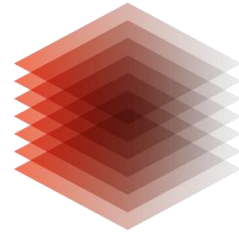

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

The Research Core Dataset (KDSF) In The Linked Data Context

Tatiana Walther, Christian Hauschke
15. June 2018, Umeå, Sweden
CRIS2018

Agenda

1. Background

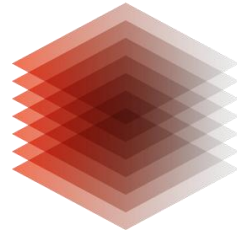
- Reporting processes in Germany
- KDSF
- VIVO

2. Implementation

- Implementation of the KDSF in VIVO
- Ontology analysis and comparison
- Matching
- Transformation

3. Conclusion / challenges / future plans

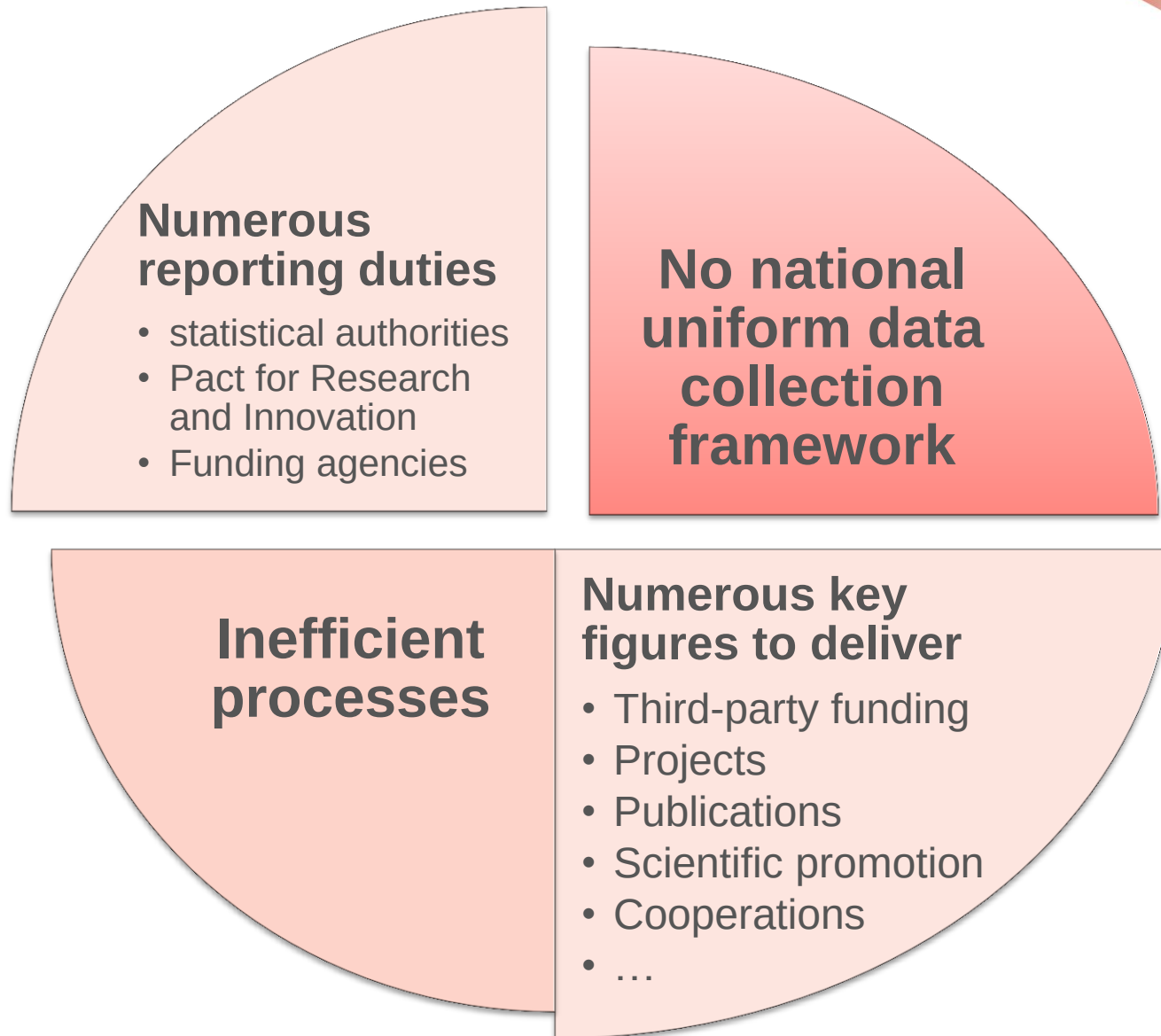
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

1. Background

Reporting Processes in Germany



The Research Core Dataset (KDSF)

- The Research Core Dataset (KDSF) for the German academic system – a specification recommended by the German Council of Science and Humanities (Wissenschaftsrat) for a uniform, standardized collection, and provision of data on research activities [1].



<https://kerndatensatz-forschung.de>

KDSF in CRIS

- Implementation cases across German institutions
- Implementation within a number of Current Research Information Systems (CRIS)
- Available in a CERIF-based XML format and in RDFS/RDF/OWL – Linked data compliant?
- VIVO is mentioned as one of the potential systems for implementation [2].

- CRIS, originally developed at the Cornell University Library, USA
- Open source software
- Data aggregation, representation, and visualization [3]



Heller, Lambert

Positions ▾

Overview Publikationen

Principal Investigator On

Aufbau einer Science 2.0
Science 2.0 awarded by L

Nachnutzung von Open-A
Forschungsgemeinschaft 2

TextTransfer: Korpusgestü
wissenschaftlichen Texten
sowie Workflow zur Bereit

Conference Paper **Filter**

Year	Title
2016	Discovery and efficient r proposal 2016-03-17
2015	Preserving the Long Tail Research Libraries 2015
2015	Überblick über den aktue Standardisierung, verteil Kommunikationstechnol 35. 2015
2014	The Quest for Research

NOA
TIB Research Subject

Overview Zugehörigkeit

Collaborators

CrossRef Organisation
Hochschule Hannover University
Wikimedia Deutschland Foundation

Projektmitarbeiter

Heller, Lambert; Warten, Christian; Raster,
Köpke, Johannes

Abstract

Entwicklung eines Verfahrens zur automa
Bereitstellung von Open-Access-Abbildungen
mittels der Infrastruktur von Wikimedia Com

Technische Informationsbibliothek (TIB) Leibniz-Informationszentrum Technik und
Naturwissenschaften und Universitätsbibliothek

How do you want to compare?
by Publications

What do you want to compare?
Organizations | People | View all...

Search

Records 1 - 6 of 6

Entity label	Grant Count	Entity Type
Programmbereich C - Forschung und Entwicklung	457	Programme Section
Programmbereich A - Bestandsentwicklung und Metadaten	68	Programme Section
Programmbereich B - Kommunikation und Entwicklung	42	Programme Section
Leibniz Association	20	Administrative Entity
Direction	8	Administrative Entity
Zentraleinheit Vorbereitung und Administration	0	Administrative Entity

Comparing publications of Organizations and People in Technische
Informationsbibliothek (TIB) Leibniz-Informationszentrum Technik und
Naturwissenschaften und Universitätsbibliothek

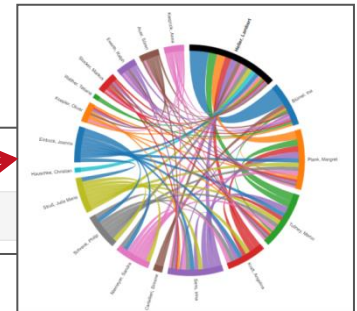
Number of publications

Year

Total Number of publications

You have selected 3 of a maximum 10 organizations and people

☒ Assoc. Stabs
☒ Execut. Stabs
☒ Stock. Marku



VIVO-Implications

- VIVO - based on Linked Data technologies
- Information encoded in RDF triples

add new statement		refresh list	
Subject	Predicate	Object	actions
Technische Informationsbibliothek (TIB) Leibniz- Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek	located in	Hannover	Edit Delete
Technische Informationsbibliothek (TIB) Leibniz- Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek	has contact info	n2824	Edit Delete

- Each item – a URI-named Individual - once recorded, multiply reusable

Admin Panel

[Edit this individual](#)

Verbose property display is **off** | [Turn on](#)

Resource URI: <http://d-nb.info/gnd/1080328793>

Technische Informationsbibliothek (TIB) Leibniz-Informationszentrum Technik und
Naturwissenschaften und Universitätsbibliothek 

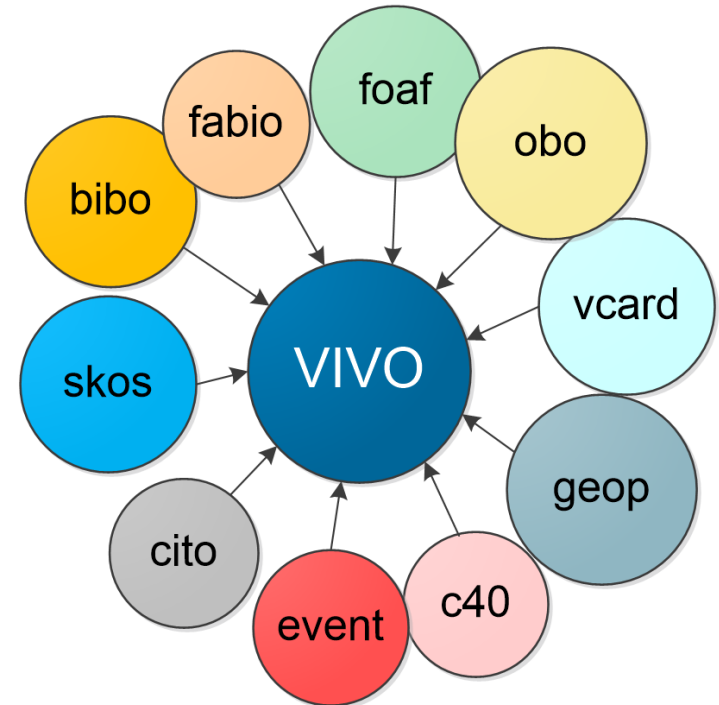
[Bibliothek](#) 

Websites 

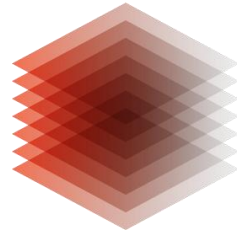
 [Temporal Graph](#)

VIVO-Implications

- Data - classified by means of established Linked Data vocabularies
- The inbuilt reasoner + axioms in vocabularies - derivation of implicit knowledge



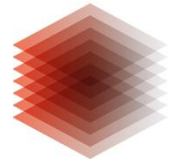
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

2. Implementation

Implementation of KDSF in VIVO at the TIB



TIB-FIS-Discovery

- Profiles
- Representation on the Web



KDSF-VIVO

- Reporting
- Data mantainance



Implementation of KDSF in VIVO

A large, solid orange shape on the left side of the slide, resembling a stylized 'C' or a large bracket. It contains two concentric white circles. The right side of the slide features a vertical stack of five rectangular boxes with orange borders. The top three boxes contain text, while the bottom two are empty.

Integration of KDSF into VIVO

**Alignment of KDSF and VIVO
ontologies**

**Matching of VIVO and KDSF
ontologies**

Implementation on the ontology level

**Analysis and
comparison of
the ontologies**

**Ontology
matching**

**Transformation
and extension
of the KDSF**

Analysis and comparison of VIVO and KDSF

- Domain of interest
- Terminology
- Syntax
- Structure
- Granularity

Similar domain, but different purposes:

- **VIVO**: representation of **scholarship** and expertise; no particular domain of knowledge; support of „evaluation and impact assessment of individual people or group of people“[4]
- **KDSF**: exchange of information on **research activities, research staff** and **their products**; no particular domain of knowledge; focus on versatile **core data** and not on assessment of scientific performance [1]

Terminology and syntax

- Differences in terminology, as terms are in different natural languages:
 - VIVO – English
 - KDSF – German
- Same Syntax: RDFS/RDF/OWL
- Different Structure – number of axioms, types of relations

Comparison and analysis of VIVO and KDSF – Entity Metrics

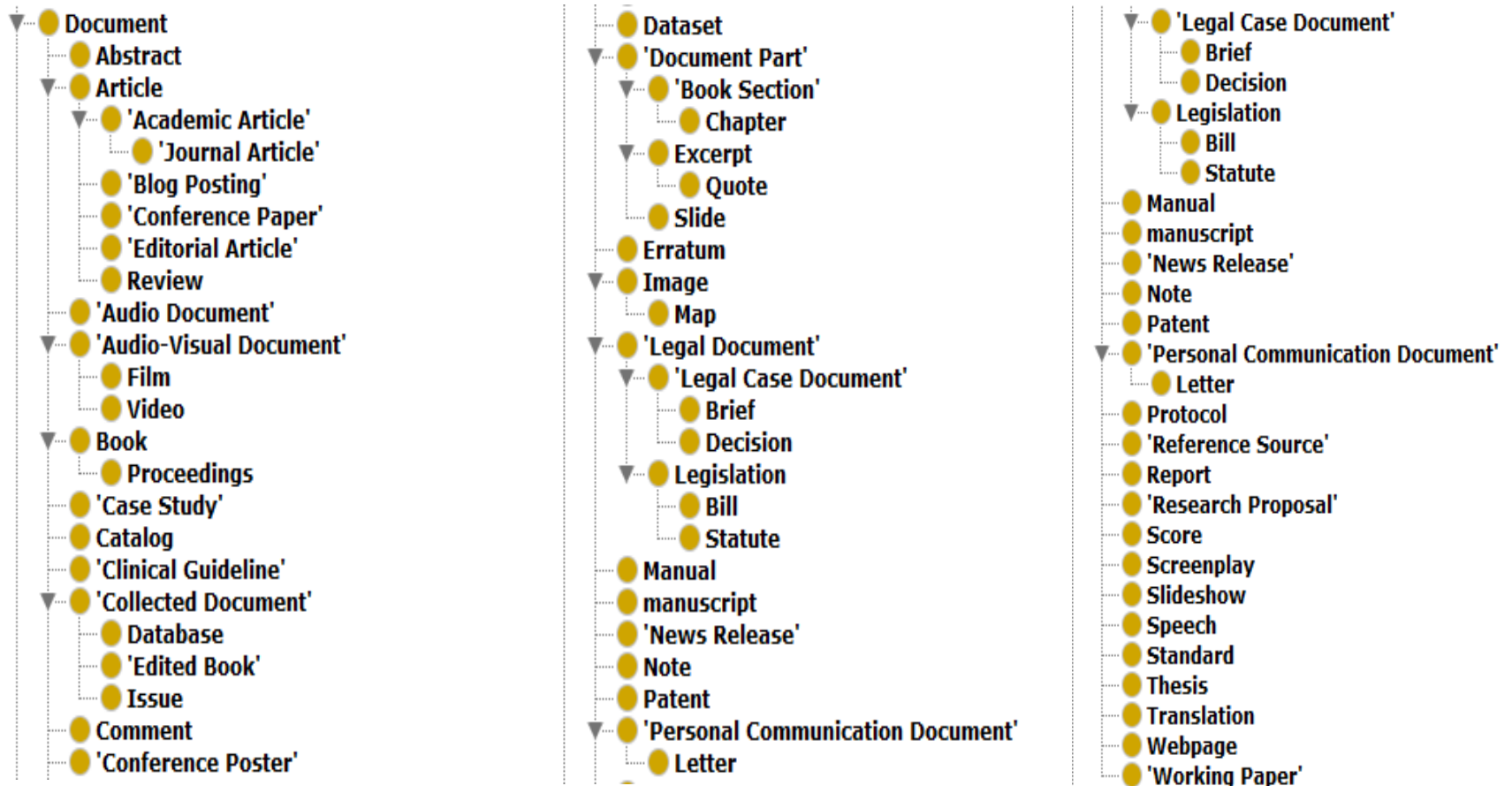
	VIVO Ontology	KDSF
Classes	414	123
Object properties	198	36
Datatype properties	217	53
Instances	14	0

Comparison and analysis of VIVO and KDSF - Axiom Metrics

	VIVO Ontology	KDSF
Class hierachies	623	96
Class equivalences	40	1
Class disjunctness	1	1
Property hierachies	65	0
Functional relations	2	12
Transitive relations	2	0
Symmetric relations	1	0
Inverse relations	57	0
Domain intersections	16	4
Range intersections	0	2
Restrictions	9	0
Enumerated datatypes	0	3

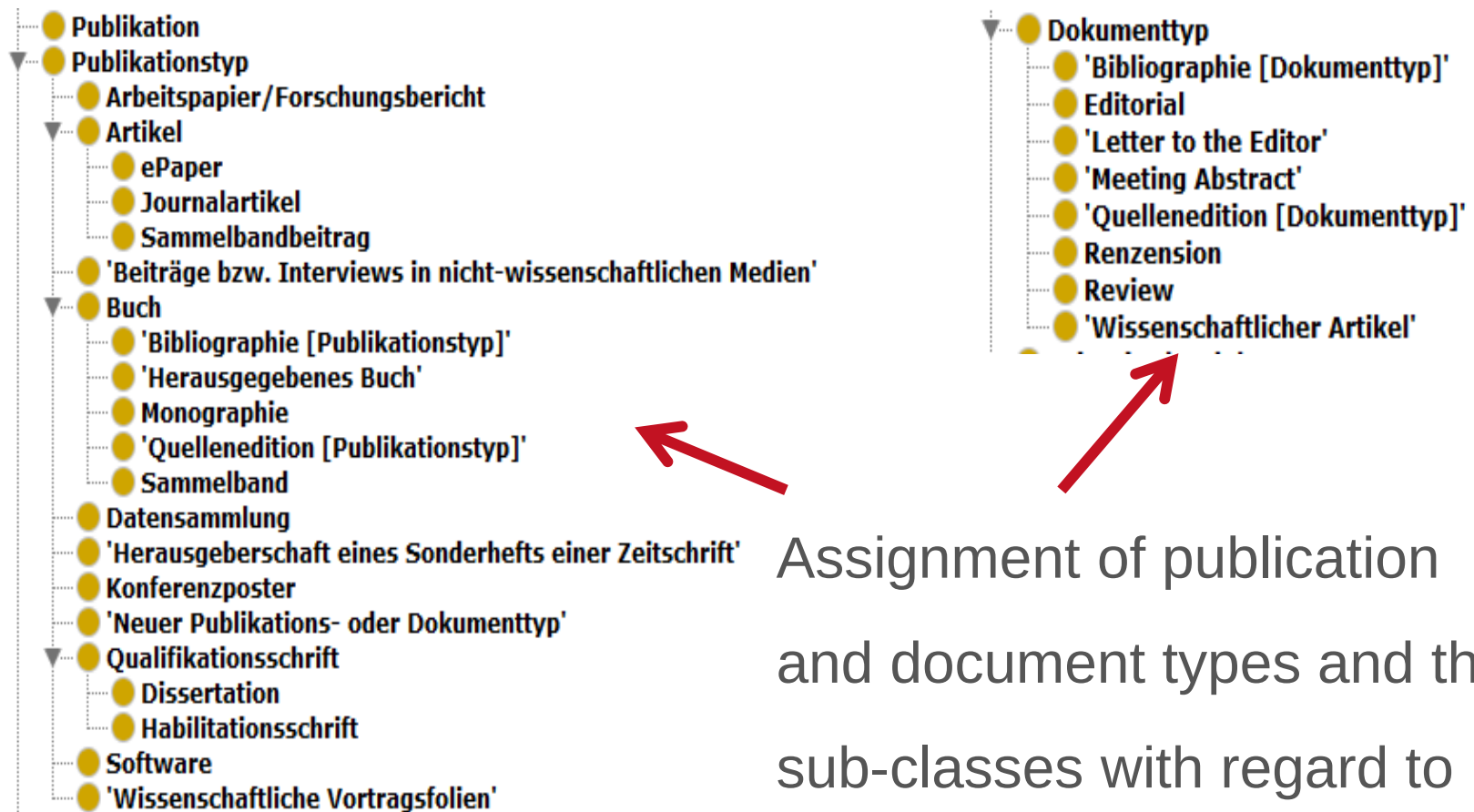
Comparison and analysis of VIVO and KDSF

- Granularity of publication-related part of VIVO



Comparison and analysis of VIVO and KDSF

- Granularity of publication-related part of KDSF



Assignment of publication and document types and their sub-classes with regard to form and content

Ontology matching

- Computational matching attempt with LogMap
- Manual matching
 - Wholesale matching
 - Partial matching

Computational Matching with LogMap

- **LogMap** - a scalable ontology matching device developed at the Oxford University
- Outputs - mappings of entities parsed on their lexical and structural similarity
- Inference check
- User involvement into the process of matching [5]

Computational Matching with LogMap

- Results: 4 mappings with equivalent correspondences

kdsf:Patent $\equiv$ *bibo:Patent*

kdsf:Review $\equiv$ *vivo:Review*

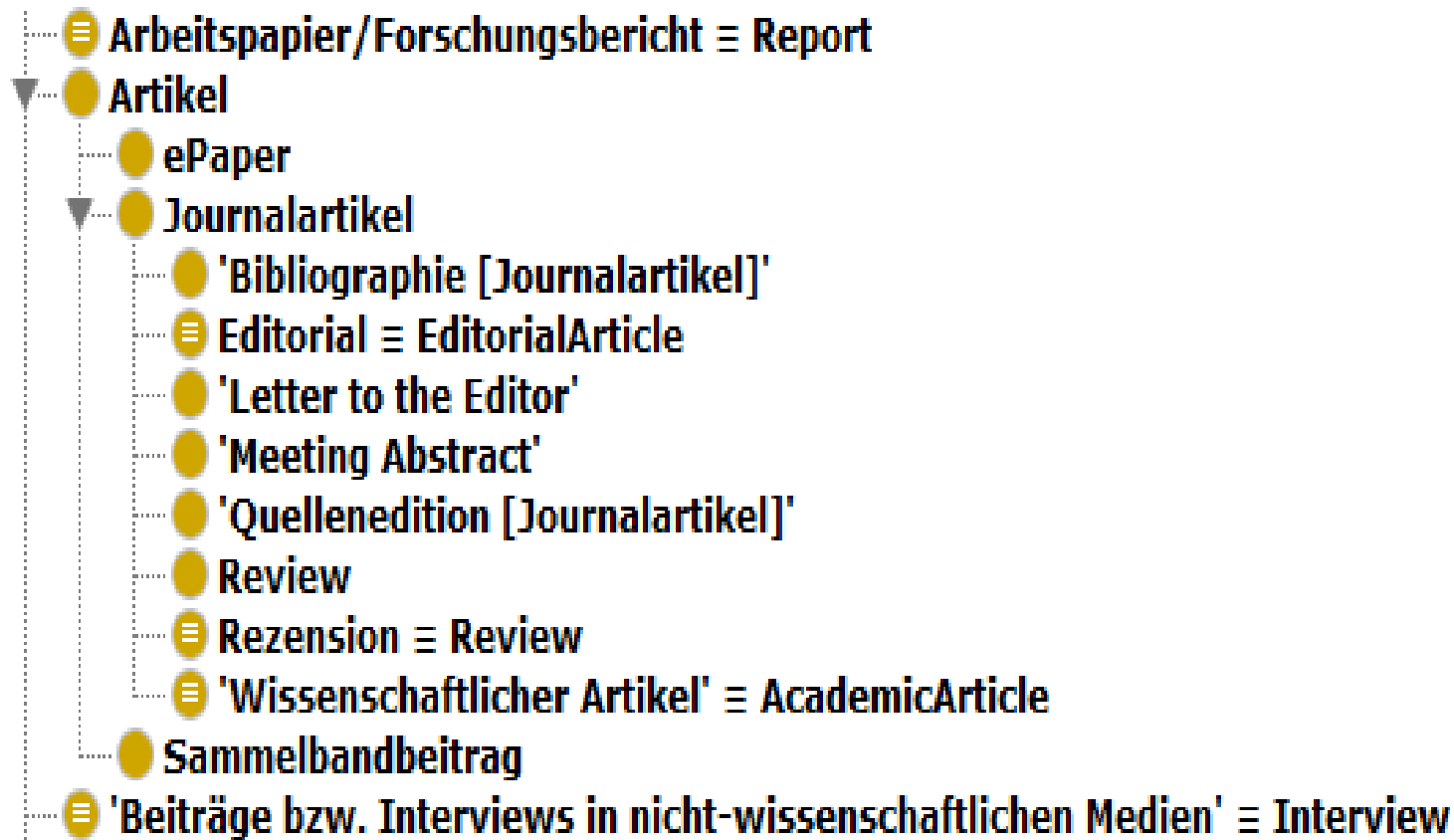
kdsf:Person $\equiv$ *foaf:Person*

kdsf:Software $\equiv$ *obo:ERO_0000071 (software)*

- Matching decision – obviously based on class labels
- Not detected** e.g. *vivo:ConferencePoster* $\equiv$ *kdsf:Konferenzposter*, *bibo:Book* $\equiv$ *kdsf:Buch*

Manual matching

- Initial approach - matching of entire ontology through **equivalence** and **subsumption** correspondences



Manual matching

- Initial approach - matching of entire ontology through **equivalence** and **subsumption** correspondences
- Equivalence correspondences =>
 - equivalent classes have the same set of sub-classes, individuals, and properties [6]
 - a lot of unintended inferences
 - undesirable effects on the applicational level

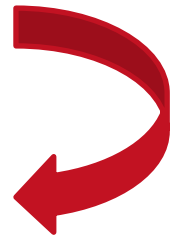
**NOT EXACTLY
WHAT WE WANT!**

Manual matching

- 2nd. approach – implementation of the KDSF concepts which **can not be represented by VIVO concepts** using subsumption correspondences

kdsf:BibliographiePublikationstyp $\subseteq$ kdsf:Buch
kdfs:Buch $\equiv$ bibo:Book

kdsf:BibliographiePublikationstyp $\subseteq$ bibo:Book



Transformation and extension

- Better applicability of KDSF in VIVO -> transformation required
- Types of transformation:
 - Classes to instances
 - Datatype properties (attributes) to object properties (relation)
 - Omissions
 - Additional axioms
- Changed entities + new entities -> extension of KDSF

Classes to instances

- Classes - groups of entities with similar properties [6].
- Some KDSF classes – real objects with no any further specialization possible -> transformed to instances

kdsf:DFG (Deutsche Forschungsgemeinschaft, in Engl.: German Research Foundation) => ***#DFG***

kdsf:EU (European Union) => ***#EU***

Data type properties to object properties

- Some datatype properties (attributes) – more valuable as object properties (relations)
- Ranges required - new classes + new instances

kdsf:Zugriffsrechte (access rights) => ***vivo-kdsf:hatZugriffsrechte***

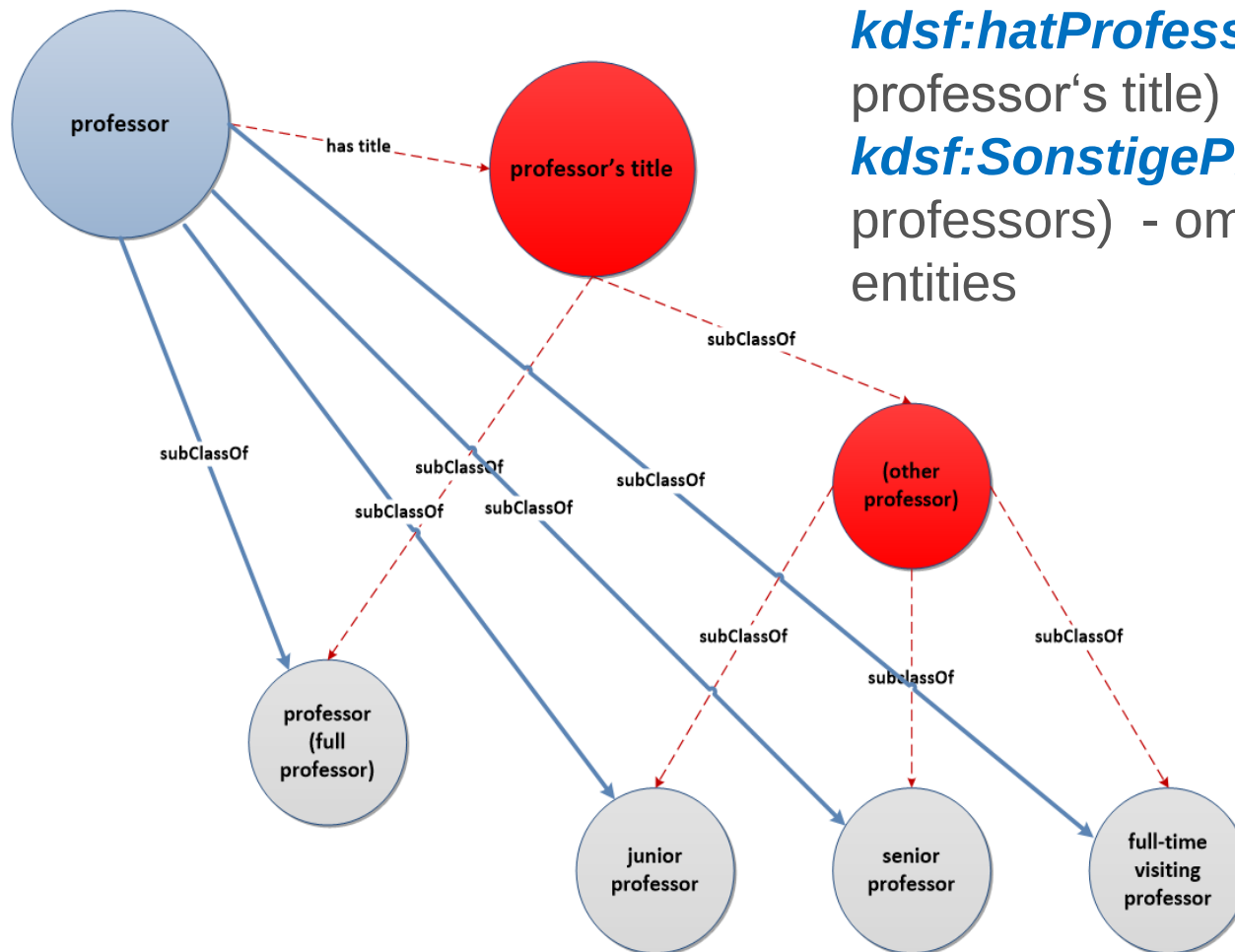
+ range: ***kdsf-vivo:Zugriffsrechte***

+ named Individuals: ***open access, restricted access, embargoed access, metadata only access***

- Reuse of the URIs from Controlled Vocabulary for Access Rights (Draft V1), Confederation of Open Access Repositories (COAR)

Omissions

- Intermediate classes + related properties



kdsf:Professorenbezeichnung

(professor's title)

kdsf:hatProfessorenbezeichnung (has professor's title)

kdsf:SonstigeProfessoren (other professors) - omitted as unnecessary entities

Additional Axioms



- Inversion – relations in both directions [6]
- E. g. persons contribute to projects $\longleftrightarrow$ projects have contributors
- 1 entry -> 2 records

TIB VIVO [BETA]

Browse ▾

A study of institutional spending on open access publication fees in Germany

Academic Article TIB Publication

<https://doi.org/10.7717/peerj.2323>

Overview Identität Additional Document Info Kdsf

Beteiligte Personen

Jahn, Najko; Tullney, Marco

Authors

Tullney, Marco

marco.tullney@tib.eu

Websites

ORCID

Visualizations

Co-author Network

Co-investigator Network

Tullney, Marco

Positions ▾

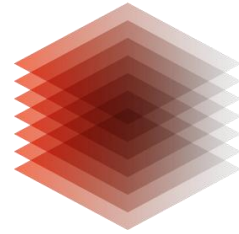
Publikationen Forschung Kontakt Identität

Selected Publications

Academic Article Filter

Year	Title	
2016-10-25	The Effect of Gender in the Publication Patterns in Mathematics. <i>PLOS ONE</i> . 11:e0165367. 2016-10-25	105
2016-10-10	Neue Marktkonzentration? Eine Analyse der Open-Access-Kosten in Deutschland. <i>Forschung & Lehre</i> . 2016:886-887. 2016-10-10	
2016-08-09	A study of institutional spending on open access publication fees in Germany. <i>PeerJ</i> . 4:e2323. 2016-08-09	176
2014	Wissen auf einen Klick. <i>DUZ-Magazin</i> . 2014:69-71. 2014	8

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

3. Conclusion / challenges / future plans

Conclusions

- Results: the 1. draft of KDSF-VIVO-Alignment and KDSF-VIVO-Extension
- Computational matching doesn't deliver satisfactory results when matching KDSF and VIVO
- Only a matching of the VIVO and KDSF – not sufficient for applicability of KDSF in VIVO; transformation and are required

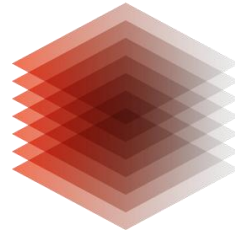
Challenges and future plans

- Further work on the alignment and extension -> an iterative process -> there are **A LOT OF** various options, how to design one and the same part of information
- Collaborative work in constant exchange with the German VIVO community -> learning from multiple perspectives
- Publication of KDSF-VIVO-Alignment and KDSF-VIVO-Extension and their versions on the Web according to the best practices of ontology publication and versioning à la W3C
- Keeping track of the changes: Evolving of the source ontologies -> Keeping KDSF-VIVO-Alignment up-to-date

References

1. Wissenschaftsrat. Empfehlungen zur Spezifikation des Kerndatensatz Forschung. [Online] January 22, 2016. [Cited: June 11, 2018.] <https://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>.
2. Institut für Forschungsinformation und Qualitätssicherung (iFQ); Fraunhofer-Institut für Angewandte Informationstechnik; Geschäftsstelle des Wissenschaftsrates. Kerndatensatz Forschung : Dokumentation der Ergebnisse ; 13 Oktober 2015. [Online] http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf.
3. Duraspace. VIVO. About VIVO. [Online] [Cited: Februar 28, 2018.] <http://vivoweb.org/info/about-vivo>.
4. Conlon, Mike. VIVO Ontology Domain Definition. GitHub. [Online] December 4, 2017. [Cited: June 2, 2018.] <https://github.com/openrif/vivo-isf-ontology/wiki/VIVO-Ontology-Domain-Definition>.
5. Jiménez-Ruiz, Ernesto and Cuenca Grau, Bernardo. LogMap: Logic-based and Scalable Ontology Matching. [Online] n.d. [Cited: June 7, 2018.] https://www.cs.ox.ac.uk/isg/projects/LogMap/papers/paper_ISWC2011.pdf.
6. W3C. OWL Web Ontology Language. Reference. W3C Recommendation. [Online] February 10, 2004. [Cited: June 5, 2018.] <https://www.w3.org/TR/owl-ref/>.

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Thank you for your attention!

Contact

Tatiana Walther

tatiana.walther@tib.eu

Christian Hauschke

christian.hauschke@tib.eu