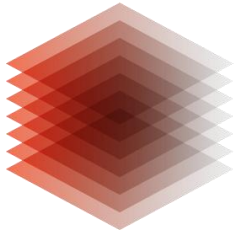
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

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

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

Agenda

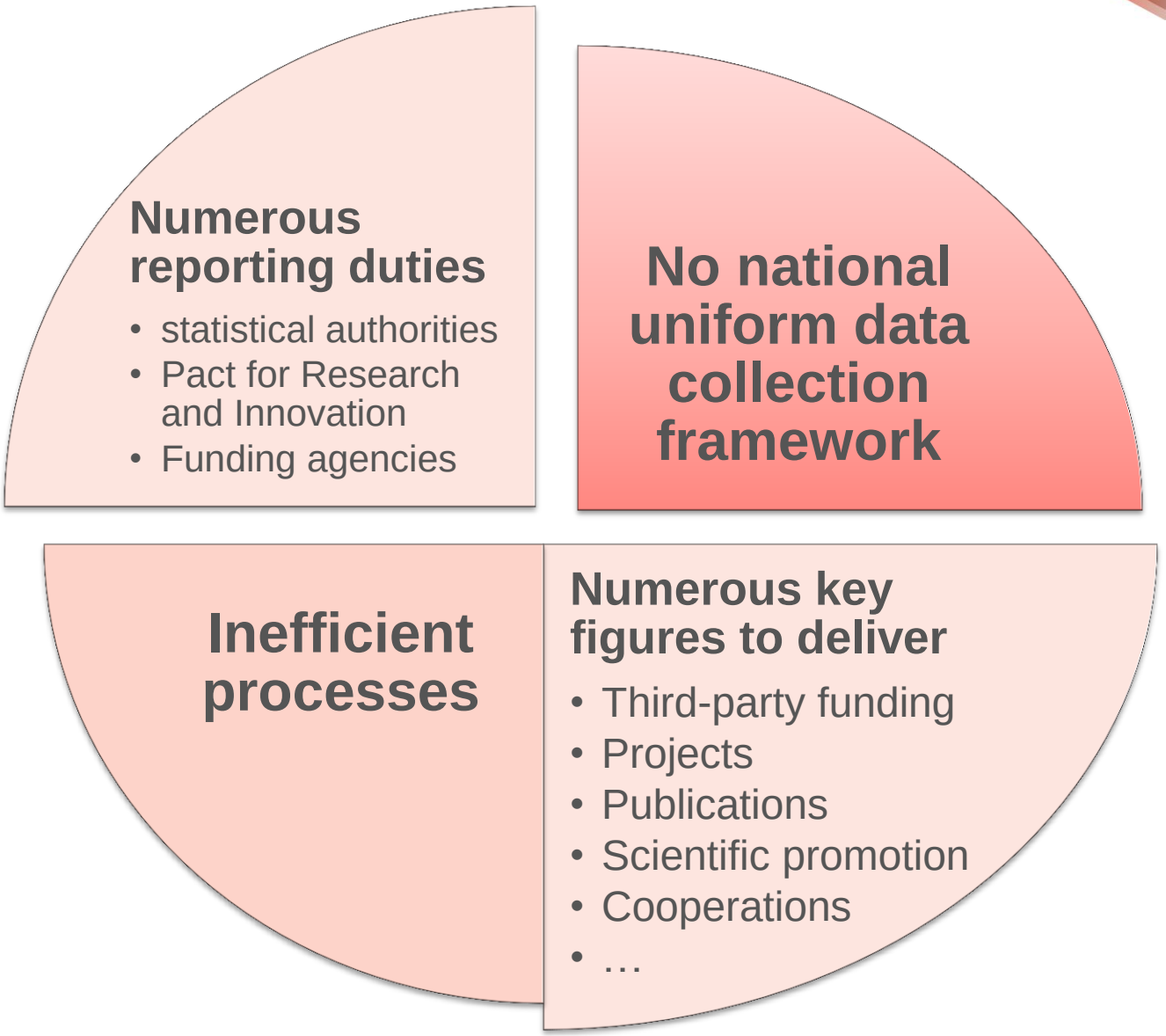
1. Background
   - Reporting processes in Germany
   - KDSF
   - VIVO
2. Implementation
   - Implementation of the KDSF in VIVO
   - Ontology analysis and comparison
   - Matching
   - Transformation
3. Conclusion / challenges / future plans

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

1. Background

Reporting Processes in Germany



The Researsch 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].

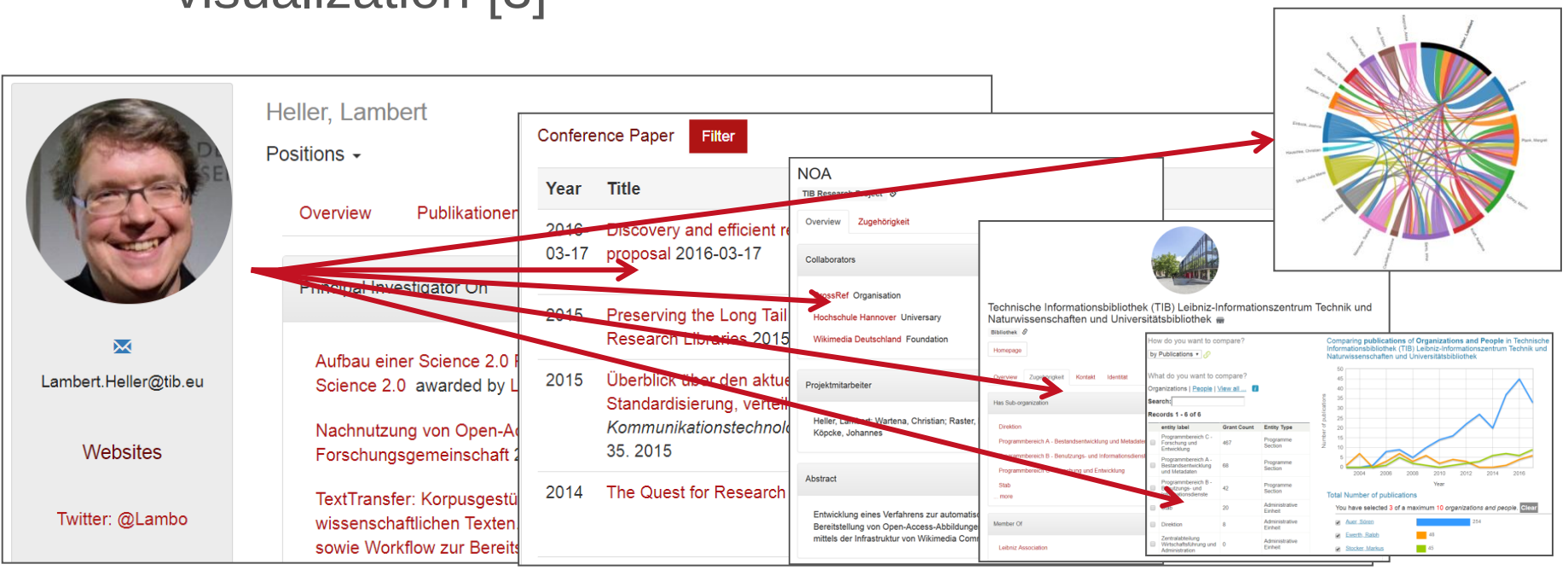
https://kerndatensatz-forschung.de

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].

VIVO

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



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

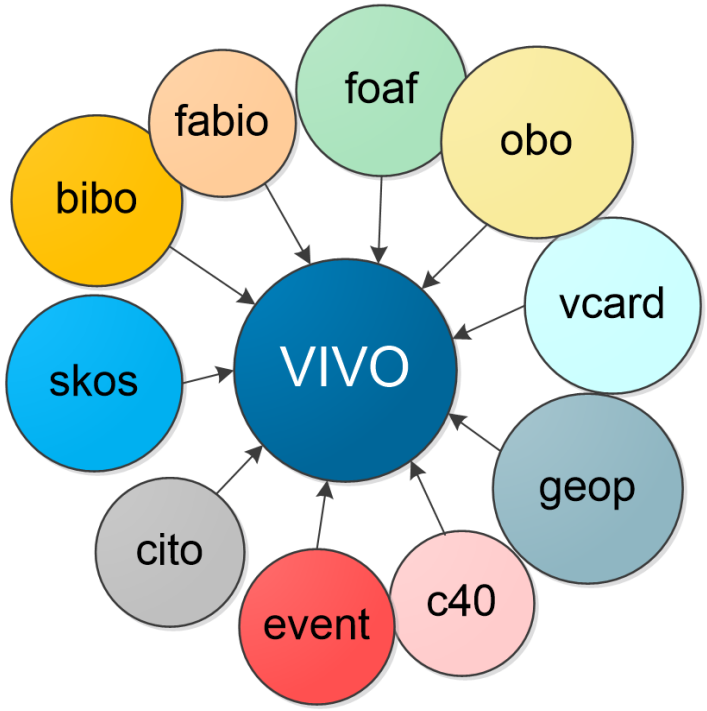
Bibliothek

Websites

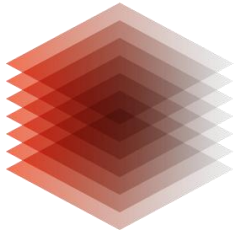
Temporal Graph

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

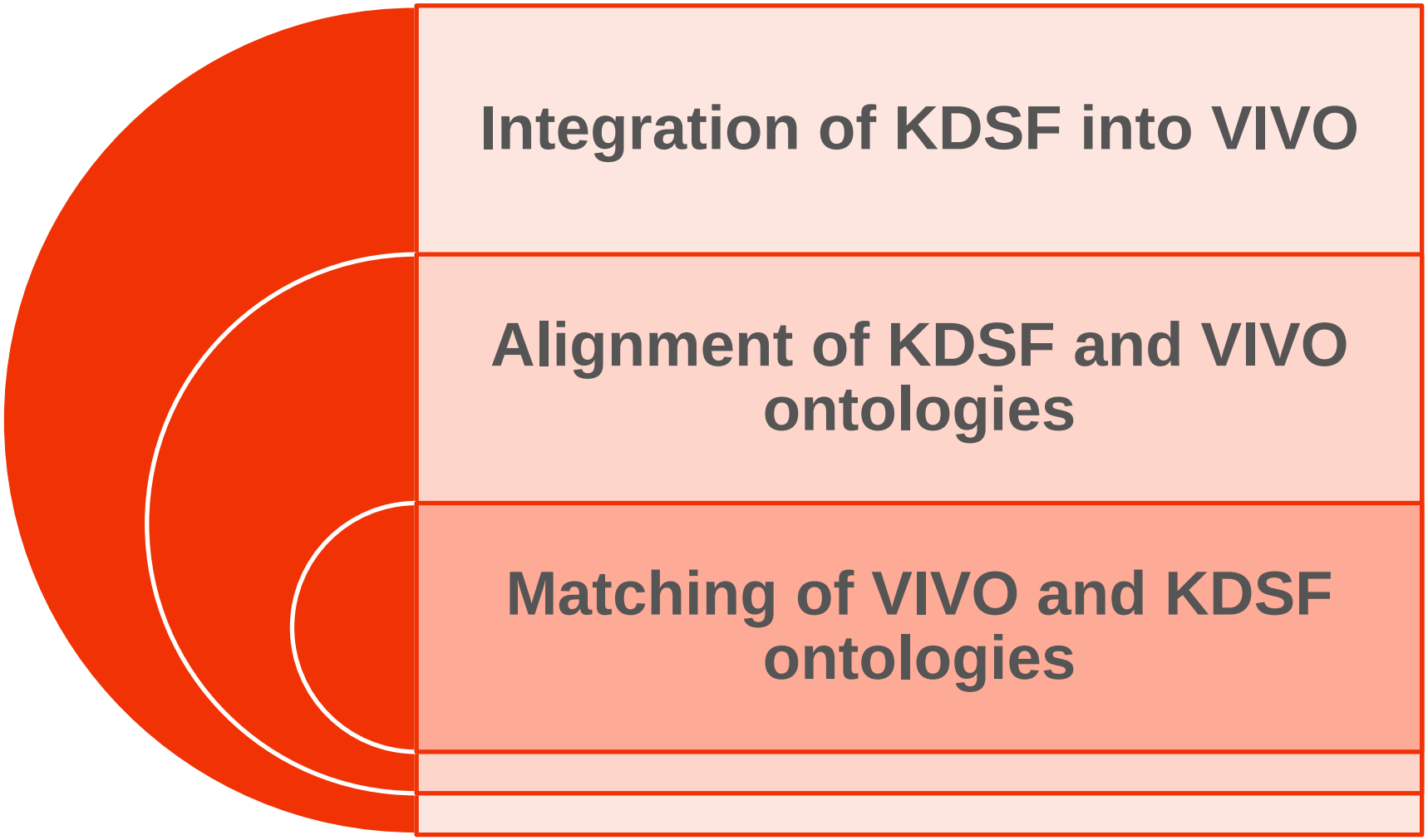
TIB-FIS-Discovery

- Profiles
- Representation on the Web

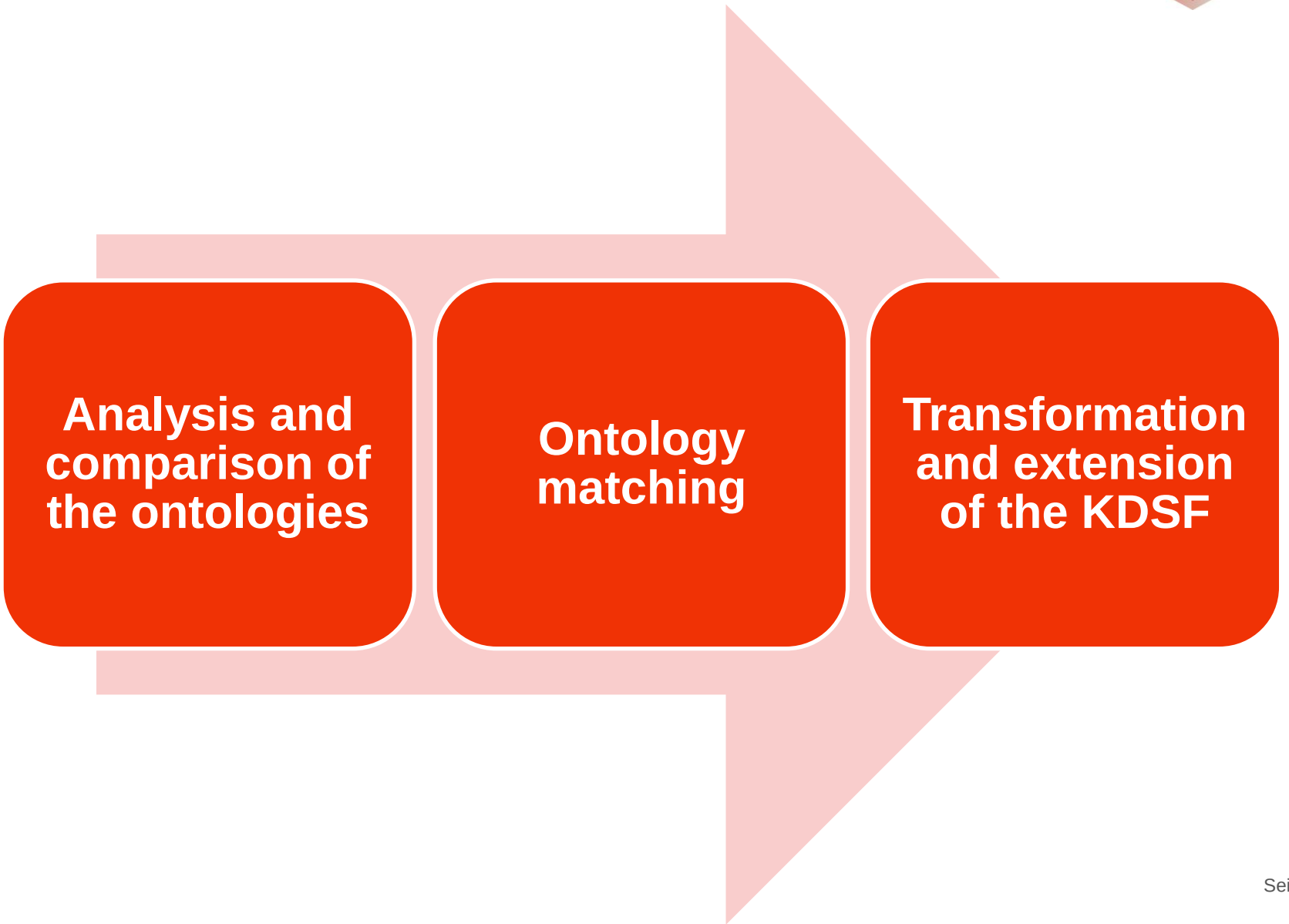
KDSF-VIVO

- Reporting
- Data mantainance

Implementation of KDSF in VIVO



Implementation on the ontology level



Analysis and comparison of VIVO and KDSF

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

Domain of interest

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

- Differencies 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

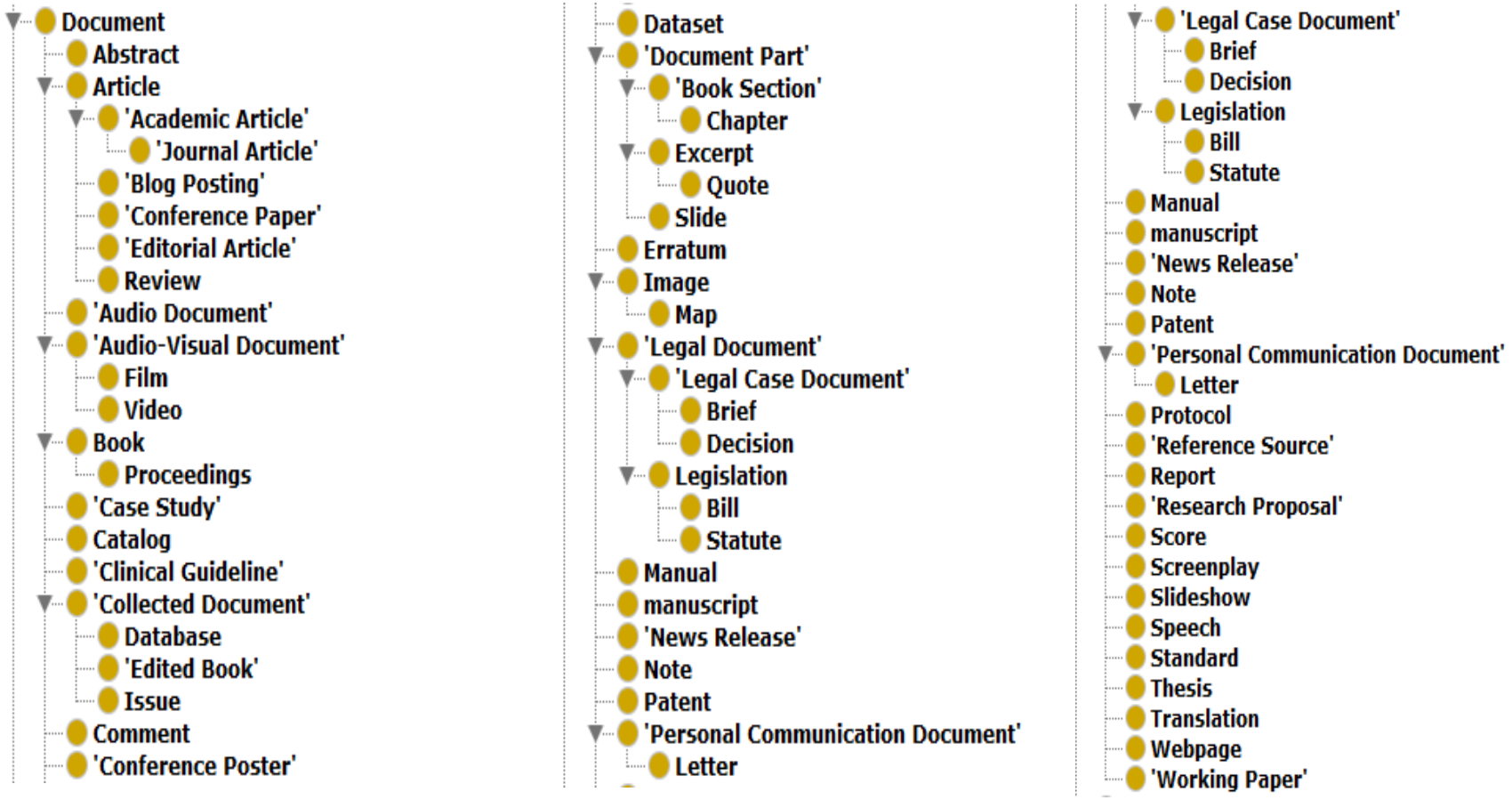
	VIVO Ontology	KDSF
Classes	414	123
Object properties	198	36
Datatype properties	217	53
Instances	14	0

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

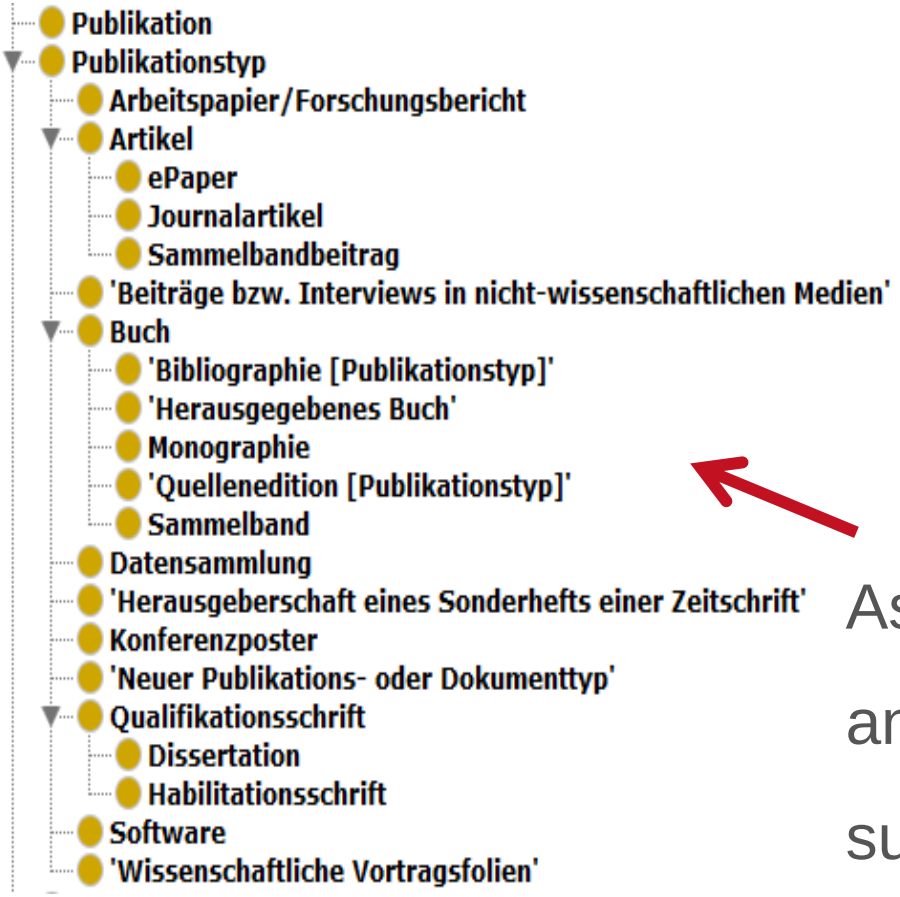
Comparison and analysis of VIVO and KDSF

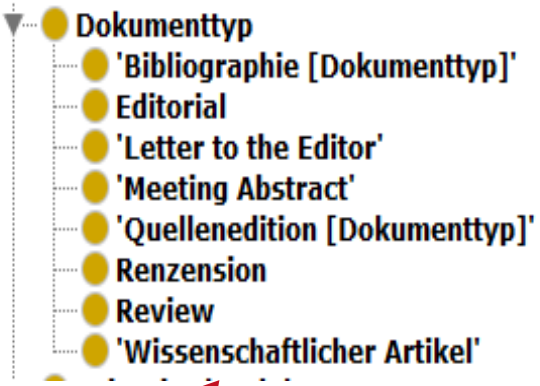
- Granularity of publication-related part of VIVO



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 scalabale 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

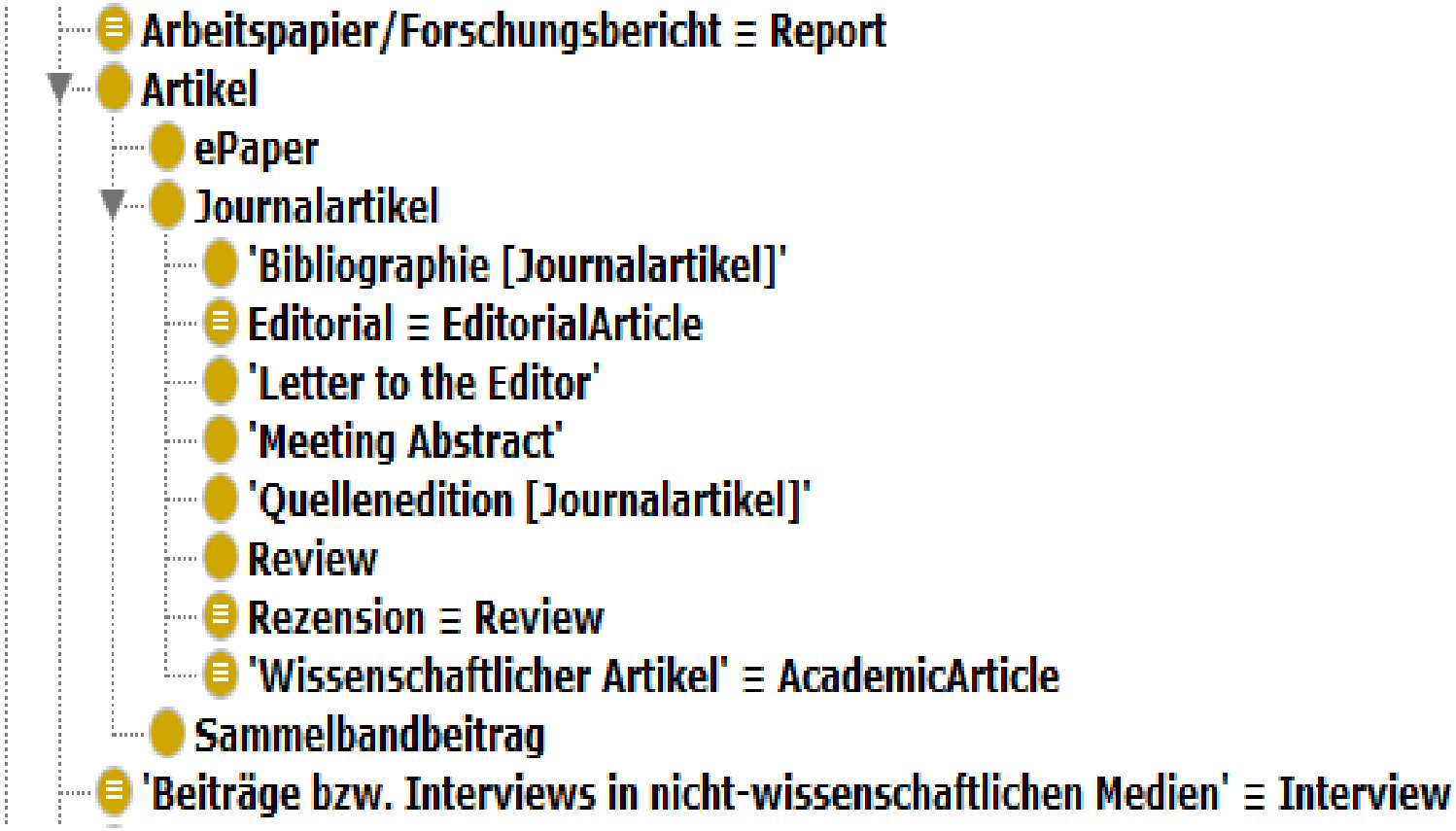
- Results: 4 mappings with equivalent correspondences

kdsf:Patent ≡ bibo:Patent
kdsf:Review ≡ vivo:Review
kdsf:Person ≡ foaf:Person
kdsf:Software ≡ obo:ERO_0000071 (software)

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

Manual matching

- Intial 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 unitended 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 ⊆ kdsf:Buch
kdfs:Buch ≡ bibo:Book

kdsf:BibliographiePublikationstyp ⊆ 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 specializiation 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***

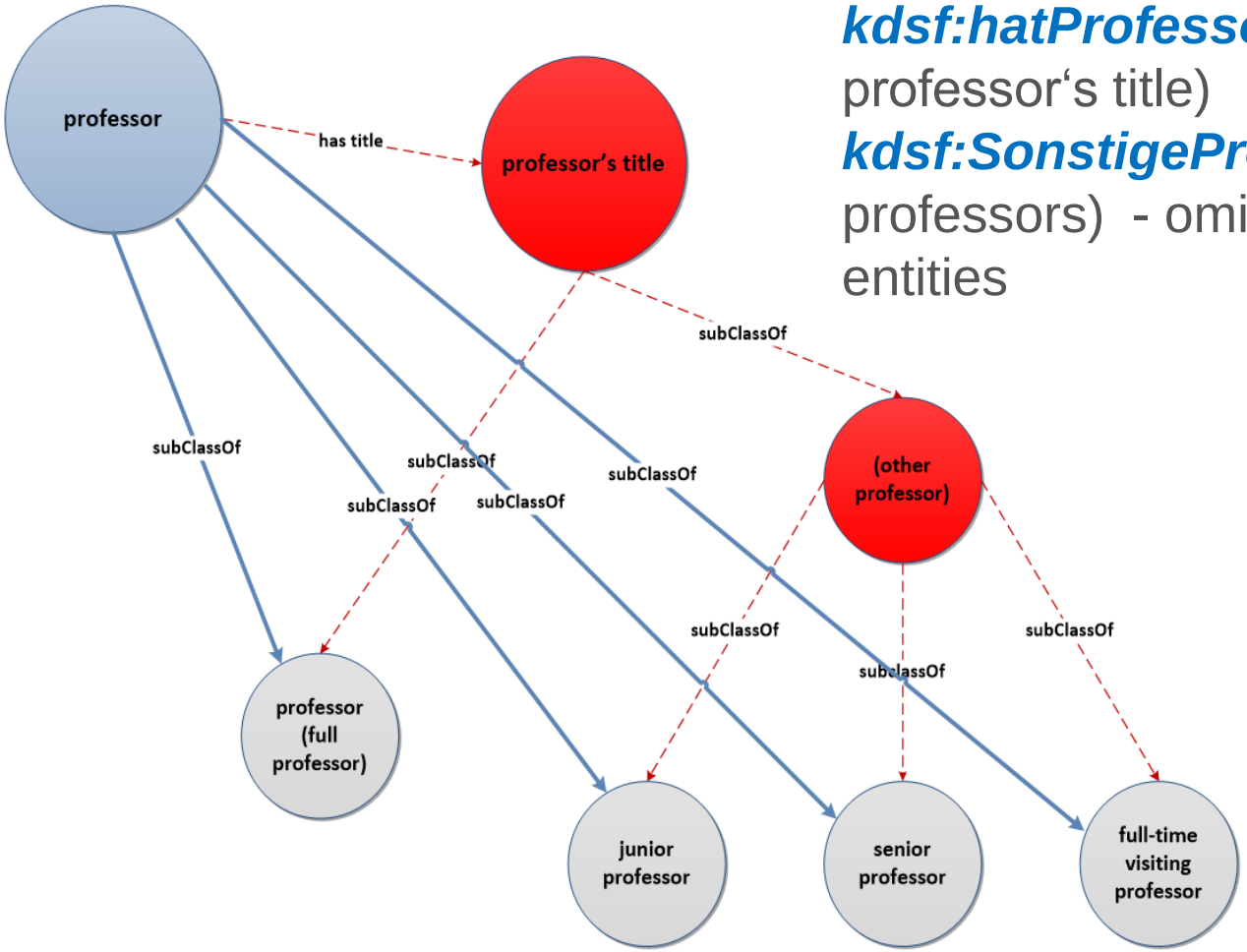
+ range: ***kdsf-vivo:Zugriffsrechte***

+ named Indviduals: ***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

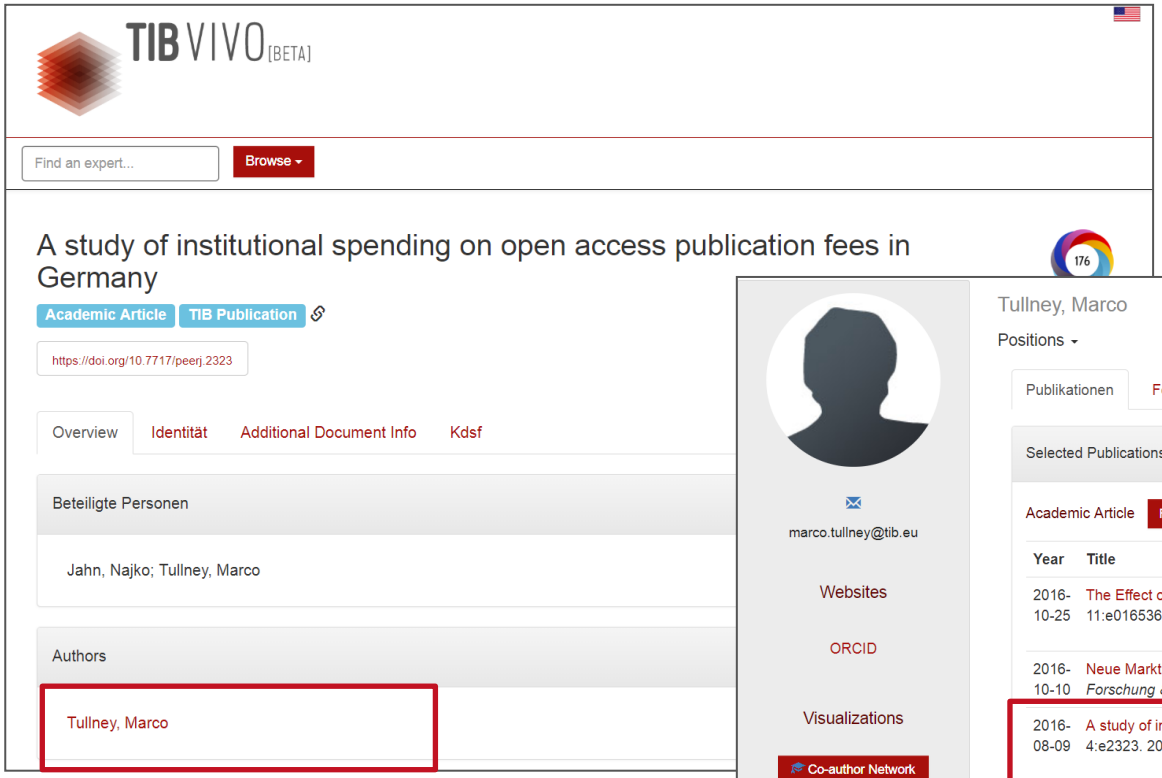
- Intermediate classes + related properties

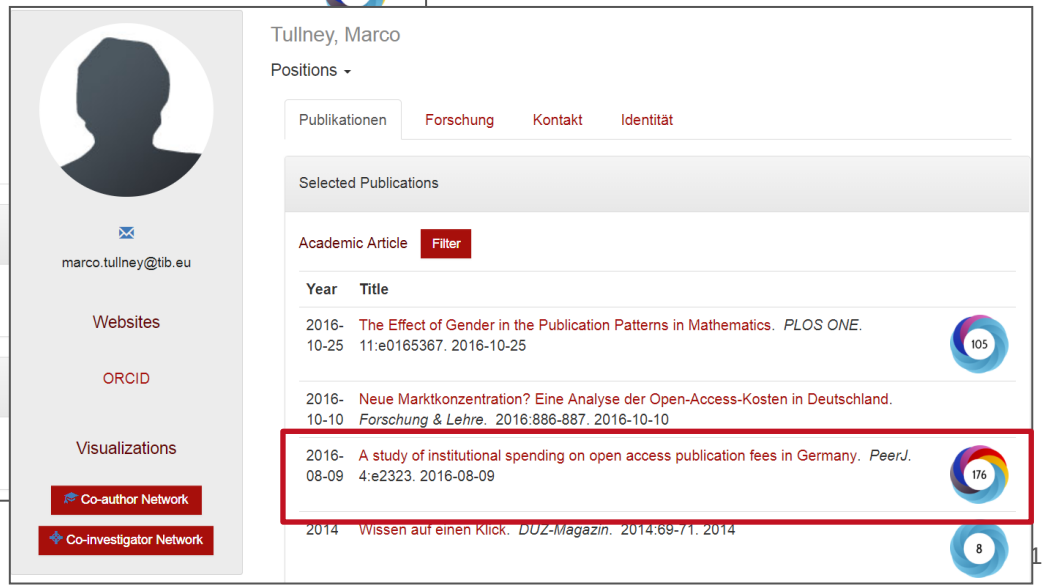


kdsf:Professorenbezeichnung (professor's title)
kdsf:hatProfessorenbezeichnung (has professor's title)
kdsf:SonstigeProfessoren (other professors) - omitted as unnecessary entities

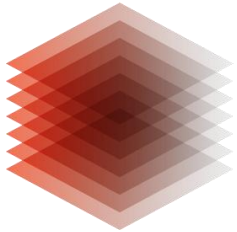
Additional Axioms

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





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 extention -> 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