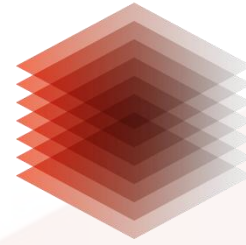

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Terminology Services

Dr.-Ing. Felix Engel
NFDI4Ing, CC-41 Community Meeting
March 03 2022

Agenda

- **Research Data Management**
 - Motivation
 - Reproducibility issues
 - Metadata in RDM
- **Terminology Service**
 - Introduction
 - Application
 - S-3 architecture
 - Outlook

Research Data Management (I/VIII)

Introduction



Current situation in Germany (*)

- Data is stored decentrally
- Stored temporarily
- **Non-standardised metadata**
- Varied quality

Nationale Forschungsdateninfrastruktur (NFDI) address need for high qualitative RDM

Funded October 2020: *DataPlant, GHGA, KonsortSWD, NFDI4BioDiversity, NFDI4Cat, NFDI4Chem, NFDI4Culture, NFDI4Health, NFDI4Ing*

TIB involvement in metadata standardization @NFDI:



NFDI4Ing (<https://nfdi4ing.de/>): *NFDI4Ing brings together the engineering communities. It offers a unique method-oriented and user centered approach in order to make engineering research data FAIR – findable, accessible, interoperable, and re-usable.*

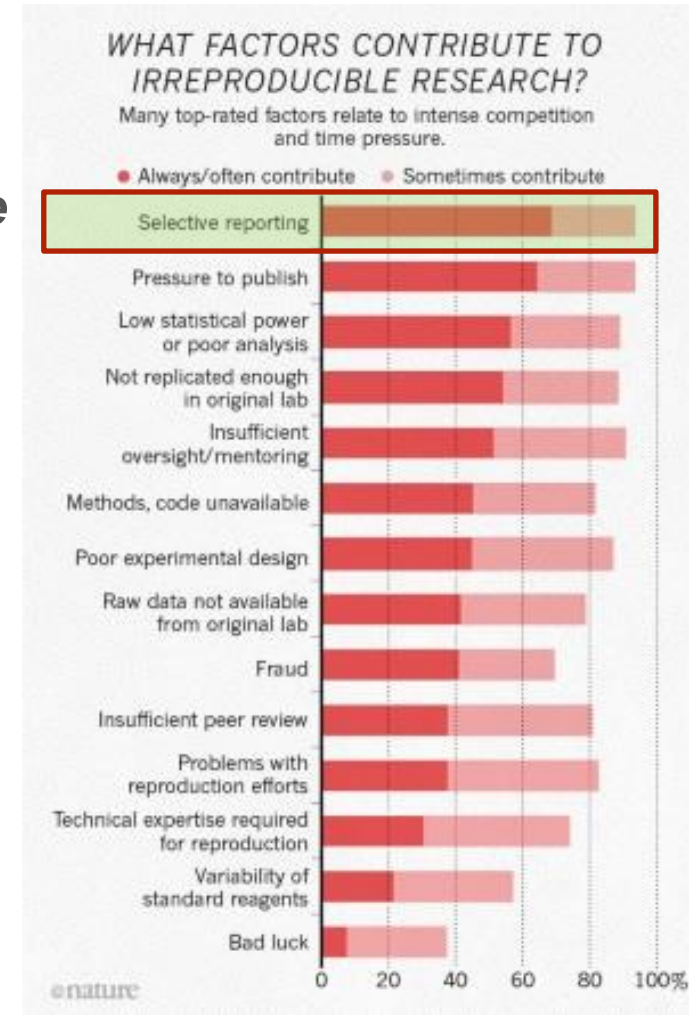
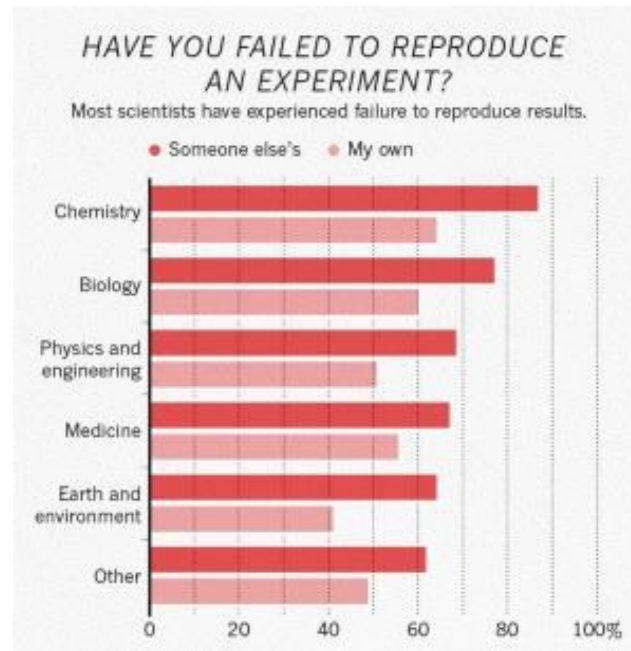


NFDI4Chem (<https://www.nfdi4chem.de/>): *The vision of NFDI4Chem is the digitalisation of all key steps in chemical research to support scientists in their efforts to collect, store, process, analyse, disclose and re-use research data*

Research Data Management (II/VIII)

Reproducibility issues: In numbers

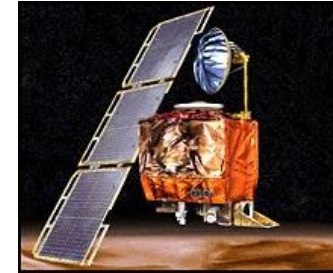
- A supporting *Nature* publication in 2016 [*]
- Survey of 1,576 researchers who took a brief **online questionnaire on reproducibility in research**



[*] <https://www.nature.com/articles/533452a>

Research Data Management (III/VIII)

Reproducibility issues: Impact of selective reporting



[**]

- **1999 Mars Climate Orbiter accident**
 - *The **disastrous accident** of the 1999 Mars Climate Orbiter quickly became part of the lore of technology failures. It occurred when firing durations for a guidance jet were **mis-communicated between engineering teams**, one of which assumed **English units** and the other assumed **metric units**. [*]*
 - **Cost:** 125 Mill. USD [**]
- **Trans-disciplinary project result discussion** with technical terms and indicators. Speaking about **micro- and macronutrients** [***]
 - **Soil scientists:** *nitrogen, phosphorus, potassium, calcium, magnesium and sulphur*
 - **Nutritionists:** *carbohydrates, protein and fat*

[*] <https://tdan.com/mars-orbiter-disaster-a-failure-of-data-management/5128>

[**] <https://www.washingtonpost.com/wp-srv/national/longterm/space/stories/orbiter100199.htm#:~:text=Washingtonpost.com%3A%20Space%20Exploration&text=NASA's%20Mars%20Climate%20Orbiter%20was,Martian%20surface%2C%20investigators%20said%20yesterday.>

[***] Jordan, I., Heil, E., & Keding, G. (2021). Coming to terms with terminology in agriculture-nutrition research projects: an interactive glossary. *Ernährungsumschau*, 68(10), 198-203.

Research Data Management (IIII/VIII)

Reproducibility issues: Introduction of FAIR principles [*]

Consortium of scientists introduce widely accepted FAIR principles in 2016

- **Findability** can be ensured by using a unique or persistent identifier with sufficient description of data and their characteristics in machine-readable metadata and storage of data in archives or repositories.
- **Accessibility** requires that metadata are always available via standardized, universally implementable communication protocols and corresponding datasets have clearly defined access conditions.
- **Interoperability** needs data and **metadata** kept in common, published standards of data formats, variables or ontologies in order to enable their integration into existing applications or workflows.
- **Reusability** is the ultimate aim of research. **Detailed description** of characteristics **according to community standards** with clear conditions enable the reuse of research data for future endeavours.

Research Data Management (V/VIII)

Metadata in RDM

- **What is Metadata or Metainformation?** ... ensure that the information can remain understandable ... [*].
- **What is metadata used for in RDM?**
 - **Unambiguous communication**
 - **Discovery**
 - Sorting (bring resources together that belongs) together
 - Search and Browsing
 - **Reuse.** E.g. Information about
 - applied methodology, involved data and software
 - Authors, licences

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation

Research Data Management (VI/VIII)

Metadata in RDM: implementation in NFDI4Ing

- **Core constituents**
- **Terminologies** [*]
 - *the technical or special terms used in a business, art, science, or special subject*
 - *E.g. a lexicon covering the terminologies of several scientific fields*
- NFDI4Ing makes use of **Ontologies**
 - Part of the Semantic Web tool stack
 - *... a formal description of knowledge [**]*
 - *... ontologies express relationships and enable users to link multiple concepts to other concepts in a variety of ways [**]*
- In delimitation, a view point from the **Philosophy**: *... the study of general and fundamental questions, such as those about existence, reason, knowledge, values, mind, and language [***]*

[*] <https://www.merriam-webster.com/dictionary/terminology>

[**] <https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies/>

[***] <https://en.wikipedia.org/wiki/Philosophy>

Research Data Management (VII/VIII)

Metadata in RDM: implementation in NFDI4Ing: Additional value of ontologies

- Building a **communication framework**, a common language and understanding of interrelations between terms.
- Make metadata machine readable and understandable
- Provide mappings between ontologies
- Automated validation of metadata

Research Data Management (VIII/VIII)

Metadata in RDM: General characteristics

- **Domain specific** (Engineering, Culture, Chemistry, ...) and **community specific**
- **Evolving continuously and dynamically over time**
- Must be **accepted, developed and maintained** by a **designated community (avoid isolated solution!)**. Includes a.o.
 - a) promotion (make community aware of its existence)
 - b) aligned with further metadata initiatives (moving away from silos)
 - c) applicable in RDM (in RDM practise)

NFDI supports terminology development and use through **introduction of community specific Terminology Services for Ontologies**

Terminology Service (I/XII)

- **In general:** A Terminology Service is a web based platform that support take-up and standardisation of terminologies
- **Used as**
 - **Entry Point** to **prepare research data** for effective **later reuse** (e.g. according to a DMP).
 - **Tag/index** with **keywords** from a controlled vocabulary (e.g. content indexation for retrieval)
 - Consolidate terminology: use the same keywords
 - Trace changes meaning
 - **Search:** Query reformulation, term suggestion, ...
 - **Hub,** that fosters **awarness** and **alignement**
 - Bundles ontologies of a domain
 - Provides meta data and statistics
 - Search for and within ontologies
 - Makes alignments visible

Terminology Service (II/XII)

TIB TS



TIB Central TS



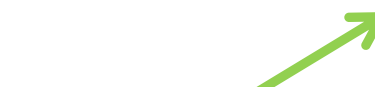
Domain specific TS's



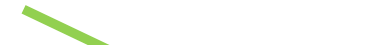
All research communities share the RDM challenge

Terminology Service

Apache 2.0 Licence



NFDi4ing

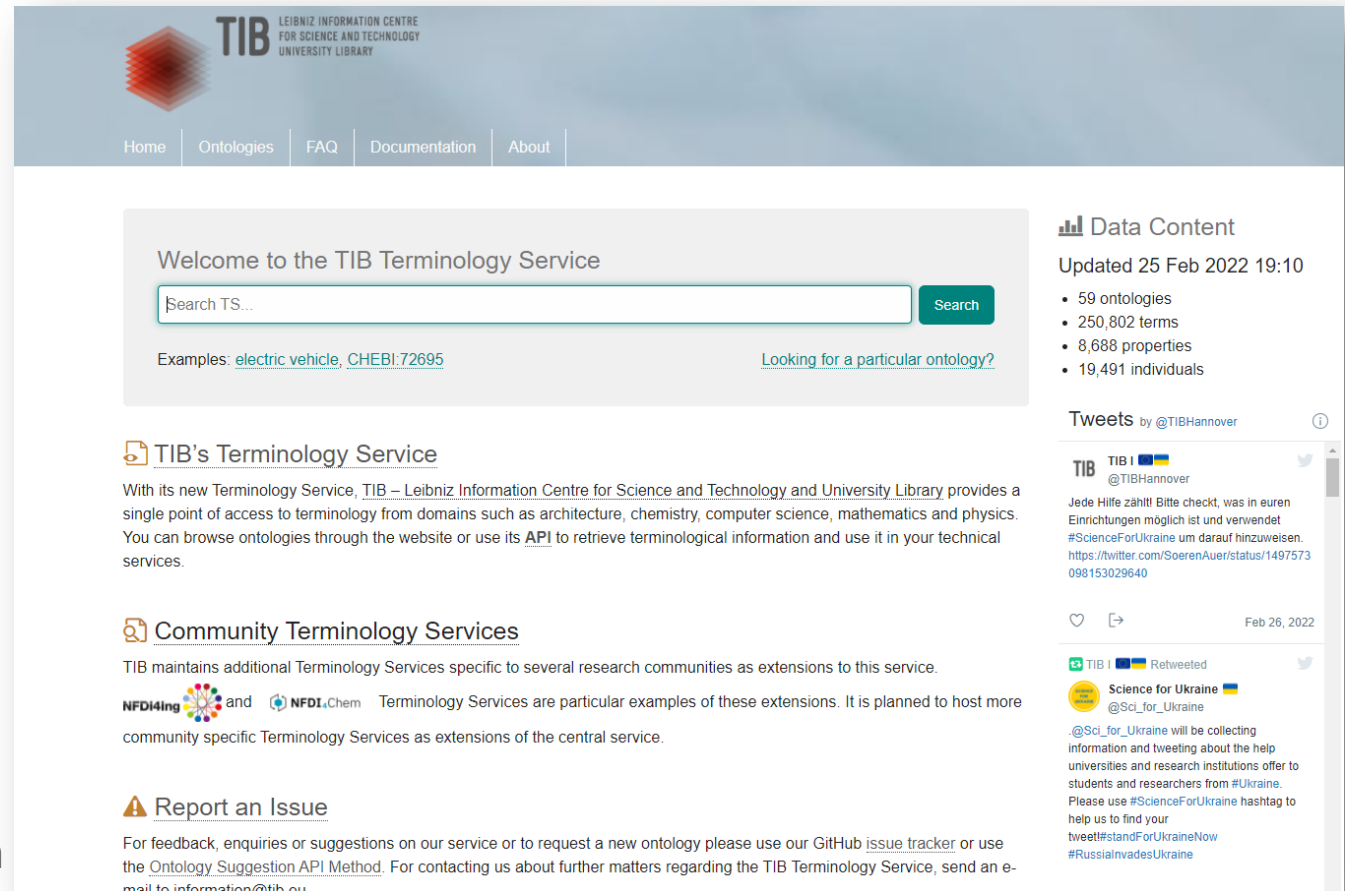


... further initiatives

Terminology Service (III/XII)

TIB TS Central

- **First version released July '21:** <https://service.tib.eu/ts4tib/index>
- **Central service instance holds all ontologies**
- **Some statistics**
 - 59 ontologies
 - 8,688 properties
- **Functional service offer**
 - Freetext search searching (for- and within ontologies)
 - Browsing and filtering
 - Visualisation
 - Issue tracker
 - Machine to machine communication (REST interfaces)

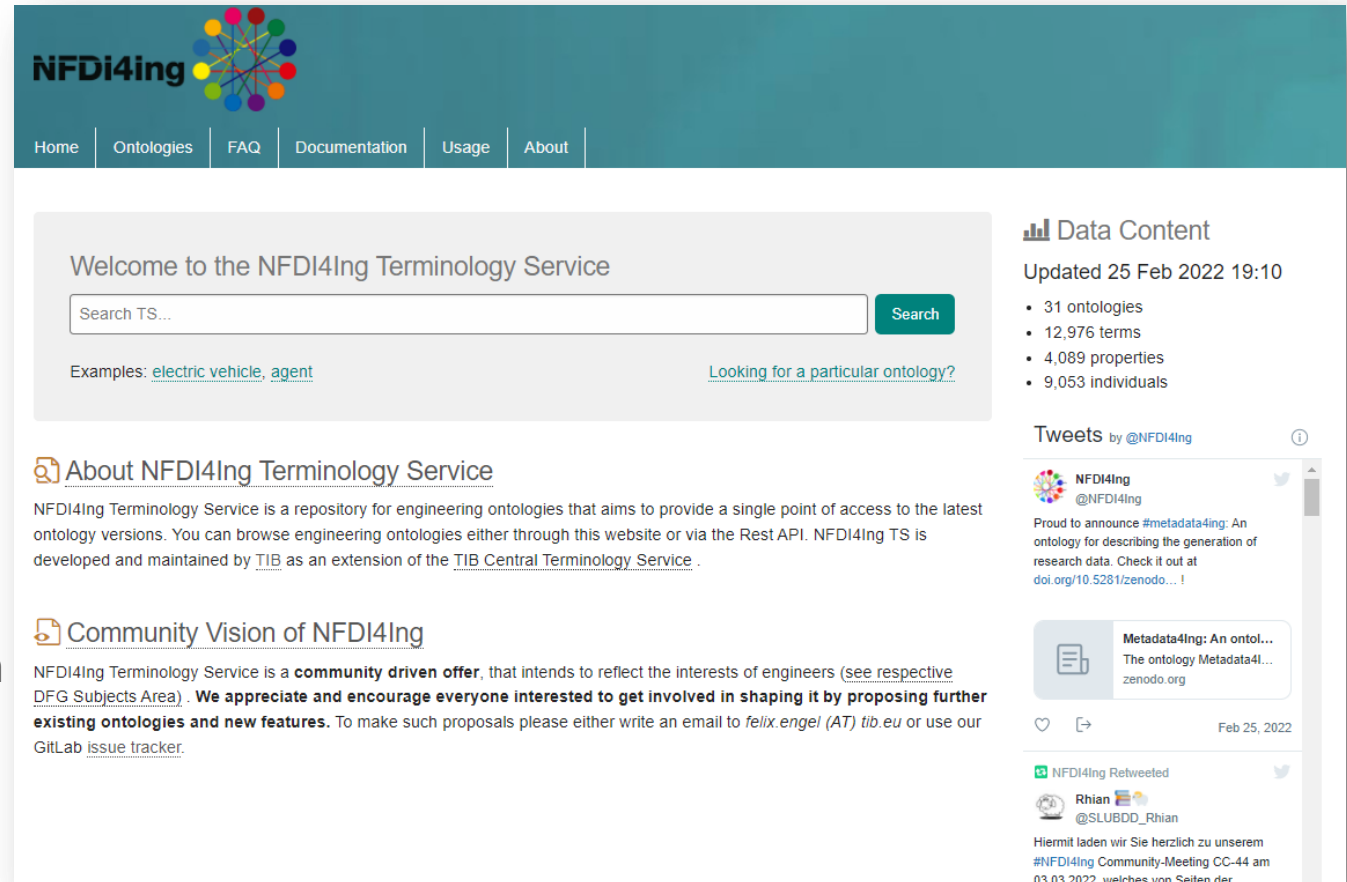


The screenshot shows the TIB Terminology Service website. At the top, there is a navigation bar with links for Home, Ontologies, FAQ, Documentation, and About. Below the navigation bar is a search interface with a text input field labeled "Search TS..." and a "Search" button. Below the search field, there are examples of search terms: "electric vehicle" and "CHEBI:72695", and a link "Looking for a particular ontology?". To the right of the search interface is a "Data Content" section with a bar chart icon, updated on 25 Feb 2022 at 19:10. It lists statistics: 59 ontologies, 250,802 terms, 8,688 properties, and 19,491 individuals. Below the statistics is a "Tweets" section by @TIBHannover, featuring a tweet from @Sci_for_Ukraine about collecting information and tweeting about the help universities and research institutions offer to students and researchers from #Ukraine. The tweet includes the hashtag #ScienceForUkraine and a link to a tweet: <https://twitter.com/SoerenAuer/status/1497573098153029640>. The tweet is retweeted by TIB | @TIBHannover. At the bottom of the screenshot, there is a "Report an Issue" section with a warning icon, providing instructions for feedback, enquiries, or suggestions, and a link to the GitHub issue tracker.

Terminology Service (IIII/XII)

TIB TS NFDI4Ing

- **NFDI4Ing TS:** <https://terminology.nfdi4ing.de>
- **Some statistics**
 - 31 ontologies
 - 4,089 properties
- **Functional service offer**
 - Freetext search searching (for- and within ontologies)
 - Browsing and filtering
 - Visualisation
 - Issue tracker
 - Machine to machine communication (REST interfaces)



The screenshot shows the NFDI4Ing Terminology Service website. The header features the NFDI4Ing logo and a navigation menu with links for Home, Ontologies, FAQ, Documentation, Usage, and About. The main content area includes a search bar with the text "Search TS..." and a "Search" button. Below the search bar, there are examples of search terms: "electric vehicle" and "agent", and a link for "Looking for a particular ontology?". The right sidebar contains a "Data Content" section with a bar chart icon, updated on 25 Feb 2022 at 19:10, listing 31 ontologies, 12,976 terms, 4,089 properties, and 9,053 individuals. Below this is a "Tweets" section by @NFDI4Ing, featuring a tweet from NFDI4Ing announcing the #metadata4ing ontology. The main content area also includes sections for "About NFDI4Ing Terminology Service" and "Community Vision of NFDI4Ing".

Terminology Service (V/XII)

TIB TS Functional service offer: Browsing and Searching

Exact match Obsolete terms

Term type
 Filter by type

- class 110
- individual 94
- property 15

Search results for electric vehicle

Previous
Showing 1 to 10 of 219 results
Next

electric vehicle OEO:00000146

http://openenergy-platform.org/ontology/oeo/OEO_00000146

Ontology: [Repository for the Open Energy Ontology \(OEO\)](#) OEO

Electric Vehicle ElectricVehicle

<http://schema.mobivoc.org/#ElectricVehicle>

Ontology: [MobiVoc Open Mobility Vocabulary](#) MV

Electric Vehicle Charging Electric_Vehicle:Charging

http://www.w3id.org/ecsel-dr-PWR#Electric_Vehicle_Charging

Ontology: [Digital Reference](#) dr

List of all ontologies in the Terminology Service

Show

Search:

entries

Ontology Name	Short name	Description	Loaded	Action
Basic Formal Ontology	BFO	The upper level ontology upon which OBO Foundry ontologies are built.	Thu Oct 28 16:55:15 GMT 2021	Search Terms Properties Individuals
Bioinformatics operations, data types, formats, identifiers and topics	EDAM	EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics, including types of data and data identifiers, data formats, operations and topics. EDAM provides a set of terms with synonyms and definitions - organised into an intuitive hierarchy for convenient use.	Thu Oct 28 16:57:03 GMT 2021	Search Terms Properties Individuals
Building Topology Ontology	bot	The Building Topology Ontology (BOT) is a minimal ontology for describing the core topological concepts of a building.	Thu Oct 28 16:55:17 GMT 2021	Search Terms Properties Individuals
Core Ontology for Robotics and Automation (CORA)	CORA	This is the OWL implementation of CORA in IEEE 1872-2015. It only includes the taxonomy of concepts and relations, with some few axioms regarding disjointness, property characteristics and property ranges/domains. The OWL implementation is an underspecified version of the SUO-KIF implementation in IEEE 1872-2015. That is, the set of allowed models of the SUO-KIF implementation is a proper subset of the allowed models by the OWL implementation.	Thu Oct 28 16:56:10 GMT 2021	Search Terms Properties Individuals
Data Privacy Vocabulary (DPV)	dpv	The Data Privacy Vocabulary (DPV) provides terms (classes and properties) to describe and represent information related to processing of personal data based on established requirements such as for the EU General Data Protection Regulation (GDPR). The DPV is structured as a	Thu Oct 28 16:56:45 GMT 2021	Search Terms Properties Individuals

Terminology Service (VI/XII)

TIB TS Functional service offer: Collections

Bundle ontologies of a Designated Community. E.g. a project or a research area

[Bioinformatics operations, data types, formats, identifiers and topics](#)

EDAM

EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics, including types of data and data identifiers, data formats, operations and topics. EDAM provides a set of terms with synonyms and definitions - organised into an intuitive hierarchy for convenient use.

collection: NFDI4ING - NFDI4CHEM -

Fri Feb 25
18:54:54
GMT 2022

[Search](#)
[Terms](#)
[Properties](#)
[Individuals](#)
[Download](#)

[Building Topology Ontology](#)

bot

The Building Topology Ontology (BOT) is a minimal ontology for describing the core topological concepts of a building.

collection: NFDI4ING - NFDI4CULTURE - FID BAUdigital -

Fri Feb 25
18:33:42
GMT 2022

[Search](#)
[Terms](#)
[Properties](#)
[Individuals](#)
[Download](#)

[CHEBI Integrated Role Ontology](#)

CHIRO

CHEBI provides a distinct role hierarchy. Chemicals in the structural hierarchy are connected via a 'has role' relation. CHIRO provides links from these roles to useful other classes in other ontologies. This will allow direct connection between chemical structures (small molecules, drugs) and what they do. This could be formalized using 'capable of', in the same way Uberon and the Cell Ontology link structures to processes.

collection: NFDI4CHEM -

Fri Feb 25
18:52:13
GMT 2022

[Search](#)
[Terms](#)
[Properties](#)
[Individuals](#)
[Download](#)

[Chemical Entities of Biological Interest](#)

CHEBI

A freely available dictionary of molecular entities focused on 'small' chemical compounds. The term 'molecular entity' refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity. The molecular entities in question are either products of nature or synthetic products used to intervene in the processes of living organisms.

collection: NFDI4CHEM -

Fri Feb 25
18:51:24
GMT 2022

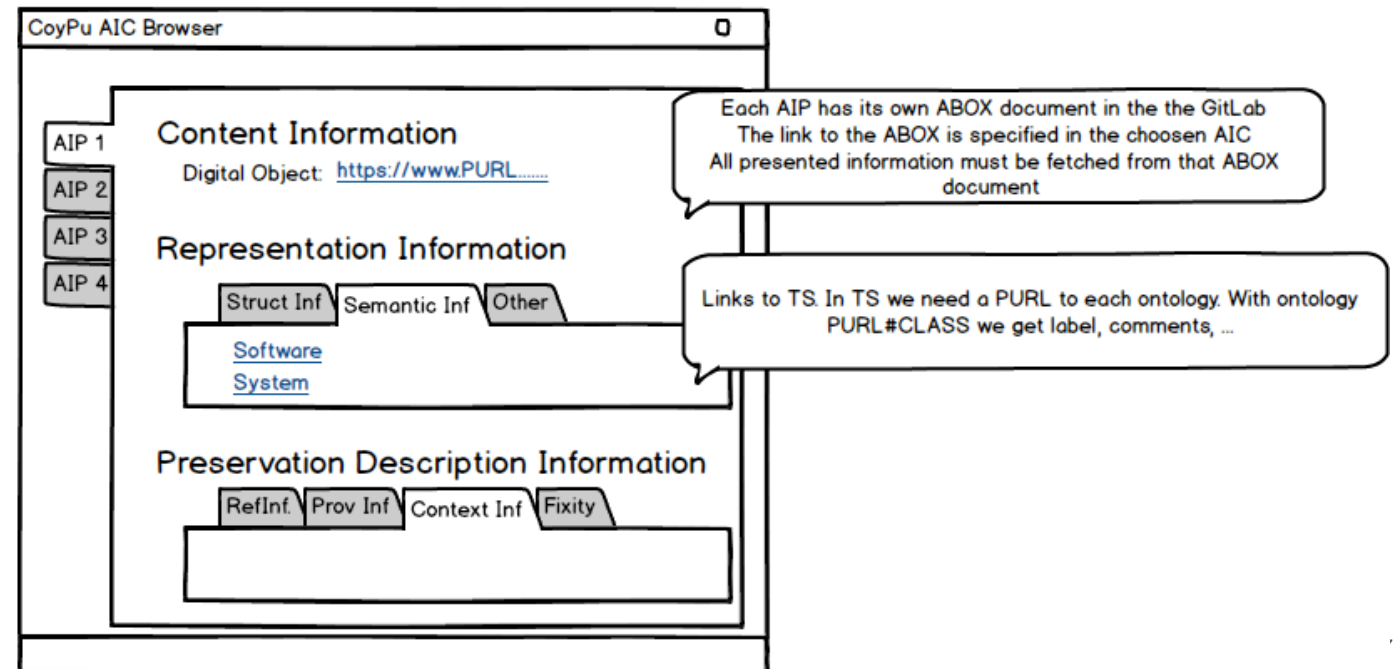
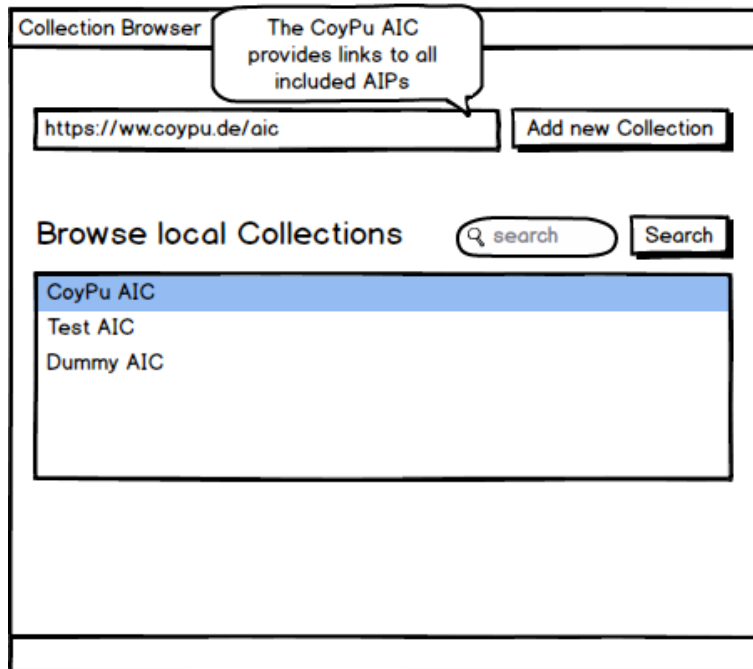
[Search](#)
[Terms](#)
[Properties](#)
[Individuals](#)
[Download](#)

Terminology Service (VII/XII)

Example applications for researcher (... work in Progress)

Data Collection Management

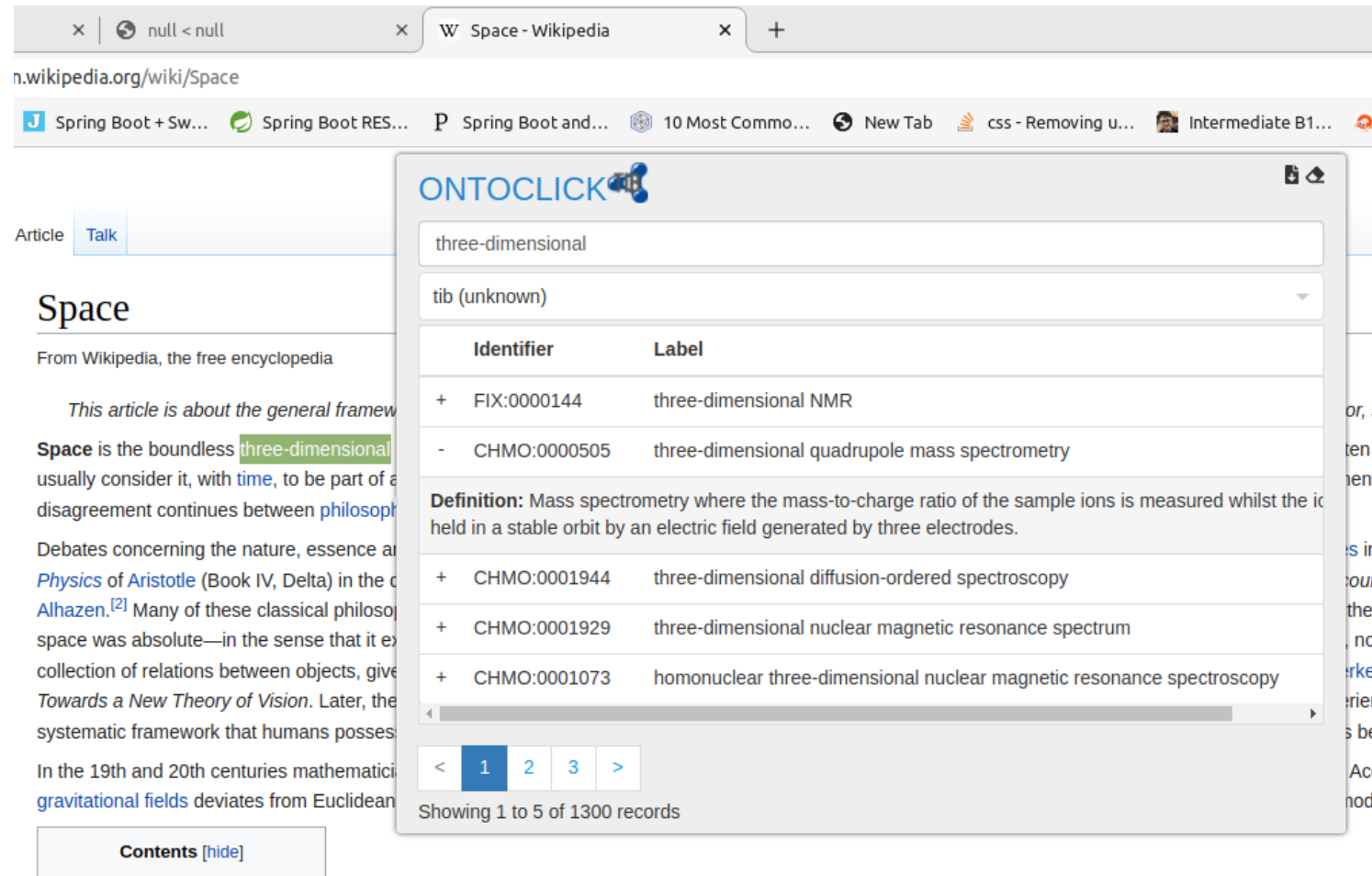
- Provide an ontology that specifies the **structure of a collections** (e.g. a set of software repositories)
- Add descriptive metadata from TS



Terminology Service (VIII/XII)

Example applications for researcher (... work in Progress)

- **Chrome TS Plugin**



The screenshot shows a Chrome browser window with the Wikipedia article for "Space" open. An ONTOCLICK plugin is overlaid on the page, displaying search results for the term "three-dimensional". The plugin includes a search bar, a dropdown menu, and a table of results.

Identifier	Label
+ FIX:0000144	three-dimensional NMR
- CHMO:0000505	three-dimensional quadrupole mass spectrometry
Definition: Mass spectrometry where the mass-to-charge ratio of the sample ions is measured whilst the ions are held in a stable orbit by an electric field generated by three electrodes.	
+ CHMO:0001944	three-dimensional diffusion-ordered spectroscopy
+ CHMO:0001929	three-dimensional nuclear magnetic resonance spectrum
+ CHMO:0001073	homonuclear three-dimensional nuclear magnetic resonance spectroscopy

Showing 1 to 5 of 1300 records

Terminology Service (VIII/XII)

Example applications for researcher (work in progress)

- **AIMS Project**
- Creation and sharing of metadata standards as so-called **application profiles**.
- An application profile is a set of requirements for subject and use-case specific metadata and represented in RDF and SHACL.
- Within the frontend, users can search and drag vocabulary terms into their application profile as properties. The TIB Terminology Service is used to retrieve these vocabulary terms, by automatically querying its **REST Interface**.

Terminology Service (X/XII)

Example applications for researcher (planned)

Bioportal ANNOTATOR

Annotator

Get annotations for biomedical text with classes from the ontologies [?](#)

Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye.

[insert sample text](#)

Match longest only
 Match partial words
 Include mappings
 Exclude numbers
 Exclude synonyms

Select ontologies

Medical Subject Headings (MESH) ✕

[clear selection](#) [select from list](#)

Select UMLS semantic types

Start typing to select UMLS semantic types

Include ancestors up to level

none ▼

Get Annotations

Annotations

total results 4 (direct 4 / ancestor 0 / mapping 0)

CLASS	filter	ONTOLOGY	filter	TYPE	filter	CONTEXT	MATCHED CLASS	filter	MATCHED ONTOLOGY
Melanoma		Medical Subject Headings		direct		Melanoma is a malignant ...	Melanoma		Medical Subject Headings
Melanocytes		Medical Subject Headings		direct		... tumor of melanocytes which are found ...	Melanocytes		Medical Subject Headings
Skin		Medical Subject Headings		direct		... predominantly in skin but also in ...	Skin		Medical Subject Headings
Eye		Medical Subject Headings		direct		... and the eye	Eye		Medical Subject Headings

Terminology Service (XI/XII)

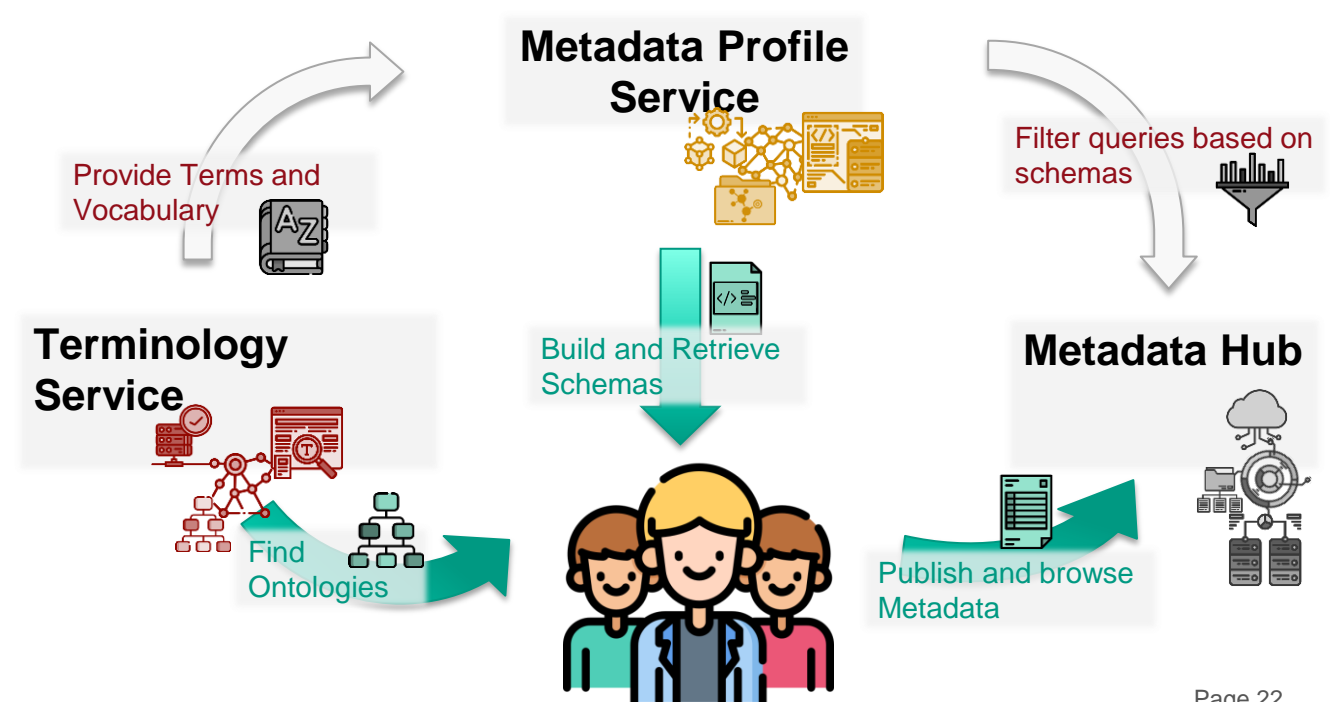
Outlook

- **Classification tagging** of ontologies
 - Top Level ontologies, domain ontologies, ...
 - Domain (e.g. in NFDI4Ing): Production Technology, Material Engineering, Construction and Architecture, ...
- Provision of **trust measures**
 - Usage frequency (download rate, ...)
 - Curation activities (e.g. versions like releases)
- Ontology alignment: **similarity measures**
- Collaborative ontology **creation** and **updating**
- **Visualisation**
- **... and many more is about to come!**

Terminology Service (XII/XII)

- **Base Services Measure S-3: Metadata and terminology services (TIB, TUDA, RWTH, KIT)**
- **Provision of ... services to facilitate the creation of subject and application-specific standardised metadata and their integration into engineering workflows**

<p>Application Profiles</p> <ul style="list-style-type: none"> • Task S-3-1 • TUDA, RWTH, KIT 	<p>Terminology Service</p> <ul style="list-style-type: none"> • Task S-3-2 • TIB 	<p>Metadata Hub</p> <ul style="list-style-type: none"> • Task S-3-3 • TUDA, RWTH, KIT
<p>Aim: Flexible modelling of user specific schemes with maximum re-usability and interoperability</p> <p>Solution: Flexible application-specific selection of suitable elements from controlled terminologies → Application Profiles</p>	<p>Aim: Build a technical infrastructure for terminology management and access</p> <p>Solution: Provision of a web based service, that enables access, curation, and subscription to domain specific terminologies</p>	<p>Aim: Data Management of metadata documents</p> <p>Challenge:</p> <ul style="list-style-type: none"> • Techn. diversity • Inconsistency of the functionalitie <p>Solution: Repository turntable: Uniform access through uniform API</p>



THANKS!