen Master, so dass jede Änderung direkt sichtbar wird und **keine Synchronisierungsprobleme** entstehen
- Automatisierte Prozesse im Hintergrund dürfen die Arbeit **nicht behindern** und es sollte **keine Kenntnis über deren Funktionsweise** vonnöten sein



Existierende Systeme: WebGND (Eurospider)

4006851-1



Biologie

Sachbegriff

S

GND-Nummer : 4006851-1

Untergliederung : [Allgemeinbegriff \[saz\]](#)

Systematik : 23.1a [Biologie allgemein](#)

DDC-Notation : 570

Quelle : M

Varianten : [Allgemeine Biologie](#)

Oberbegriffe : [Biowissenschaften](#)

Weitere Aktionen

[Weitere Beziehungen](#)

[Hierarchie anzeigen](#)

[MarcXML anzeigen](#)

[Marc21 anzeigen](#)

[Marc21-Text anzeigen](#)

[Ansetzung kopieren](#)

[In DNB öffnen](#)

[In NEBIS suchen](#)

[In GVK suchen](#)

[Wikidata](#)

[Wikipedia \(de\)](#)

WebGND: Ein paar Kritikpunkte

- Autovervollständigungsliste wird auf 12 Vorschläge **gekapp**t
- Bei Vorschlagsannahme wird **Begriff** durch **ID** ersetzt → keine Nachkorrektur
- „**Varianten**“ in der Anzeige ist irritierend → „**Synonyme**“
- Trennung zwischen **Datensatz** und „**Weitere Beziehungen**“ ist irritierend
- wünschenswert: Anzeige des **kompletten Teilbaumes**, bzw. Parametrisierung der Anzeige zur Steuerung der Anzeigetiefe/-breite
- wünschenswert: **Anzeige des Oberbegriffstyps** (Relationscode)
- keine **schnelle Wiederverwendbarkeit** (Übernahme)

Wissenschaft

Naturwissenschaften

Biowissenschaften

Biologie

Weitere Aktionen

[Diesen Begriff anzeigen](#)

[Weitere Beziehungen](#)

Konstitutionsbiologie

Kosmobiologie

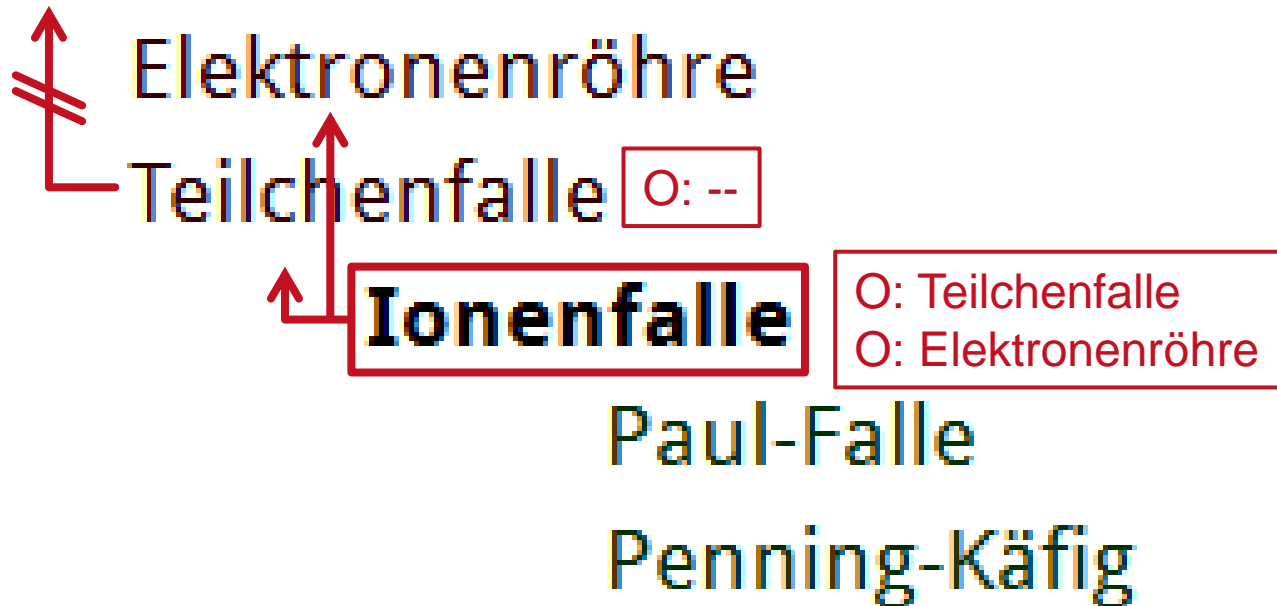
... und 15 weitere

WebGND: Probleme mit Polyhierarchie



WebGND: Probleme mit Polyhierarchie II

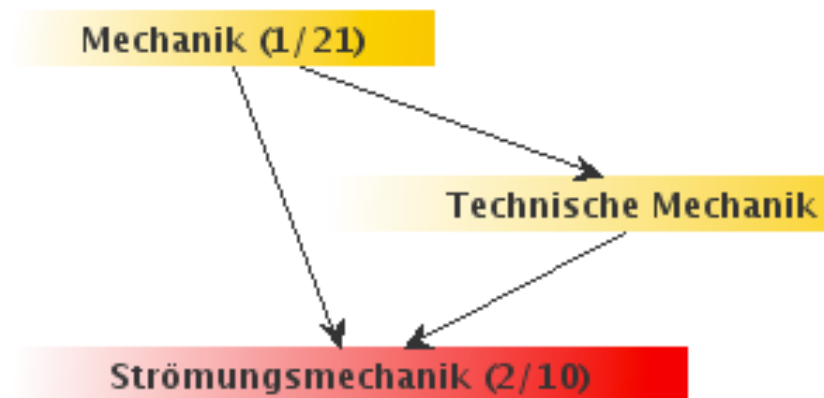
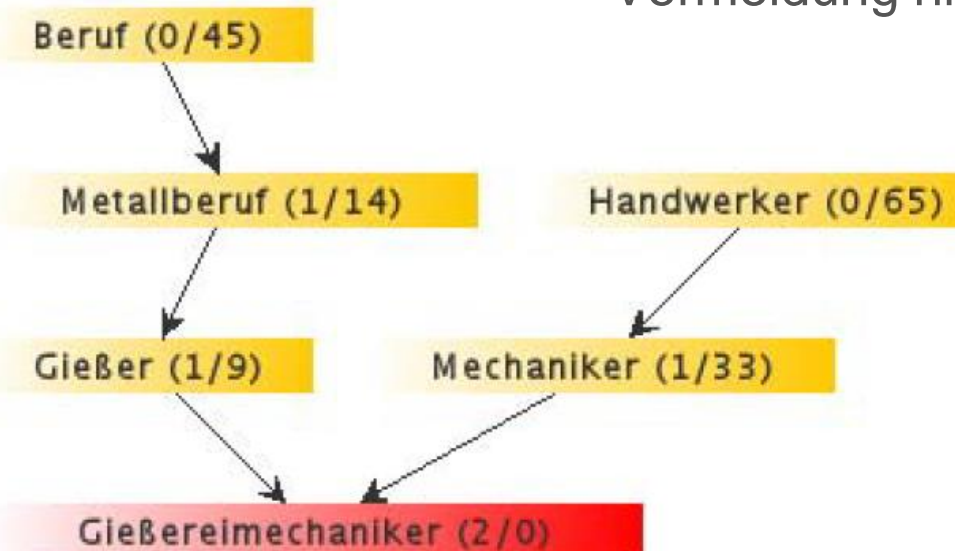
Elektronisches Bauelement



Der SWD-Viewer [Maas 2010] – Funktionalitäten

- Erweiterung der trunkierten Suche in der WinIBW
→ erweiterte Boolesche Suche, reguläre Ausdrücke
- Visualisierung mehrerer Schlagworte gleichzeitig

– Vermeidung hierarchiebezogener Fehler:



- mausgesteuerte Navigationsmöglichkeit entlang der Hierarchien
- Ansetzung in Analogie durch Copy&Paste

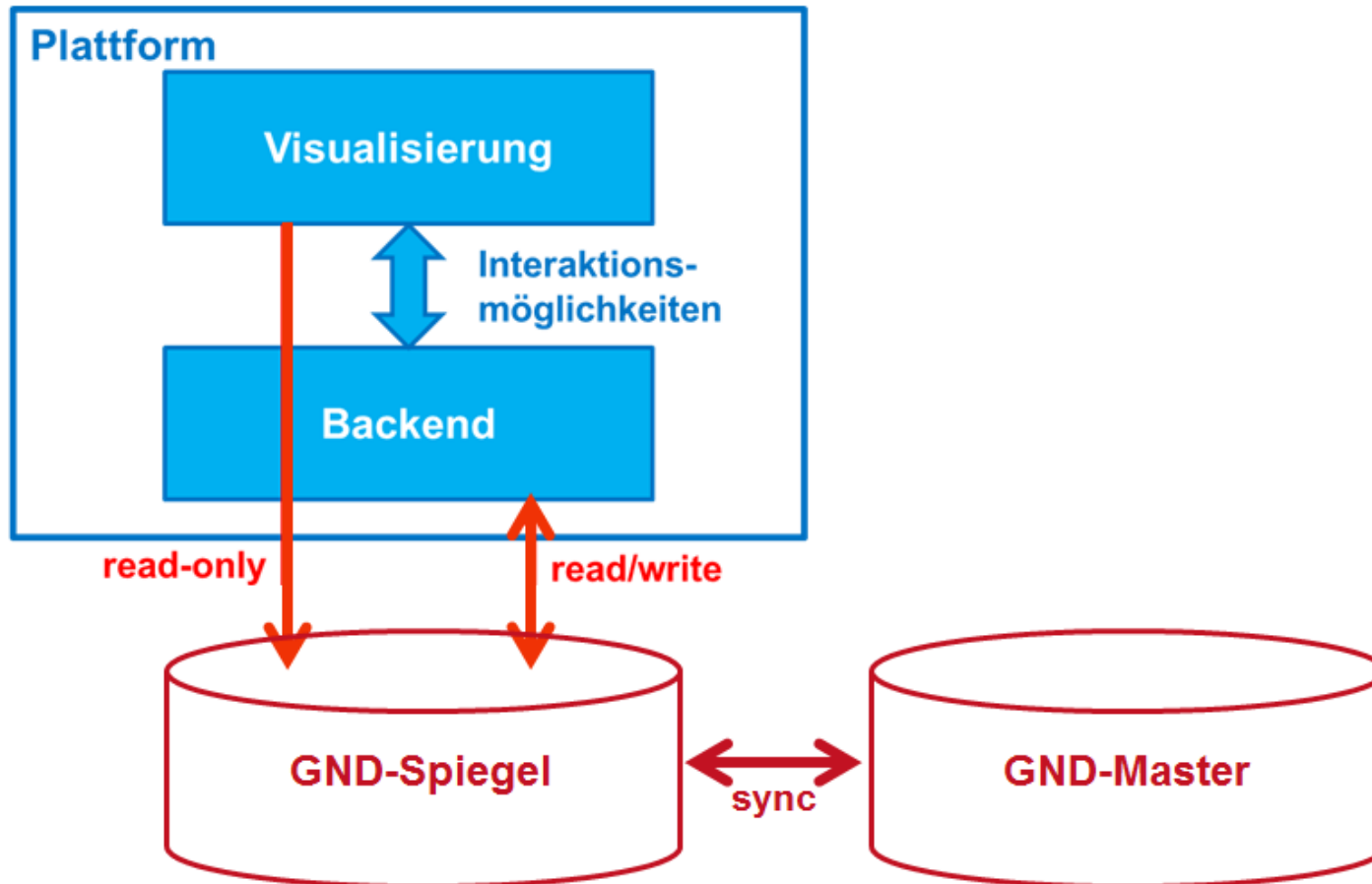
Vorschlag für eine Assistenzplattform

- **Visualisierung**: graphbasiert (mehrere Schlagworte im **Kontext**), individuell anpassbar, direkte Übernahme bei Navigation
- direkt angeschlossene **Editierfunktion**
- **intelligente Vorschlagshilfen** (z.B. auch Wikipedia-Anbindung)
- **Strukturprüfungen** beim Anlegen und auf Anstoß
- **webbasiert**
- **modular**
- ergonomischer Workflow



dadurch **lückenlos** in gängige
Bibliothekssysteme integrierbar

Vorschlag für eine Assistenzplattform – Architektur



Vorschlag für eine Assistenzplattform – potentielle Mehrwerte I

Fachreferat:

- im Fachreferat produktiv genutzte Anwendung mit **ergonomischem, effizientem Gesamtworkflow**
- durch zukünftige Live-Anbindung an den GND-Master sind **Änderungen sofort verfügbar**



Bibliotheksgemeinschaft:

- durch automatisierbare Strukturprüfung und Live-Anbindung **verbessert sich die Gesamtstruktur der GND** direkt und kontinuierlich (zusammenhängend, dubletten- und zyklfrei, ...)
- **generische Schablone** (inkl. organisatorischen und technischen Aspekten, APIs, Standards, etc.) für **community-getriebene Entwicklung von Open-Source-Werkzeugen**

Vorschlag für eine Assistenzplattform – potentielle Mehrwerte II

Endnutzer:

- z.B. attraktive und intuitive Visualisierungen für einen **explorativen Einstieg** in die GND und assoziierte Bestände/Fachgebiete, mit Wikipedia-Anbindung o.Ä.
- verbesserte Erschlossenheit → **hochqualitativere Suchergebnisse**

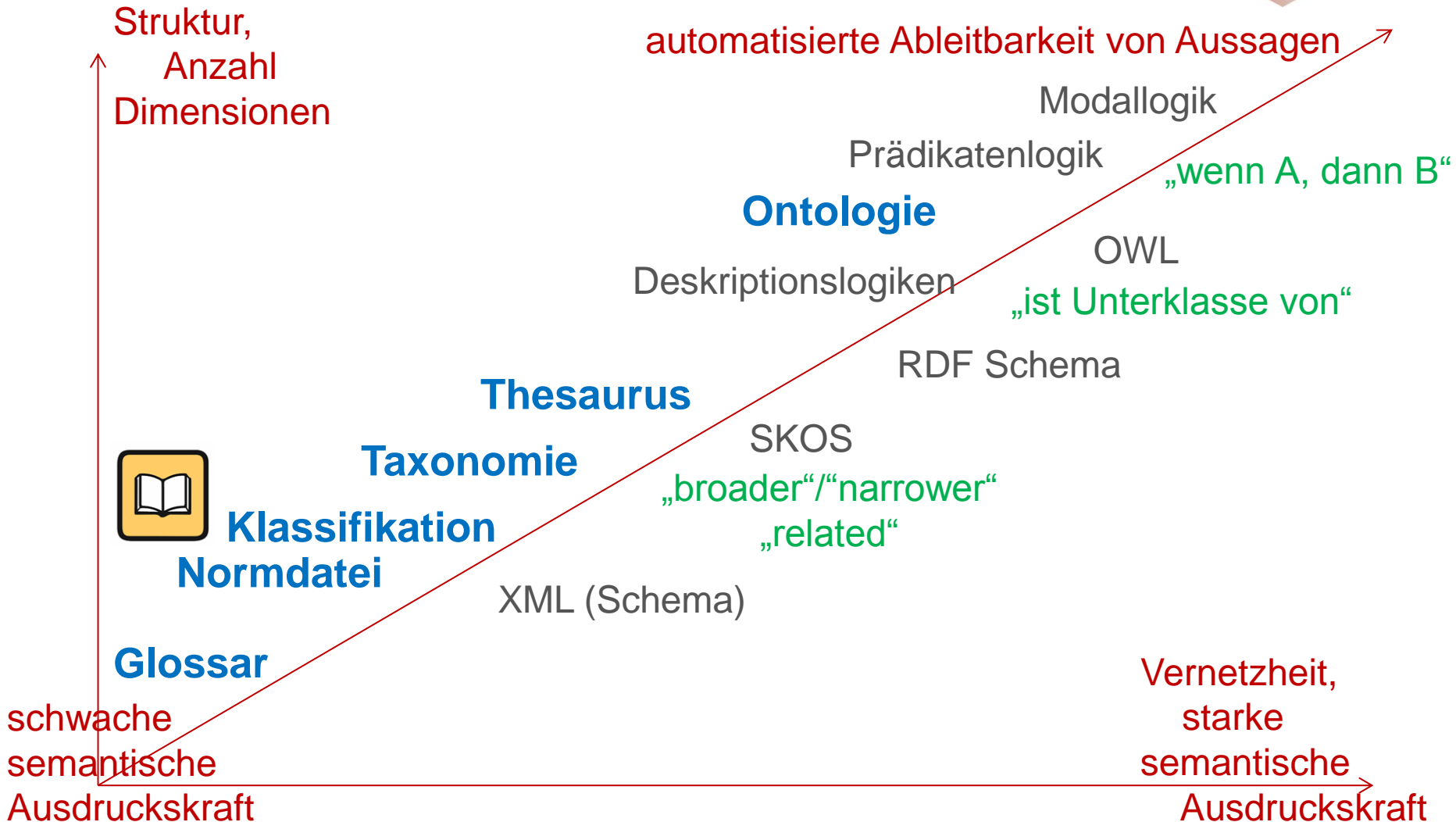


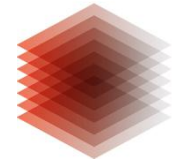
Knowledge Engineers:

- hochqualitatives, **nachnutzbares Wissensorganisationssystem** (für Fachcommunities außerhalb des Bibliothekswesens)
- **aussagekräftige Trainingsdaten** für Machine-Learning-Anwendungen

**Von der Normdatei zum Thesaurus,
und vom Thesaurus zur Ontologie –
strukturelle Stolpersteine**

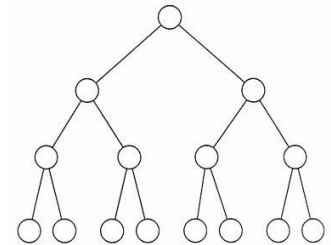
Wissensorganisationssysteme – das Spektrum

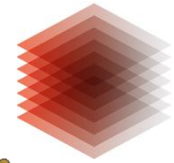




Glossar, Taxonomie, Thesaurus, Ontologie

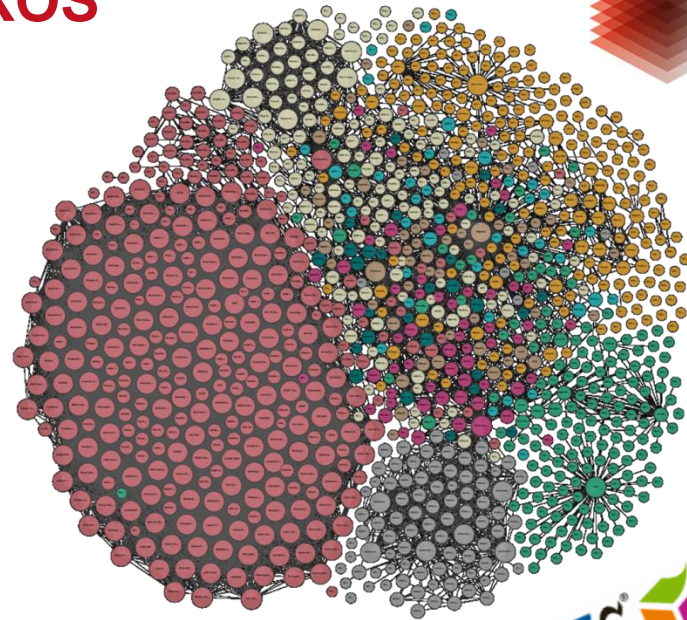
- **Glossar**: Liste von Termen, ggf. mit Definitionen und weiteren Zusatzinformationen [**linear**]
- **Taxonomie**: Konzepte sind hierarchisch angeordnet mit Hilfe von Ober-/Unterbegriffsbeziehungen [**Baumstruktur**]
- **Thesaurus**: (idealerweise) Trennung in Konzepte und Terme, zusätzliche Beziehungen wie Synonymie und andere semantische Relationen [**Baum oder Graph, Polyhierarchie möglich**]
- **Ontologie**: Klassen, Relationen & Eigenschaften plus Regeln, die logisches Schließen erlauben [**Mengentheorie**]





Thesauri im Semantic Web: SKOS

- Simple Knowledge
Organisation System
- Basis: RDF (Tripel)
Alle Aussagen nehmen die Form
Subjekt – Prädikat – Objekt ein
- Einfach, flexibel, erweiterbar, maschinenlesbar
→ Veröffentlichung im World Wide Web
→ ermöglicht Austausch zwischen Softwareanwendungen
- Pragmatischer Kompromiss zur stark logisch
ausgerichteten Web Ontology Language OWL



Semantische Beziehungen in SKOS

- skos:broader

```
ex:Säugetiere rdf:type skos:Concept;  
skos:prefLabel „Säugetiere“@dt;  
skos:broader ex:Tiere.
```

- skos:narrower

```
ex:Tiere rdf:type skos:Concept;  
skos:prefLabel „Tiere“@dt;  
skos:narrower ex:Säugetiere.
```

- skos:related

```
ex:Vögel rdf:type skos:Concept;  
skos:prefLabel „Vögel“@dt  
skos:related ex:Ornithologie.
```

SKOS: Label

- skos:prefLabel

```
ex:Tiere rdf:type skos:Concept;  
skos:prefLabel „Tiere“@dt;  
skos:prefLabel „animals“@en.
```

- skos:altLabel

```
ex:Tiere rdf:type skos:Concept;  
skos:prefLabel „animals“@en;  
skos:altLabel „creatures“@en.
```

- skos:hiddenLabel

```
ex:Tiere rdf:type skos:Concept;  
skos:prefLabel „Tiere“@dt;  
skos:hiddenLabel „Tire“@dt.
```

SKOS-XL: Thesauri gemäß ISO 25964

SKOS-XL: Label als eigene Klasse!

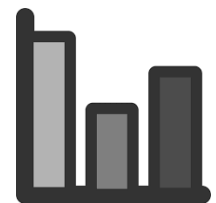
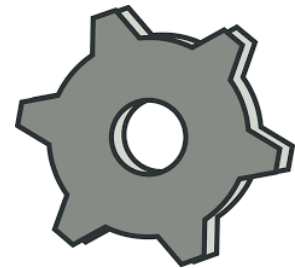
Label können so mit Metadaten versehen werden.

```
ex:C_Tiere rdf:type skos:Concept;  
           skosxl:prefLabel ex:Tiere;  
           skosxl:prefLabel ex:Animals.  
  
ex:Tiere rdf:type skosxl:Label;  
         skosxl:literalForm „Tiere“@dt.  
  
ex:Animals rdf:type skosxl:Label;  
           skosxl:literalForm „animals“@en.
```

Der TEMA-Thesaurus: „Technik und Management“

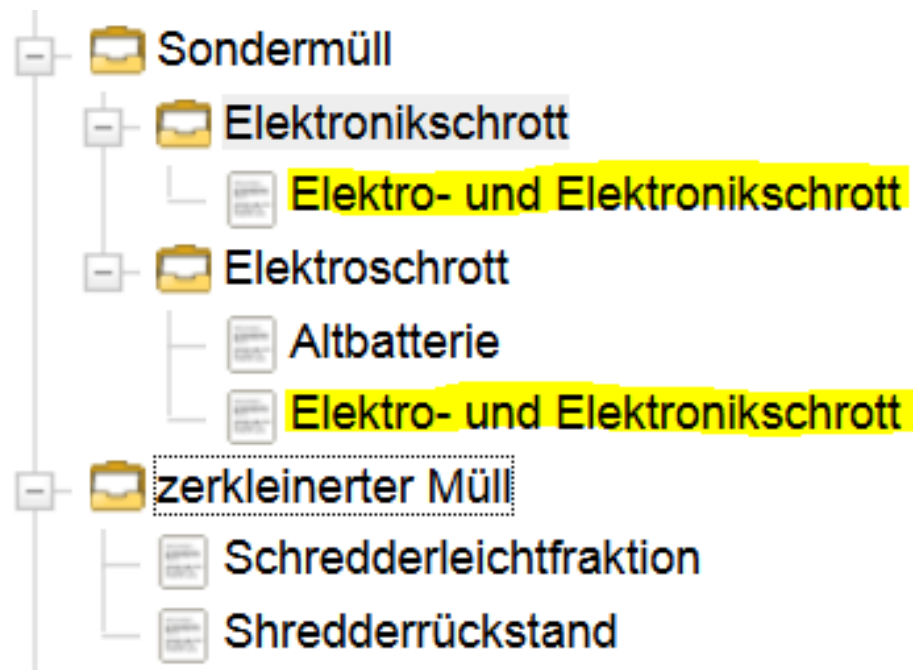


- rund 57.000 Konzepte, 195.000 Terme
- bilingual: deutsch und englisch
- Fachgebiete:
 - Ingenieurwissenschaftliche Grundlagen
 - Maschinenbau, Fertigungstechnik
 - Elektrotechnik, Elektronik
 - Mess- und Regelungstechnik
 - Betriebsführung und -organisation
 - Umwelttechnik
 - Informationstechnik
 - Fahrzeugbau
 - Textiltechnik
 - Bergbau
 - Werkstoffe
 - Energietechnik
 - Medizintechnik
- Historisch zusammengewachsen aus sechs Print-Thesauri (!)
- Momentan gepflegt auf der Averbis Terminology Platform (ATP)

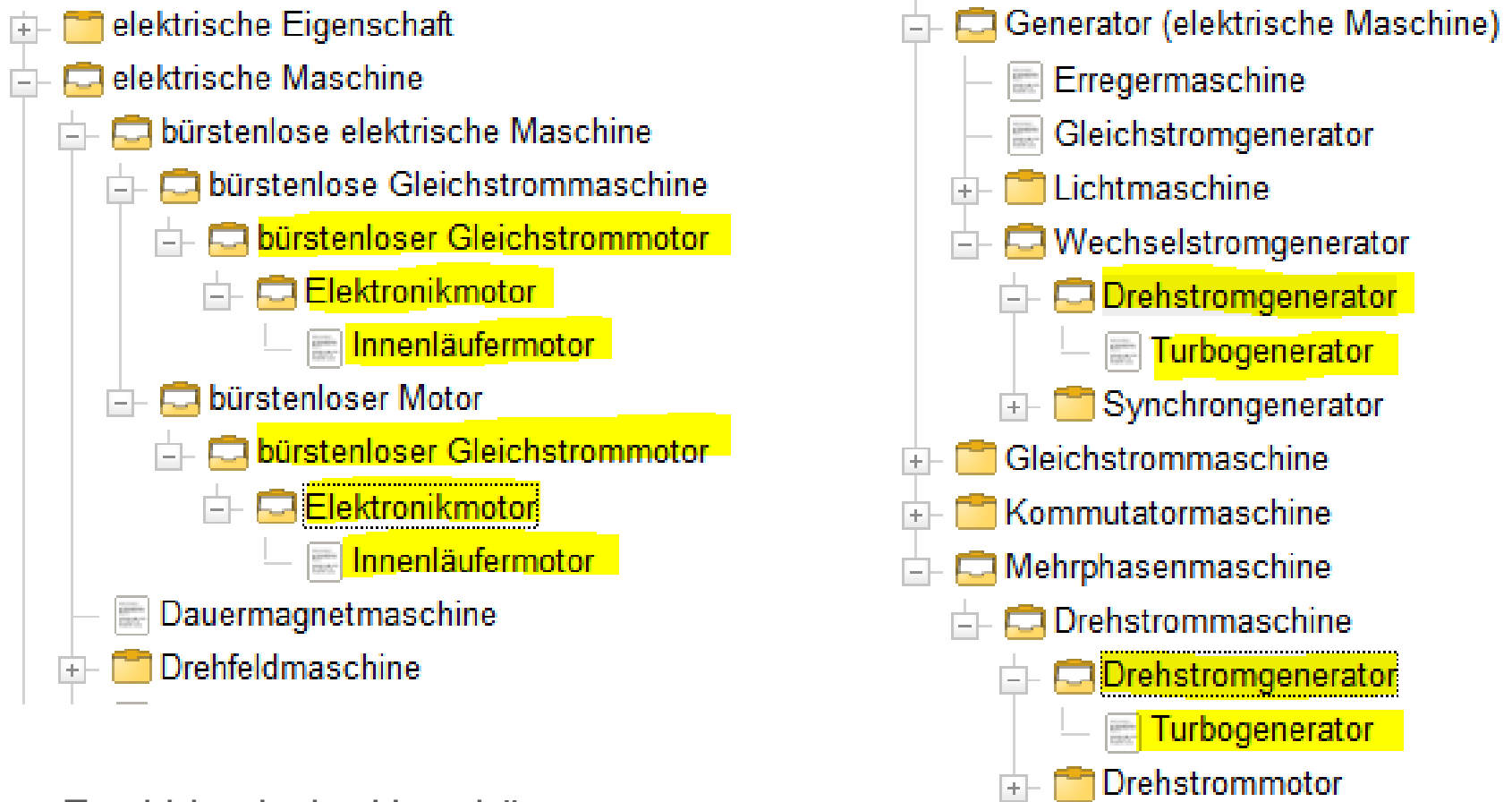


Kuriositäten und Inkonsistenzen – „Müll“

„Tektonische Verwerfungen“ durch händische Zusammenführung



Anzeige des TEMA-Thesaurus auf der ATP: Baumartefakte durch Polyhierarchie



→ Zwei identische Unterbäume

Anzeige zeigt auch: Fehlende „Dachstruktur“

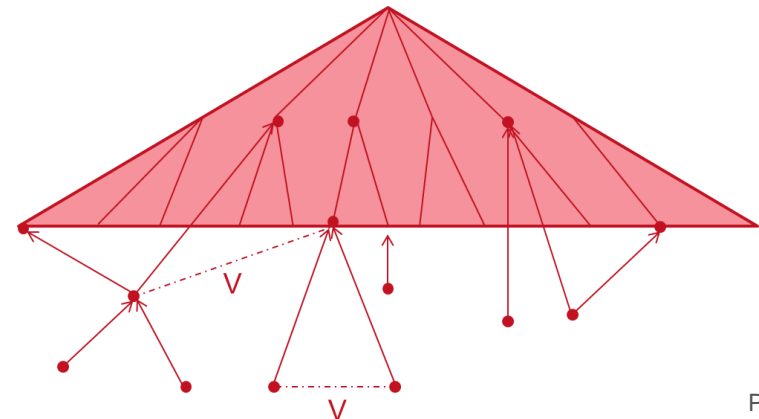
- TEMA 1.0
 - + Abschirmung
 - + Abtastung
 - + Abweichung
 - + Achse
 - + Adhäsion
 - + Aktivierung
 - + Analogschaltung
 - + Analysator
 - + Analyse (allgemein)
 - + Anpassung
 - + Anschliff
 - + Antenne
 - + Antriebstechnik
 - + Antwort
 - + Anwendung (allgemein)
 - + Anwendungsgebiet
 - + Apparat

Oberste Ebene,
unter einem künstlichen
Mutterknoten („TEMA 1.0“):

Alle Begriffe, die im Thesaurus
keinen Oberbegriff haben

Deshalb:

**Wir brauchen eine „Dachstruktur“,
die einen graduellen Einstieg vom
Allgemeinen zum Spezifischen erlaubt!**



Projekt: Eine „Dachstruktur“ für TEMA

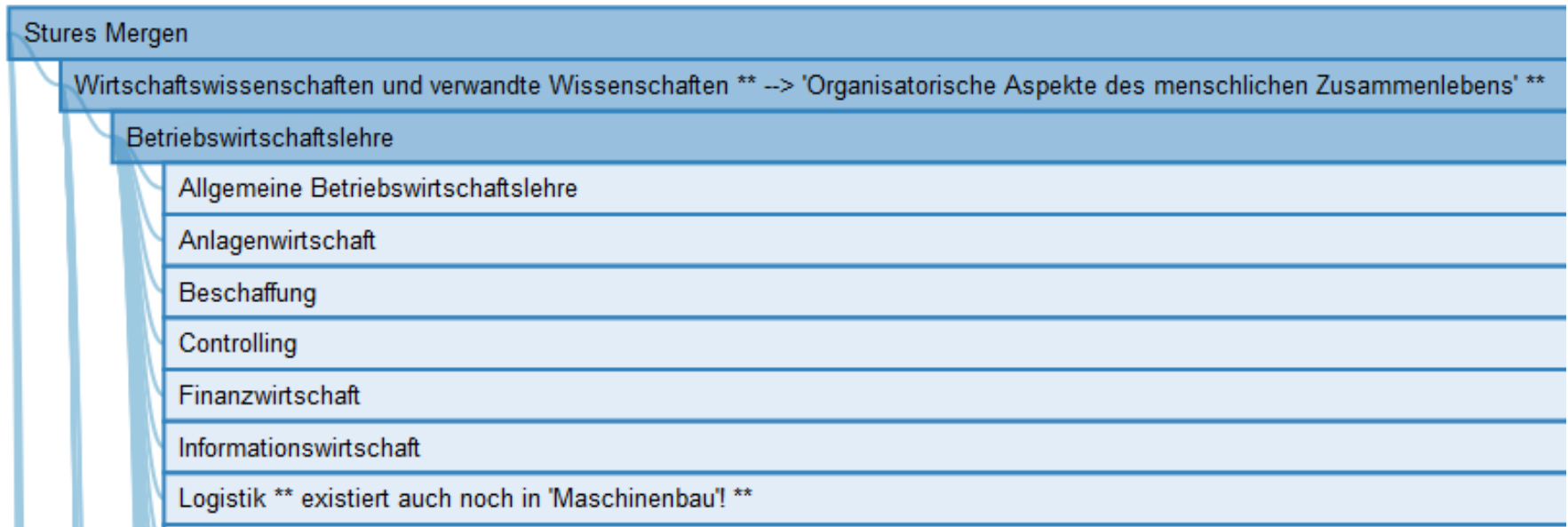
- Evaluierung verschiedener Kandidaten, z.B.:
Patentklassifikation (IPC), DDC, GND-Systematik, TEMA-Fachordnung
- Visualisierung über JSON und D3
- Kollaboratives Erarbeiten der Struktur durch die Fachreferenten der WTI

Alles ausgeklappt

Ebene 1 ausgeklappt

Ebene 2 ausgeklappt









Ebene 3 ausgeklappt



Relationen – Stand und Desiderate

Radionuklidbatterie

Xrefs (1)
Linked Data
Images
Settings

	Type	Label	Concept ID	
	V	Energiedirektumwandlung	000979	
	V	Isotopentechnik	001652	
	V	radioaktives Nuklid	004840	
	V	Rankine-Wandler	403132	

Verschiedenste Arten von „related“...

Relationen – Stand und Desiderate

- Maschinenteil
 - Angussentnahmegerät
 - Anker (elektrische Maschine)
 - Antriebskopf
 - Aufhänger
 - Aufsteckleiste
 - Ausgleichselement
 - + **Bremse**
 - Drehmaschinenbett
 - Düsenring
 - Düsenstock
 - Elevatorbecher
 - + Fadenkreuz
 - + Feder (Federungselement)
 - Federungsbauteil
 - Flügelarm
 - Flügelblatt

Facetten: Hilfskonstruktionen wie „Maschinenteil“

<input type="radio"/>	Maschinenteil
<input type="radio"/>	Zentriereinrichtung

Relationen – Lösungsansatz: Intelligente Verwendung von SKOS concept schemes / SKOS collections



SKOS concept schemes:

- Sammlung von Konzepten, die thematisch zusammengehören
- Relation *hasTopConcept* für Hierarchiebildung

SKOS collections:





- Konzeptgruppierungen nach bestimmten Aspekten, **quer zur Hierarchie**
- Beispiel im SKOS Primer:

```
<milk by source animal>
  cow milk
  goat milk
  buffalo milk
  _:b0 rdf:type skos:Collection;
      skos:prefLabel "milk by source animal"@en;
      skos:member ex:cowMilk;
      skos:member ex:goatMilk;
      skos:member ex:buffaloMilk.
```

A thick red curved arrow starts from the left side of the text, pointing towards the right side, specifically towards the SKOS collection definition. It indicates the relationship between the list of milk types and the collection that groups them.

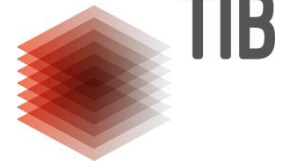
GND als Wissensorganisationssystem (KOS) im Sinne des Semantic Web? – Eine Inventur

Strukturelle Gegebenheiten:

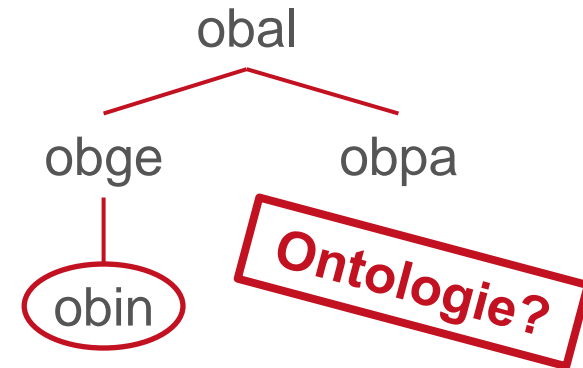
- traditionell: **Bestandsorientierung** ⇔
Knowledge Organization System <Semantic Web>
- **polyhierarchisch** (Bsp.: „Neurobiologie“)  
- **Relationen nicht durchgängig** eingepflegt
(auch **Oberbegriffsrelationen!** Bsp.: „Teilchenfalle“)  
- laut Fachreferat (Mathematik):
teilweise **fachlich unzutreffende Relationen**
(Unterbegriffe, die nebengeordnet sein sollten, u.Ä.)

Ontologie?

Gegebenheiten mit Potential: GND-Relationscodes



- obal: unspezifizierte Oberbegriffsrelation
- obge: „untergeordneter Begriff besitzt alle Merkmale des übergeordneten plus eins“
- obin: Instantiierung einer Gattung/Klasse
- obpa: Teil-Ganzes-Beziehung



Inventur (kleine Stichprobe):

Sachbegriff:

Zapfsäule

Oberbegriff:

Tankstelle [Oberbegriff partitiv]

Sachbegriff:

Schimmel <Pferd>

Oberbegriff:

Pferd [Oberbegriff allgemein]

Sachbegriff:

Schimmel <Spiel>

Bell and Hammer

Oberbegriff:

Würfelspiel [Oberbegriff generisch]

Sachbegriff:

Schach

Oberbegriff:

Brettspiel [Oberbegriff allgemein]

Sachbegriff:

Mensch ärgere dich nicht

Oberbegriff:

Brettspiel [Oberbegriff instantiell]

Würfelspiel [Oberbegriff instantiell]

Gegebenheiten mit Potential: GND-Relationscodes



Weitere Relationen:

- vbal – verwandter Begriff (allgemein)
- rela – Relation (allgemein) [Personen]
- beru – Beruf
 - berc – Beruf, charakteristisch
- beza – Beziehung, Bekanntschaft, Freundschaft
 - bezb – Beziehung beruflich
 - bezf – Beziehung familär
- etc. ...

Weitere KOS im Umfeld der GND

GND-Systematik

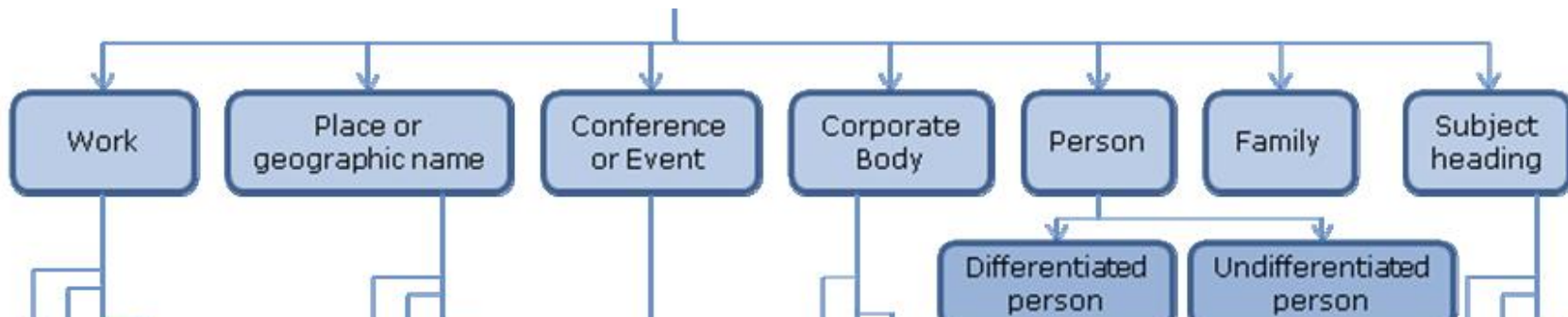
soll „systematischen Zugang zum Vokabular ermöglichen und fachliche Ausschnitte oder Teilausgaben bilden“

GND Ontology

Beschreibungssprache: RDFS/OWL

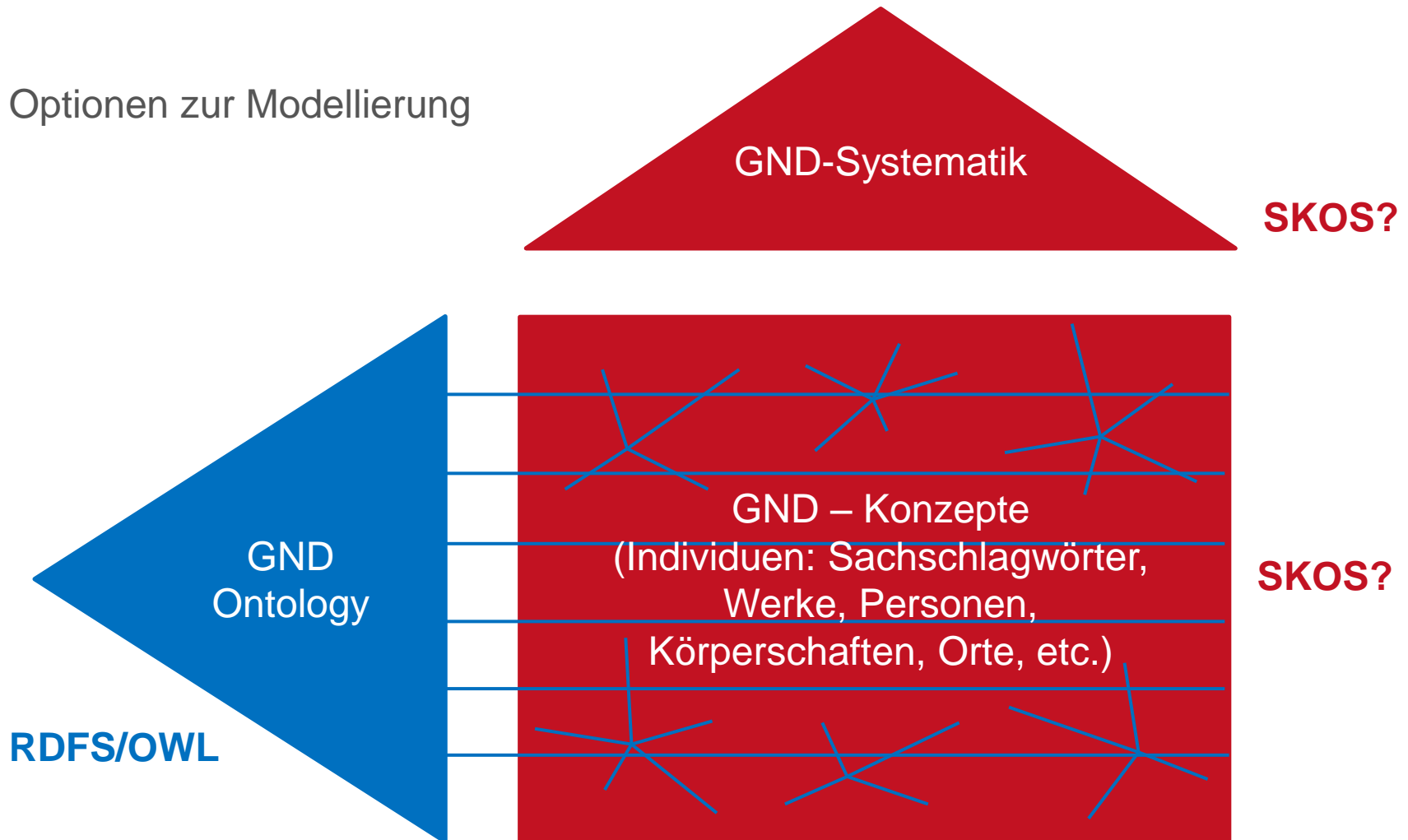
beschreibt den Inhalt der GND auf einer **Metaebene**:

- Klassen von Normdatenentitäten/GND-Elementen
- Relationen zwischen diesen und zu Datenwerten/Annotationen (Basis: GND-Relationscodes)
- ein AllDisjointClasses-Axiom für die oberste Ebene



Versuch einer Visualisierung der Inventur

Optionen zur Modellierung

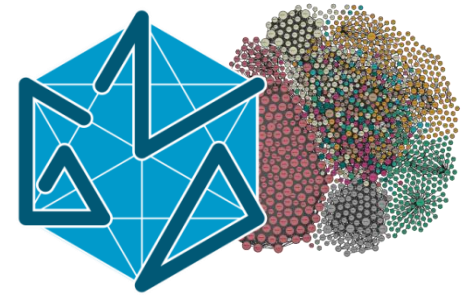


GND als KOS im Sinne des Semantic Web? Fazit



Gedanke 1: GND als Ontologie? (Konzepte zu Klassen)

- zu groß, zu unscharf
- nicht Primärzweck der GND
- denkbar: Ausschnitte säubern und schärfen



Gedanke 2: GND in SKOS

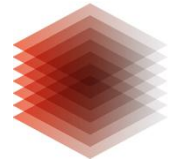
- Experiment der ZBW (2016): Relationen aus der GND Ontology, Einarbeitung der GND-Systematik, keine SKOS-XL-Unterstützung
- Zeit für den nächsten Schritt? z.B. SKOS-XL, *skos:Collection*, SKOS-Erweiterung iso-thes für volle Konformität mit ISO 25964

Gedanke 3: Wie weiter mit der GND Ontology?

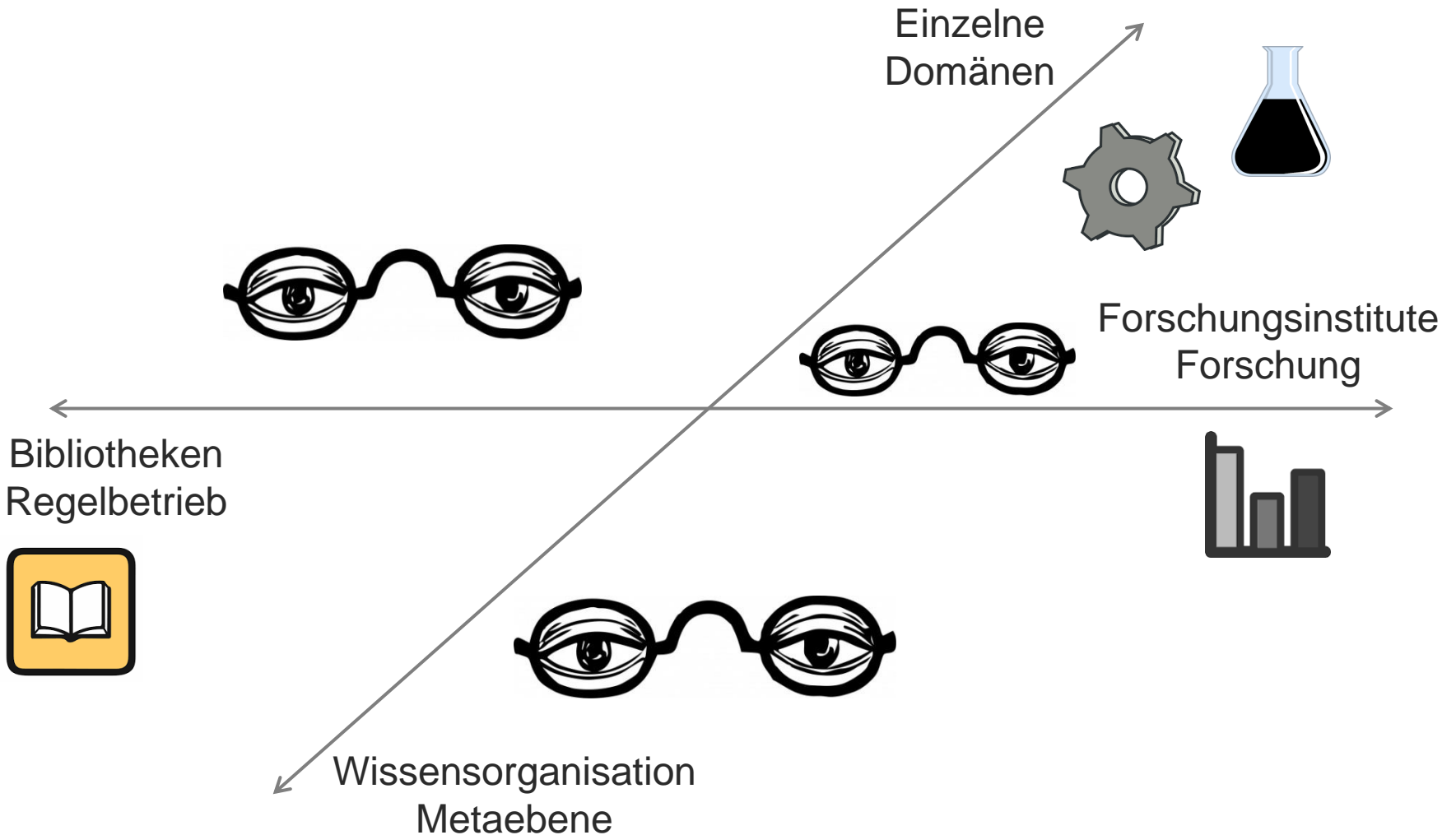
- Anwendung z.B. für Konsistenzprüfungen in der GND
 - Axiome ausbauen, weitere OWL-Konstrukte ausnützen

Plädoyer für RDF, SKOS und Co.

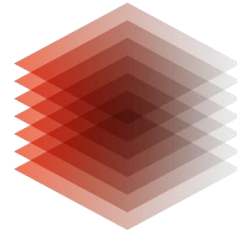
- Bibliothekarische Wissensorganisationssysteme (KOS) sind wertvoll, auch für andere Fachcommunities, aber dafür müssen sie **sichtbar** und **interoperabel** gemacht werden
 - Durch **Interaktion mit Fachcommunities (über Linked Open Data)**, verbessert sich auch die strukturelle Qualität und semantische Ausdruckskraft der GND (inhaltlicher Input, aktuelle Forschungstrends)
 - Durch Einspeisen hochqualitativer intellektueller semantischer Daten (sorgfältig erstellte bibliothekarische KOS und Metadatensätze) **verbessern sich semantische Technologien**, denn: *garbage in, garbage out!*
- ➔ sowohl Bibliothekswelt als auch Data-Science-Welt und andere Fachcommunities profitieren



In welchem Bereich bewegen wir uns? Verschiedene Perspektiven



LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

DANKE!

www.tib.eu/de/forschung-entwicklung/

Kontaktdaten

Dr. Anna Kasprzik

T 0511 762-14219, anna.kasprzik@tib.eu



Creative Commons Namensnennung 3.0 Deutschland
<http://creativecommons.org/licenses/by/3.0/de>

RDF und Linked Open Data *in a nutshell*

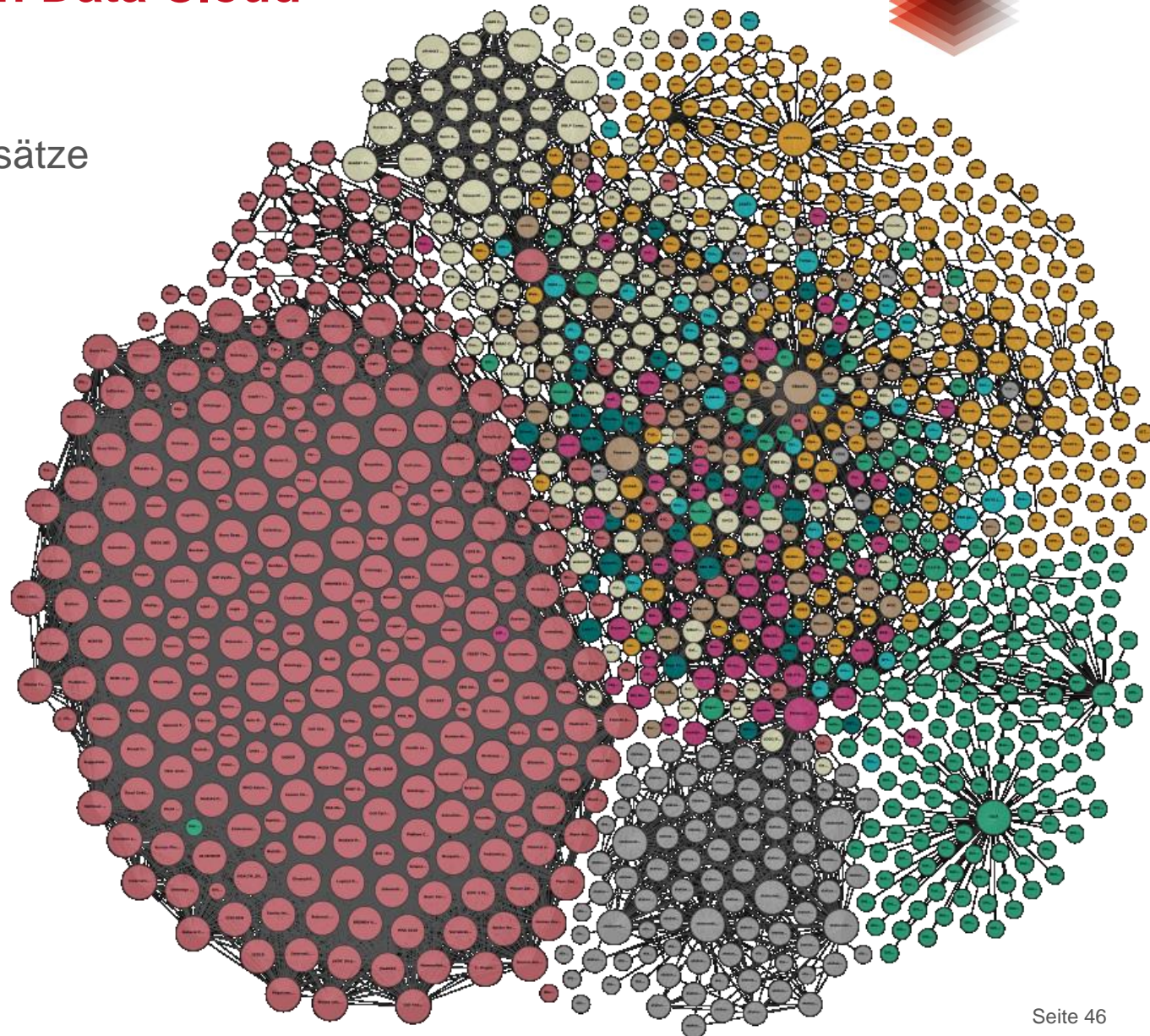
- Das **RDF (Resource Description Framework) Modell**:
ALLE Aussagen werden als Tripel von
 Subjekt – Prädikat – Objekt
repräsentiert, also als Relationen zwischen zwei Entitäten
oder zwischen einer Entität und einem Datenwert
- **Linked Open Data**: ALLE Entitäten, ob digital oder physisch, bekommen
Unified Resource Identifiers (URIs), die im Idealfall im WWW auffindbar,
mit Informationen zu dieser Ressource versehen, und vernetzt sind
- Da Publikationen auch immer mehr im WWW enthalten sind,
lässt sich mit diesen Instrumenten so ziemlich alles abdecken,
was die Sacherschließung in Bibliotheken abdeckt...



– ODER ??

Die Linked Open Data Cloud

- ca. 10.000 Datensätze
- **150.000.000.000 RDF-Tripel**



- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated

Grundlagen

- **Konzept \approx Begriff** (von lat. *cum* + *capere*, “zusammen-fassen”)
Das abstrakte Ding, sprachunabhängig
- **Term \approx Label**
Natürlichsprachige Bezeichnung für ein Konzept,
z.B. “Schimmel”, “weißes Pferd”, “white horse”, ...
- **Relation**
Semantische (oder andere) Beziehung zwischen Einheiten,
kann zwischen Konzepten, zwischen Termen
und zwischen Konzepten und Termen bestehen
z.B. “ist Unterbegriff von”, “ist Abkürzung von”,
“ist altertümliche Bezeichnung für”

Relationen und Facetten

Facetten sind in einem Suchsystem NACH der initialen Suche angebotene Filtermöglichkeiten („drill-down“).

Vorschlag: WENN das jeweilige System das kann, lassen sich aus definierten Relationen, die im Thesaurus vergeben sind, Facetten für das entsprechende Suchsystem extrahieren.

Beispiel: Verwende die Relation „ist Anwendung von“ (z.B. *Strömungsmechanik*), um eine Facette „ist Anwendung der Strömungsmechanik“ anzubieten.

Die einfachere Option:

Lege einige Querschnittsaspekte fest und biete nur diese als Facetten an.

Die komplexere Option:

Biete dem Nutzer die Möglichkeit, Relationen und die dadurch verbundenen Konzepte beliebig abzufragen – sehr anspruchsvoll (SPARQL-Queries).