# Research information platform with VIVO

Rolf Guescini

Berlin University Alliance

6. VIVO-WORKSHOP 15-16.06 2022

Berlin University Alliance

Freie Universität Berlin

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN

# UNDER DEVELOPMENT

- Use of Data classification schemas (disciplinary and interdisciplinary)
- Text Mining (Title, Abstracts) to classify works
  - 3 Clusters of Excellence:
    - Matters of Activity
    - Neurocure
    - Science on Intelligence
- Ontology to link structures of the BUA member organisations

# Modular, Reusable and Extensible

- Taking example from the research and insights made by the HERCULES project in Spain, we strive for
  - **Reusability**
    - re-modelling any concept that could be represented by any other ontology should be avoided
  - **Extensibility**
    - development of a core ontology which can be extended by specific sub-modules
  - **Maintainability**
    - **modular design** seeks an easier maintainability of the ontology

# Follow FAIR principles

- **Findable -** through a stable, persistent identifier and including metadata
  - use of IDs that are permanently assigned to a resource even if the location the resource changes over time, such as https://w3id.org
  - Multilingual
    - labels and descriptions of both classes and properties of the ontology should be expressed in at least German and English
- **Accessible -** through the universal HTTP protocol, API's and SPARQL endpoints
- **Interoperable -** using widely adopted vocabularies
- **Reusable -** published using user licenses that promote reusability

# ONTOLOGY EXTENSIONS

- Extension of the VIVO ontology for BUA-specific academic entities.
- Ontology for disciplinary data classification using EUDAT-B2FIND list
- Ontology for interdisciplinary data classification using the list of Research Core Dataset.
- Ontology for BUA alliance organisations

# VIVO ONTOLOGY EXTENSIONS

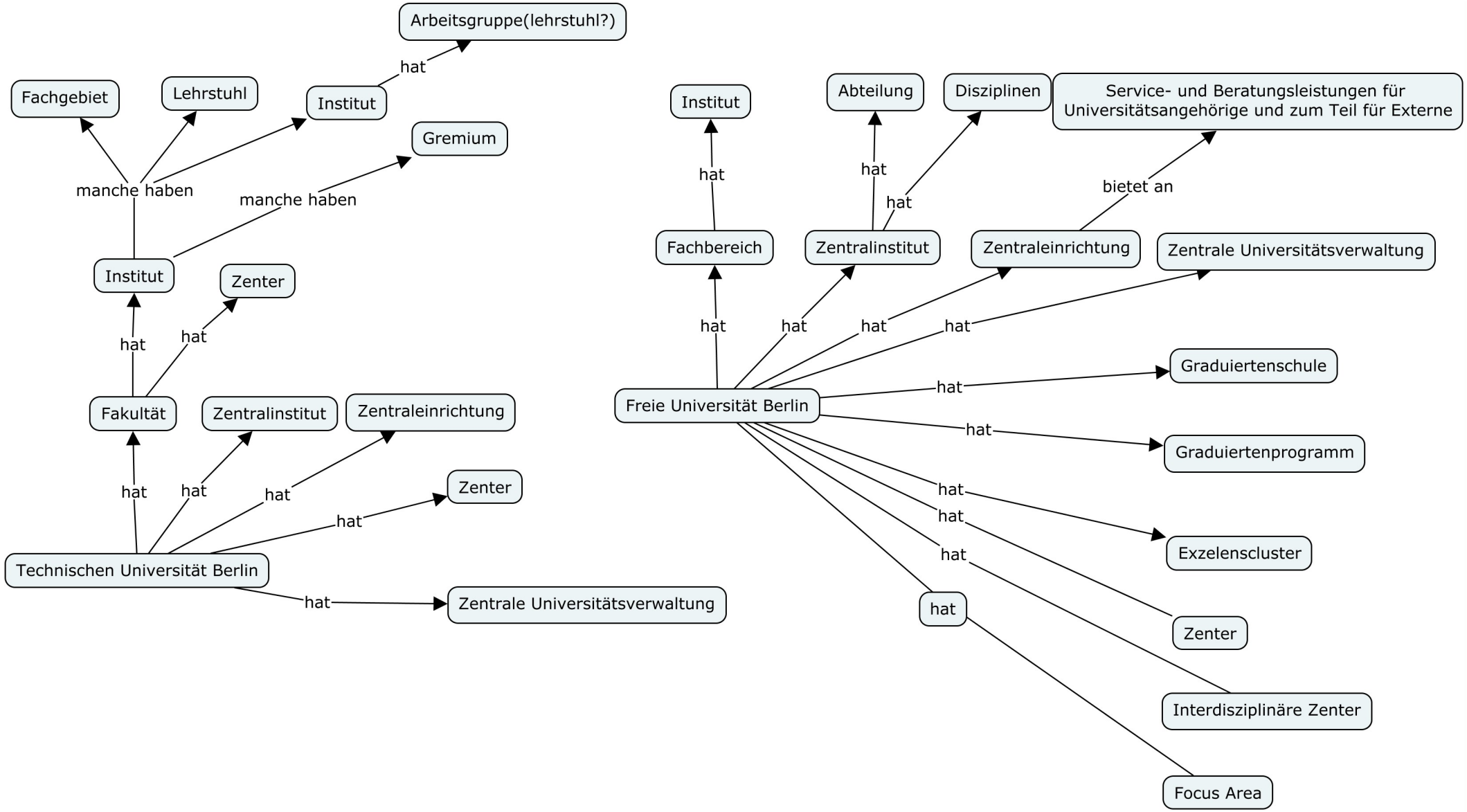
## VIVO.DE

- ERO_0000014
  - **SonderforschungsbereichDFG**
- foaf:Organization
  - **Fakultaet**
  - **Graduiertenkolleg**
  - **Graduiertenschule**
  - **Lehrstuhl**
  - **ServiceUnit**
  - **ZentraleEinrichtung**
- Grant
- IAO_0000030
- Project
  - **Habilitation**
  - **Promotion**
- Thesis
  - **Dissertation**
  - **Habilitationsschrift**

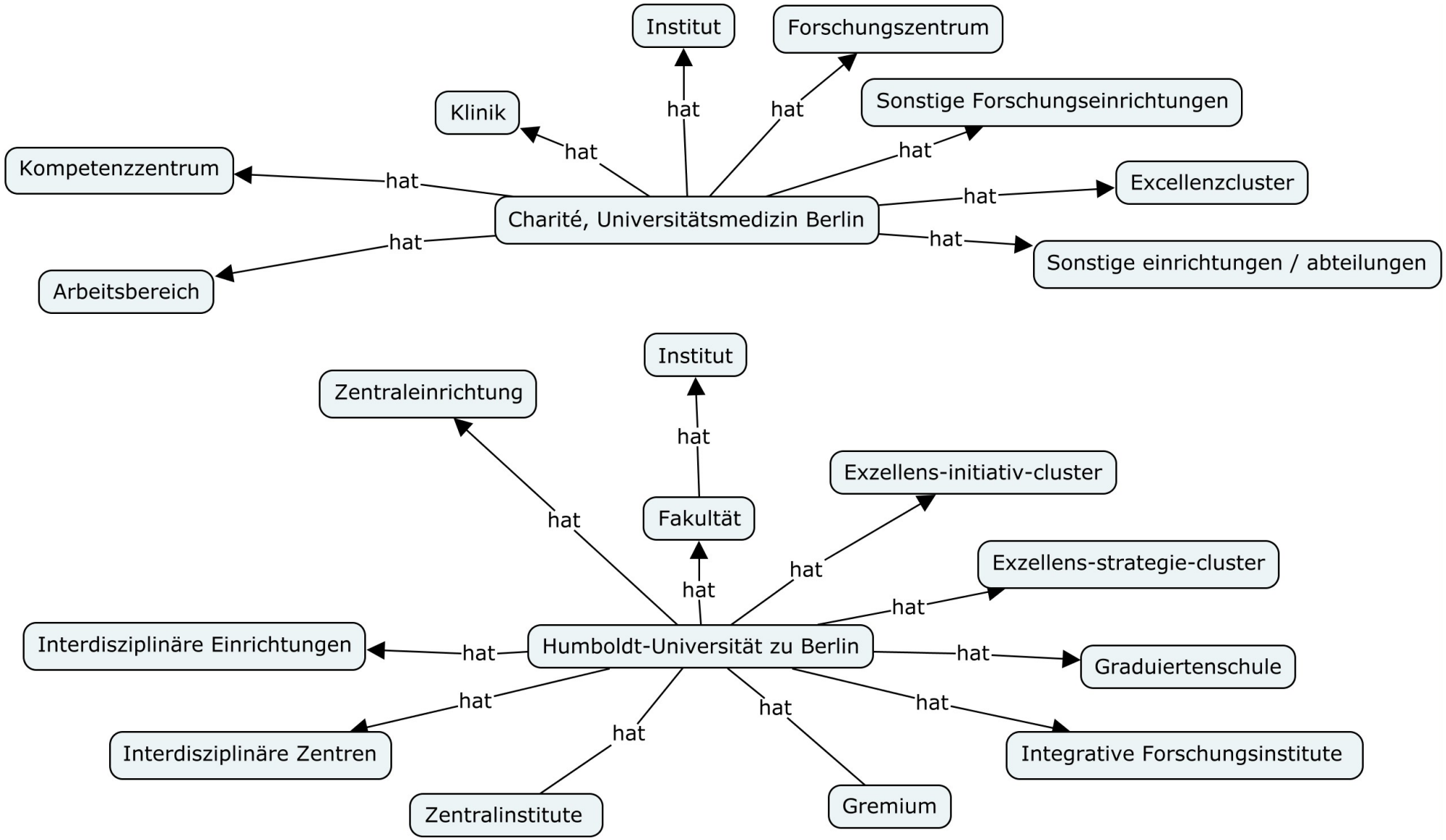
## VIVO-BUA

- AcademicDepartment
  - **Fachbereich**
  - **Fachgebiet**
- AcademicUnit
  - **Arbeitsgruppe**
- Center
  - **Forschungszentrum**
    - **SonstigeForschungseinrichtung**
  - **Interdisziplinaerezentrum**
  - **Kompetenzzentrum**
- ClinicalOrganization
  - **Klinik**
- Committee
  - **Gremium**
- Department
  - **Arbeitsbereich**
  - **forschungsallianz**
- Document
  - **ContributionInEditedVolume**
  - **EditedVolume**
  - **SoftwarePublication**
- foaf:Group
  - **Gremium**
- foaf:Organization
  - **Arbeitsbereich**
  - **Arbeitsgruppe**
  - **Exzellenzcluster**
  - **Exzellenzinitiativcluster**
  - **Exzellenzstrategiecluster**
  - **Fachbereich**
  - **Fachgebiet**
  - **forschungsallianz**
  - **Forschungszentrum**
  - **Gremium**
  - **Integrativeforschungsinstitut**
  - **Interdisziplinaereeinrichtung**
  - **Interdisziplinaerezentrum**
  - **Klinik**
  - **Kompetenzzentrum**
  - **SonstigeForschungseinrichtung**
  - **Zentralinstitut**
- foaf:ResearchOrganization
  - **Exzellenzcluster**
- Institute
  - **Integrativeforschungsinstitut**
  - **Zentralinstitut**
- ResearchOrganization
  - **Integrativeforschungsinstitut**
  - **Interdisziplinaerezentrum**
- ServiceUnit
  - **Interdisziplinaereeinrichtung**

# BUA ORGANISATION ONTOLOGY



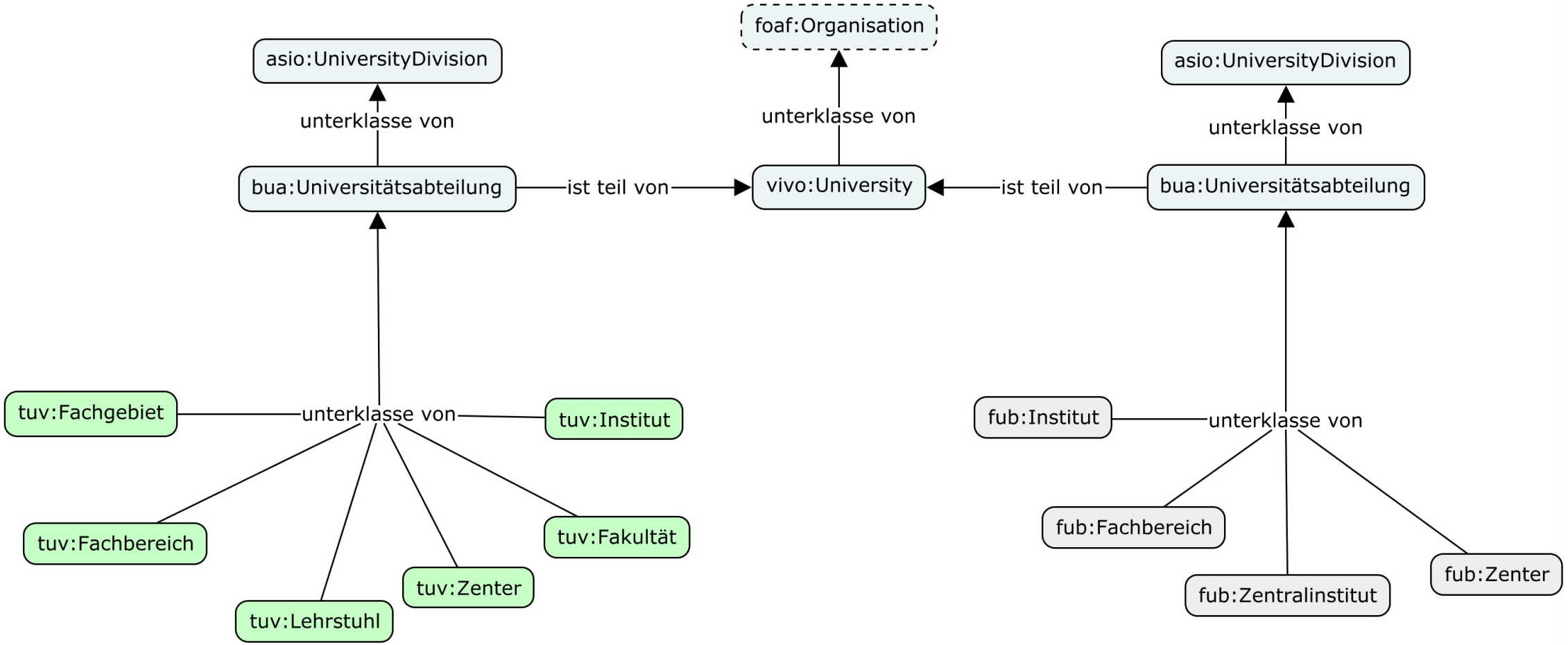
# BUA ORGANISATION ONTOLOGY



# Open World Assumption and scalability

- Academic organisational hierarchies although seemingly similar in structure, have a high variation in both structure and semantics, usually using different schemes to structure and classify their entities
- at outset, the the initial process of expressing an organisation ontology to describe the alliance provided much insight into the variegated nature of not only the structures between universities, but also within universities them selves
- trying to harmonise the organisational structures into a **closed world assumption**, highlights the challenges posed by semantic differences of the terms as well as the hierarchies used for organisational entities
- The **Open World Assumption** both states that Knowledge is never complete and always contextual, and points in the direction of modularisation and polymorphism
- Priority must been given to the flexibility and easy extensibility of the ontologies to ensure completeness and scalability

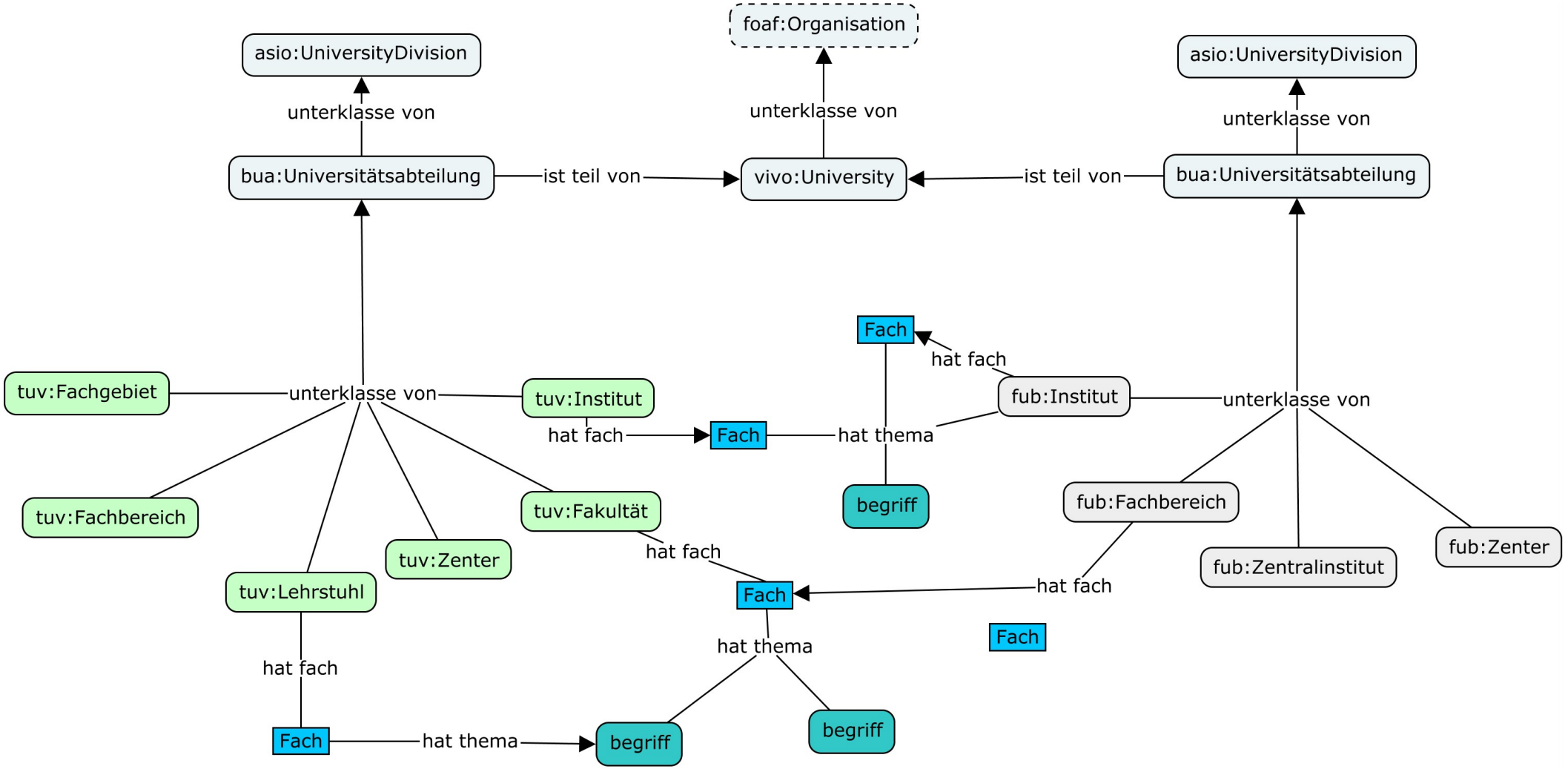
# BUA ORGANISATION ONTOLOGY
## - Local extensions



# The collocation challenge

- a modular ontology design, allows any organisational entity (university) to develop its own sub ontology describing local particularities and semantic variation
- the openness and flexibility in the inherent modularity of the RDF standard and the open world assumption comes with the price of a heightened complexity when wanting to collocate semantically equivalent or similar entities
- classification with common, open vocabularies is the first obvious way to bridge the gap and provide filters to diminish ambiguity
  - one could envision several topic vocabularies which could represent the different domains within the alliance as well as the use of national or global ones
  - A classification with subject matter could be another step which could help narrow down ambiguities during collocation as they are common across institutions

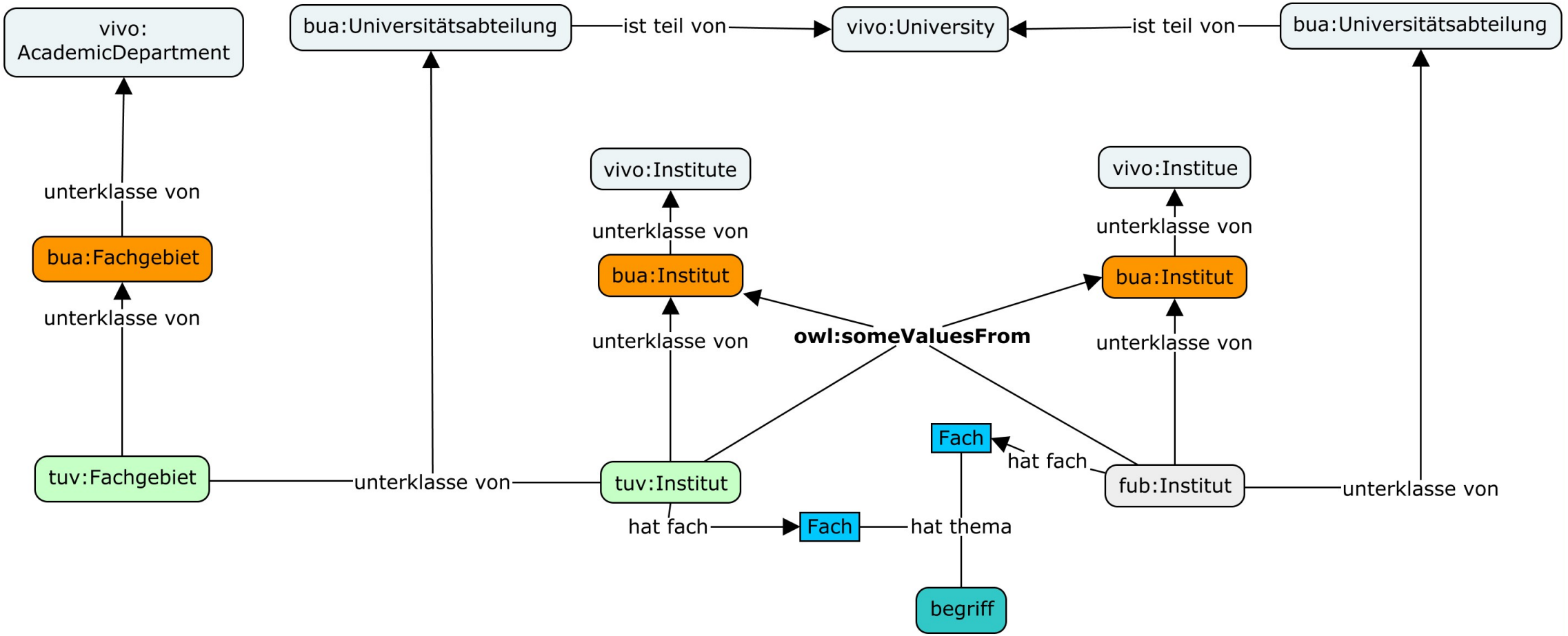
# BUA ORGANISATION ONTOLOGY
## - subject matters and topics



# Import, inherit and constrain

- importation of and heredity from a generic class provides the opportunity to express the correct semantics for each inherited entity, but provides few clues or possibilities to refine and narrow amount of ambiguity when collocating entities
- A step further towards less ambiguity  is the definition of a set of classes which inherit the generic classes, defined with OWL constraints indicating for instance
  - obligatory data properties
  - obligatory object properties
  - ranges
- using **owl:allValuesFrom** and **owl:someValuesFrom** properties express the premises for each entity

# BUA ONTOLOGY:Inheritance



# Import, inherit and constrain

- a further step is to enrich the of a set of classes which inherit the generic classes, with OWL constraints observed by the OWL reasoner to classify a instance on basis of met restrictions

# BUA ORGANISATION ONTOLOGY: Further Constraints

**Description: Institute**

SubClass Of (Anonymous Ancestor)

- 'has knowledge area' only Concept
- title some xsd:string
- 'has contact info' some Organization
- provides only Service
- 'has accreditation' only Accreditation
- 'located in' only Feature
- description some xsd:string
- homepage only owl:Thing
- 'predecessor organization' only Organization
- 'has role' only Role
- title only xsd:string
- abbreviation only rdfs:Literal
- 'has part' only Organization
- 'date/time interval' only 'Date/Time Interval'
- participates only Activity
- 'has reservable' only Reservable
- 'successor organization' only Organization
- identifier some xsd:string
- keyword only xsd:string
- keyword only xsd:string
- 'related by' only Relationship

Instances

Target for Key

Disjoint With

- 'Research Group', Center, Company, Foundation, 'Government Agency', University, 'Ethics Comitee', Department, Consortium

**Description: Funding Organization**

Equivalent To

- Organization
  and (promotes some
  ('Funding Program' or 'Funding Source'))
  and (funds only Funding)
  and (promotes only
  ('Funding Program' or 'Funding Source'))

# Further classification schemes

- with subject matters and domain vocabularies in place we already have a to bridge specialised entities
- adding disciplinary and interdisciplinary classification not only provides classified action for research artefacts in the graph as a whole, but can also provide added strength of expression when connected to the domain vocabularies and subject matters
- for the **research discipline vocabulary,** we have chosen to work with metadata from the EUDAT-B2FIND service, from the EUDAT Collaborative Data Infrastructure (or EUDAT CDI) is one of the largest infrastructures of integrated data services and resources supporting research in Europe
- For the **interdisciplinary vocabulary** we have chosen to use the **Classification for interdisciplinary research fields from KDSF** (Standard für Forschungsinformationen in Deutschland)
- The classifications are both expressed as **skos:Concept** / **skos:ConceptSchemes** in both German and English

# DISCIPLINES ONTOLOGY

- ✚ Engineering Science
- ✚ Humanities
- ✚ Life Sciences
- ✚ Natural Sciences
- ✚ Social and Behavioural Sciences

- ○ ━ Engineering Science
  - ▪ ✚ Computer Science, Electrical and System Engineering
  - ▪ ✚ Construction Engineering and Architecture
  - ▪ ✚ Materials Science and Engineering
  - ▪ ✚ Mechanical and industrial Engineering
  - ▪ ✚ Thermal Engineering/Process Engineering
- ○ ━ Humanities
  - ▪ ✚ Ancient Cultures
  - ▪ ✚ Fine Arts, Music, Theatre and Media Studies
  - ▪ ✚ History
  - ▪ ✚ Linguistics
  - ▪ ✚ Literary Studies
  - ▪ ✚ Philosophy
  - ▪ ✚ Social and Cultural Anthropology
  - ▪ ✚ Theology and Religion Studies
- ○ ━ Life Sciences
  - ▪ ✚ Agriculture, Forestry, Horticulture and Veterinary Medicine
  - ▪ ✚ Biology
  - ▪ ✚ Medicine
- ○ ━ Natural Sciences
  - ▪ ✚ Chemistry
  - ▪ ✚ Mathematics
  - ▪ ✚ Physics
- ○ ━ Social and Behavioural Sciences
  - ▪ ✚ Economics
  - ▪ ✚ Education Sciences
  - ▪ ✚ Jurisprudence
  - ▪ ✚ Psychology
  - ▪ ✚ Social Sciences

- ━ Engineering Science
  - ▪ ━ Computer Science, Electrical and System Engineering
    - ▪ ━ Computer Science
      - ▪ Artificial Intelligence, Image and Language Processing
      - ▪ Computer Architecture, Computer Engineering and Embedded Systems
      - ▪ Information Science
      - ▪ Operating, Communication, Library and Information Systems
      - ▪ Research Data Management
      - ▪ Software Technology
      - ▪ Theoretical Computer Science
    - ▪ ✚ Electrical Engineering
    - ▪ ✚ Systems Engineering
  - ▪ ✚ Construction Engineering and Architecture
  - ▪ ✚ Materials Science and Engineering
  - ▪ ✚ Mechanical and industrial Engineering
  - ▪ ✚ Thermal Engineering/Process Engineering
- ━ Humanities
  - ▪ ━ Ancient Cultures
    - ▪ Ancient History
    - ▪ Archaeology
    - ▪ Classical Archaeology
    - ▪ Classical Philology
    - ▪ Egyptology and Ancient Near Eastern Studies
    - ▪ Prehistory
  - ▪ ✚ Fine Arts, Music, Theatre and Media Studies
  - ▪ ✚ History
  - ▪ ✚ Linguistics
  - ▪ ✚ Literary Studies
  - ▪ ✚ Philosophy
  - ▪ ✚ Social and Cultural Anthropology
  - ▪ ✚ Theology and Religion Studies

# INTERDISCIPLINARY RESEARCH FIELDS

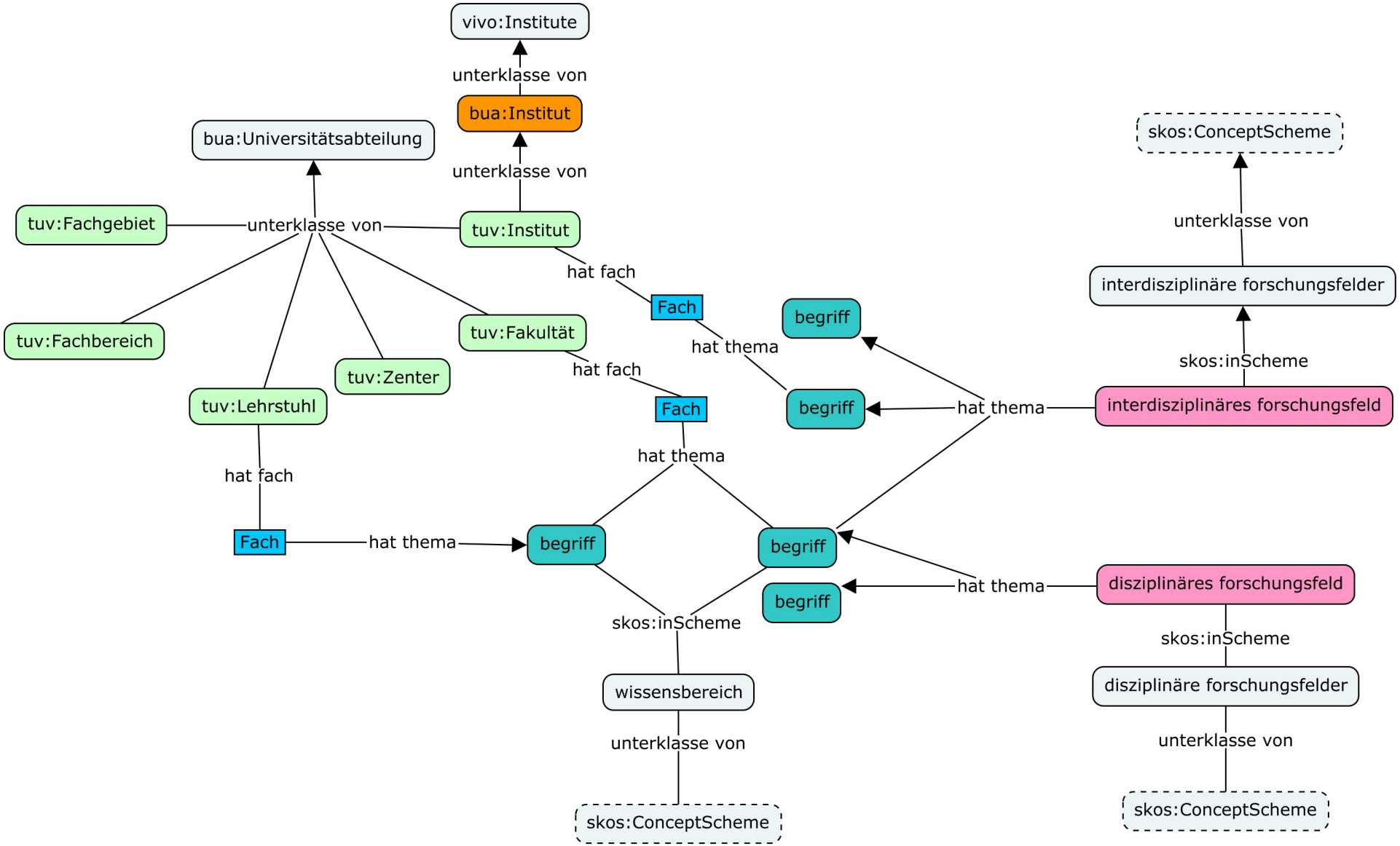
- ✚ Work and Economy
- ✚ Earth and cosmos
- ✚ Globalization and sustainablili
- ✚ Industry
- ✚ Information technology
- ✚ Infrastructure
- ✚ Cognition and knowledge
- ✚ Culture
- ✚ Life and wellbeing
- ✚ Materials
- ✚ People and Society
- ✚ Nature and environment
- ✚ Technology
- ✚ Science

- ━ Work and Economy
  - Working world and design
  - Work and Economy - General
  - Digital Economy
- ━ Earth and cosmos
  - Earth and cosmos - General
  - Matter
  - Planet Earth
  - Regions of the world
  - Space
- ━ Globalization and sustainablility
  - Development Cooperation
  - Globalization and Sustainability - General
  - Migration
  - Sustainability
- ━ Industry
  - Disposal and recycling
  - Industry - general
  - Smart Production
  - Robotics
- ━ Information technology
  - Information security
  - Information systems
  - Information technology
  - Information technology - general
  - Internet of things
  - Artificial intelligence and big data
  - Simulation research
  - Metropolitan areas and urban development
  - Building and living
  - Infrastructure - general
  - Infrastructure and networks
  - Mobility, transport and traffic
- ━ Cognition and knowledge
  - Innovation
  - Cognition and Knowledge - General
  - Learning and learning processes
  - Human brain
  - Language and language learning
  - Knowledge transfer and knowledge representation
- ━ Culture
  - Creativity and performance
  - Culture - general
  - Cultural goods and heritage
  - Media
- ━ Life and wellbeing
  - Artificial or synthetic life
  - Life and Wellbeing - general
  - Living beings
  - Disease prevention
  - Therapy and healing
  - Cells and genes
- ━ Materials
  - Materials - General
  - Surfaces and interfaces
  - Polymers
  - Elementary Materials

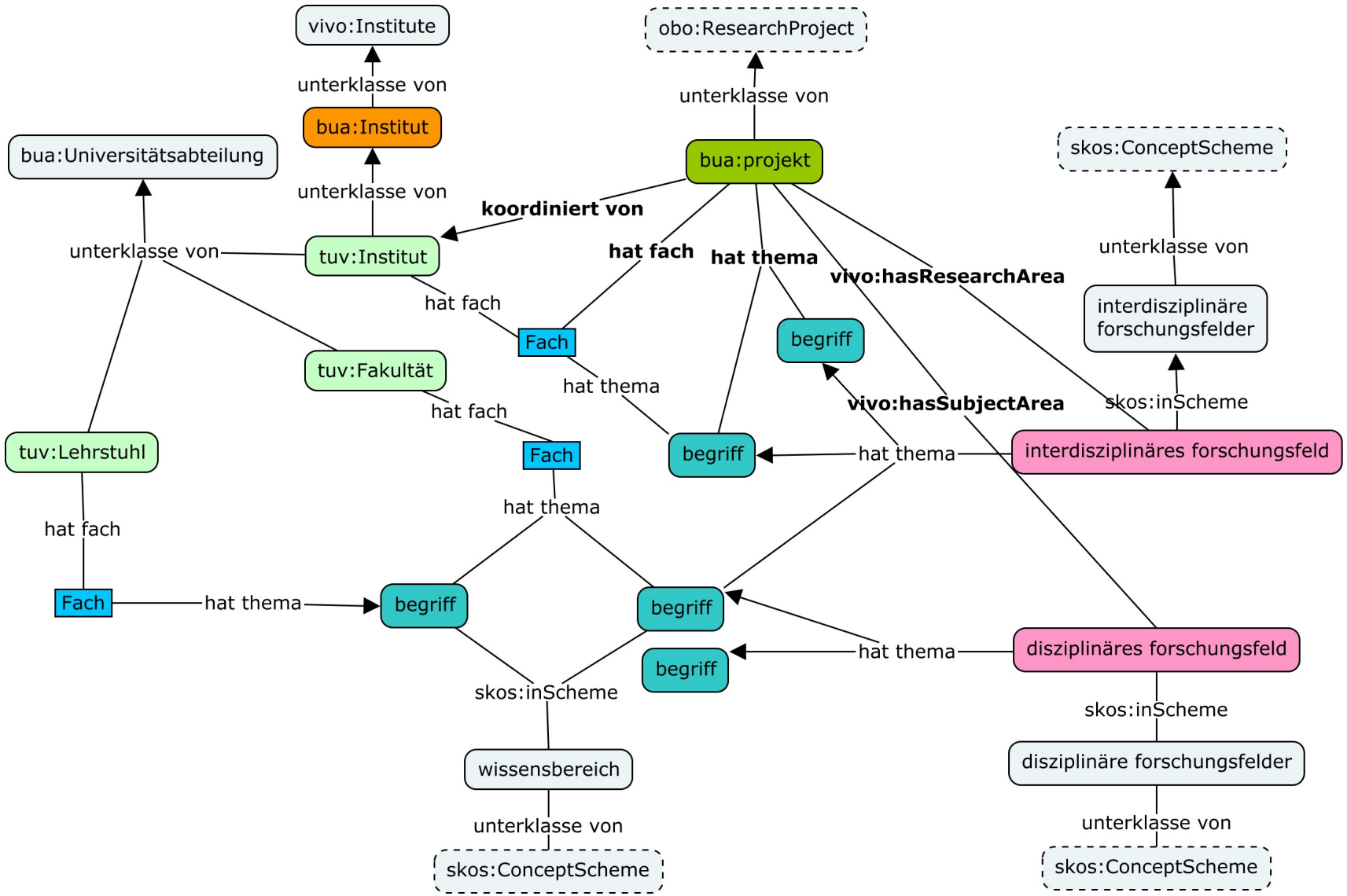
# Classification and Linking Concept schemes

- Using the disciplinary and interdisciplinary vocabularies we can classify entities in our graph  with the corresponding properties from the vivo ontology:
  - **vivo:hasSubjectArea**
  - **vivo:hasResearchArea**
- linking the disciplinary vocabulary to the interdisciplinary vocabularies could be an option that could heighten the expressivity of the graph, one which requires manual work
- an additional opportunity lies in linking the terms to the subject matter concept scheme and the domain specific concept schemes
- this provides an additional dimension of data to consider when trying to minimise ambiguity while collocating between graph entities

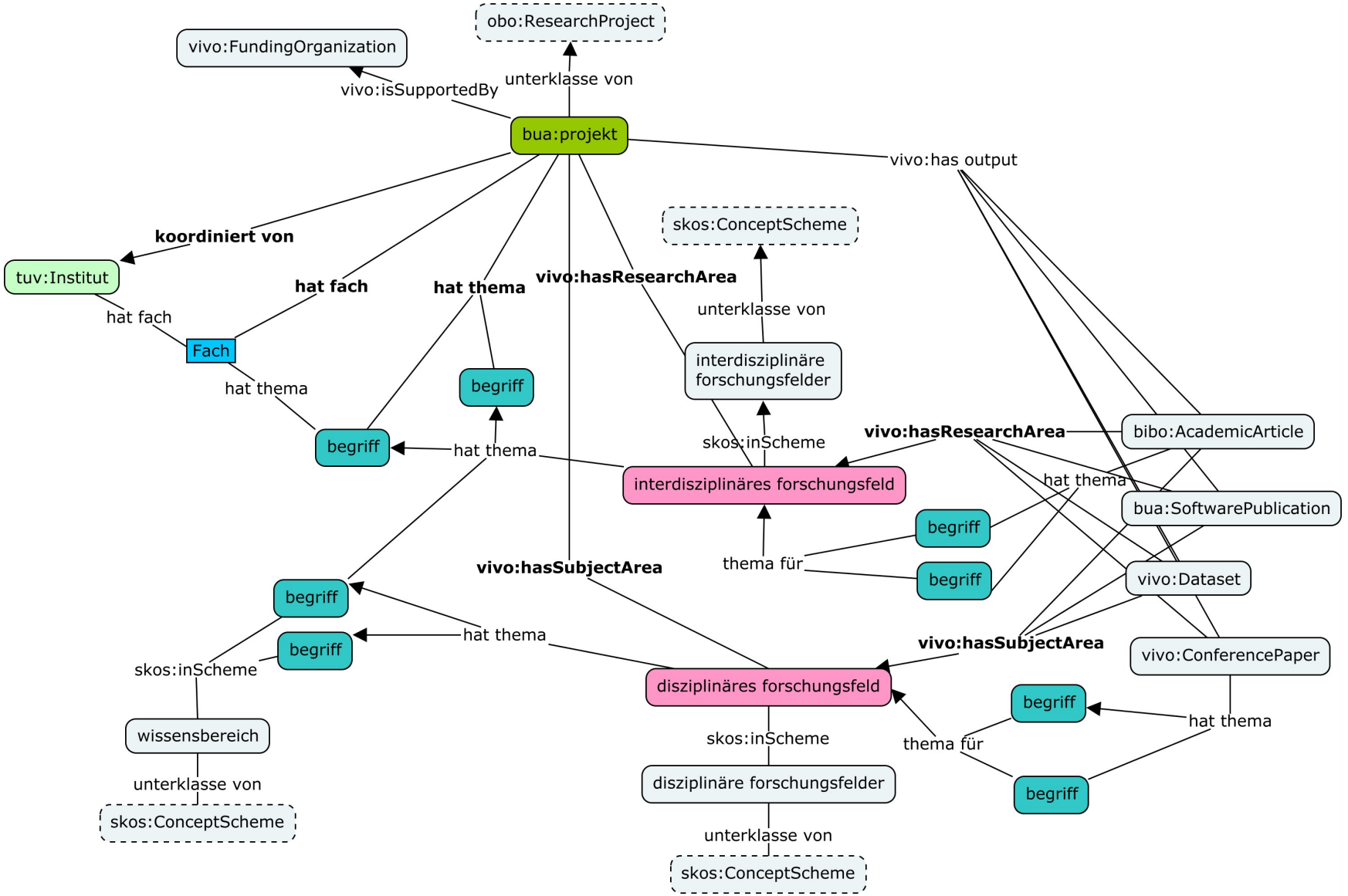
# BUA ONTOLOGY
## Classification and Linking Concept schemes



# BUA ONTOLOGY
## Classification and Linking Concept schemes



# BUA ONTOLOGY
## Classification and Linking Concept schemes



# NEXT STEPS :
# Steps towards classification

- Classification of Excellence clusters / SFBs / projects according to the Interdisciplinary (KDSF) vocabulary with metadata from the Excellence clusters / SFBs / projects or other sources
- Processing of project and publication data to prepare a training corpus for classification
  - we have 2 options
    - supervised learning would be ideal, for this we need tagged publications (if possible tagged by the authors)
    - unsupervised learning would be an option (clustering by KDFS keywords), but then we still need to find researchers to confirm the clusters/suggest improvements at the end

# NEXT STEPS

- Integration of more data within the BUA
- Development of stable Pipelines
- Multiple Frontends for different Alliance members
- Handling of general data protection regulations
- Multilingual Support

Danke!