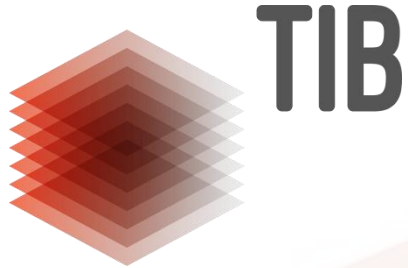
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

*Fachtagung **„Data Science in Studium und Lehre “***

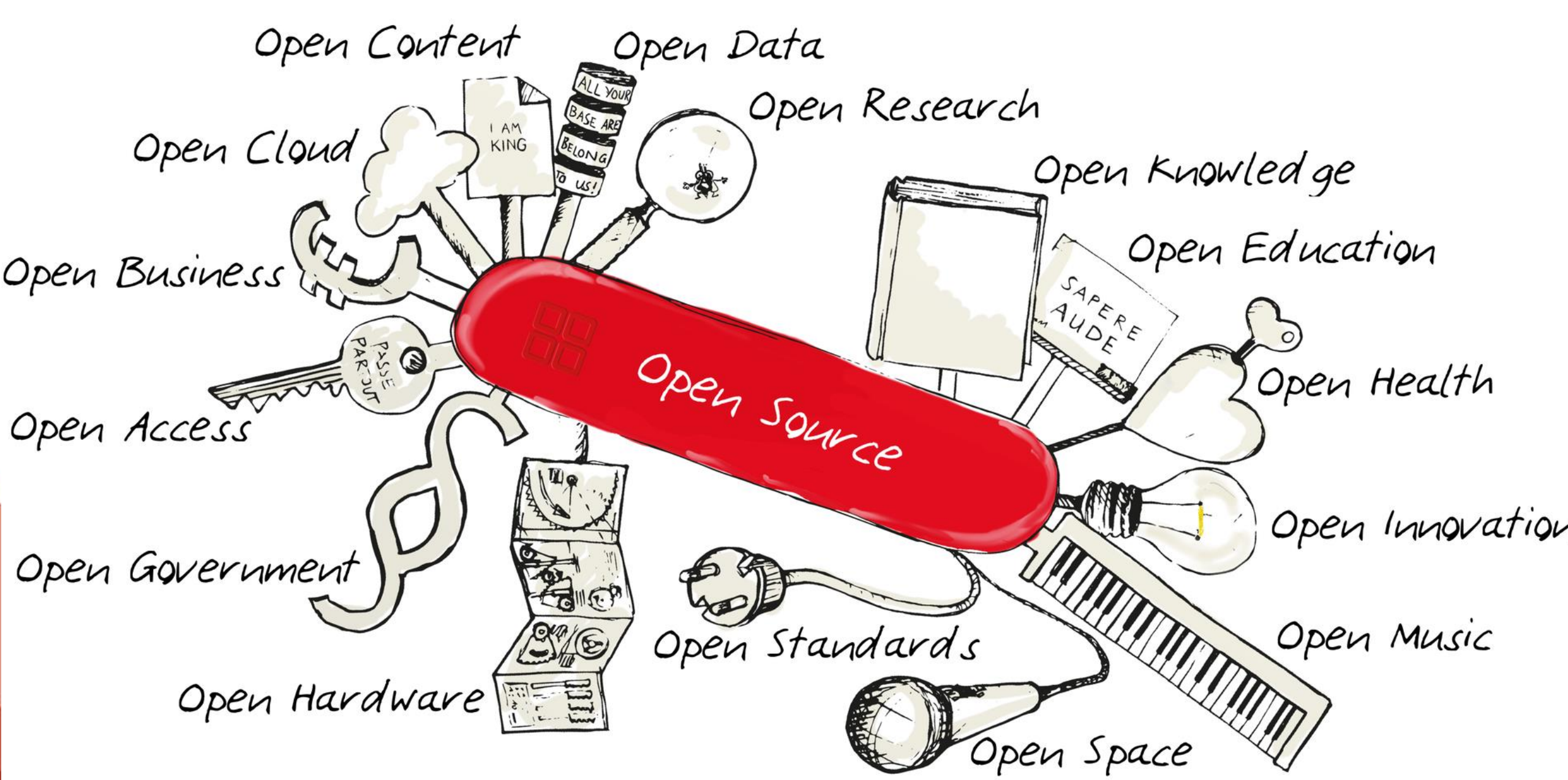
Open Knowledge, Open/FAIR Data, Open Science & Open Education

Prof. Dr. Sören Auer
TIB Technische Informationsbibliothek &
Forschungszentrum L3S, Leibniz Universität

19. September 2018

Leibniz Gemeinschaft

Culture of Open



Open Culture Factors

Digital technologies, networking

Communities, Culture

Collaboration / Crowdsourcing

Zero marginal costs

Economic

Need for transparency

Role models: Linux, Wikipedia, OpenStreetMaps

The Open Definition

The **Open Definition** sets out principles that define "openness" in relation to **data and content**.

It makes **precise** the meaning of "open" in the terms **"open data"** and **"open content"** and thereby ensures **quality** and encourages **compatibility** between different pools of open material.

It can be summed up in the statement that:

> "Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness)."

Put most succinctly:

> "Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**"

Read the full Open Definition »

THE OPEN DEFINITION IN YOUR LANGUAGE

العربية | Беларуская | Български | Català | Czech | Dansk | Deutsch | Eesti | Ελληνικά | English | Español | Euskara | Suomi | Français | Galego | עברית | हिन्दी | Croatian | Magyar | Bahasa Indonesia | Íslenska | Italiano | 日本語 | ಕನ್ನಡ | 한국어 | македонски јазик | नेपाली | Norsk (bokmål) | Polszczyzna | Português Brasileiro | Português | Русский | Shqip | Српски | Svenska | తెలుగు | Türkçe | Українська | 简体中文 | 繁體中文

Important:
- **No discrimination of commercial use**
- **No restriction regarding access or modification**

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

Open/FAIR Research Data

Leibniz Gemeinschaft

Ausgangslage

Wissenschaften im Wandel

Reproduzierbarkeitskrise, Flut von Publikationen (Verdopplung in den letzten 10 Jahren), Peer-Review Crisis, Digitalisierung, Monopolisierungs-bestrebungen kommerzieller Akteure (DEAL), Zunehmende Inter-/Transdisziplinarität, ...

+

Zentrale Rolle von Forschungsdaten

Rat für Informations-infrastrukturen regt Gründung einer Nationalen Forschungsdaten-infrastruktur (NFDI) an, Stärkung der digitalen Kompetenz junger Forscher
European Open Science Cloud (EOSC)

+

Dezentralität & Heterogenität in jeder Hinsicht

Daten: Formate, Datenstrukturen, Metadaten, Lizenzen/ Nutzungs-bedingungen, Anwendungen, Identifikationssysteme, …

Akteure: Forschungsorganisationen, Fachgesellschaften, Wissenschaftler, Universitäten, Bürger (Citizen/Open Science), Politik, Verlage/Unternehmen, ...

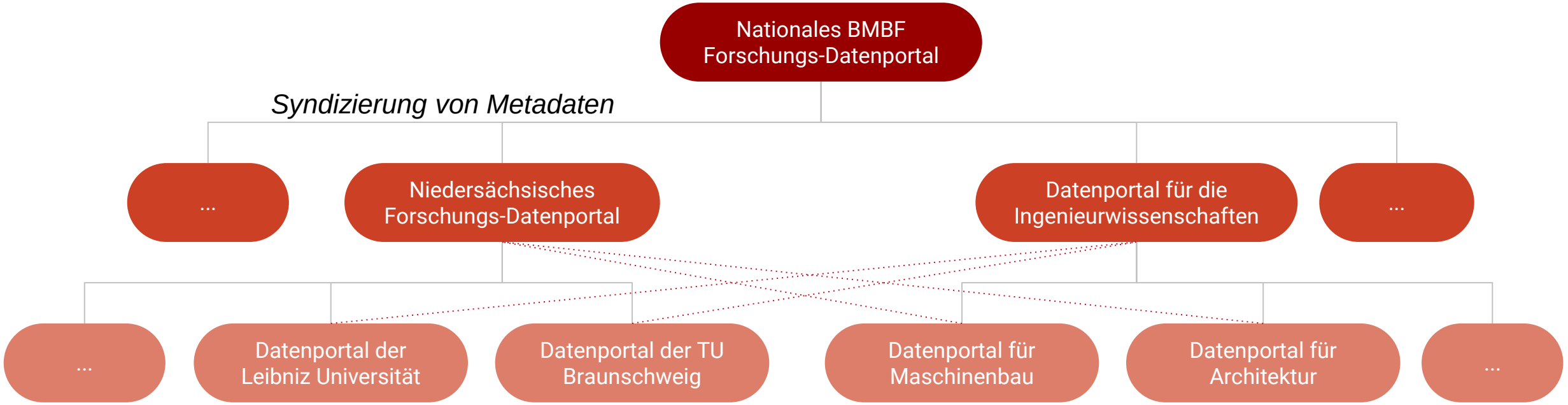
Ziel: Eine Netzwerk von integrierten Forschung-Datenportalen

Nationale Forschungsdateninfrastrukturen müssen der Dezentralität und Heterogenität Rechnung tragen und aber die Vernetzung, Integration und Austausch über Organisations-, Fach- oder Regionsgrenzen effektiv unterstützen

→ Vernetzung von Organisations-, domänen- und regionspezifischen Datenportalen mit direkten Mehrwerten für Forscher

Metadaten werden an “upstream” Datenportale (Aggregatoren) und European Open Science Cloud(EOSC) syndiziert

- Wissenschaftler und andere Akteure können auf Forschungsdaten über verschiedene Einstiegspunkte zugreifen
- DOIs, Metadaten und Vokabulare stellen eindeutigen Zugriff und Vernetzung/Integration sicher



Lösungsansatz

Die sich etablierenden nationalen Forschungsdateninfrastrukturen müssen der Dezentralität und Heterogenität der Forschung Rechnung tragen

Interoperabilität zwischen Dateninfrastrukturen sollte realisiert werden über:

- **Semantisch vernetzte Datenportale** mit DCAT, FAIR Principles, W3C Data on the Web Best Practices
- **Etablierung eines gemeinsamen Verständnisses der Daten** durch Mappings auf Domänen-Vokabulare
- **Agile, iterative Interoperabilität und Weiterentwicklung** der Vokabulare, Mappings, Metadatenstandards mit kooperativen Governancestrukturen

Ergebnis:

- Heterogene Daten aus verschiedenen Domänen können effektiv (in NFDIs und EOSC) integriert werden
- Verschiedene Akteure können agil und effizient zusammenarbeiten ohne in eine zentrale Plattform gezwungen zu werden (“cooperate on standards, compete on implementations”)
- Gänzlich neue Perspektiven für die Wissenschaften: automatisierte Hypothesengenerierung, Maschinelles Lernen, Open Science, ...

Initiativen zur dezentralen, semantischen Datenvernetzung

	Web/Ecommerce	Digitale Bibliotheken	Lebens-wissenschaften	Industrie
Vokabulare	schema.org	Europeana Data Model	DCAT, DC, PROV-O, FOAF, VoiD	DCAT, IDS Vocabulary
Teilnehmer	Ca. 30% der Webseitenbetreiber	Gedächtnisinstitutionen (2000 in D)	Pharmaunternehmen	80 Unternehmen (SAP, Siemens, Telekom, PWC)
Lizenz *Governance*	CC-BY-SA GitHub, Google, Yahoo, Microsoft, Yandex	CC0 Europeana Association	CC-BY-SA	IDS Association
Anwendungen	Google Knowledge Graph (Produktsuche)	DDB.de, Europeana.eu	OpenPhacts.org	Industrial Data Space

Bausteine zur Realisierung

1. CKAN als offene Plattform für Forschungsdaten
2. DCAT-AP Vokabular zum Austausch von Metadaten
3. Erschließung und Integration mit Metadaten und Vokabularen
4. Agile, Gemeinsame Kuratierung von Vokabularen mit VoCol
5. Daten-Portabilität, Reproduzierbarkeit, Datenschutz und Souveränität durch Forschungsdaten-Container
6. Kooperative Governance

CKAN als offene Plattform für Forschungsdaten

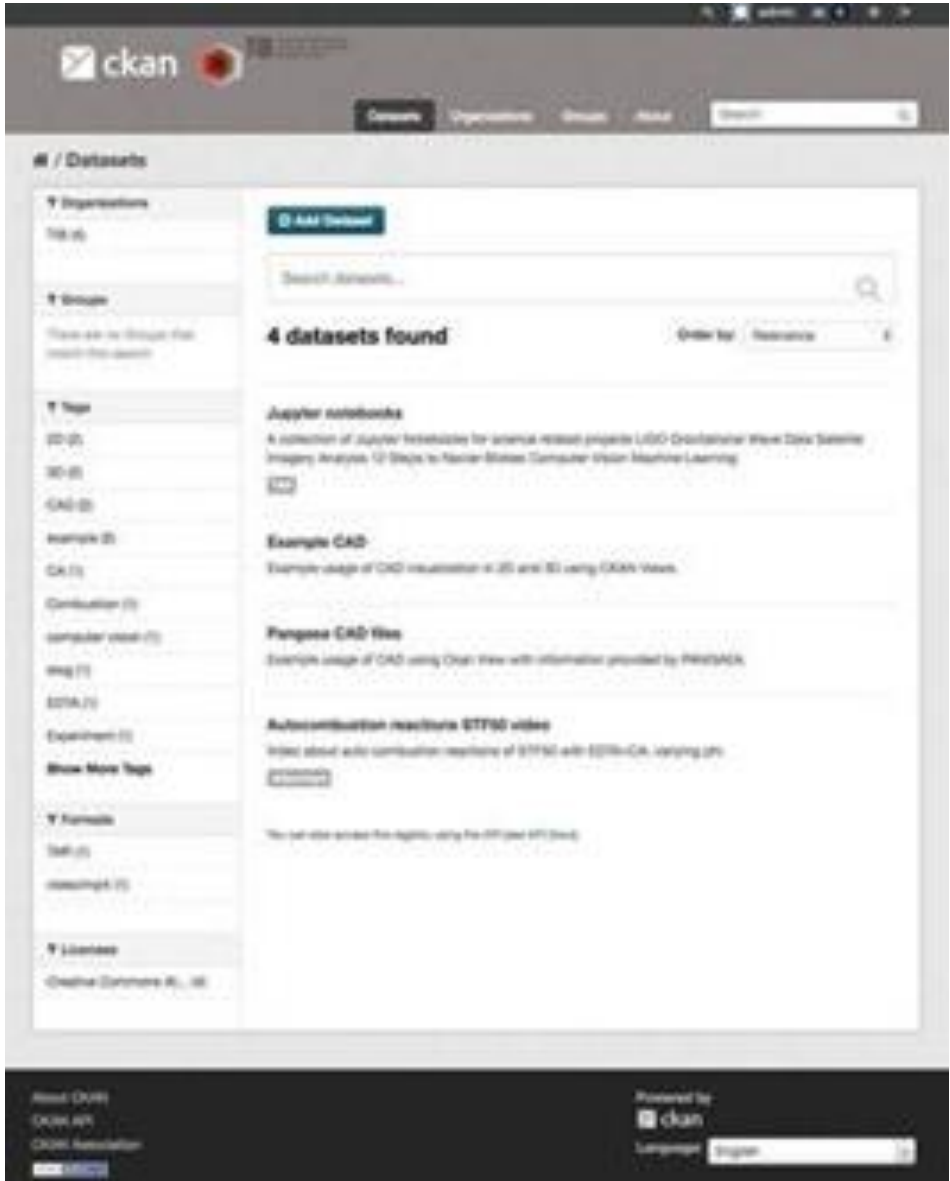
CKAN Data-Repository - wird seit über 10 Jahren für Open Government Data Portale eingesetzt (data.gov, data.gov.uk, govdata.de)

- Open-source, mit vielen Erweiterungen und aktiver Community
- Unterstützung semantischer Metadaten mit DCAT und DCAT-AP
- Out-of-the-box Vernetzung und Syndizierung von CKAN-Instanzen

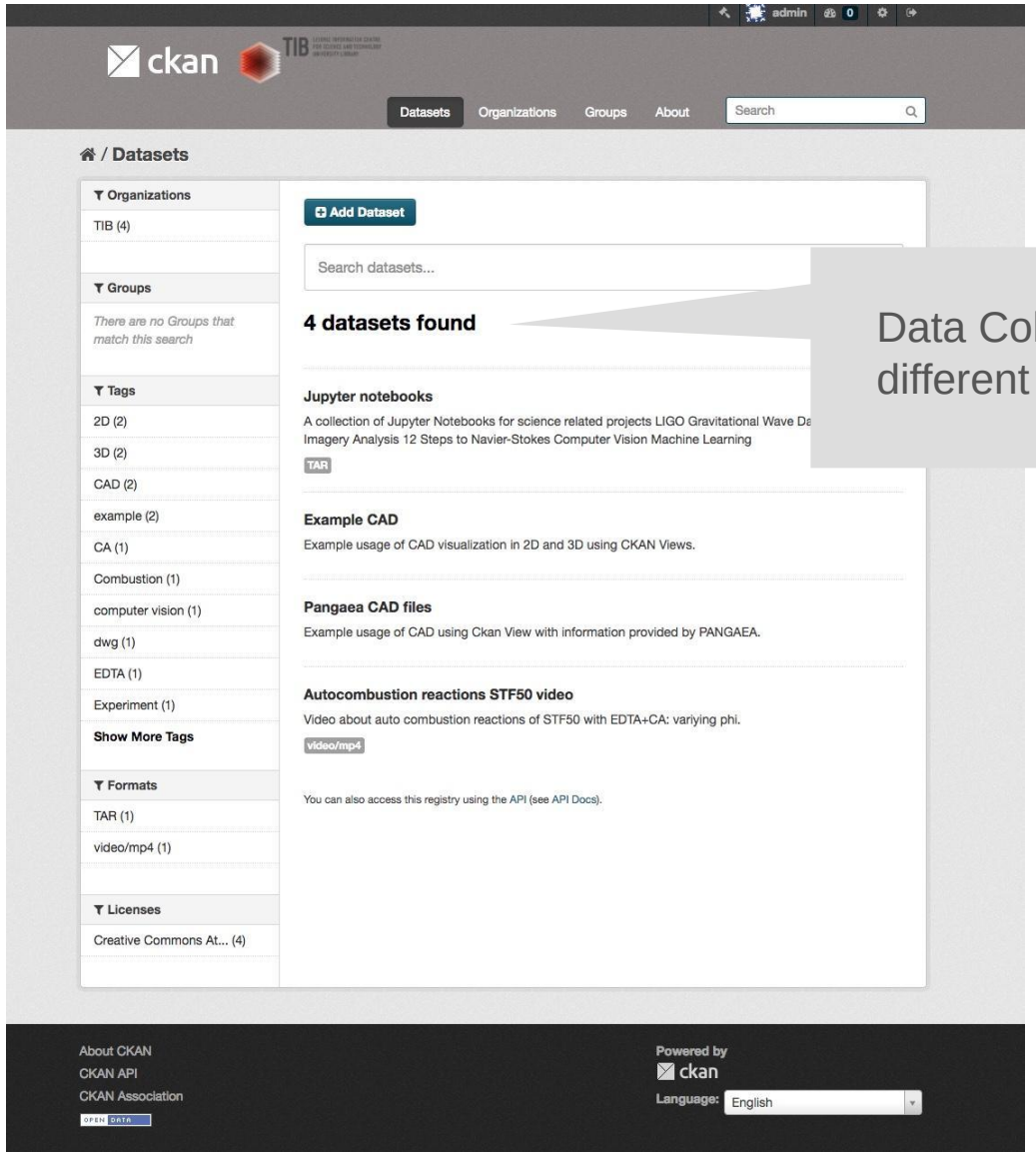
Leibniz Datamanager (http://datamanager.tib.eu) - spezifisch angepasste CKAN Distribution für Forschungsdatenmanagement

- Integration mit DOI und DataCite
- Viewer für verschiedene Arten von Forschungsdaten
- Unterstützung für Jupyter Notebooks

Mit CKAN/Leibniz Datamanager können Forschungsdatenportale effizient realisiert und automatisch mit anderen vernetzt werden

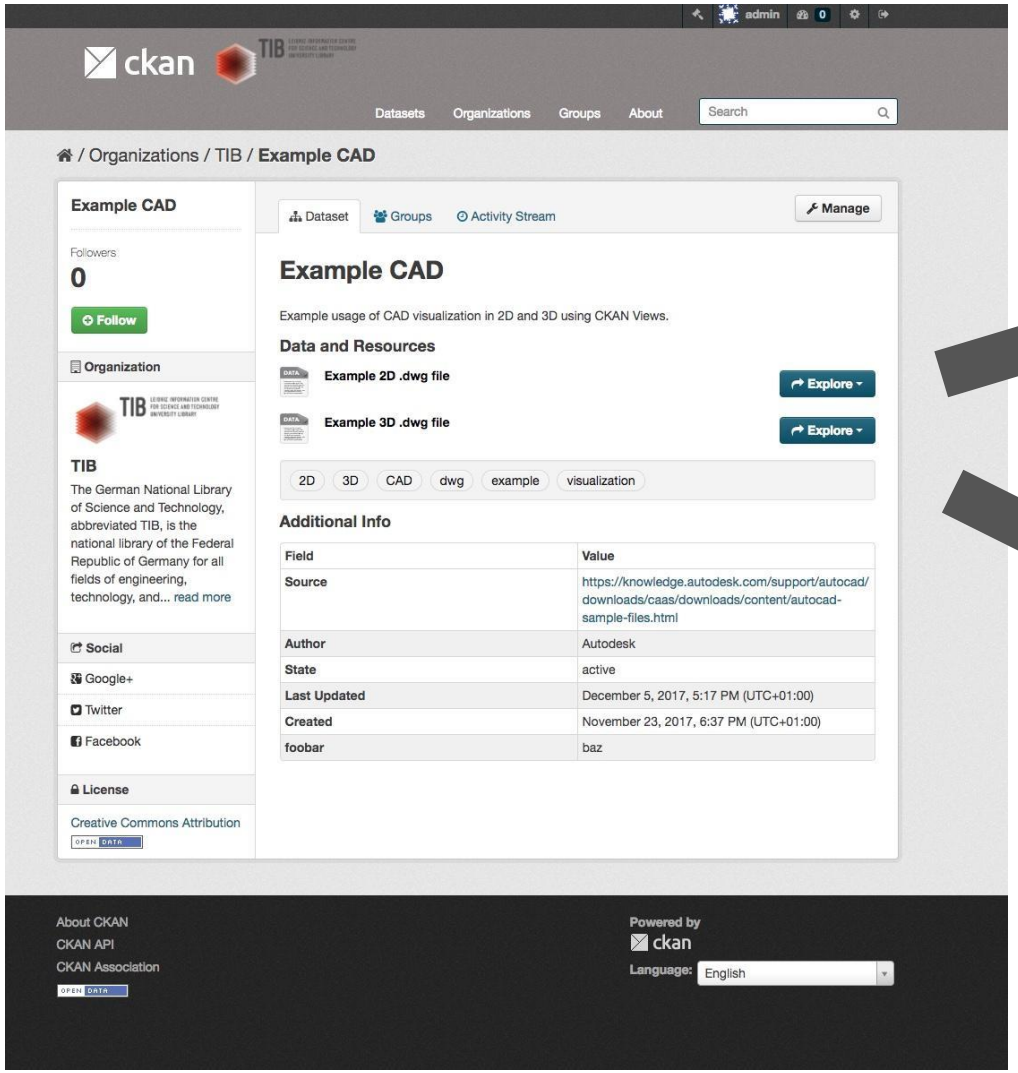


CKAN: A Repository for Heterogeneous Data Collections

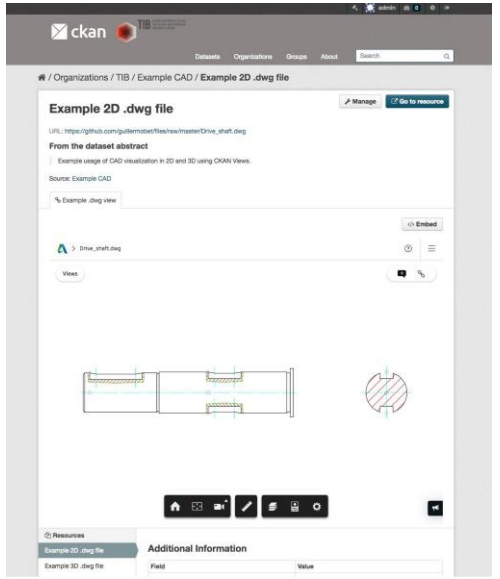


Data Collections with different formats

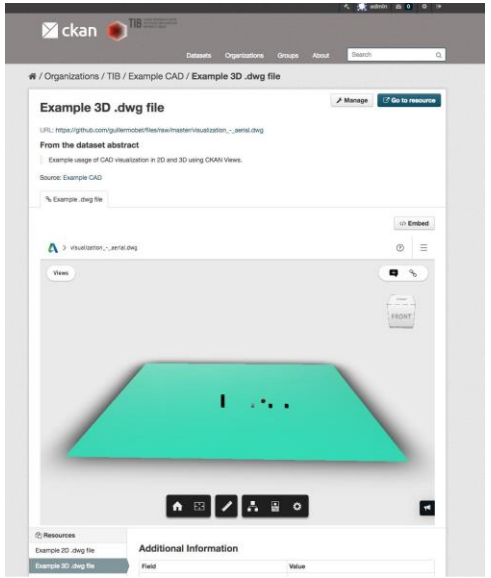
CKAN: Different Views of the Same Data Collections



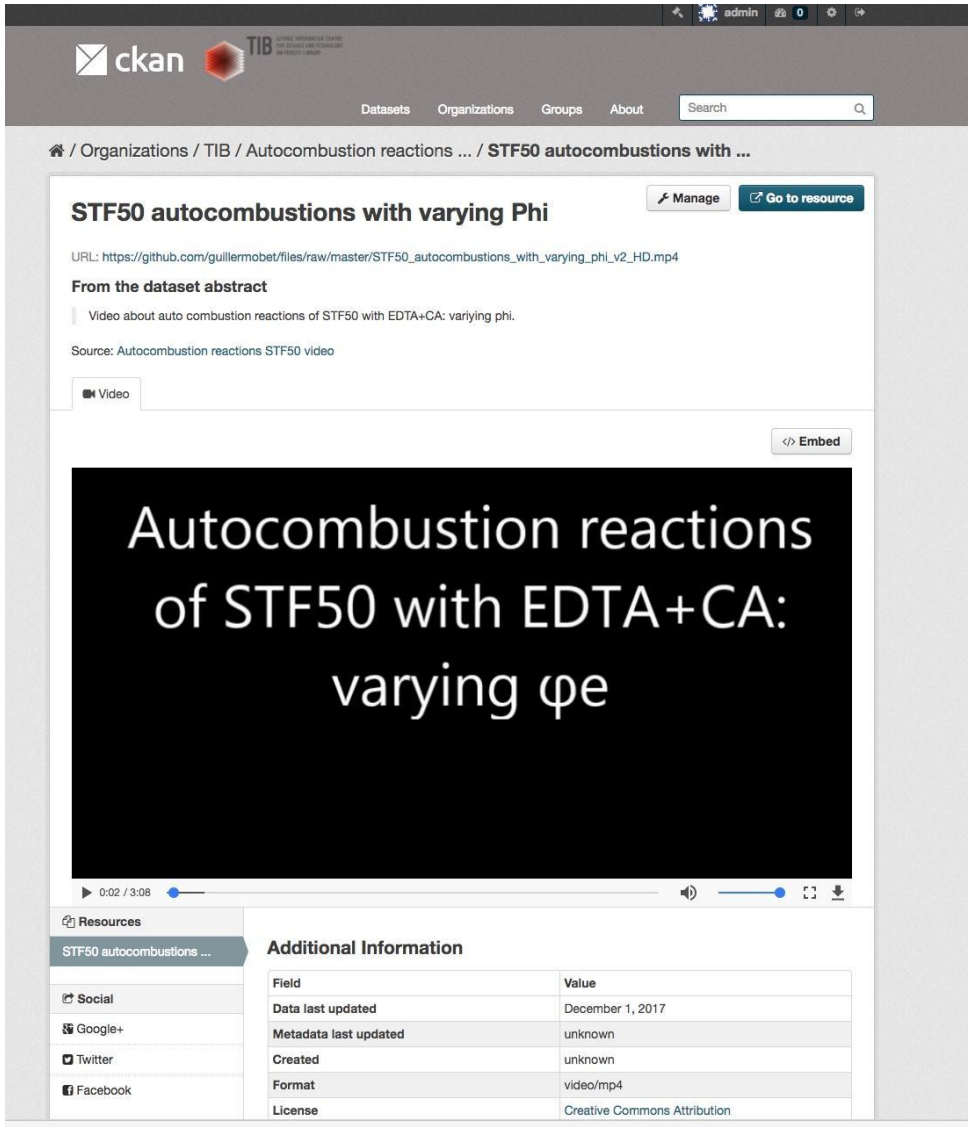
2D View

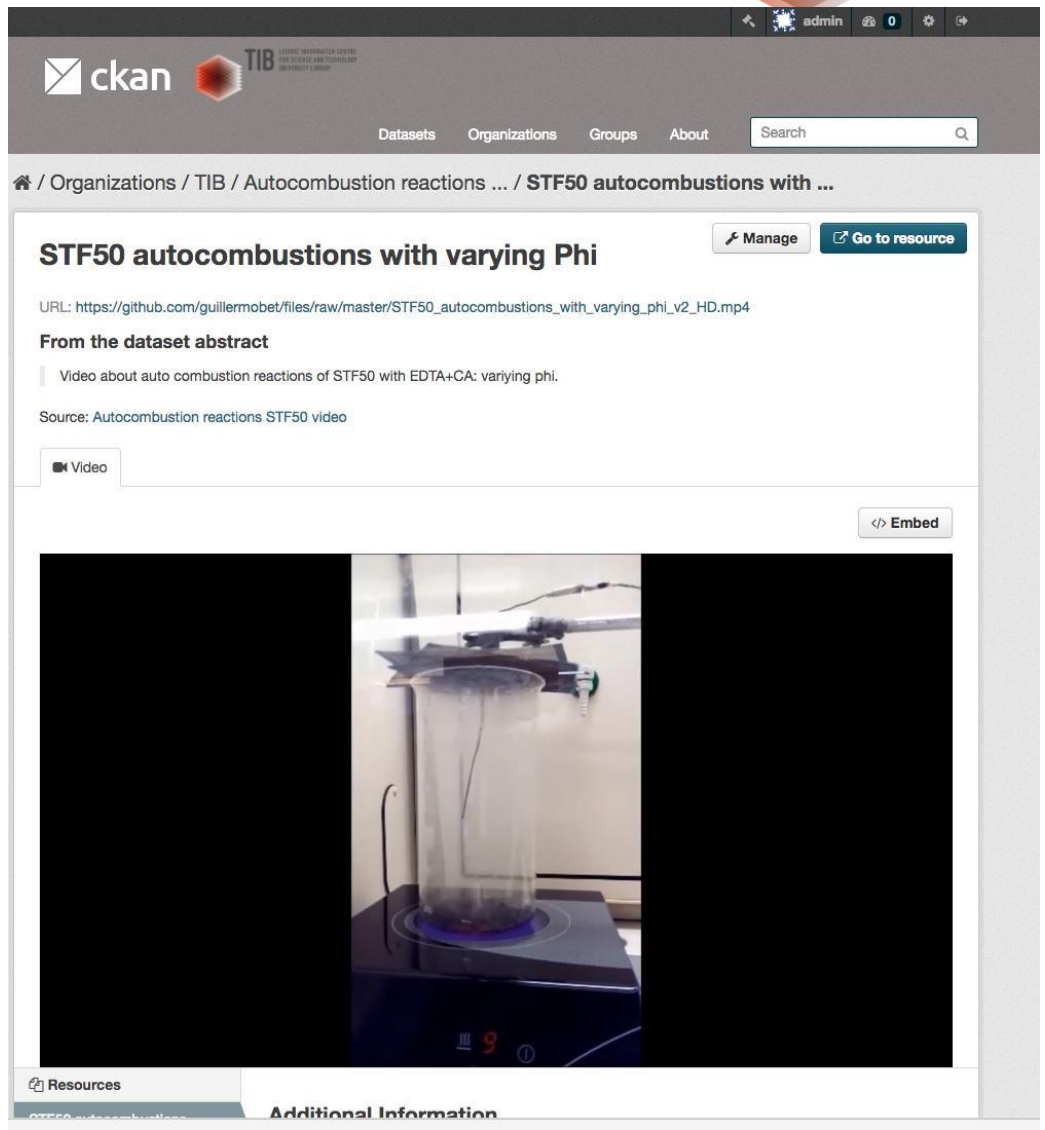


3D View

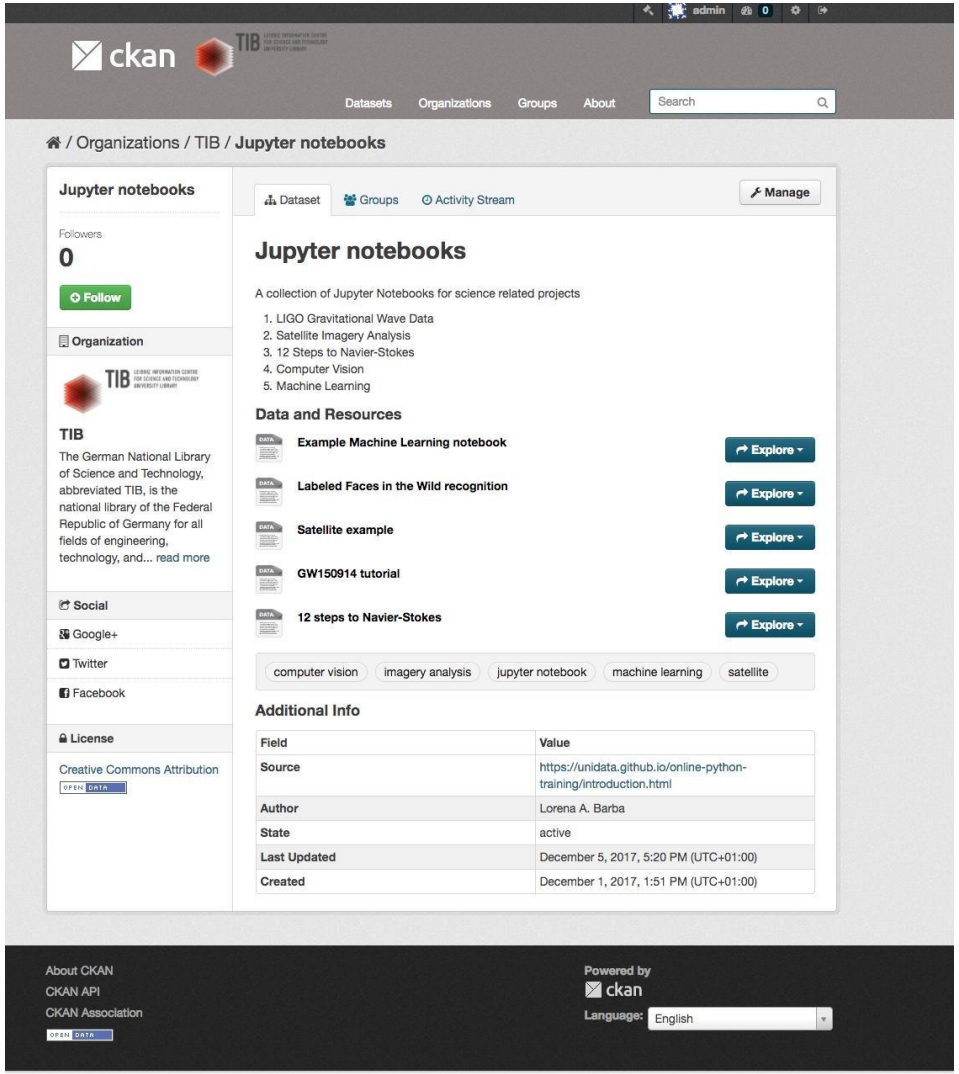


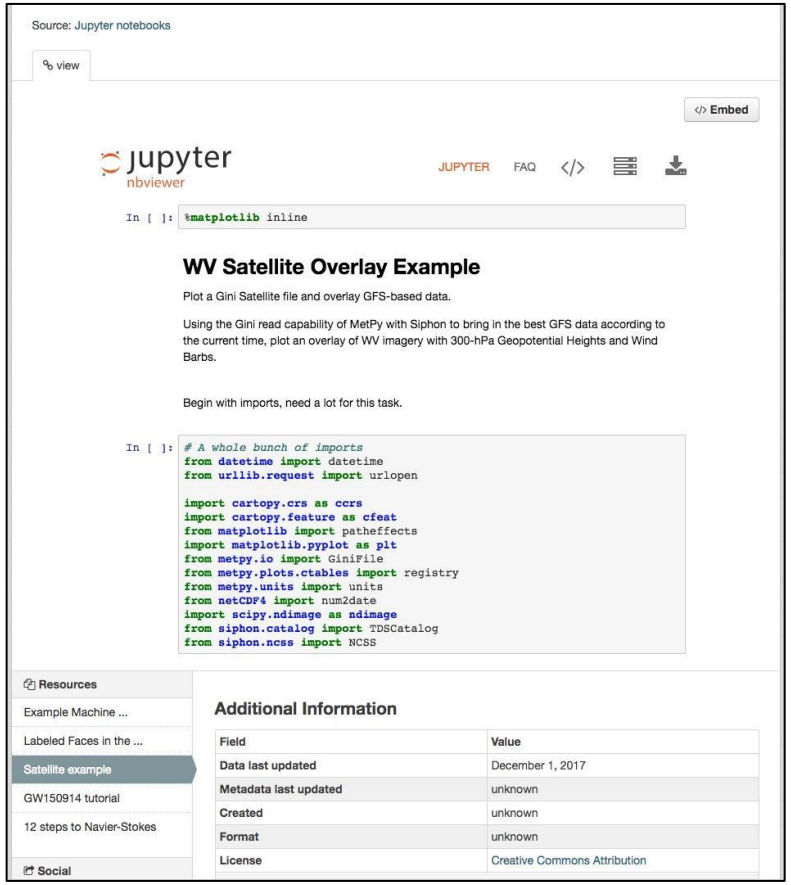
CKAN: Playing a Video



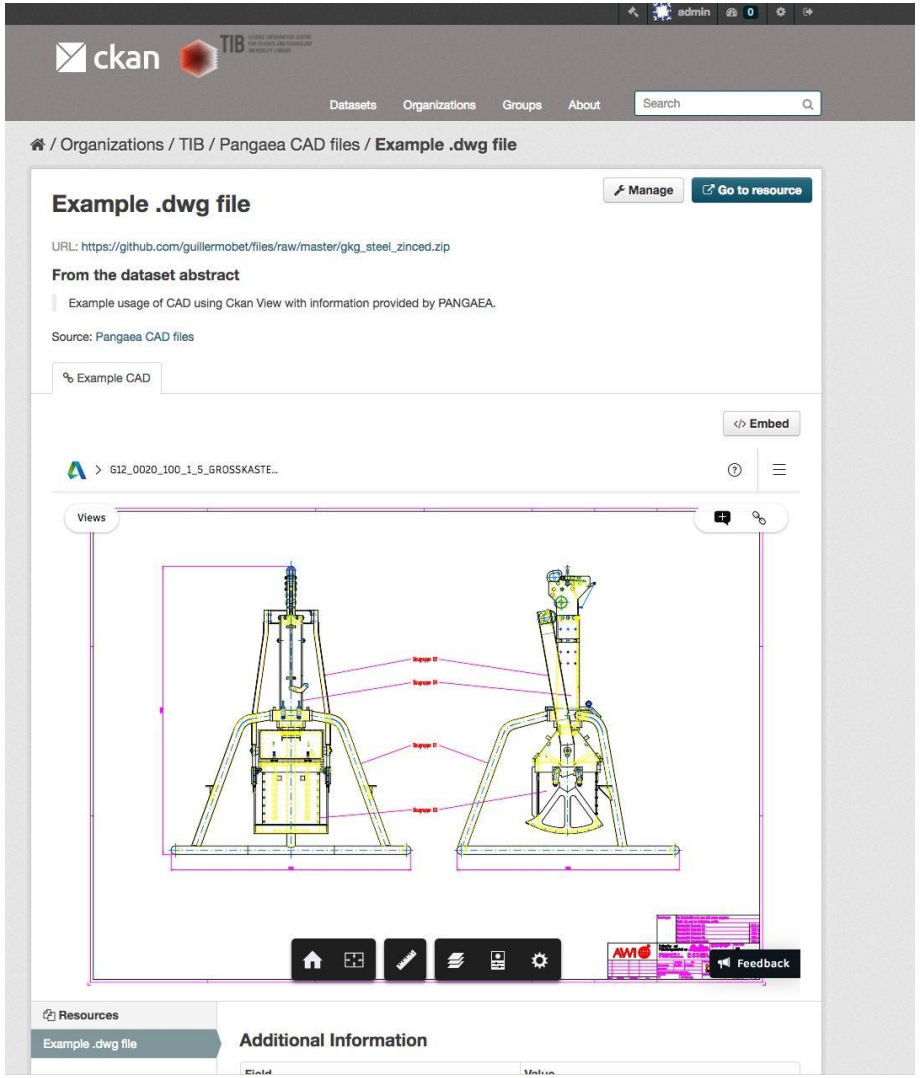


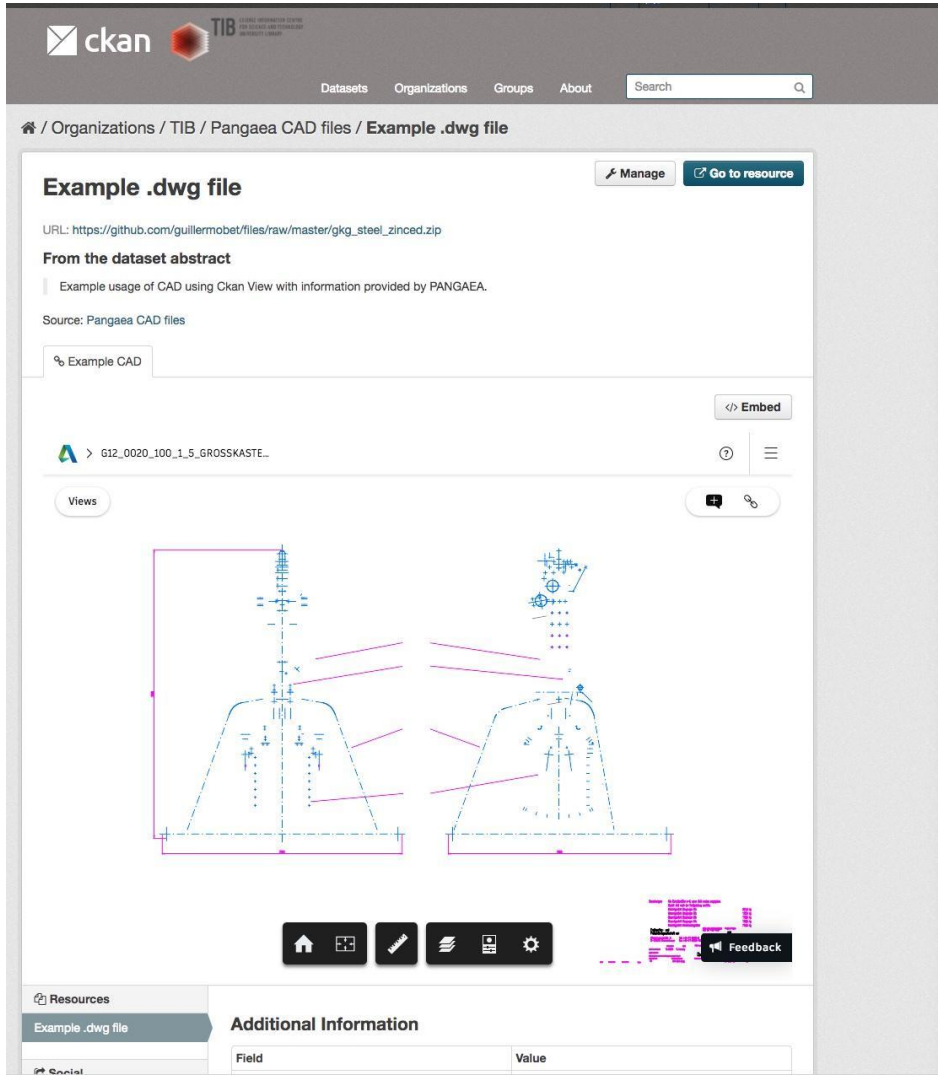
CKAN: Jupyter Notebooks for Demonstrating Live Code



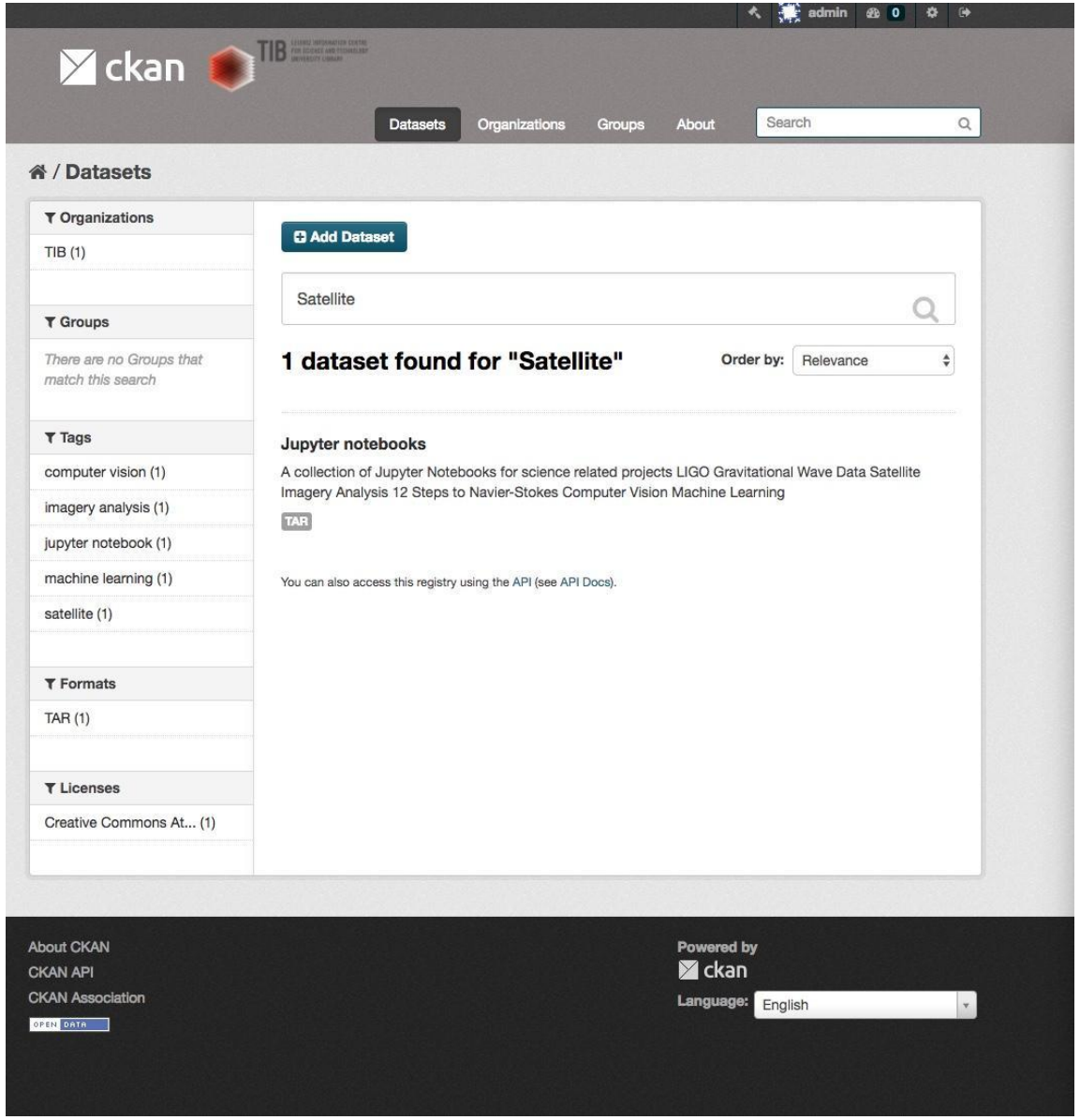


CKAN: Visualizations of Data Collections using Auto CAD

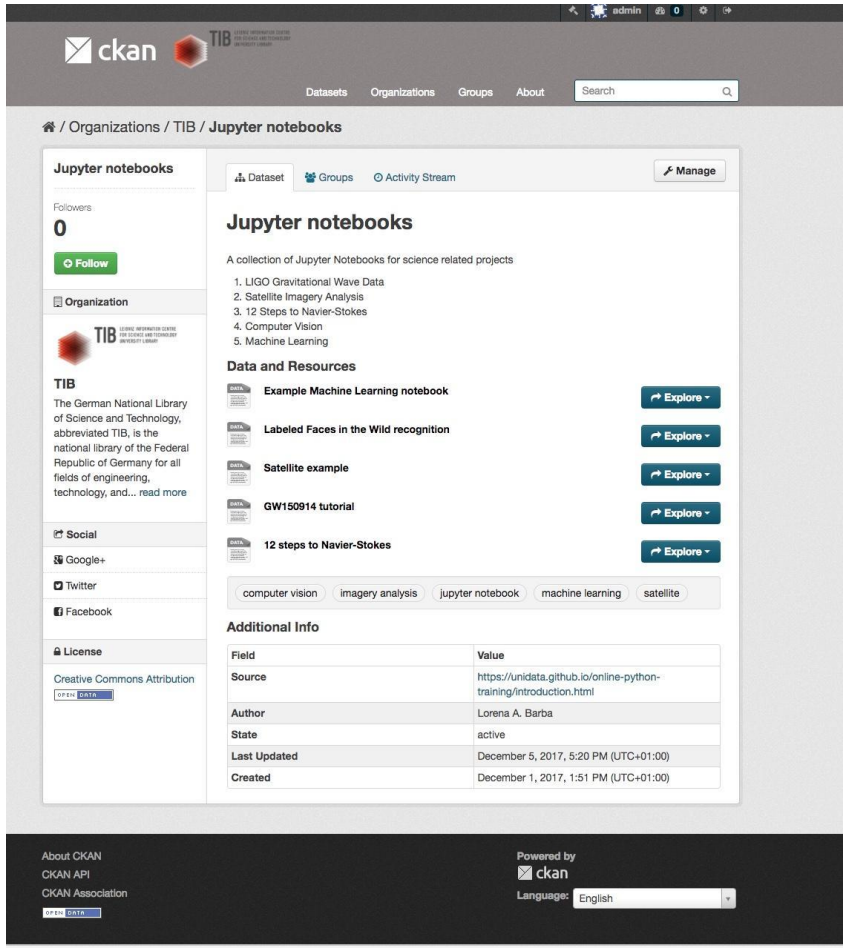




CKAN: Searching Data Collections



CKAN: RDF Description of Data Collections




```
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix gsp: <http://www.opengis.net/ont/geosparql#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix time: <http://www.w3.org/2006/time> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e> a dcat:Dataset ;
    dct:description """A collection of Jupyter Notebooks for science related projects\r
\r
1. LIGO Gravitational Wave Data\r
2. Satellite Imagery Analysis\r
3. 12 Steps to Navier-Stokes\r
4. Computer Vision\r
5. Machine Learning""" ;
    dct:identifier "labefb2e-6a83-4004-b7db-74c34b545d2e" ;
    dct:issued "2017-12-01T12:51:12.218503"^^xsd:dateTime ;
    dct:modified "2017-12-05T16:20:26.498874"^^xsd:dateTime ;
    dct:publisher <http://194.95.157.196:5000/organization/0c5362f5-b99e-41db-8256-3d0d7549bf4d> ;
    dct:title "Jupyter notebooks" ;
    dcat:contactPoint [ a vcard:Organization ;
            vcard:fn "Lorena A. Barba" ] ;
    dcat:distribution <http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/036bcac0-c857-4bf0-bc71-1c78ed35d93a>,
        <http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/1e335b61-123e-4ba4-9c5b-9d1d6309dba9>,
        <http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/4577e551-96f8-4e13-ac81-012a866d00ac>,
        <http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/e4cc8bf6-5e32-4c1f-b22e-109d47340c96>,
        <http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/ec1c5422-b8ab-4401-96fb-0792dacb8e40> ;
    dcat:keyword "computer vision",
        "imagery analysis",
        "jupyter notebook",
        "machine learning",
        "satellite" ;
    dcat:landingPage <https://unidata.github.io/online-python-training/introduction.html> .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/036bcac0-c857-4bf0-bc71-1c78ed35d93a> a dcat:Distribution ;
    dct:title "Labeled Faces in the Wild recognition" ;
    dcat:accessURL <https://raw.githubusercontent.com/ogrisel/notebooks/master/Labeled%2520Faces%2520in%2520the%2520Wild%2520recognition.ipynb> ;
    dcat:byteSize 717993.0 .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/1e335b61-123e-4ba4-9c5b-9d1d6309dba9> a dcat:Distribution ;
    dct:title "Example Machine Learning notebook" ;
    dcat:accessURL <https://raw.githubusercontent.com/rhiever/Data-Analysis-and-Machine-Learning-Projects/master/example-data-science-notebook/Example%20Machine%20Learning%20Notebook.ipynb> ;
    dcat:byteSize 703819.0 .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/4577e551-96f8-4e13-ac81-012a866d00ac> a dcat:Distribution ;
    dct:title "GW150914 tutorial" ;
    dcat:accessURL <https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.ipynb> ;
    dcat:byteSize 2683661.0 .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/e4cc8bf6-5e32-4c1f-b22e-109d47340c96> a dcat:Distribution ;
    dct:title "Satellite example" ;
    dcat:accessURL <http://unidata.github.io/python-gallery/_downloads/Satellite_Example.ipynb> ;
    dcat:byteSize 7216.0 .

<http://194.95.157.196:5000/dataset/labefb2e-6a83-4004-b7db-74c34b545d2e/resource/ec1c5422-b8ab-4401-96fb-0792dacb8e40> a dcat:Distribution ;
    dct:format "TAR" ;
    dct:title "12 steps to Navier-Stokes" ;
    dcat:accessURL <https://github.com/guillermobet/files/raw/master/12%20steps%20to%20Navier-Stokes.tar.gz> ;
    dcat:byteSize 5708395.0 ;
    dcat:mediaType "application/x-tar" .

<http://194.95.157.196:5000/organization/0c5362f5-b99e-41db-8256-3d0d7549bf4d> a foaf:Organization ;
    foaf:name "TIB" .
```

RDF Description of the Jupyter Notebooks

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

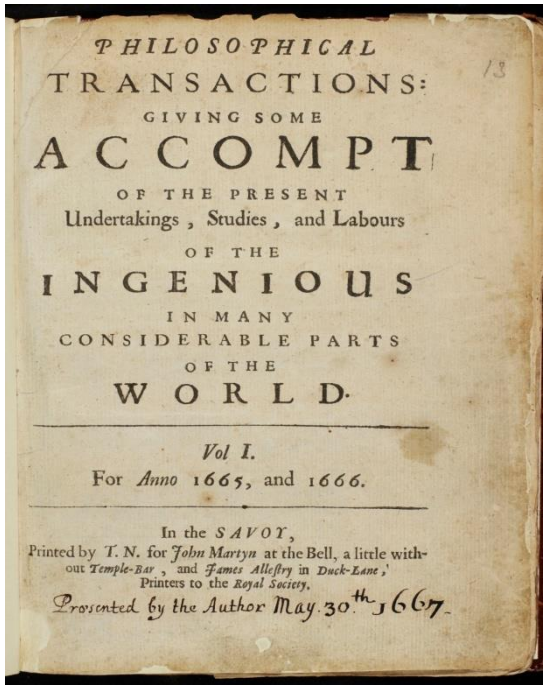
TIB

Open Science

Leibniz Gemeinschaft

Scholarly Communication has not changed (much)

17th century

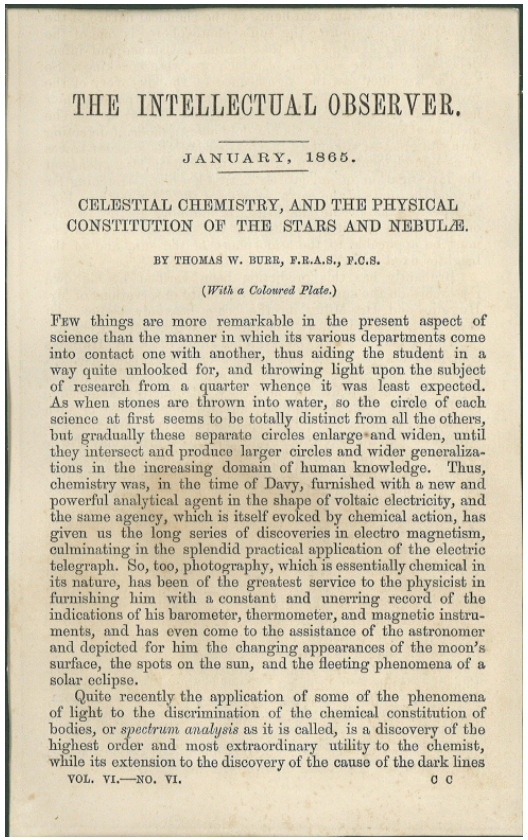


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.
Presented by the Author May. 30th 1667.

19th century



THE INTELLECTUAL OBSERVER.

JANUARY, 1865.

CELESTIAL CHEMISTRY, AND THE PHYSICAL CONSTITUTION OF THE STARS AND NEBULÆ.

BY THOMAS W. BURR, F.R.A.S., F.C.S.

(*With a Coloured Plate.*)

Few things are more remarkable in the present aspect of science than the manner in which its various departments come into contact one with another, thus aiding the student in a way quite unlooked for, and throwing light upon the subject of research from a quarter whence it was least expected. As when stones are thrown into water, so the circle of each science at first seems to be totally distinct from all the others, but gradually these separate circles enlarge and widen, until they intersect and produce larger circles and wider generalizations in the increasing domain of human knowledge. Thus, chemistry was, in the time of Davy, furnished with a new and powerful analytical agent in the shape of voltaic electricity, and the same agency, which is itself evoked by chemical action, has given us the long series of discoveries in electro magnetism, culminating in the splendid practical application of the electric telegraph. So, too, photography, which is essentially chemical in its nature, has been of the greatest service to the physicist in furnishing him with a constant and unerring record of the indications of his barometer, thermometer, and magnetic instruments, and has even come to the assistance of the astronomer and depicted for him the changing appearances of the moon's surface, the spots on the sun, and the fleeting phenomena of a solar eclipse.

Quite recently the application of some of the phenomena of light to the discrimination of the chemical constitution of bodies, or *spectrum analysis* as it is called, is a discovery of the highest order and most extraordinary utility to the chemist, while its extension to the discovery of the cause of the dark lines

20th century

A Relational Model of Data for Large Shared Data Banks

E. F. Codd
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity
CR CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

1. Relational Model and Normal Form

1.1. Introduction

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a paper by Childs [1], the principal application of relations to data

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for noninferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the "connection trap").

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

1.2. Data Dependencies in Present Systems

The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed *without logically impairing some application programs* is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need to be removed are: ordering dependence, indexing dependence, and access path dependence. In some systems these dependencies are not clearly separable from one another.

1.2.1. *Ordering Dependence.* Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each

21th century

AGDISTIS - Graph-Based Disambiguation of Named Entities using Linked Data

Ricardo Usbeck[1,2], Axel-Cyrille Ngonga Ngomo[1], Michael Röder[1,2], Daniel Gerber[1], Sandro Athaide Coelho[3], Sören Auer[4], and Andreas Both[2]

[1] University of Leipzig, Germany , [2] R & D, Unister GmbH, Germany, [3] Federal University of Juiz de Fora, Brazil, [4] University of Bonn & Fraunhofer IAIS, Germany
email: {usbeck|ngonga}@informatik.uni-leipzig.de

Abstract. Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

1 Introduction

The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.[1] Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings `Barack Obama` and `Washington, D.C.`. A high-quality DBpedia-based named entity disambiguation (NED) approach would use these already recognized named entities and map the strings

[1] Data gathered from `http://www.worldwidewebsize.com/` on January 4th, 2014.

Meanwhile other information intense domains were completely disrupted: mail order catalogs, street maps, phone books, …

We need to rethink the way how research is represented and communicated

Challenges we are facing:

Digitalisation of Science	Monopolisation by commercial actors	Reproducibility Crisis	Proliferation of publications	Deficiency of Peer Review
• Data integration and analysis • Digital collaboration	• Publisher look-in effects • Maximization of profits [1]	• Majority of experiments are hard or not reproducible [2]	• Publication output doubled within a decade • continues to rise [3]	• Deteriorating quality [4] • Predatory publishing

[1] http://thecostofknowledge.com, https://www.projekt-deal.de
[2] M. Baker: *1,500 scientists lift the lid on reproducibility*, *Nature,* 2016.
[3] *Science and Engineering Publication Output Trends*, National Science Foundation, 2018.
[4] J. Couzin-Frankel: *Secretive and Subjective, Peer Review Proves Resistant to Study.* Science, 2013.

AUSLEIHEN & BESTELLEN | RECHERCHIEREN & ENTDECKEN | LERNEN & ARBEITEN | PUBLIZIEREN & ARCHIVIEREN | FORSCHUNG & ENTWICKLUNG | DIE TIB

TIB

CRISPR

Nur im Bibliothekskatalog der TIB suchen

Search for CRISPR:
>4.000 Results

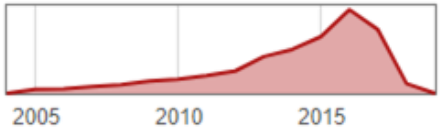
Seite 1 von 4.373 Ergebnissen

Sortieren nach: Relevanz | Aktualität | Titel

« ‹ 1 2 3 4 5 6 7 ›

Treffer erschließen

Erscheinungsjahr



- Ok

Medientyp

- Aufsatz (Zeitschrift) (3.961)
- Patent (205)
- Hochschulschrift (93)
- Aufsatz (Konferenz) (34)
- Sonstige (30)

+ Weitere

Datenquelle

- British Library Online Contents (1.369)
- CiteSeerX (558)

CRISPR Critters and CRISPR Cracks

Charo, R. Alta / Greely, Henry T. | Taylor & Francis Verlag | 2015
This essay focuses on possible nonhuman applications of **CRISPR**/Cas9 that are likely to be widely overlooked because they are unexpected

...

CRISPR BIOLOGY CRISPR-Cas: Adapting to change

Jackson, S. A. | British Library Online Contents | 2017

CRISPR decoys: Competitive inhibitors of CRISPR immunity

Maniv, I. / Hatoum-Aslan, A. / Marraffini, L.A. | British Library Online Contents | 2013

CRISPR-Cas

Das Immunsystem der Prokaryoten
Marchfelder, Anita / Maier, Lisa-Katharina / Heidrich, Nadia et al. | Wiley | 2013

Google Scholar

CRISPR

Search for CRISPR:
>163.000 Results

Articles

About 163,000 results (0.04 sec)

My profile

My library

Any time
Since 2018
Since 2017
Since 2014
Custom range...

Sort by relevance
Sort by date

include patents
include citations

Create alert

[PDF] **CRISPR**-P: a web tool for synthetic single-guide RNA design of **CRISPR**-system in plants

[PDF] researchgate.net

IS Palindromic Repeats **CRISPR**-associated - 2014 - researchgate.net

Dear Editor, Precise and efficient genome editing is very important for gene functiona
characterization. In recent years, sequence-specific DNA nucleases have been deve
to increase the efficiency of gene targeting or genome editing in animals and plants.

Cited by 149 Related articles All 10 versions

[HTML] Multiplex genome engineering using **CRISPR**/Cas systems

L Cong, FA Ran, D Cox, S Lin, R Barretto... - ..., 2013 - science.sciencemag.org

Functional elucidation of causal genetic variants and elements requires precise geno
editing technologies. The type II prokaryotic **CRISPR** (clustered regularly interspaced
palindromic repeats) adaptive immune system has been shown to facilitate RNA-guid

Cited by 6299 Related articles All 48 versions

How good is CRISPR (wrt. precision, safety, cost)?

What specifics has genome editing with insects?

Who has applied it to butterflies?

[HTML] **CRISPR** provides acquired resistance against viruses in prokaryotes

[HTML] sciencemag.org

R Barrangou, C Fremaux, H Deveau, M Richards... - ..., 2007 - science.sciencemag.org

Clustered regularly interspaced short palindromic repeats (**CRISPR**) are a distinctive feature of the genomes of most Bacteria and Archaea and are thought to be involved in resistance to bacteriophages. We found that, after viral challenge, bacteria integrated new spacers ...

Cited by 2815 Related articles All 28 versions

Efficient genome editing in zebrafish using a **CRISPR**-Cas system

[HTML] nih.gov

WY Hwang, Y Fu, D Reyon, ML Maeder, SQ Tsai... - Nature ..., 2013 - nature.com

In bacteria, foreign nucleic acids are silenced by clustered, regularly interspaced, short palindromic repeats (**CRISPR**). **CRISPR**-associated (Cas) systems. Bacterial type II **CRISPR**

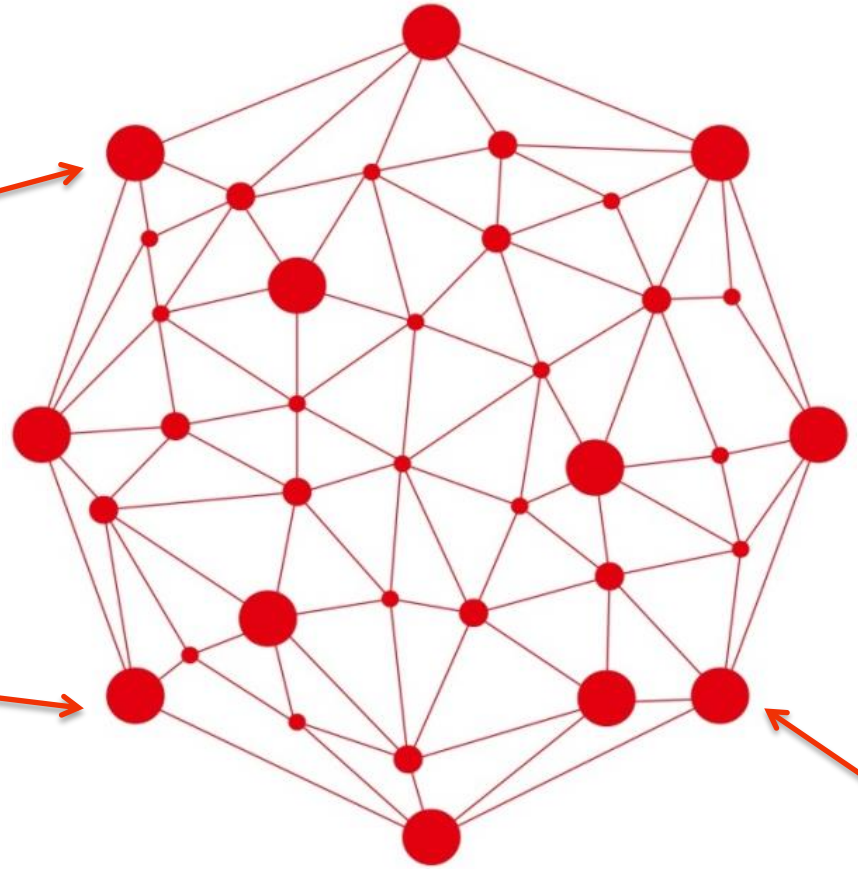
Concepts

Overarching Concepts

- Research problems
- Definitions
- Research approaches
- Methods

Artefacts

- Publications
- Data
- Software
- Image/Audio/Video
- Knowledge Graphs / Ontologies



Domain specific Concepts

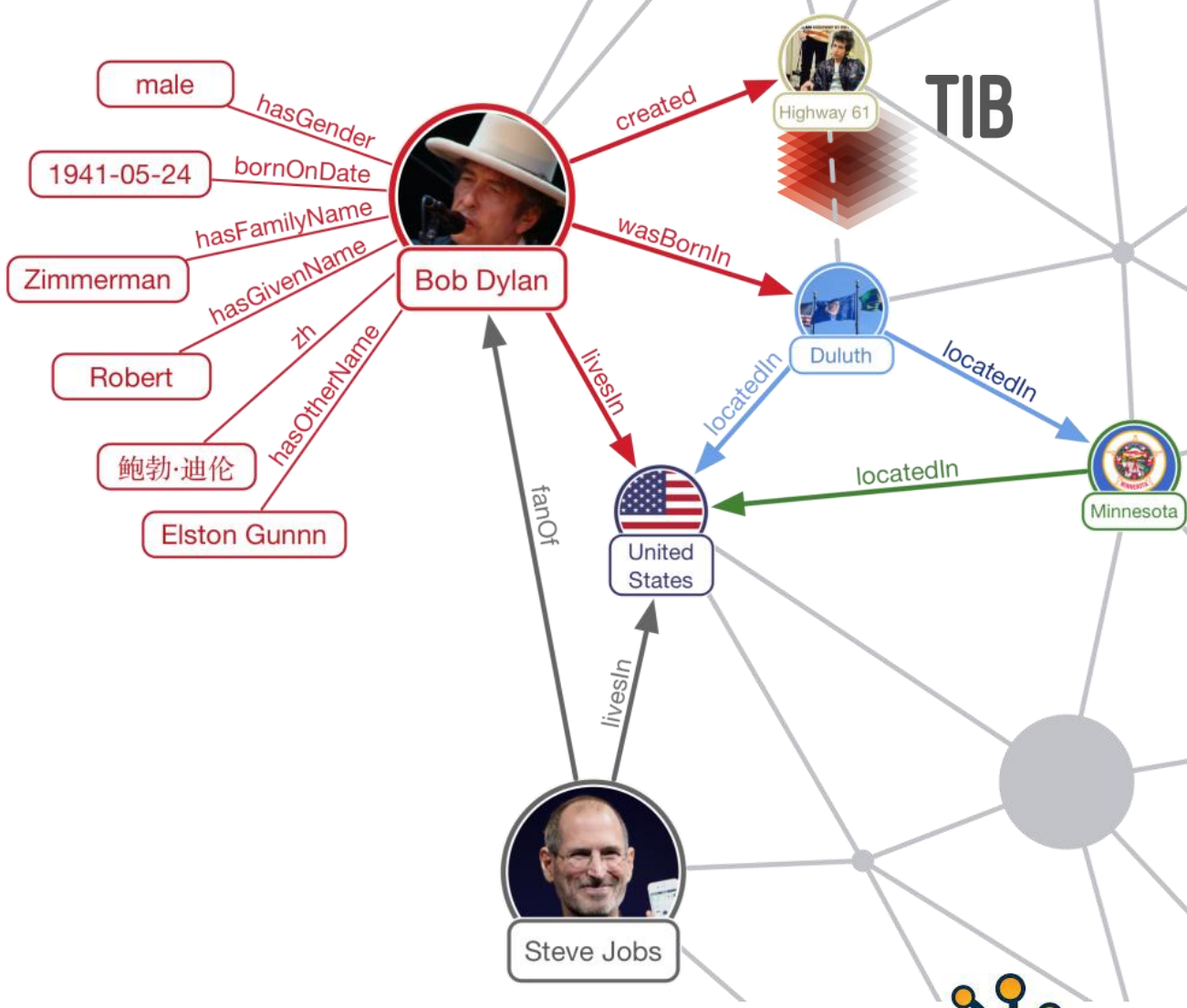
Mathematics	Physics	Chemistry	Computer Science	Technology	Architecture
• Definitions • Theorems • Proofs • Methods • …	• Experiments • Data • Models • …	• Substances • Structures • Reactions • …	• Concepts • Implemen-tations • Evaluations • …	• Standards • Processes • Elements • Units, Sensor data	• Regulations • Elements • Models • …

Cognitive Knowledge Graphs for scholarly knowledge

KGs are *proven to capture factual knowledge* [1]

Research Challenge: Manage
- Uncertainty & disagreement
- Varying semantic granularity
- Emergence, evolution & provenance
- Integrating existing domain models

But maintain *flexibility and simplicity*



DBpedia

[1] S Auer et al.: *DBpedia: A nucleus for a web of open data.* 6th Int. Semantic Web Conf. (ISWC)
cf. also knowledge graphs from: WikiData, BBC, Google, Bing, Thomson Reuters, AirBnB, BNY Mellon …

Chemistry Example: CRISPR Genome Editing

New Results

Previous Next

A practical guide to CRISPR/Cas9 genome editing in Lepidoptera

Linlin Zhang, Robert Reed

doi: https://doi.org/10.1101/130344

Posted June 22, 2017.

Now published in *Diversity and Evolution of Butterfly Wing Patterns* doi: 10.1007/978-981-10-4956-9_8

Abstract | Info/History | Metrics | Preview PDF

Download PDF | Share | Email | Citation Tools

Tweet | Like 0 | G+

Abstract

CRISPR/Cas9 genome editing has revolutionized functional genetic work in many organisms and is having an especially strong impact in emerging model systems. Here we summarize recent advances in applying CRISPR/Cas9 methods in Lepidoptera, with a focus on providing practical advice on the entire process of genome editing from experimental design through to genotyping. We also describe successful targeted GFP knock-ins that we have achieved in butterflies. Finally, we provide a complete, detailed protocol for producing targeted long deletions in butterflies.

Subject Area

Genetics

Subject Areas

All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics

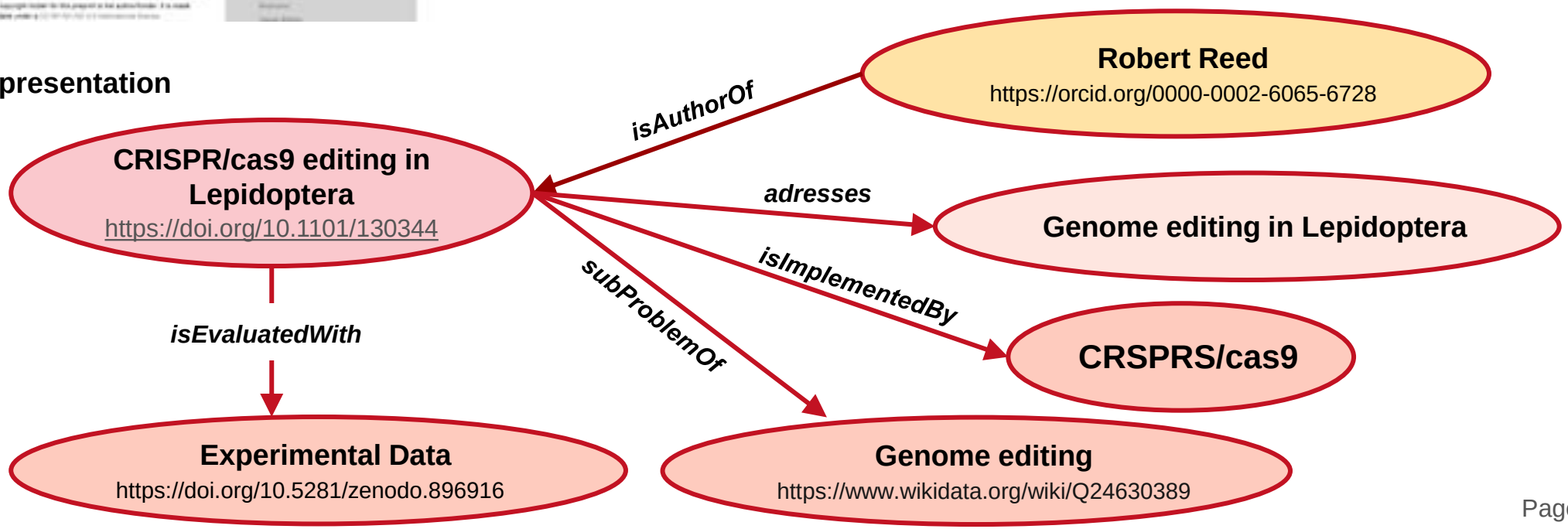
Chemistry Example: Populating the Graph

1. Original Publication

2. Adaptive Graph Curation & Completion

Author	Robert Reed
Research Problem	Genome editing in Lepidoptera
Methods	CRISPR/cas9
Applied on	Lepidoptera
Experimental Data	https://doi.org/10.5281/zenodo.896916

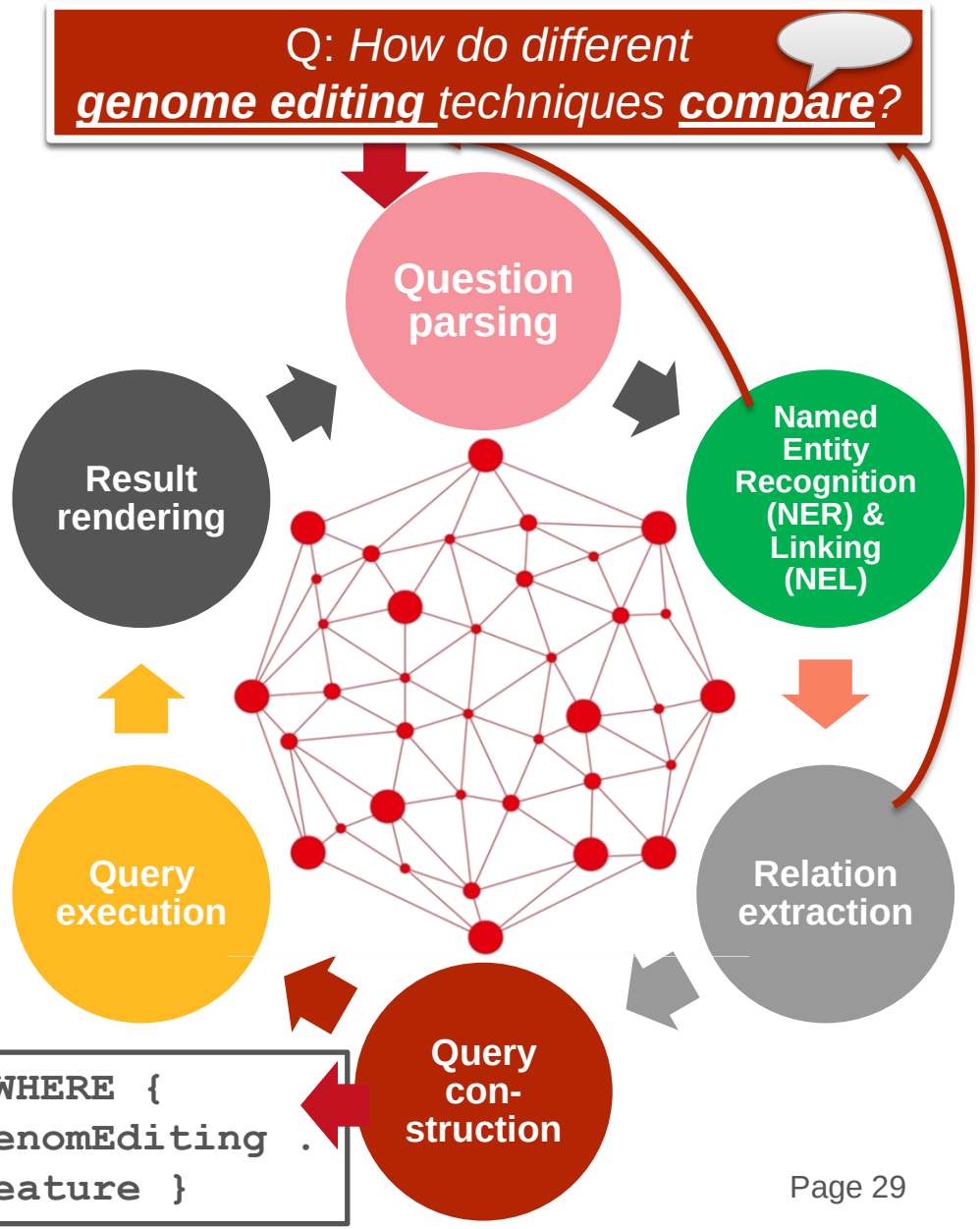
3. Graph representation



WP4 Exploration and Question Answering

Research Challenge:

- Intuitive exploration leveraging the rich semantic representations
- Answer natural language questions
- Juxtaposition of approaches




```
SELECT Approach, Feature WHERE {
  Approach  adresses     GenomEditing .
  Approach  hasFeature   Feature }
```

[1] K. Singh et al: *Why Reinvent the Wheel? Let's Build Question Answering Systems Together.* The Web Conference (WWW 2018).

Result: Automatic Generation of Comparisons/Surveys

Engineered Nucleases	Site-specificity	Safety	Ease-of-use / costs/ speed
zinc finger nucleases (ZFN)	++ 9-18nt	+	-- $$$: screening, testing to define efficiency
transcription activator-like effector nucleases (TALENs)	+++ 9-16nt	++	++ Easy to engineer 1 week / few hundred dollar
engineered meganucleases	+++ 12-40 nt	0	-- $$$ Protein engineering, high-throughput screening
CRISPR system/cas9	++ 5-12 nt	-	+++ Easy to engineer few days / less 200 dollar

Open Science Graph Outlook

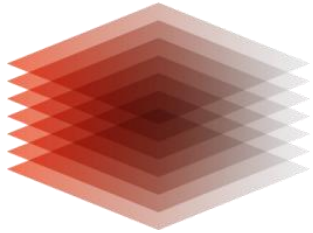
There is a lot to do:

- Equip existing services with Linked Data interfaces
- Enable the deep semantic description of research, requires
  - Good user interfaces
  - Scalable storage and search facility
  - Collaboration between scientists, libariens, knowledge engineers, machines

Stay tuned

- Mailinglist/group: https://groups.google.com/forum/#!forum/orkg
- Comming soon: Open Research Knowledge Graph: https://orkg.org
- Next workshop at TIB on November, 22nd (after DILS Conference: https://events.tib.eu/dils2018/)

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

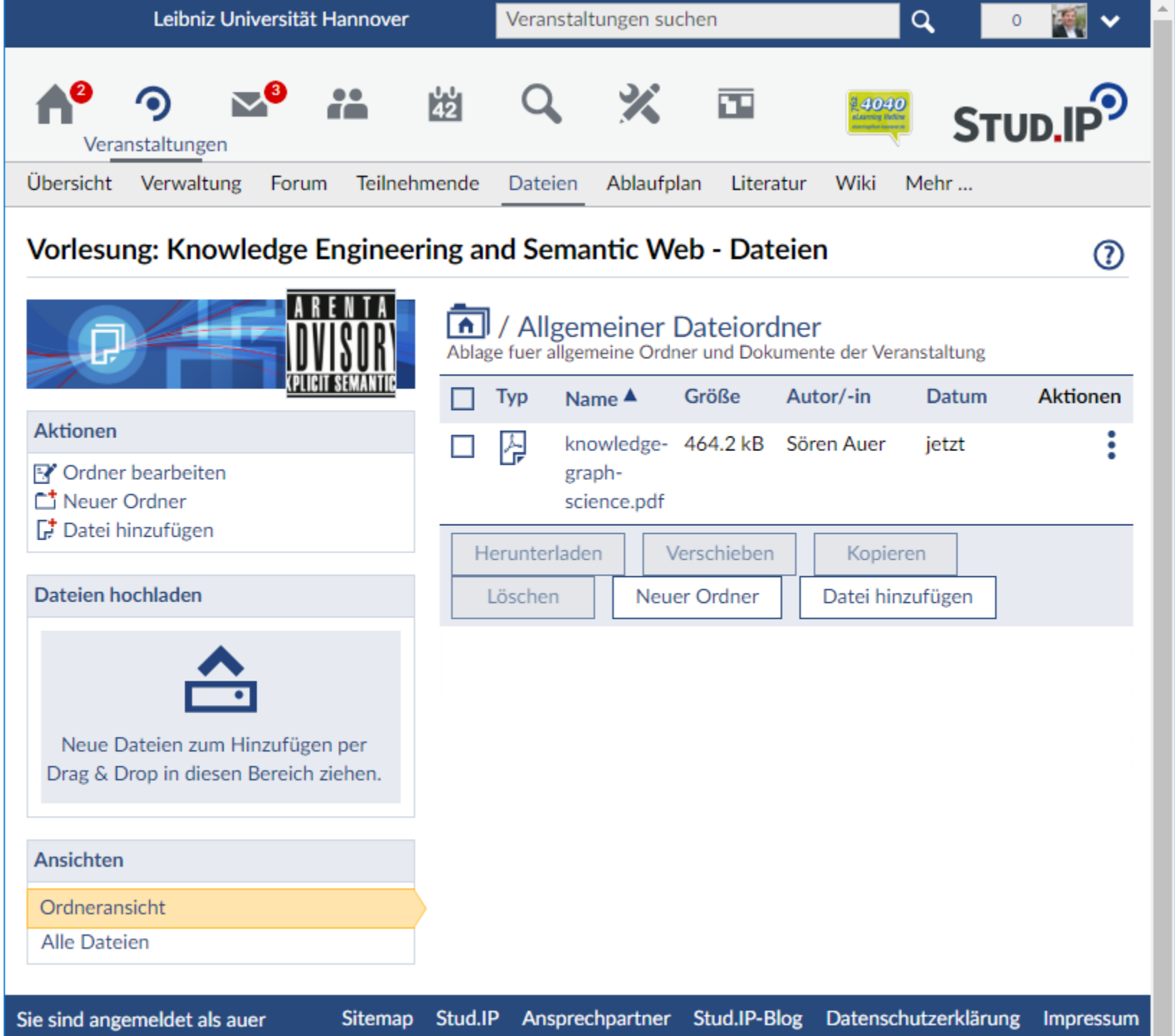


TIB

Open Education

Lehrmaterialien früher

Lehrmaterialien heute



Neue Möglichkeiten in der Digitalen Welt

Machine-readability

Semantic representation

Dynamic content, **interactive** examples

Integration of **multimedia** content

Rich **interlinking with context** (related work, calls, reviews, comments/ discussions)

Integration of **rich metadata** (provenance, licensing)

Interactive **collaboration**

…

Warum Open Education?

- Lehrende erfinden das Rad jeden Tag ***neu***
- Wenig Unterstützung für mehrsprachige Inhalte und deren Übersetzung in verschiedene Sprachen
- Mangel an hochstrukturierten, qualitativ hochwertigen freien Lehr- und Lernmaterialien (***OER***)

OCW Schwächen

Systematic analysis of 100 courses from major OCW repositories revealed weaknesses:

- **Legal Reusability:** Only 28 of 100 courses have an truly open license, majority (57 of 100) are restricting reuse to non-commercial scenarios
- **Multilinguality:** Vast majority of the courses in English, only two have been translated to other languages. 12 of 100 courses were offered in other languages.
- **Format repurposeability:** 68 courses are offered in formats supporting some form of repurposeability. 52 of the 68 courses are only available in PDF.
- **Recency:** Out of 100 courses, only 32 have been updated in the last two years.
- **Self-assessment:** only for 15 courses and separately for another 40 courses. Out of these 55 courses, just 25 provide solutions.
- **Engaging course material:** 65 of the courses have at least one example and one illustration, just 25 have more than 50 examples. Almost two thirds of the courses are objectively of low attractiveness.
- **Community Involvement:** 61 courses have been created by a single author. Only 16 courses are the result of collaborative work.

Vahdati, S.; Lange, C.; Auer, S.: OpenCourseWare observatory: does the quality of OpenCourseWare live up to its promise? 5th Int. Conf. on Learning Analytics And Knowledge, LAK '15, ACM 2015, ISBN 978-1-4503-3417-4.

SlideWiki

Projekt-Webseite*: https://slidewiki.eu/

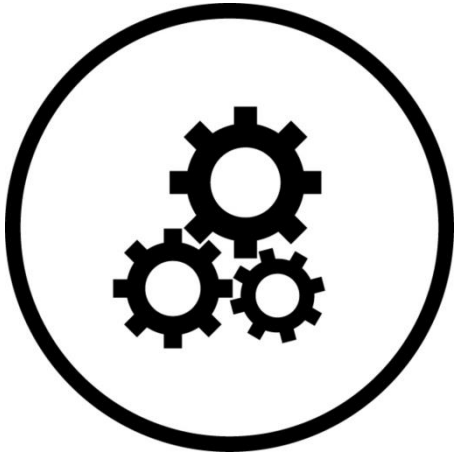
Plattform-Webseite: https://slidewiki.org/

* H2020, GA Nr. 688095.

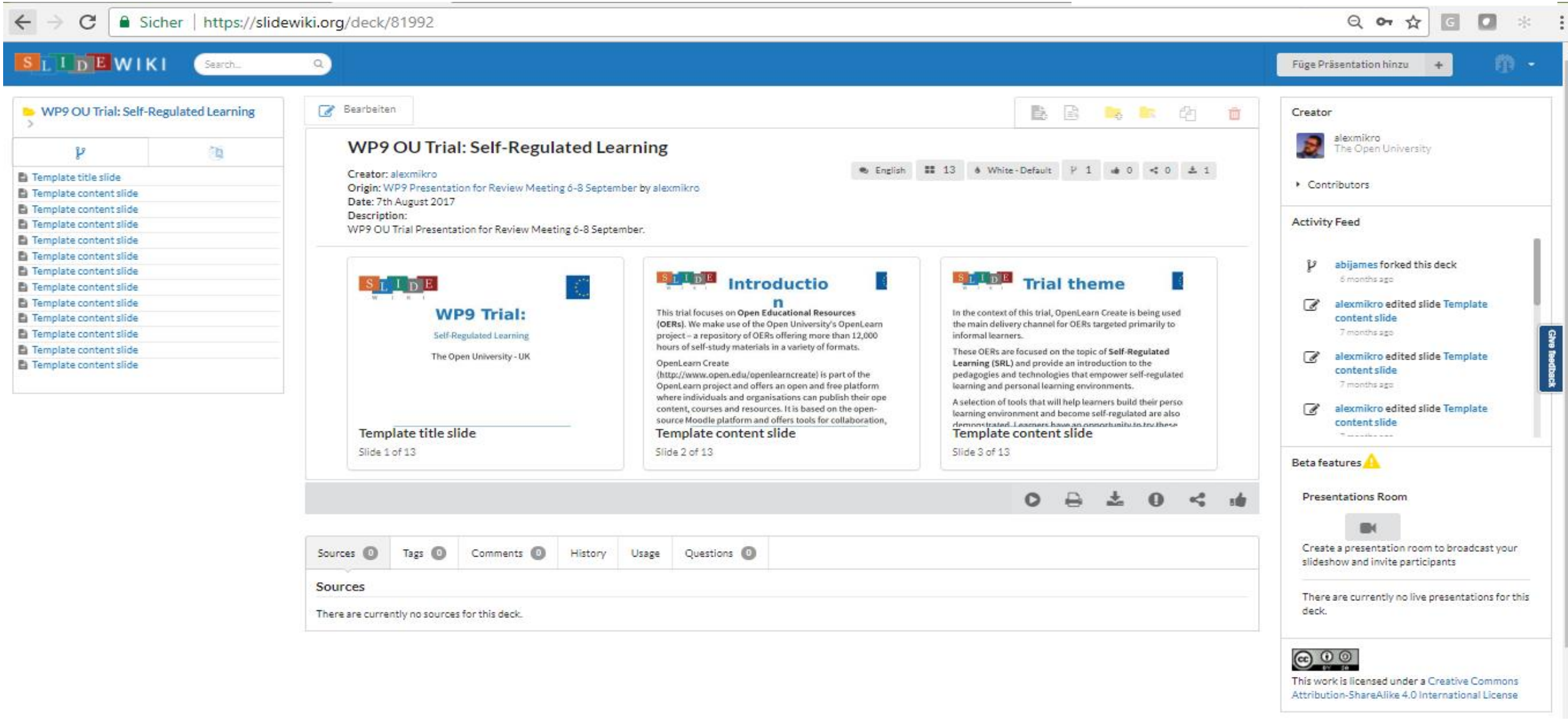
Wissenserwerb und -vermittlung effizient steigern

SlideWiki ermöglicht:

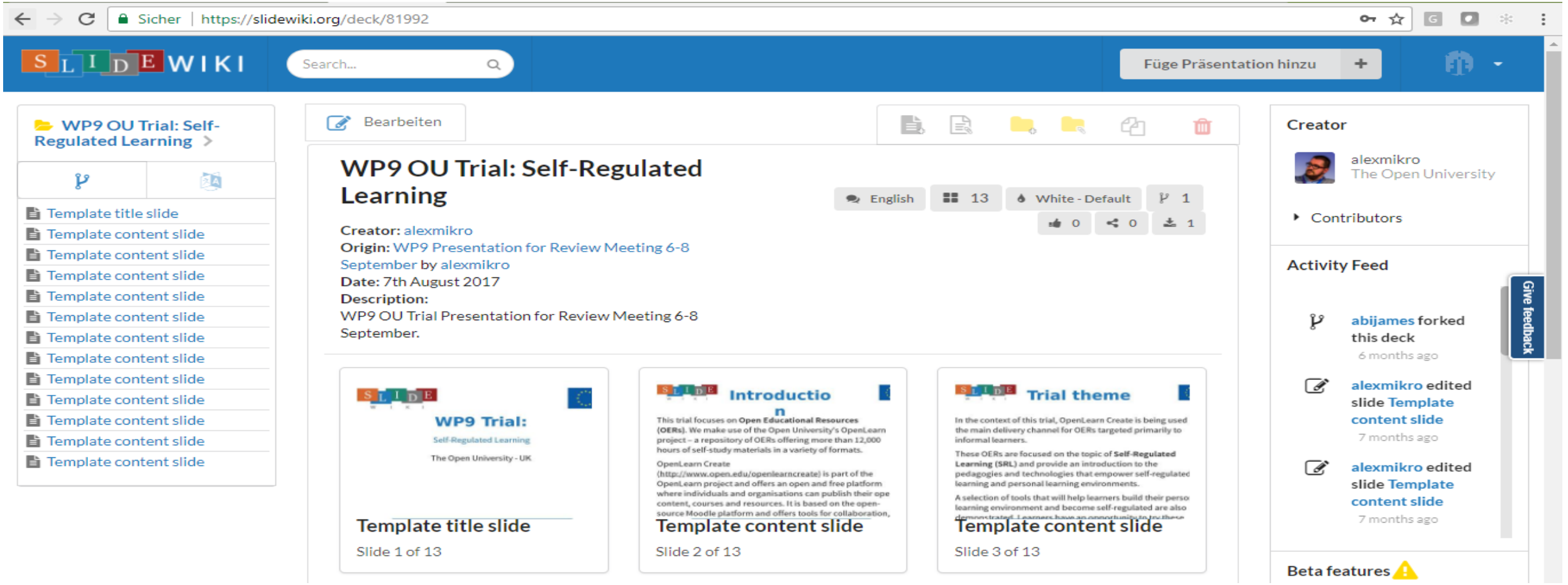
- die gemeinschaftliche Erstellung umfassender OCWs (Curricula, Folienpräsentationen, Selbstbewertungstests, Illustrationen usw.) in der Form von ***Präsentationen.***
- die ***halbautomatische Übersetzung*** auf mehr als 50 verschiedene Sprachen,
- die Steigerung von ***Effizienz, Effektivität und Qualität*** der Ausbildung und der Wissensvermittlung auf ***Crowdsourcing-Basis***.



Im Überblick…



[1/2]


Sicher | https://slidewiki.org/deck/81992
SLIDEWIKI
Search...
Füge Präsentation hinzu
WP9 OU Trial: Self-Regulated Learning
Bearbeiten
Template title slide
Template content slide
WP9 OU Trial: Self-Regulated Learning
English
13
White - Default
1
0
0
1
Creator: alexmikro
Origin: WP9 Presentation for Review Meeting 6-8 September by alexmikro
Date: 7th August 2017
Description:
WP9 OU Trial Presentation for Review Meeting 6-8 September.
WP9 Trial:
Self-Regulated Learning
The Open University - UK
Template title slide
Slide 1 of 13
Introduction
Template content slide
Slide 2 of 13
Trial theme
Template content slide
Slide 3 of 13
Creator
alexmikro
The Open University
Contributors
Activity Feed
abijames forked this deck
6 months ago
alexmikro edited slide Template content slide
7 months ago
alexmikro edited slide Template content slide
7 months ago
Beta features
Give feedback

Semantic Data Web lecture series > RDF Data Model > RDF Overview > RDF Overview

No title
Introduction
IBM 1620 data processing machine, 1962
Who is this?
The Web
The Web penetrates society
The current Web
Limitations of the Web
What Google does not find
What's the problem with the Web
Basic ingredients for the Semantic Web
Data Models, Access & Integration
LOD Cloud
LOD Cloud May 2007
LOD Cloud October 2007
LOD Cloud February 2008
LOD Cloud September 2008
LOD Cloud March 2009
LOD Cloud September 2010
LOD Cloud September 2011
LOD Cloud August 2014
The Web of Data
Map to the Semantic Web
The Semantic Data Web Stack
The Semantic Data Web Stack
... also known as "layer cake"
URIs and Unicode
Resource Description Framework – RDF

Edit

RDF Overview

- RDF = Resource Description Framework
- W3C Recommendation since 1998
  - Version 1.1 since 2014
- RDF is a data model
  - Originally used for metadata for web resources, then generalized
  - Encodes structured information
  - Universal, machine readable exchange format
- Data structured in graphs
  - Vertices, edges

36/812

Sources 0 | Comments 0 | History | Usage | Questions 0

Add source

RDF Overview

Creator

gerb

Contributors

Activity Feed

There are currently no activities for this slide.

Save Cancel

Schriftart Größe Line Hei... Stil Format

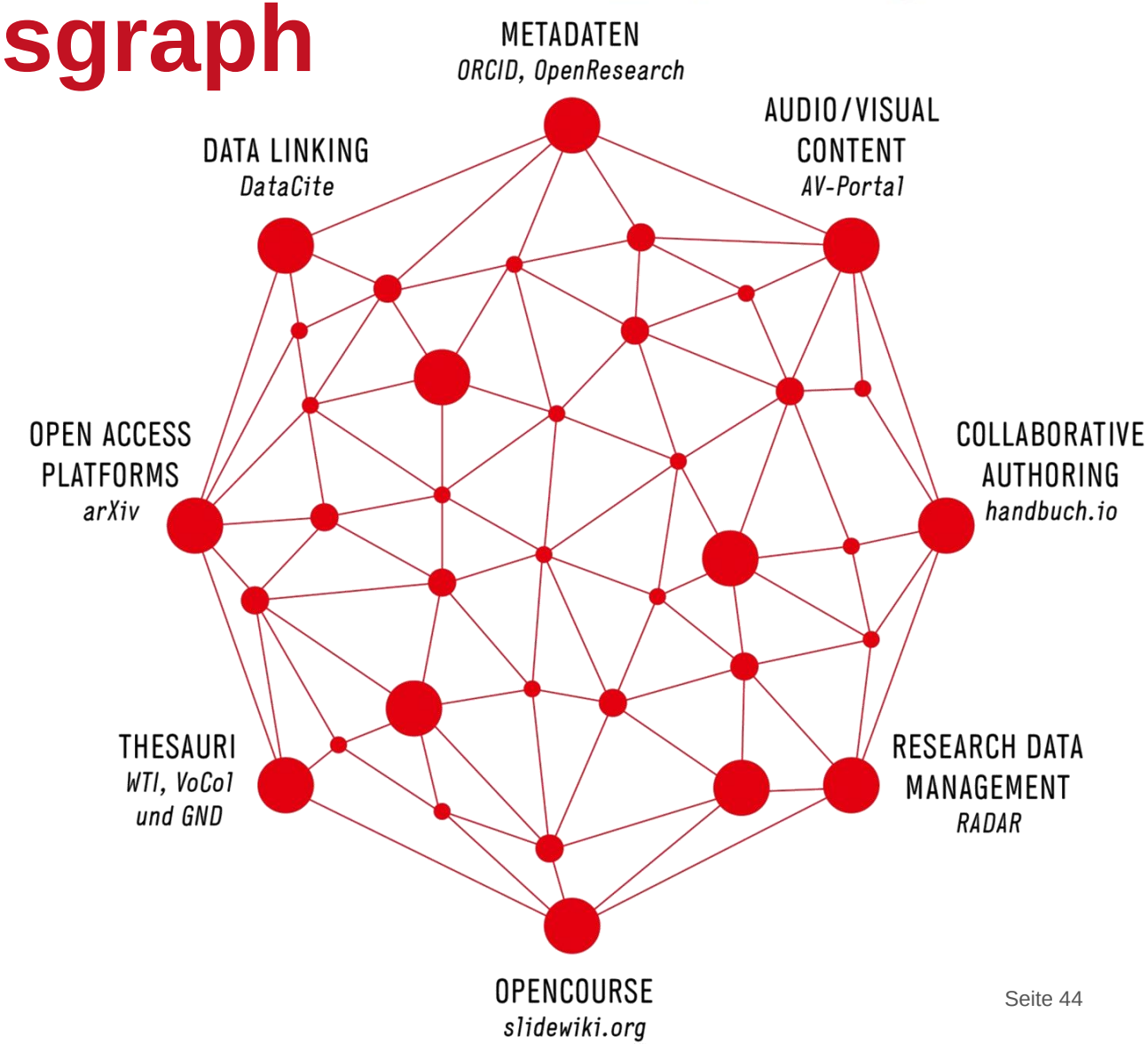
RDF OVERVIEW

- RDF = Resource Description Framework
- W3C Recommendation since 1998
  - Version 1.1 since 2014
- RDF is a data model
  - Originally used for metadata for web resources, then generalized
  - Encodes structured information
  - Universal, machine readable exchange format
- Data structured in graphs
  - Vertices, edges

Speaker notes:

Integration von OER in einen Wissensgraph

- **Semantische** Annotation, Anreicherung & Empfehlung
- Inhalt als verknüpfte Daten (**RDF-Mapping** & **SPARQL-Endpunkte**)
- **Vokabulare**



Zusammenfassung

Data Science erfordert mehr Zusammenarbeit

Open (Data|Knowledge|Source|Education) unterstützt dies

Wir brauchen

- mehr Werkzeuge zur gemeinsamen, kollaborativen Erstellung von Inhalten
- Mehr semantische Beschreibung und Vernetzung von Daten

Weitere interessante Open Education Entwicklungen: Learning Analytics, Formal-Informal Learning Kombination/Zertifizierung, Software Carpentry, …

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

https://de.linkedin.com/in/soerenauer

https://twitter.com/soerenauer

https://www.xing.com/profile/Soeren_Auer

http://www.researchgate.net/profile/Soeren_Auer

Prof. Dr. Sören Auer

TIB & Leibniz University of Hannover

Soeren.Auer@tib.eu

Leibniz Universität Hannover

Leibniz Gemeinschaft