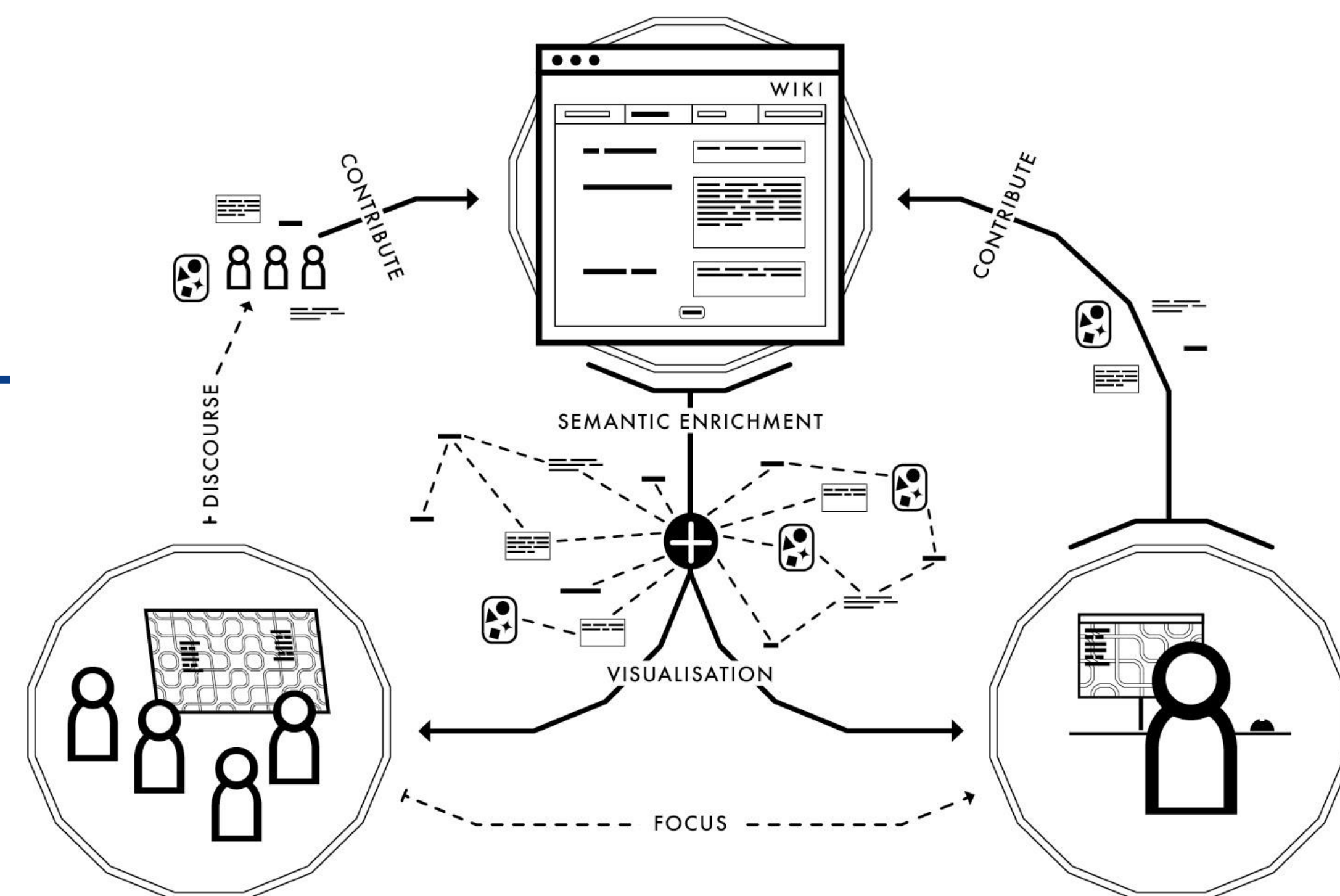


# Project IKON

## Sichtbarmachung von Wissenstransfer- potenzialen in Forschungsmuseen

Jonas Oppenländer  
Wissenschaftlicher Mitarbeiter

Institute of Computer Science  
Freie Universität Berlin



1. Ontology

2. Data

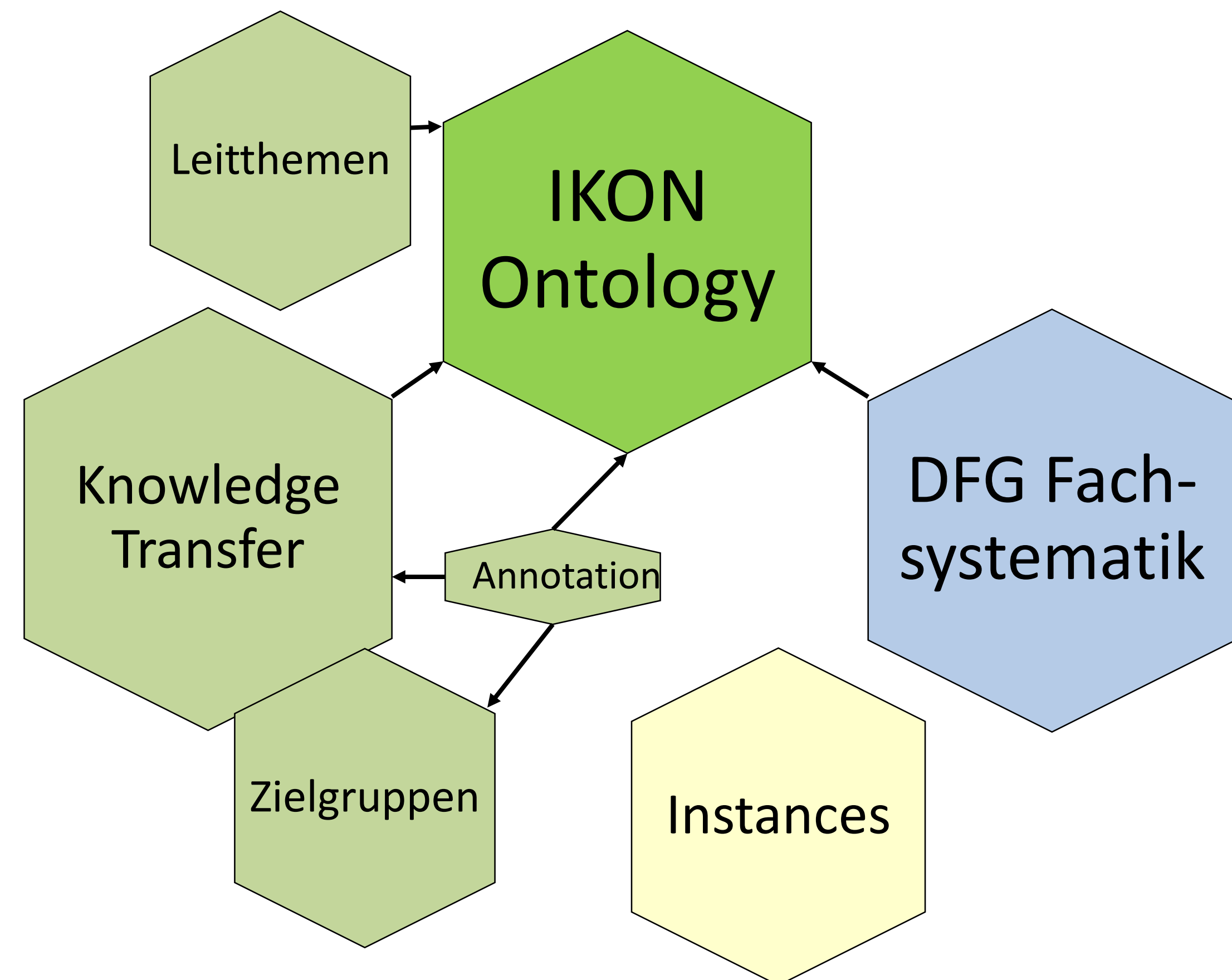
3. Application

# Ontologies

# Ontologies

► [github.com/FUB-HCC/IKON-ontology](https://github.com/FUB-HCC/IKON-ontology)

- Development: [dev.owl](#)
- Production: [prod.owl](#)
- Prod is produced by running reasoner in Protégé and exporting
- Namespaces (URIs) are currently temporary



# Wiki & Ontology

## General Principles

- ikon:mapping property identifies the wiki fields
- Named individuals populate dropdown lists
- Excel masterlist for status info and feedback
- Separate wiki pages → separate terminologies
  - Leitthemen
  - Zielgruppen
  - Knowledge transfer activities

Forschungsprojekte

Jonas Oppenländer Diskussion Einstellungen Beobachtungsliste Beiträge Abmelden

Seite Diskussion Lesen Mit Formular bearbeiten Quelltext bearbeiten Versionsgeschichte

Suche

Navigation

Hauptseite  
Aktuelle Projekte  
Projekte seit 2016  
Archiv bis 2015  
Alle Projekte im System  
Leitthemen  
Letzte Änderungen  
Projekt neu anlegen

Inhaltsverzeichnis

AG Forschungsprojekte  
Bedarfsanalyse  
Gespeicherte Anfragen  
Empfohlene Aufgaben  
Glossar  
Neu, bitte sichten  
WT-Aktivitäten

Werkzeuge

Links auf diese Seite  
Änderungen an verlinkten Seiten  
Datei hochladen  
Spezialseiten  
Druckversion  
Permanenter Link  
Seiteninformationen  
Attribute anzeigen

Überblick English Schlagworte

Verbundprojekt NeFo3: Unterstützung der Biodiversitätsforschung in Deutschland und ihre Einbindung in Wissenschafts-Politik-Prozesse, insbesondere in die Implementierung des IPBES-Arbeitsprogramms- Teil 2 -MfN Berlin

Juni 1992, Rio de Janeiro: die Vereinten Nationen kommen mit zahlreichen Vertretern von Nichtregierungsorganisationen (NGOs) zusammen. Die 12-tägige Konferenz mit parallelem NGO-Forum hat insgesamt über 20.000 Teilnehmer. Eine zentrale Erkenntnis der bahnbrechenden Konferenz über Umwelt und Entwicklung von Rio war, dass die biologische Vielfalt auf unserem Planeten eine elementare Bedeutung auch für Entwicklungsmöglichkeiten und Lebensbedingungen von uns Menschen hat - und dass akuter Handlungsbedarf besteht. Und so wurde als Ergebnis unter anderem das [Abkommen zur Biologischen Vielfalt \(CBD\)](#) verabschiedet. Diese Konvention hat drei Ziele: den Schutz der biologischen Vielfalt, die nachhaltige Nutzung ihrer Bestandteile sowie Zugangsregelung und gerechter Ausgleich von Vorteilen, die aus der Nutzung genetischer Ressourcen entstehen.

**Vernetzte Biodiversitätsforschung in Deutschland**

April 2009, Deutschland: Der Biodiversitätsforschung geht es besser als vielen Arten. Sie hat sich in den letzten Jahrzehnten immer mehr als eigenes Feld etabliert, Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Disziplinen forschen an zahlreichen Institutionen und sammeln Erkenntnisse über Ursachen und Zusammenhänge des Artensterbens. Allein in Deutschland existieren etwa 130 Institute wie das MfN. Das [Netzwerk-Forum Biodiversitätsforschung Deutschland \(NeFo\)](#) hat es sich nun zum Ziel gesetzt, diese fragmentierte Forschungslandschaft zu vernetzen und nach außen sichtbar zu machen. Auf der Plattform gibt es neben vielen Informationen zum Thema auch einen Forschungsatlas, der all die Orte verzeichnet, an denen Biodiversitätsforschung betrieben wird. Das NeFo-Team, bestehend aus Wissenschaftlerinnen und Wissenschaftlern des Helmholtz-Zentrums für Umweltforschung (UFZ) und des MfN, identifiziert außerdem aktuell relevante Themen und organisiert Veranstaltungen mit und für die Forschenden. Doch die interne Vernetzung und thematische Begleitung ist nur einer von zwei Schwerpunkten der Aktivitäten von NeFo. Als Schnittstelle zwischen Wissenschaft, gesellschaftlichen Akteuren und der Politik schlägt das Netzwerk die Brücke zwischen nationalen Forschungsaktivitäten, paneuropäischem Dialog und internationaler Umweltpolitik.

**Der Weltbiodiversitätsrat**

April 2012, Panama: Der [Weltbiodiversitätsrat \(IPBES\)](#), eine UN-Organisation zur wissenschaftlichen Politikberatung, wird gegründet. Denn so wichtig der Austausch innerhalb der nationalen Forschungsgemeinschaft ist - dass deren wertvolles Wissen dorthin gelangt, wo politische Entscheidungen getroffen werden, ist für die Lösung der drängenden Probleme essentiell. Nur dann kann es Wirkung entfalten, in dem es das Handeln von Staaten, Unternehmen und jedem Einzelnen beeinflusst. NeFo hat den Prozess zur Gründung dieses Gremiums von Anfang an begleitet und betreut bis heute inhaltlich wie strukturell den deutschen und zum Teil den europäischen Beitrag zur Arbeit des Rates. Der Bedarf zur Zusammenarbeit über Länder- und Fachdisziplinengrenzen hinweg ist größer denn je: Die biologische Vielfalt unserer Erde nimmt stetig ab, und es ist bisher nicht gelungen, diese Entwicklung aufzuhalten oder auch nur zu bremsen.



# Ontologies

## ► Imported Concepts from:

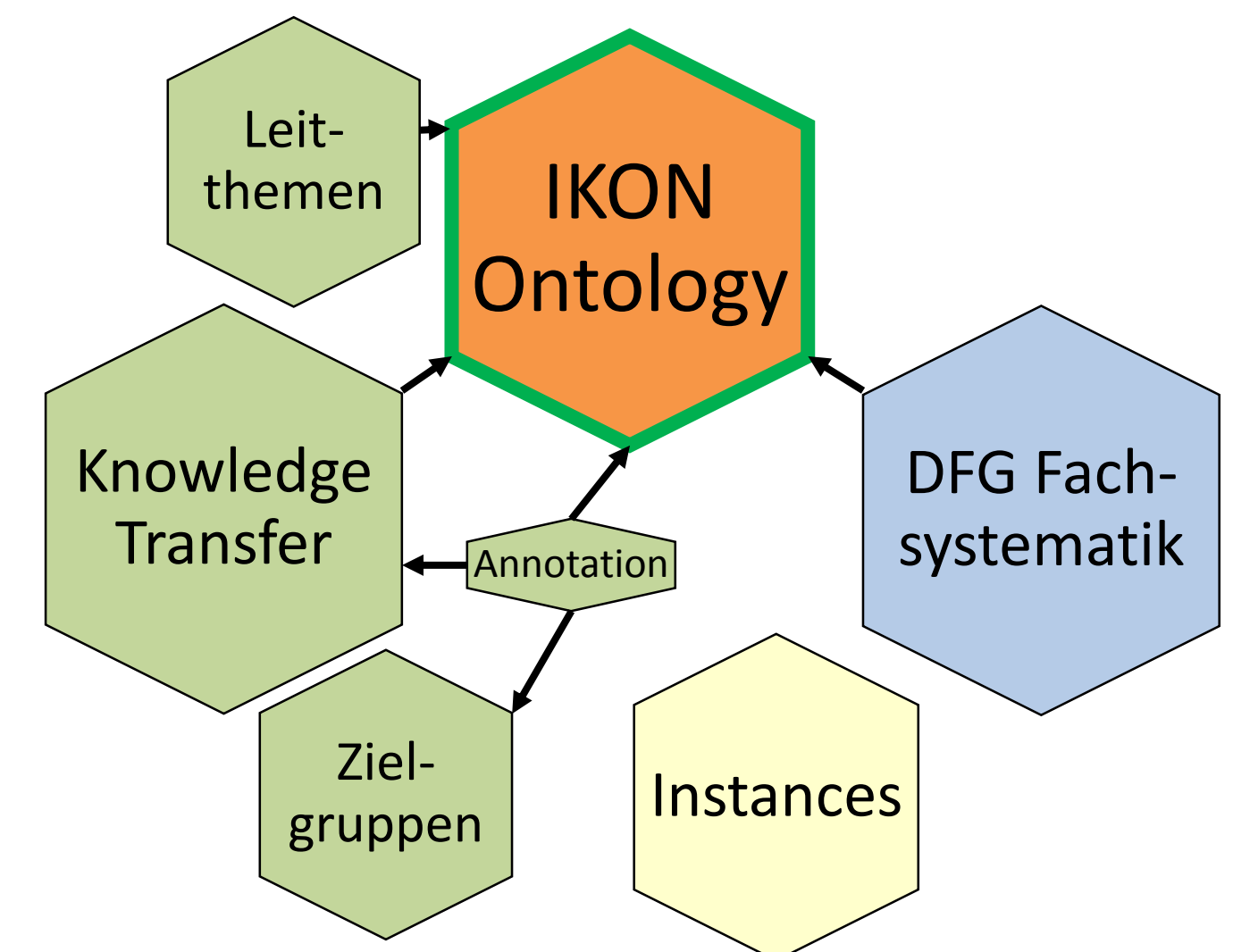
- FOAF, vCard
- Dublin Core
- FRAPO
- SKOS
- GND
- void, VANN
- ...



# IKON Ontology

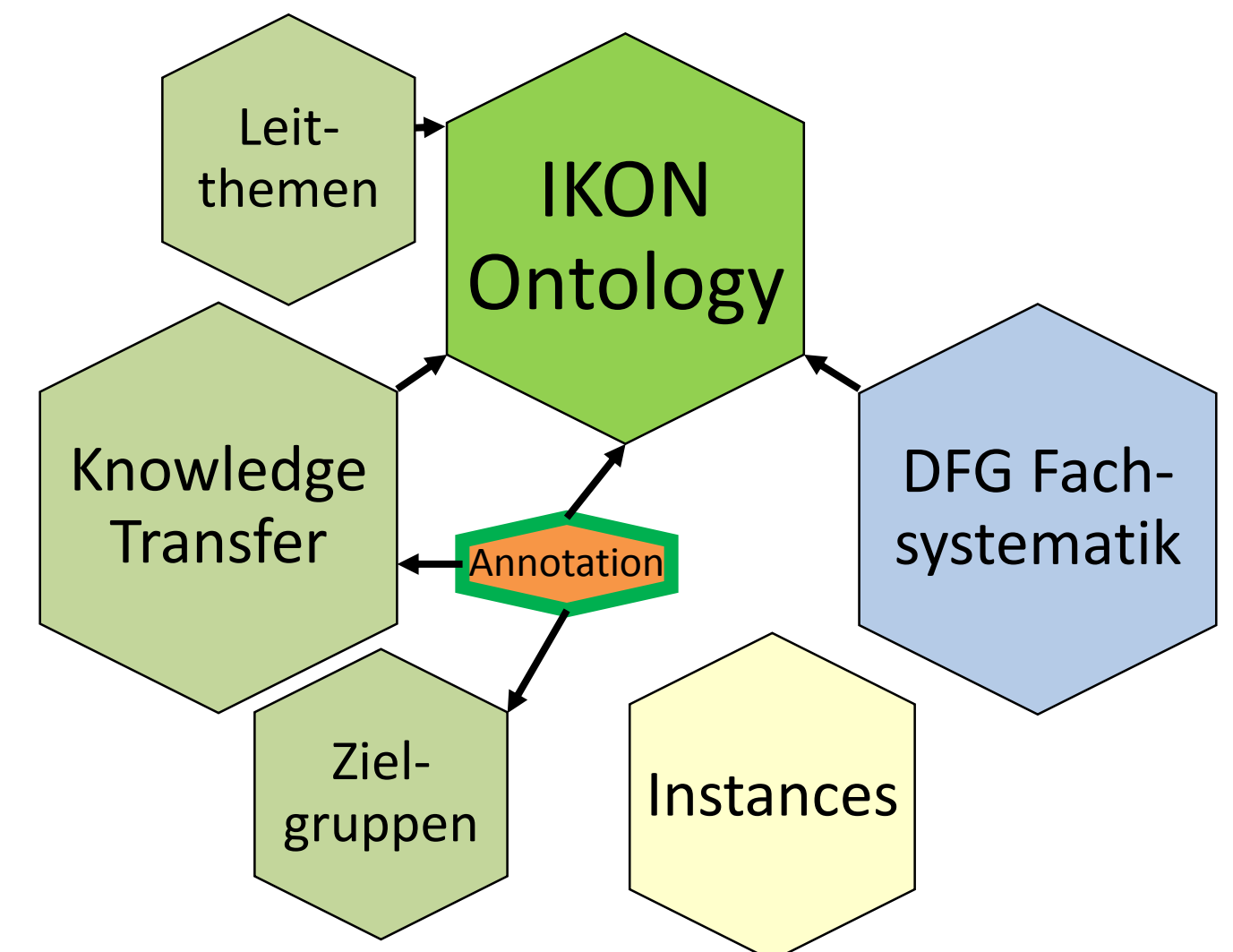
- Source: [github.com/FUB-HCC/IKON-ontology](https://github.com/FUB-HCC/IKON-ontology)
- Documentation: [docs](#) – Alternative documentation: [ontodoc](#)
- Basis: Research Core Data Set ([kerndatensatz-forschung.de](https://kerndatensatz-forschung.de))
- Imports all other terminologies
- Metrics (April 6, 2018):

Axiom	4021
Logical axiom count	1047
Declaration axioms count	654
Class count	163
Object property count	78
Data property count	71
Individual count	326
Annotation Property count	21
DL expressivity	SHOIF(D)



# Annotations

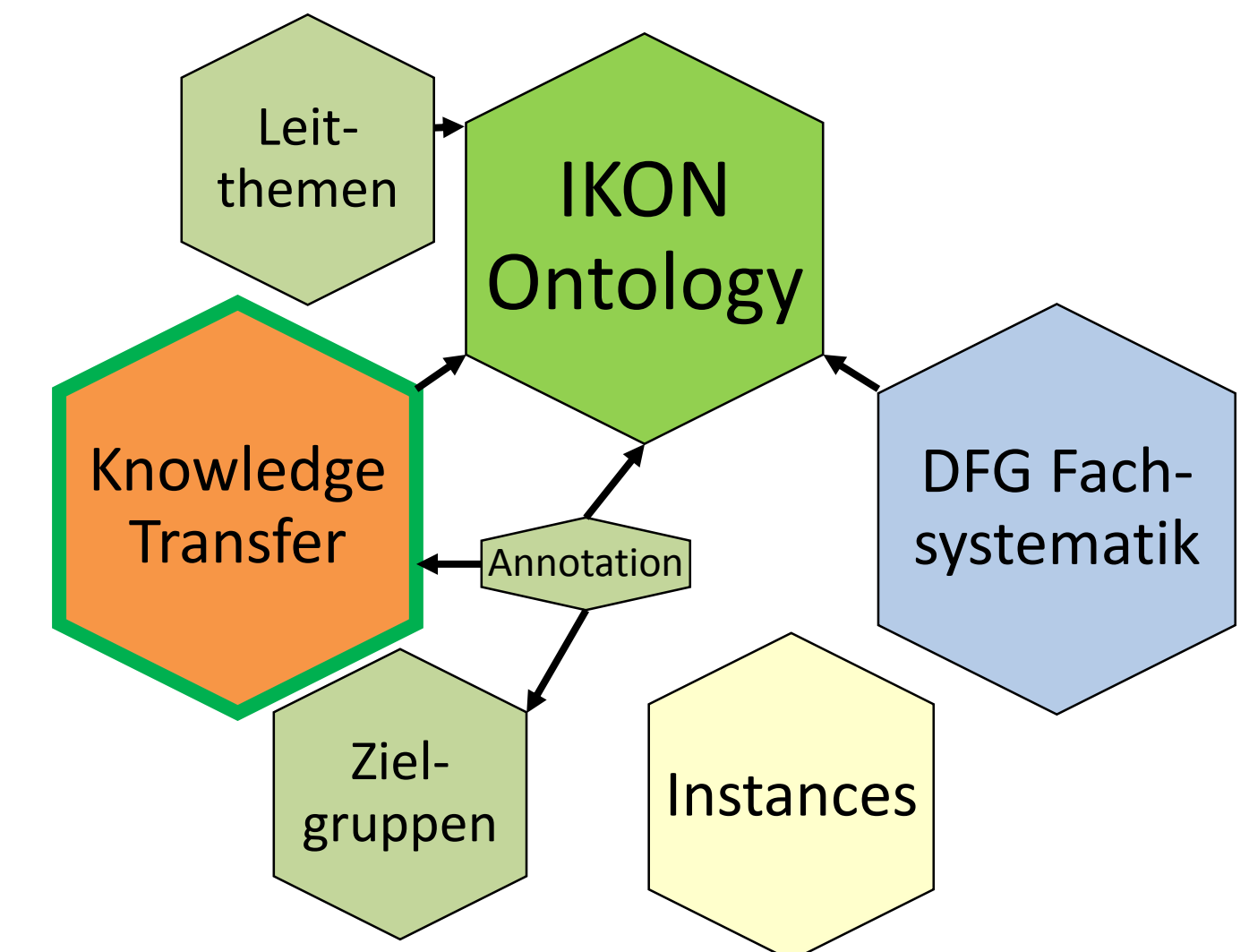
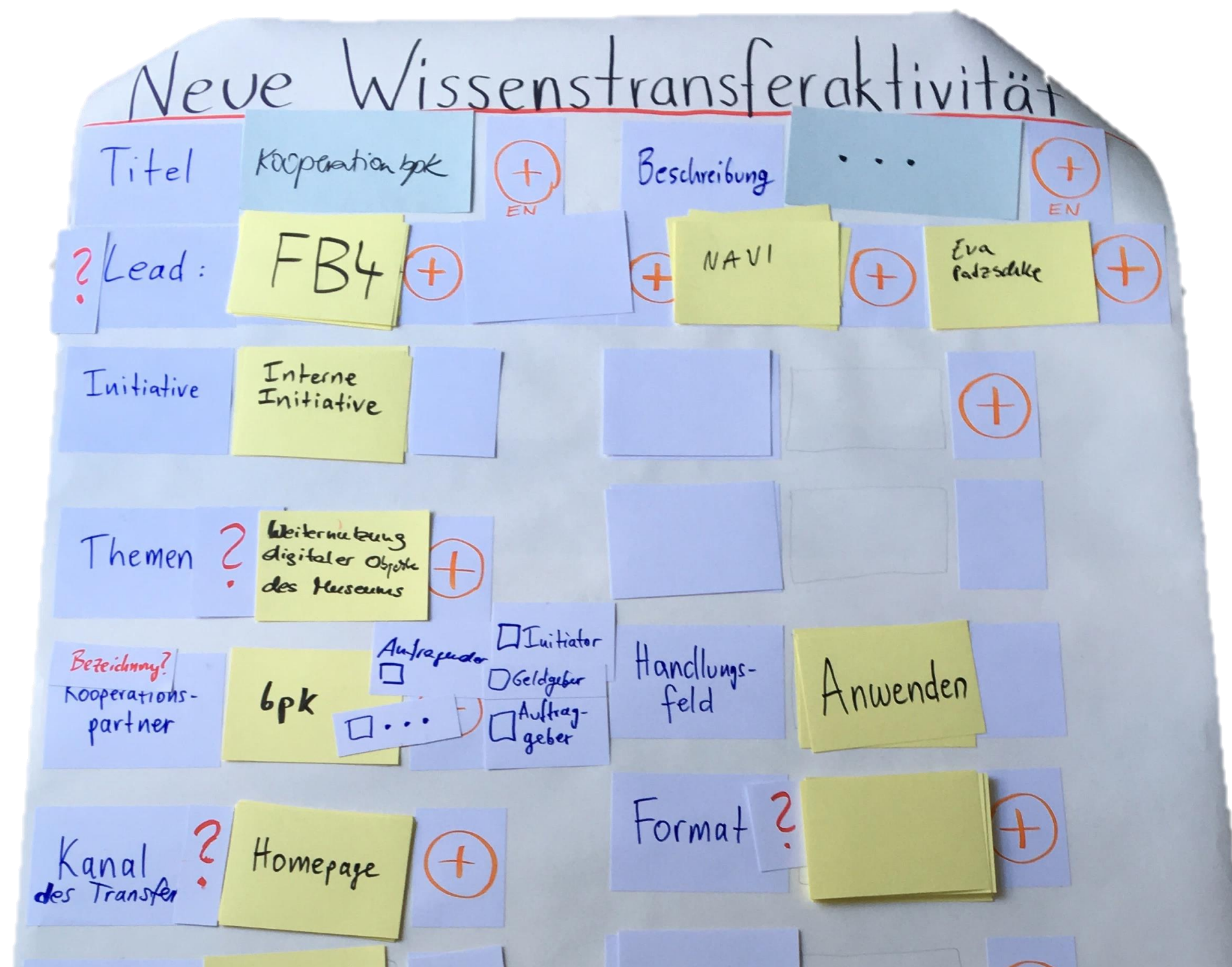
- Source: [imports/annotation](#)
- Shared annotation properties
- Imported by three other ontologies
  - KT
  - ZG
  - IKON





# Knowledge Transfer

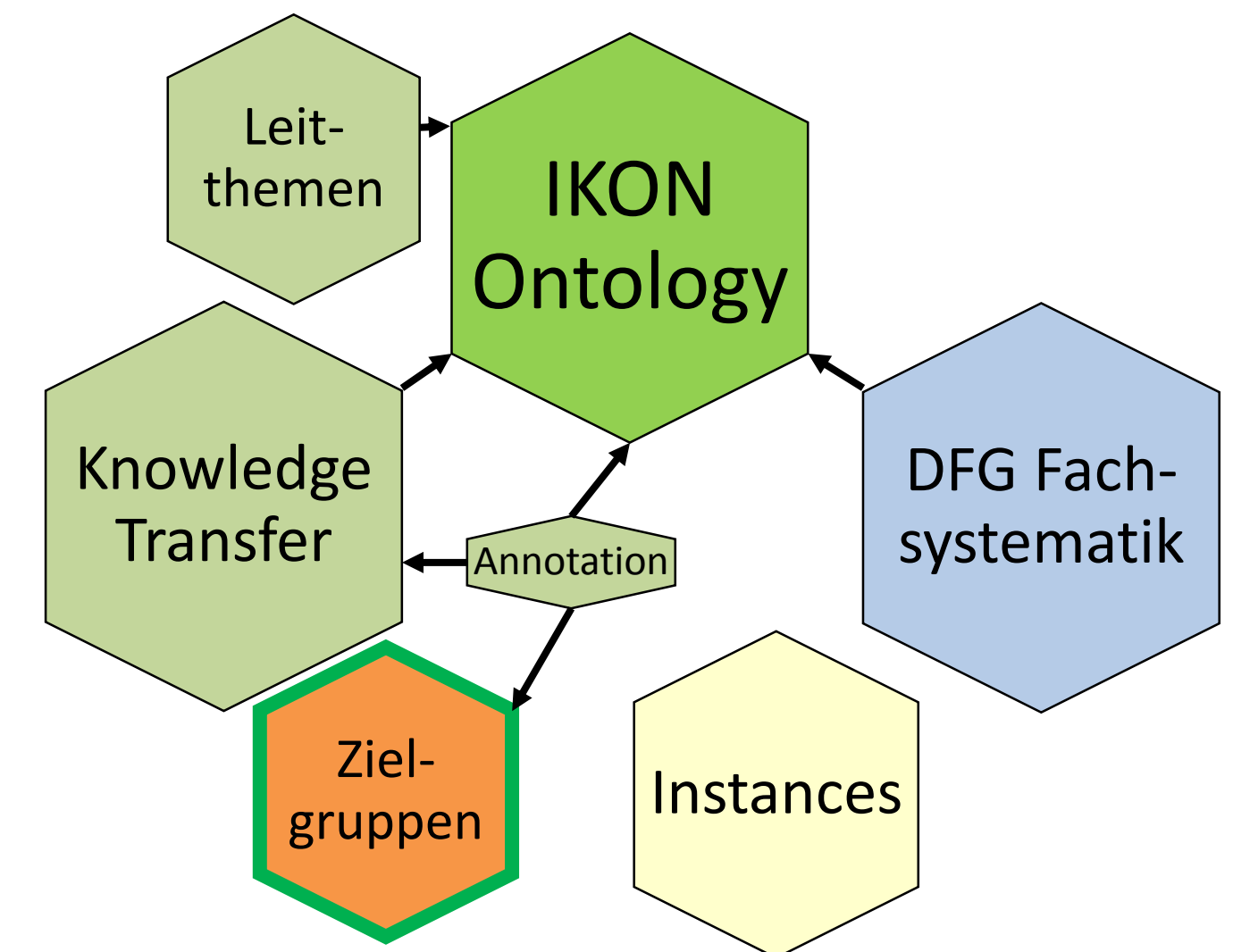
- Source: imports/wt
- Basis for formative workshop on April 5, 2018





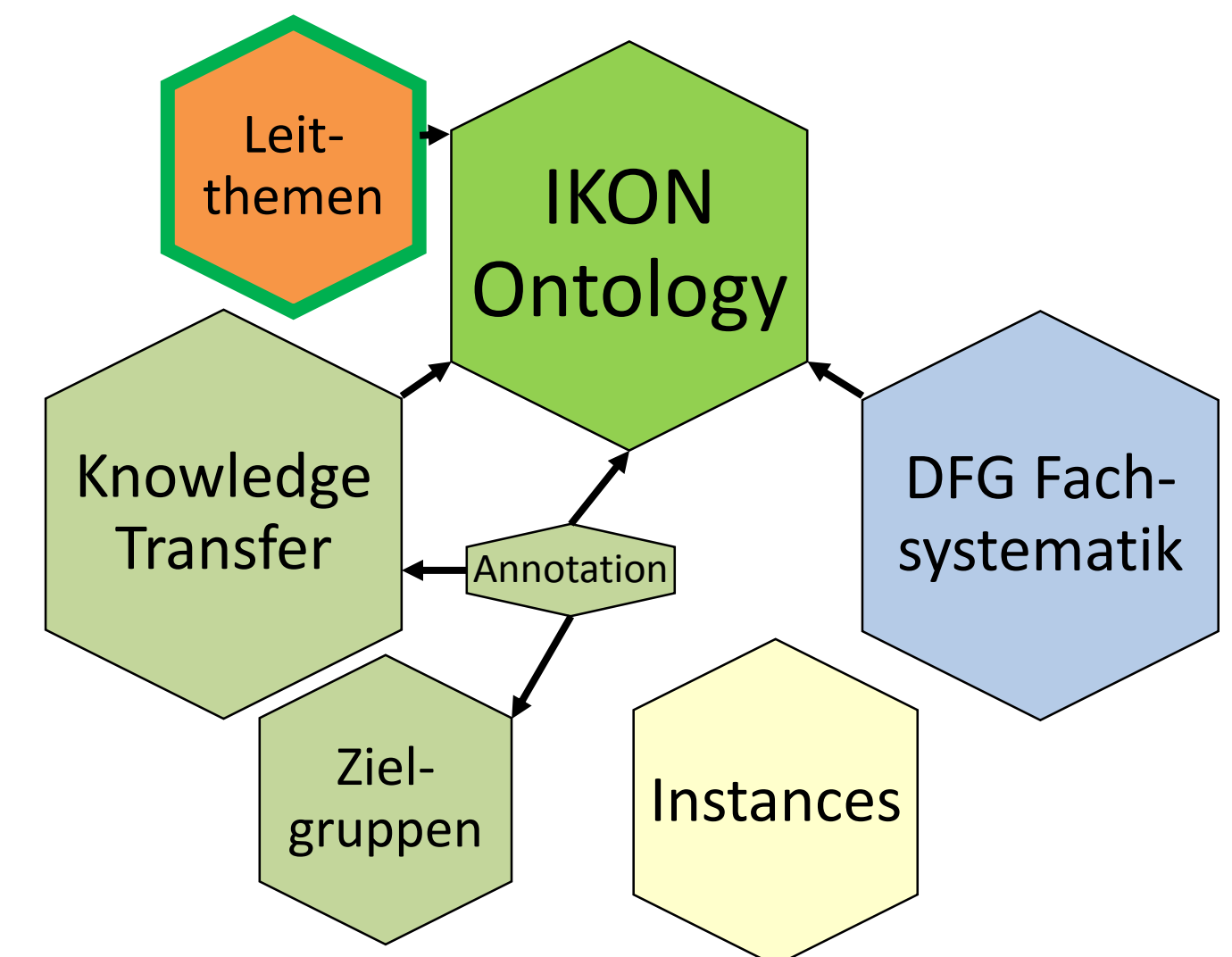
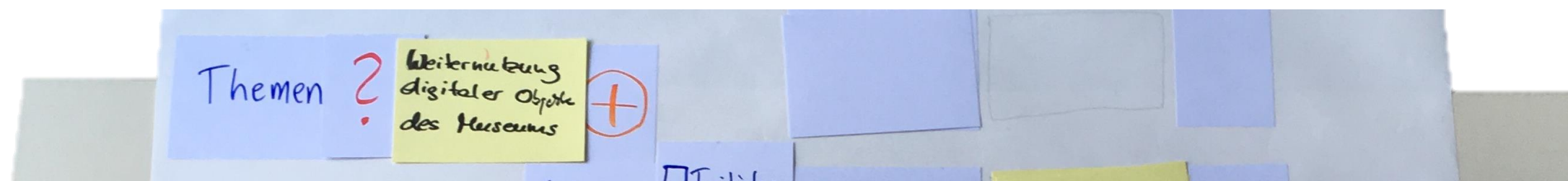
# Zielgruppen

- Source: [imports/zielgruppen](#)
- Part of knowledge transfer ontology



# Leitthemen

- Source: [imports/leitthemen](#)
- Leitthemen modelled as classes
- Named individuals located in imports/instances

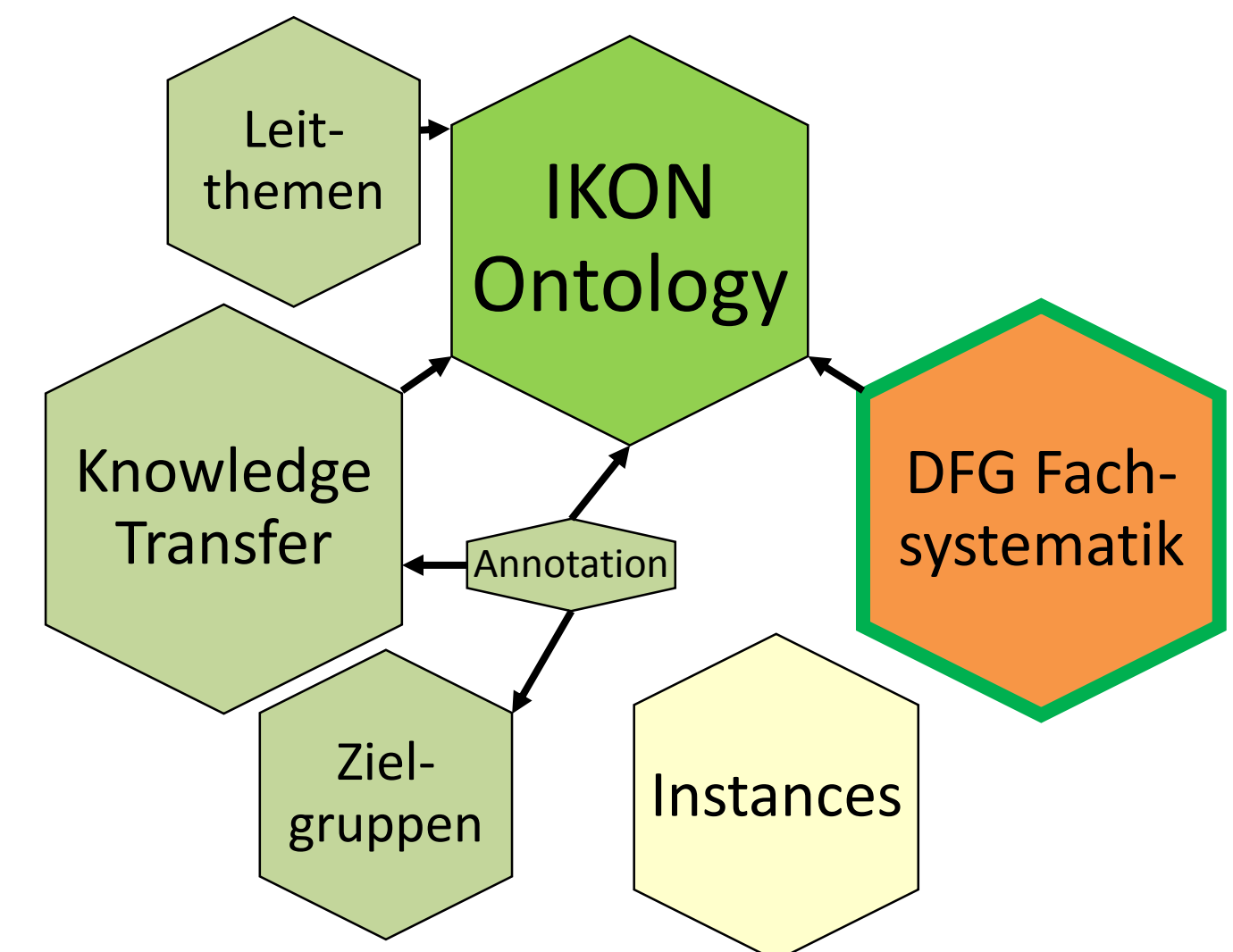


# DFG-Fachsystematik

- Source: [dfg-fachsystematik](https://github.com/joetm/dfg-fachsystematik)
- Model based on [dfg.de](https://www.dfg.de)
- Documentation: [joetm.github.io/dfg-fachsystematik](https://joetm.github.io/dfg-fachsystematik)
- Simple meronymy (3 classes)

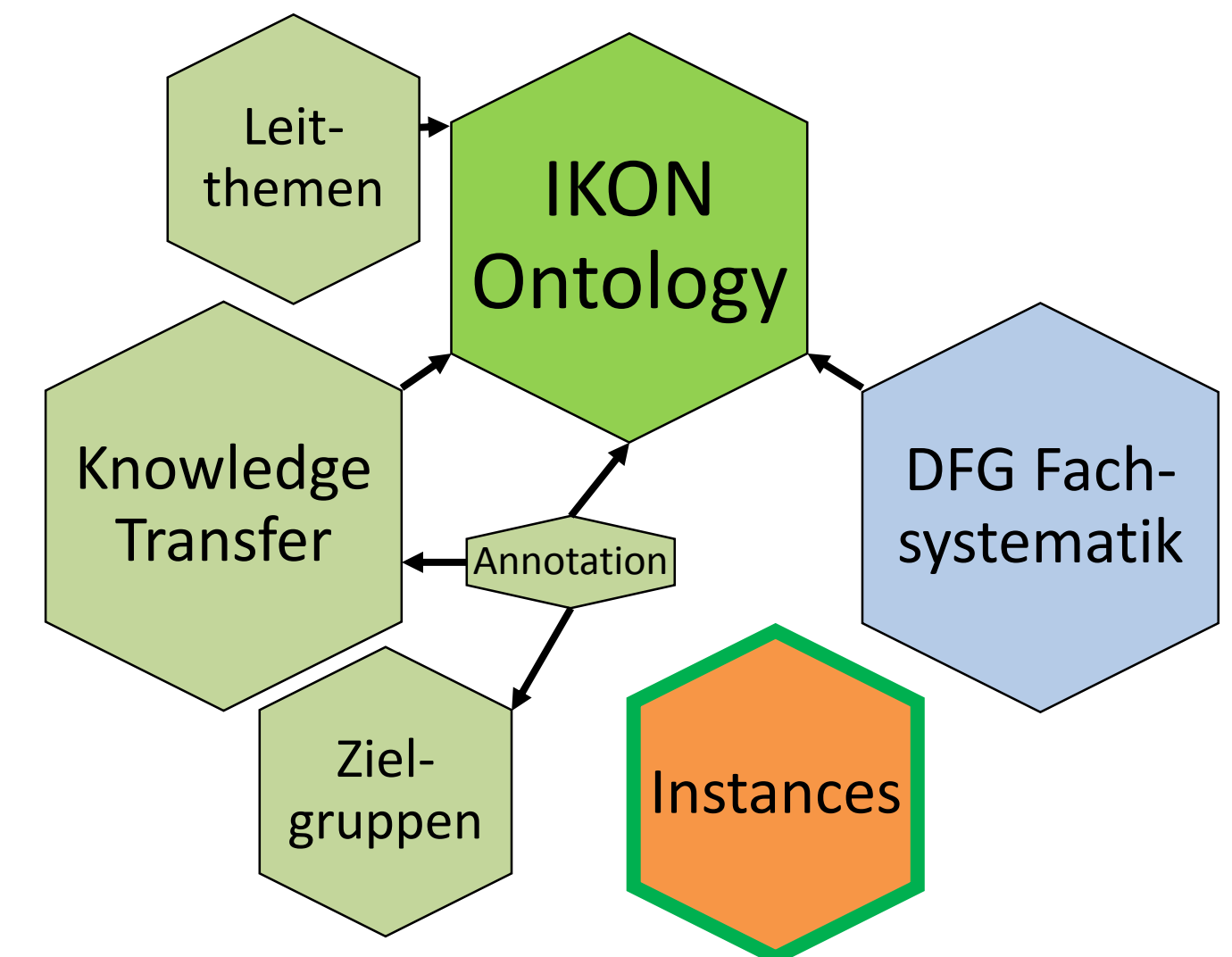


- Scraper and conversion script available (see applications)



# Instances

- Source: imports/instances
- Instances populate dropdowns in Wiki
  - Forschungsbereiche
  - Leitthemen
  - Persons
  - etc.





Data

- Employee pages (Tim)
  - Example:  
<https://www.museumfuernaturkunde.berlin/de/taxonomy/term/234/johannes.vogel>
- Projects (Nicolo)
  - Example:  
<https://www.naturkundemuseum.berlin/de/einblicke/forschung/genart>
- Departments (Nicolo)
  - Example:  
<https://www.naturkundemuseum.berlin/de/einblicke/forschung/evolution-und-geoprozesse>



**Project Lead:** Ulrike Sturm  
**Communication and Social Research:** Martin Tscholl  
**Lead Developer:** Dr. Omid Khorramshahi  
**App Developer:** Alexander Buhl  
**Web Developer:** Saskia Gennrich  
**Student Employees:** Madeleine Hallmann, Susan Karleowski, Tina Birnbach  
**Pattern Recognition:** Mario Lasseck

- Research project descriptions  
== abstracts of research  
project applications

Forschungsprojekte

Jonas Oppenländer Diskussion Einstellungen Beobachtungsliste Beiträge Abmelden

Seite Diskussion Lesen Mit Formular bearbeiten Quelltext bearbeiten Versionsgeschichte

Suche

Navigation

- Hauptseite
- Aktuelle Projekte
- Projekte seit 2016
- Archiv bis 2015
- Alle Projekte im System
- Leitthemen
- Letzte Änderungen
- Projekt neu anlegen

Inhaltsverzeichnis


- AG Forschungsprojekte
- Bedarfsanalyse
- Gespeicherte Anfragen
- Empfohlene Aufgaben
- Glossar
- Neu, bitte sichten
- WT-Aktivitäten

Werkzeuge

- Links auf diese Seite
- Änderungen an verlinkten Seiten
- Datei hochladen
- Spezialseiten
- Druckversion
- Permanenter Link
- Seiteninformationen
- Attribute anzeigen

Überblick English Schlagworte

## MEMIN II: Struktur und Formation von Strahlenkegeln in experimentellen und natürlichen Impakt-Kratern



### Zusammenfassung

Strahlenkegel sind die einzigen Einschlag- (Impact-) -Indikatoren, die leicht auf der makroskopischen Ebene identifiziert werden können. Sie sind daher ein wichtiges diagnostisches Werkzeug, um alte erodierten Einschlagkrater zu verifizieren. Dieses Projekt zielt auf eine bessere Beschränkung der Entstehungsbedingungen von Strahlenkegeln ab, durch eine strenge experimentelle Analyse und den Vergleich mit natürlichen Strahlenkegeln aus dem Steinheimer Becken in Deutschland. Es gibt noch keinen Konsens in Bezug auf die physikalischen Randbedingungen, die zu ihrer Bildung benötigt werden. Wir beabsichtigen, Stoßdruck, Drucklänge und -profil, Tensor von Spannung und Dehnung und Dehnungsrate durch die Anwendung des in dem Programm MEMIN entwickelten numerischen Modells zu bestimmen. Darüber hinaus gibt es keinen Konsens darüber, wie sich Strahlenkegel formen und es ist umstritten, ob die Bruchflächen das Ergebnis von Scher- oder Zugversagen sind.

Durch die erstmalige Anwendung von Weißlichtinterferometrie und Rasterkraftmikroskopie (AFM) führen wir eine vollständige morphometrische Analyse der Splitterkegelflächen durch. Dies und eine strenge Mikrostrukturanalyse unter Verwendung von SEM, EBSD und TEM wird genutzt, um den Bildungs- und Versagensmechanismus von Labor- und Natur-Strahlenkegeln aus dem Steinheimer Becken zu entschlüsseln. Da die energiereichen MEMIN Experimente zur Kraterbildung feste Felsen als Ziel-Materialien verwenden, sind sie besonders dazu geeignet Strahlenkegel zu formen und zu erkennen, da die Kraterbildungsprozess mit höchster Präzision überwacht werden, sich die Kratergröße im Dezimeterbereich befindet und eine detaillierte räumliche Untersuchung ermöglicht. Es ist beabsichtigt, Laborbedingungen zu schaffen, die eine reproduzierbare und vorhersehbare Bildung von Splitterkegeln ermöglichen. Ebenso ist es wichtig, die Bedingungen, unter denen Strahlenkegel sich nicht ausbilden, zu begrenzen. Das letztendliche Ziel ist es, ein einheitliches Modell für die Erzeugung von Strahlenkegeln abzuleiten, welches mit unserem makroskopischen und mikroskopischen Beobachtungen und den erhaltenen physikalischen Randbedingungen übereinstimmt.

Keine der veröffentlichten Bildungsmodelle fand bisher universelle Anerkennung und liefert eine Erklärung für die konische Form und die hierarchischen Schachtelhalm-Streifenmuster der Strahlenkegel. Die meisten Bildungsmodelle haben gemeinsam, dass eine Art von Heterogenität (Eingliederung, Porenraum, etc.) an der Spitze eines Strahlenkegels existieren muss. Diese Heterogenität verursacht einen Impedanzkontrast zu dem umgebenden Gestein und ist der Ursprung einer Kugelwelle, die sich mit der Standard-Schockwelle überlagert. In diesem Projekt werden diese Voraussetzung hinterfragt, da sich die Splitterkegel besonders gut in feinkörnigen mikritischen Kalksteinen, wie im Steinheimer Becken entwickelten, in dem jede Art von einer großen Aufnahme vermutlich selten oder nicht vorhanden ist.

Kurztitel: MEMIN II TP 9

Museum für Naturkunde Berlin  
FB 1

# Extracted Entities

Nicolo

A small icon representing a JSON file, consisting of a light gray square with a folded corner and a red rectangle containing the word "JSON" in white text.

- Source: Research project descriptions (from above)
- Stop words removed
- Tested three tools: DBPedia Spotlight, Dandelion, TextRazor
- Stemming and lemmatization in progress (IKON-213)
- TF-IDF calculation in progress (IKON-195)
- Issue: compound terms (e.g. “Great Escarpment”)
  - DBPedia detects this as a location
  - All subsequent tasks (stopword removal, TF-IDF) need to preserve these terms

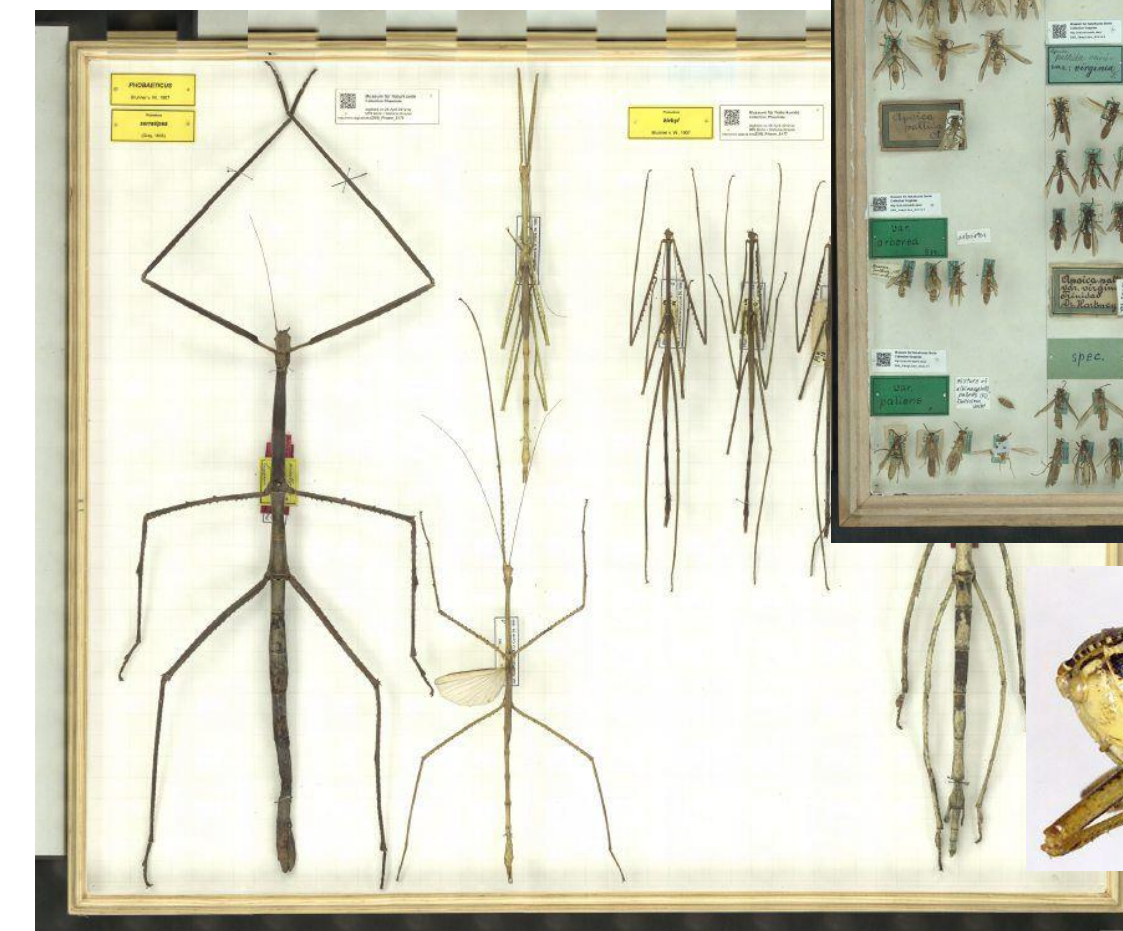
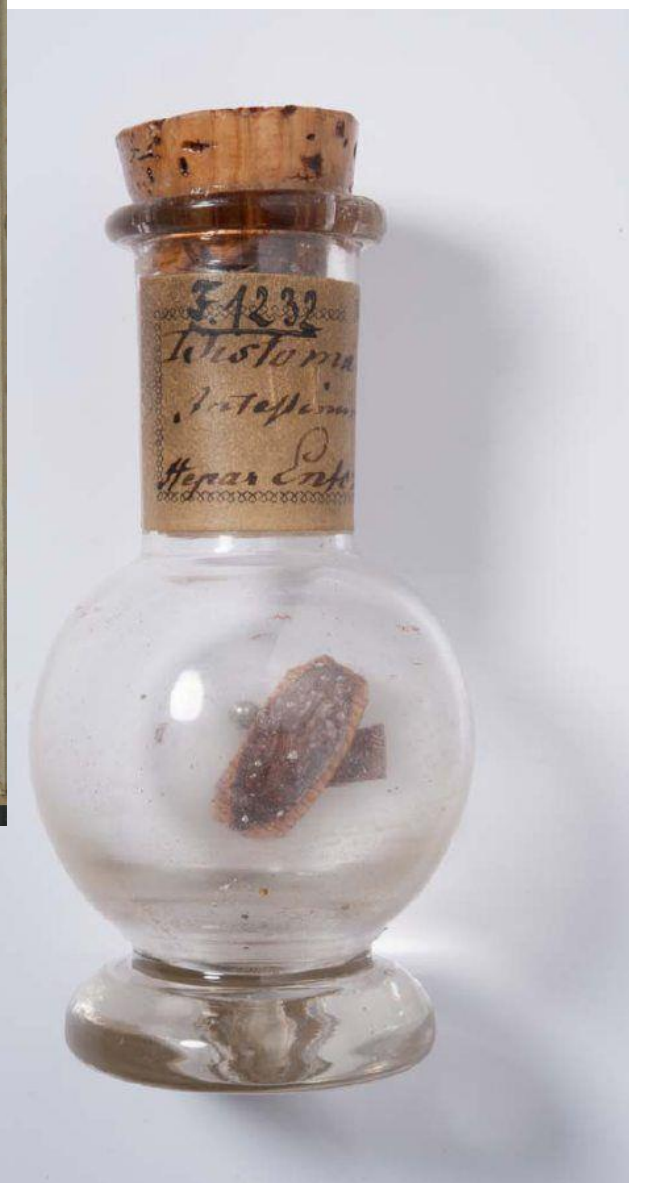
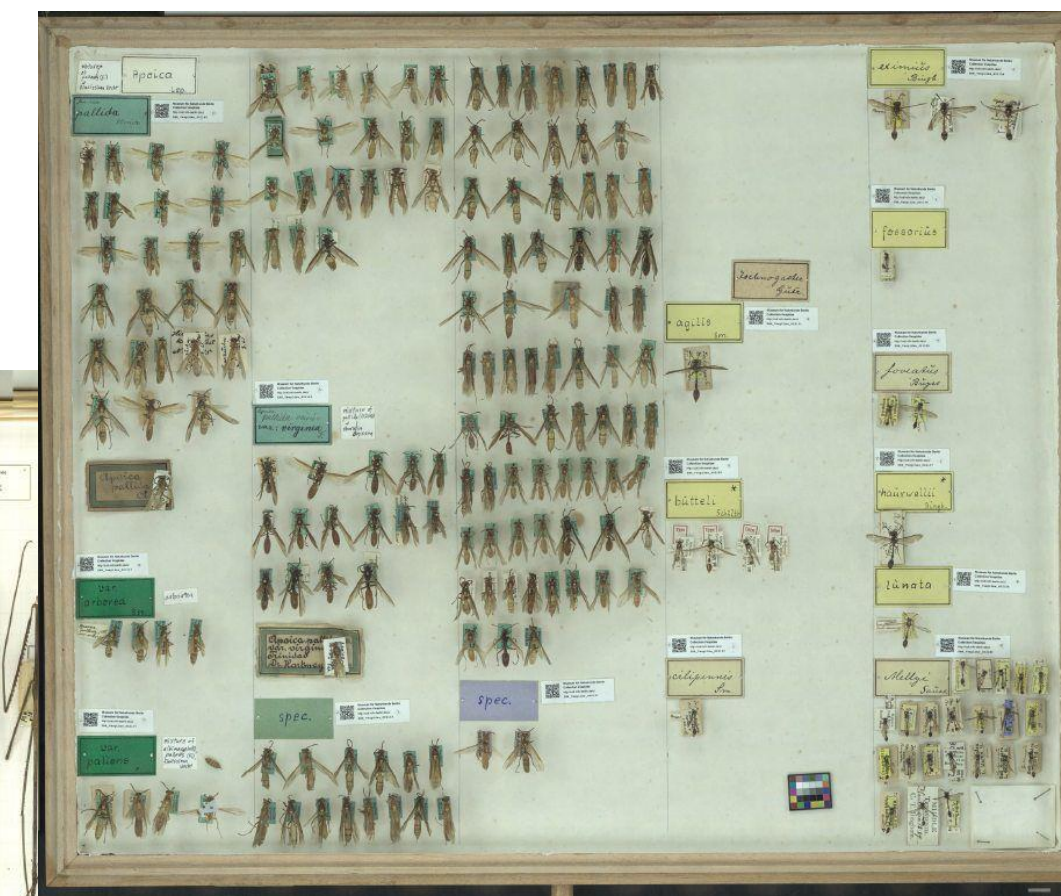
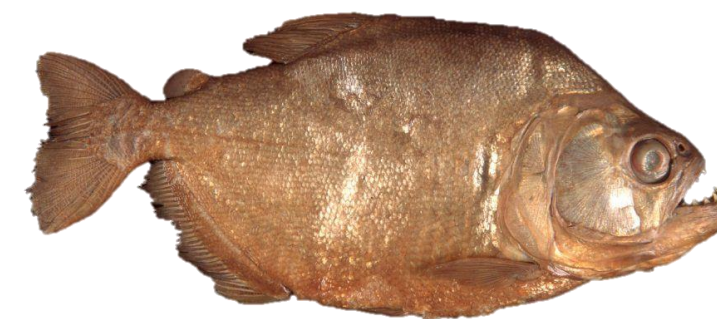


# Collection Images

Tim

JSON

- Images scraped for machine learning task
  - Basic info was available as (incomplete) RDF
- Link to collectors (historic and/or recent)





# MfN @ Gepris

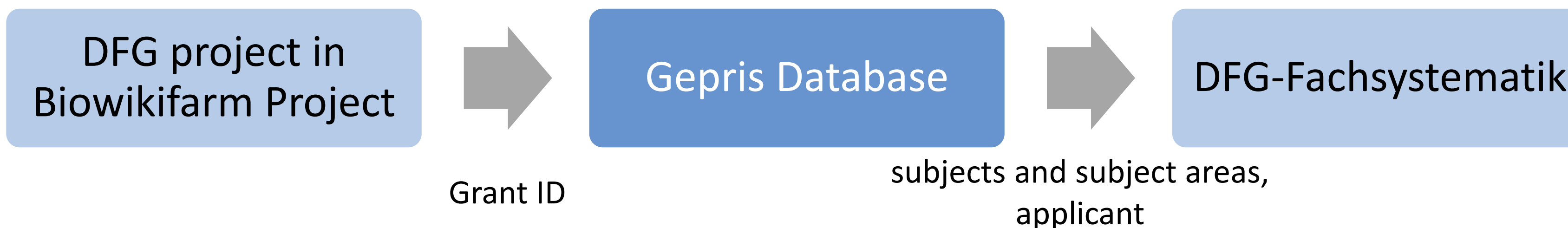
Spaude

SQL

- From Daniel Spaude
- Format: sqlite

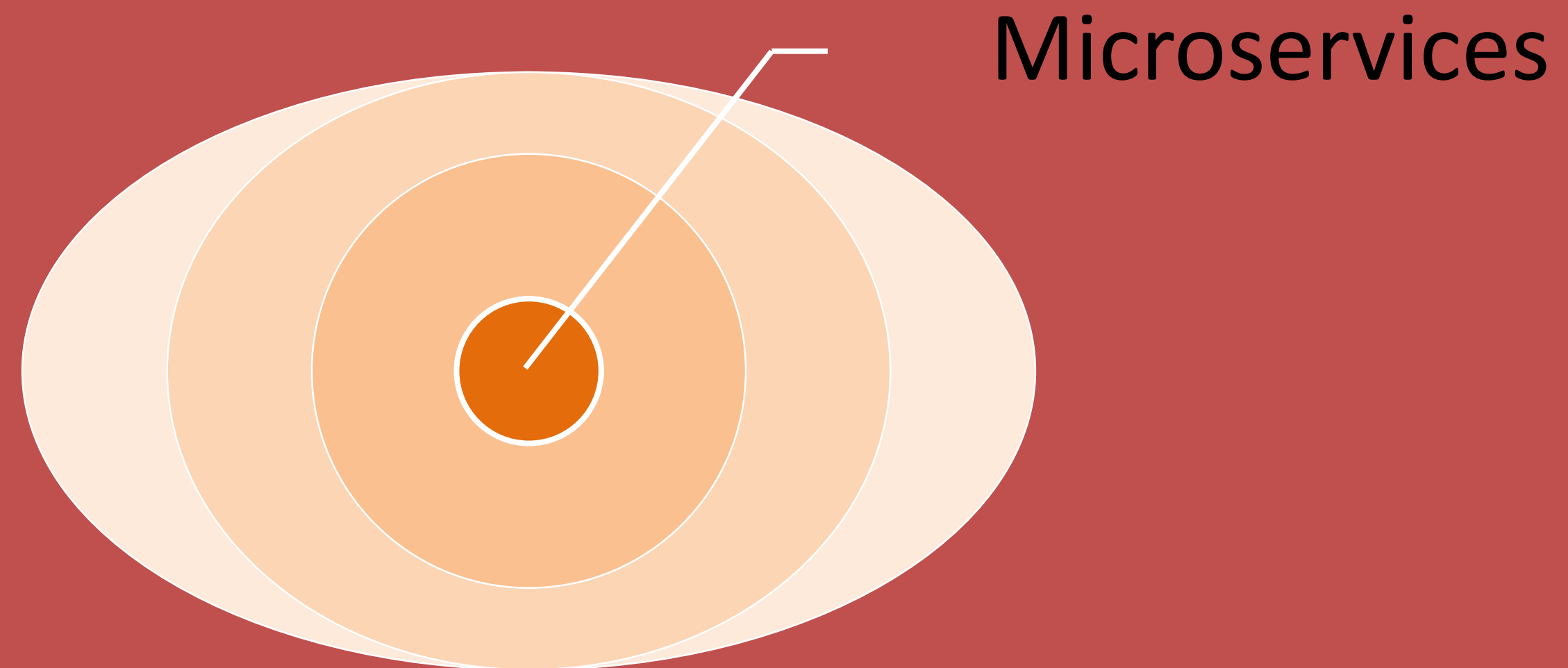
	project_id	title	rg_subject_areas	dfg_verfahren	funding_start_year	funding_end_year	applicant_id	applying_institution_id	subject_area
1	99939532	Aufbau einer virtuellen Forschun...		Wissenschaftskom...	2009	2013		12944	Asienbezogene Wisse..
2	99939522	Bewusstseinsprozesse im vegetat...		Sachbeihilfen	2009	2014	1506533		Differentielle Psychol..
3	99939395	Das deutsche Italienbild 1800-18...		Forschungsstipen...	2008	2012	99939351		Kunstgeschichte
4	99939390	Integration der Publikation von U...		Wissenschaftskom...	2009	2014	1493019		Physik und Chemie de..
5	99919608	GRK 1582: Fluor als Schlüsselele...	Chemie, Pharm...	Graduiertenkollegs	2009			10165	Molekülchemie
6	99899284	rDNA genome re-programming in...		Sachbeihilfen	2008	2013	1202712		Tierzucht, Tierhaltung
7	99898696	FOR 1209: Theoretische Grundfra...	Philosophie, Re...	Forscherguppen	2009				Geisteswissenschaften
8	99878648	The lignification of strawberry fru...		Sachbeihilfen	2009	2012	1395486		Pflanzenzüchtung
9	99877981	Modifizierung von linearem PMM...		Sachbeihilfen	2009	2014	1614775		Kunststofftechnik
	99877981	Modifizierung von linearem PMM...		Sachbeihilfen	2009	2014	1614775		Kunststofftechnik

- Associate projects with DFG-Fachsystematik





# Applications





## Data Acquisition

- MfN Website Scraper
- Biowikifarm Wiki
- MfN Collection
- DFG Gepris

## Data Conversion

- MfN Website
- Biowikifarm Wiki

## Data Interpretation

- Entity Extraction
- Publication Atomizer
- Crowdsourcing Interface
- MfN Collection Classifier

## Data Fusion

- Person Matching
- Project Matching
- Fusion with DFG Gepris





## ► MfN Website Employee Scraper

- Scrapes employee data
  - Name, title, research interests, research projects, publications, etc.
  - Including separate CV and publication pages (if available)
  - Includes basic data cleansing
- Source: [scraper/MfNwebsitescraper](#)
- Status: Update in progress (new MfN website)

## ▶ MfN Website Project/Department Scraper

JSON

- Scrapes project and department data
  - Associations of projects with departments
  - Associations of people with projects
- Manual scrape with scraping tool
- IKON/Scrapen+--+Projekte+MfN-Webseite

## ► Biowikifarm Wiki Scraper

- API is incomplete (missing fields) → Scrape project data
- Source: [scraper/projectscraper](#)
- Status: update required (url and login details changed due to new location of wiki)





# Data Acquisition

Tim

# JSON



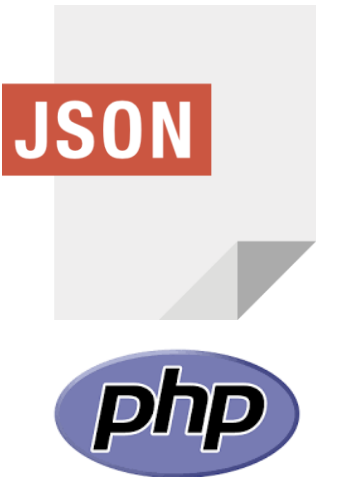
# ► MfN Collection Scraper

- Source: scraper/collectionscraper
- > 38,000 objects
- Several gigabytes of images



## ► DFG-Fachsystematik Scraper

- Source: [dfg-fachsystematik/scrapper](https://github.com/dfg-fachsystematik/scrapper)
- Scrapes [www.dfg.de/dfg\\_profil/gremien/fachkollegien/faecher/](https://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/)
- Result: [dfg-fachsystematik/scrape](https://github.com/dfg-fachsystematik/scrape)





## ► MfN Website employee data

- Using RML Mapper (rml.io)
- Status: in progress



## ► Convert scraped DFG-Fachsystematik

- Using RML Mapper (rml.io)
- <https://intern.hcc.mi.fu-berlin.de/confluence/pages/viewpage.action?pageId=33260836>



## ► Publication Atomizing

- Analysis: Anystyle vs. Grobid?

	GROBID	Anystyle.io
Precision	0.8136	0.8369
Recall	0.8205	0.7938
F1-Score	2.4615	2.3814



## ► Publication Atomizer

- Source: citation-parse
- Input: Scraped MfN website (JSON)
- Output: Atomized publication
  - Year, Authors, Title, Journal, ...
  - TEI format
- Todo: Use Pelzer's Grobid data processing pipeline for conversion of TEI output to RDF



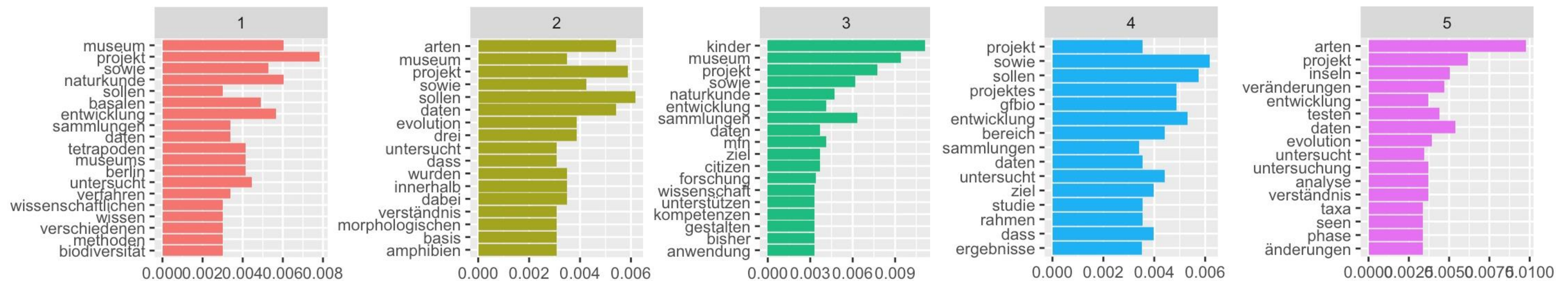
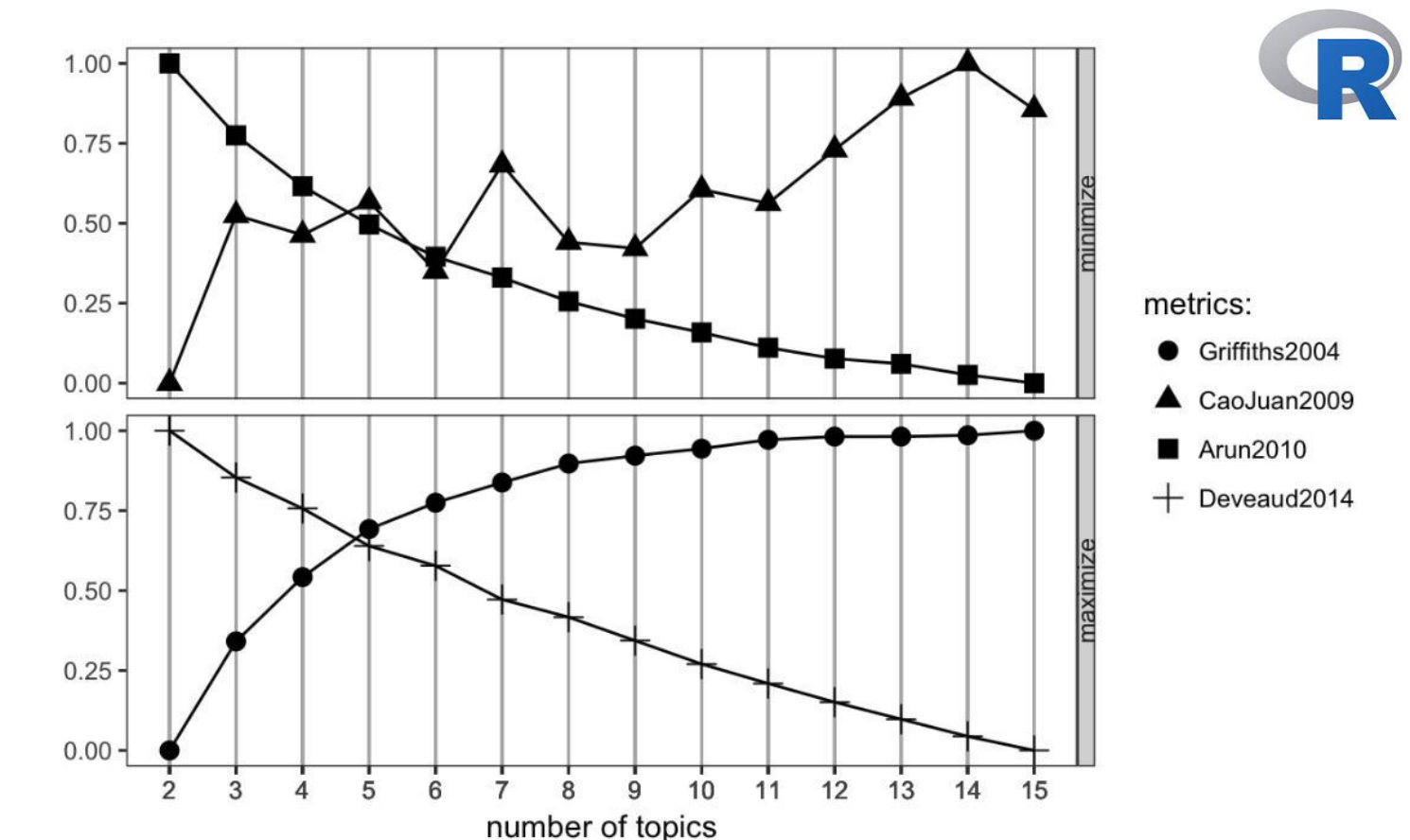
## ► Entity Extraction

- Source: scraper/projectscraper
- Extract entities from Biowikifarm project descriptions
- Removed stop words (tools/stop word filter)
- Stemming and lemmatization (in progress)
- Three tools: DBPedia Spotlight, Dandelion, TextRazor



## ► Topic Modelling (LDA)

- Data source: all Biowikifarm project descriptions
- Number of topics detected: 5-6
- Results:



- Suggestion: segment by Leitthema



## ► MfN Collection Classifier

- Classification of MfN collection images with Machine Learning (PyTorch)
  - Display case (multiple exhibits) vs. single exhibit
  - Handwritten notes
  - Insects, etc.
- Side project
- Status: in progress; images acquired

### COLLECTION



IMAGE URL: [http://coll.mfn-berlin.de/img/ZMB\\_MAM\\_091613\\_IMG\\_4663.JPG](http://coll.mfn-berlin.de/img/ZMB_MAM_091613_IMG_4663.JPG)  
This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

### Object (ZMB\_MAM\_091613)

UnitID: ZMB\_MAM\_091613; SourceID: ZMB\_MAM;  
Collection\_Catalogue\_URI: [Http://Coll.Mfn-Berlin.De/C/ZMB\\_MAM](http://Coll.Mfn-Berlin.De/C/ZMB_MAM);  
  
FullScientificNameString: Lagothrix Lagothrix (Humboldt, 1812); Author: Humboldt, 1812; Author\_VIAF: [Http://Viaf.Org/Viaf/95193235](http://Viaf.Org/Viaf/95193235); Family: Cebidae; PreferredFlag: TRUE; TypeStatus: Type;  
  
FullName: Lako; Primarycollector: TRUE; Sequence: 1;  
  
LocalityText: Canabouca, Parana Do Jacare, Brazil, South America; LatitudeDecimal: -3.1165200000; LongitudeDecimal: -68.0349000000;  
  
AreaName: Canabouca; AreaClass: Region;  
  
AreaName: Brazil; AreaClass: Country;

### Additional Specimen Images



### Related Specimens

## ▶ Name Matching

- Documentation: IKON/Name+Matching
- Match names from MfN website with Biowikifarm persons
- Results:

**≈80%** of persons can be matched

→ Use (controlled) crowdsourcing for matching

## ► Crowdsourcing Interface

- Source: [tools/matching-tool](#)
- Status: Draft (interface-only)
- Suggestion: Use pybossa framework

The screenshot displays a crowdsourcing interface for data matching. It features two identical data entry forms side-by-side. Each form contains four input fields with labels: 'id', 'title', 'href', and 'description'. Below the forms, there is a progress bar and a set of navigation buttons: 'Don't Match', 'Match', 'Previous', and 'Next'.

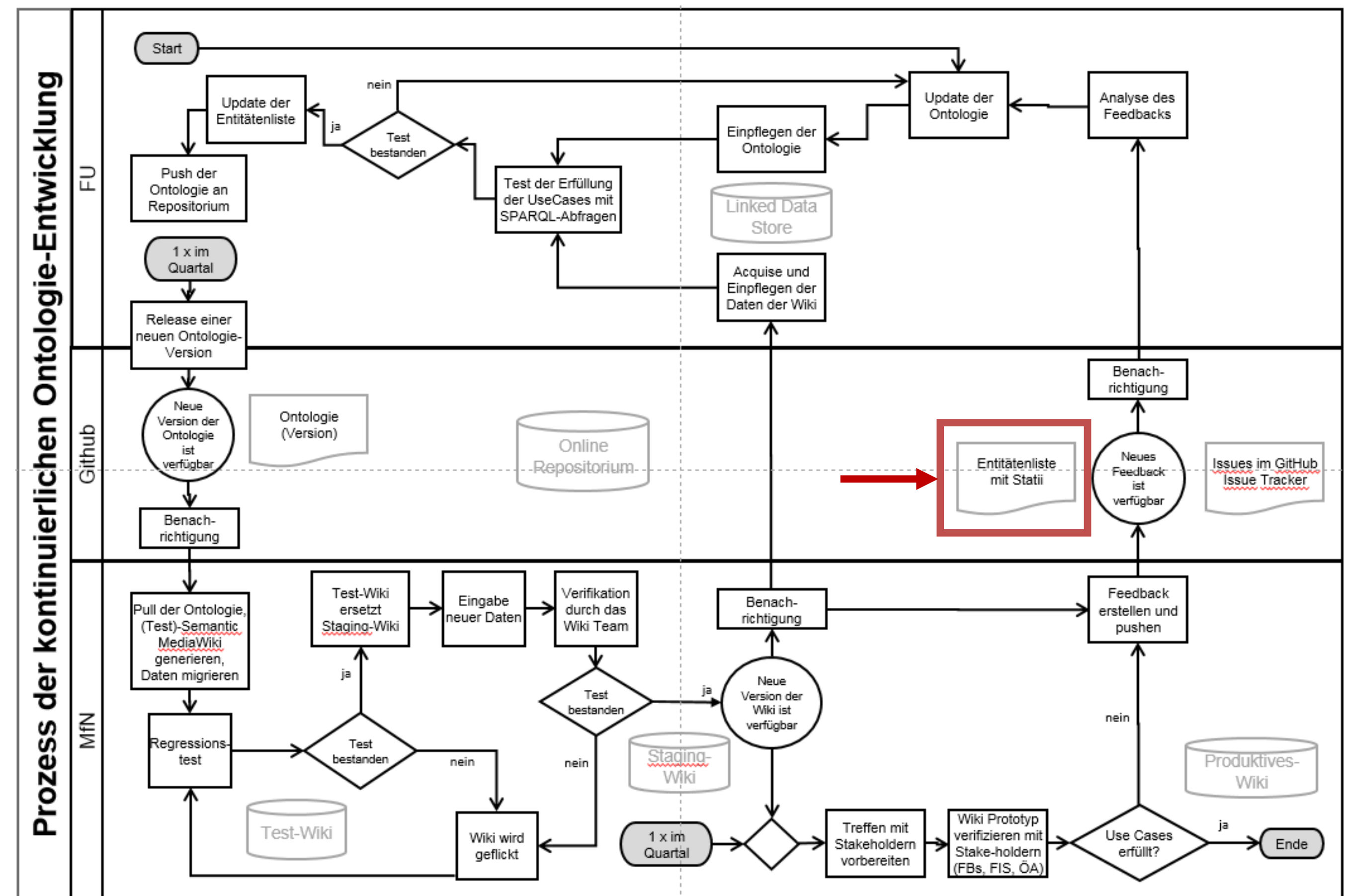
## ► NLP Verifier + Category Matcher

- Source: IKON-NLPverifier
- “Controlled Crowdsourcing”
- Two purposes
  - Verify the correctness of extracted entities
  - Assign entities to categories (research objects, regions, methods,...)
- Status: Backend (NodeJS API), Frontend (React; only framework)
- Suggestion: Use pybossa framework



## ► Masterlist Generator

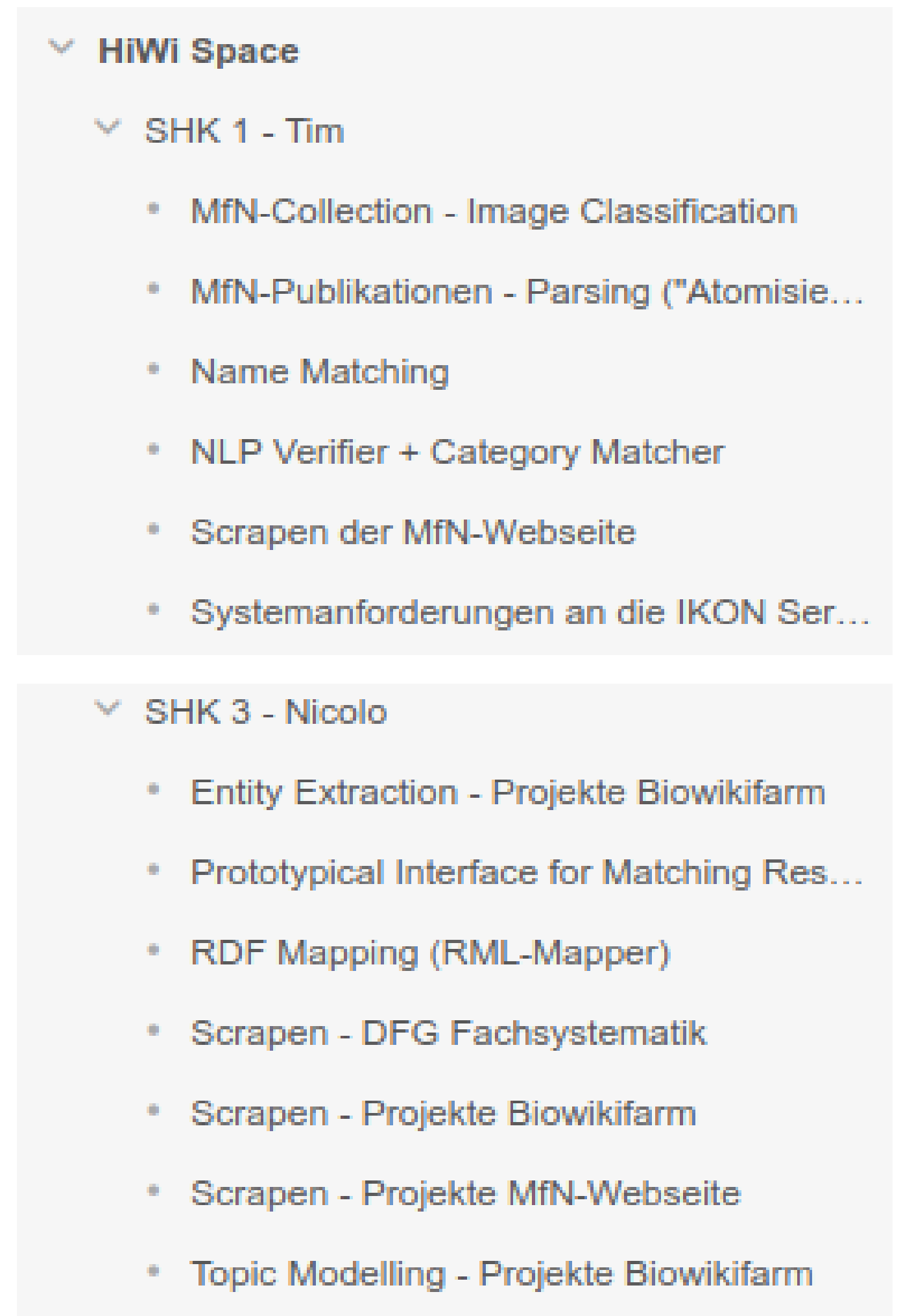
- Source:  
[IKON-ontology/masterlist](https://github.com/IKON-ontology/masterlist)
- Generates Excel Worksheet from production ontology
- Two purposes:
  - Track status of Classes and Properties
  - Communication of feedback



# Results

## ▶ See HiWi-Space on Confluence

- Results not on GitHub
- HiWi-Space on Confluence contains:
  - Documentation of Jira Tasks
  - Results
  - Smaller task-related scripts that were not checked into GitHub



# Open Jira Tasks

- **Nicolo**

- IKON-207: Topic Modelling
- IKON-213: Stemming and lemmatization
- IKON-216, IKON-195: Filter via TF-IDF to reduce complexity
- IKON-202: Convert MfN employee data to RDF/XML or RDF/TTL
- IKON-197: Convert scraped DFG-Fachsystematik to OWL
- IKON-214: Identify n-grams

- **Tim**

- IKON-113: Improve MfN Website Harvester
- IKON-139: Classification of collection images with Machine Learning
- IKON-203: Install Linux on Desktop
- IKON-205, IKON-206: MfN @ Europeana
- IKON-180: Identify technical terms in project descriptions



# Project IKON

