LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

Uni Bonn: Digital Science - Perspektiven & Herausforderungen

Daten- und wissensbasierte Informationsflüsse für Wissenschaft und Technik

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

Leibniz Gemeinschaft

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

Wie haben sich Informationsflüsse durch die Digitalisierung gewandelt?

1 **SANYO-Farbmonitor »CD 3195 C«** für Commodore »C 64«, »C 128«, Atari »800 XL« und Atari »130 XE«, 36-cm-Bildschirm, 40×25 Zeichen für Schrift- und Grafikdarstellungen. PAL-Composite, Grunschalter für Datenverarbeitung. Bitte Kabel mitbestellen.

Bestell-Nummer	726 203	598.–
Kabelf. C64/PC128	727 751	29,90
Kabelf. Atari-XE/XL	726 602	29,90

2 **Grün-Monitor »DM 6112«** mit 31-cm-Bildschirm, Lautsprecher, Auflösung 15 MHz, 25 Zeilen, 40

grammen und Daten auf Compact-Cassetten.

Bestell-Nummer	683 639	99,–

4 **COMMODORE »C 64 C«** – der Home-Computer, für den es ein riesiges Softwareangebot gibt. 64 KB-RAM, 20 KB-ROM, Mikroprozessor 6510, programmierbare Funktionstasten. Anschlüsse für: Joysticks, TV-Gerät oder Monitor, Datenrecorder, Drucker, Diskettenlaufwerk und Akustikkoppler. Grafik: in 16 Farben, hochauflösend mit 320×200 Punkten. Inkl. Netzteil, TV-Kabel, Handbuch und Betriebssystem (GEOS) auf Dis-

stems GEOS. Auch für Eigenprogramme einsetzbar.

Bestell-Nummer	839 247	129.–

6 **Joystick** zum richtigen »lospowern«. Für C 64, C 128, Atari 600/800/130.

Bestell-Nummer	[redacted] 539 577	19,90

7 **Diskettenlaufwerk »VC 1541«**. Hohe Speicherkapazität von 170 KByte je Diskette. Mit deutschem Handbuch, Verbindungskabel und Demonstrations-Disketten.

6 Monate GARANTIE	649 115	448.–

10er-Pack 5¼-Disketten »quali
track«, 2seitig/doppelte Dichte.

10 Stück	839 018	1 Pckg.

10er-Pack 5¼-BASF-Disketten, 100
geprüft, 1seitig/doppelte Dichte.

10 Stück	683 526	1 Pckg. 24,

10er-Pack, 5¼-BASF-Disketten
100% geprüft, 2seitig/doppelte Dichte.

10 Stück	649 294	1 Pckg. 34,

9 **Diskettenkasten »Mini 50«** für 5¼-Disketten. Mit Klebeetikette und Faserstift.

Bestell-Nummer	649 996

10 **Joystick »Turbo 3«** mit 3 auswechselbaren Griffen für viele Spiele und einige Anwenderprogramme. Für Commodore, Atari und Schneider geeignet.

Bestell-Nummer	649 307

11 **»Deactivators«** helfen Ihnen bei der Suche nach Bomben, die das Institut für Schwerkraft in die Luft jagen sollen. Action unter Zeitdruck.

Diskette für C64/C128	649 555	39,
Diskette für CPC6128	649 614	49,

12 **»Die Erbschaft«** – Panik in Las Vegas. Sie haben ein riesiges Vermögen geerbt, allerdings müssen Sie eine Bedingung erfüllen: 1 Mio Dollar in nur einer Nacht verdienen. Text-Adventure mit schöner Grafik.

Diskette für C64/C128	649 895	59,
Diskette für CPC6128	649 931	59,

13 **»Dragon's Lair«**. Es war einmal: Ein Drache raubte die schöne Königstochter – und wenn sie nicht gestorben sind, können Sie sie noch heute suchen.

Diskette für C64	649 828	49,
Diskette für CPC6128	649 884	49,

14 **»Werner – mach hin«**. Die Comic-Figur aus dem Norden rast mit seinem Bike über den Bildschirm, um sein Lieblingsspiel »Meiern« zu spielen. Supergrafik zum Schmunzeln.

Diskette für C64/C128	649 506	39,
Diskette für CPC6128	649 512	49,
Diskette für ATARI-ST	649 539	59,

15 **»Floppybuch – VC 1541«** mit Tipps und Tricks zum Gebrauch der Floppy.

Bestell-Nummer	839 420

16 **»C 64 Geos-Buch«**. Hinweise für das neue Betriebssystem.

Bestell-Nummer	726 221

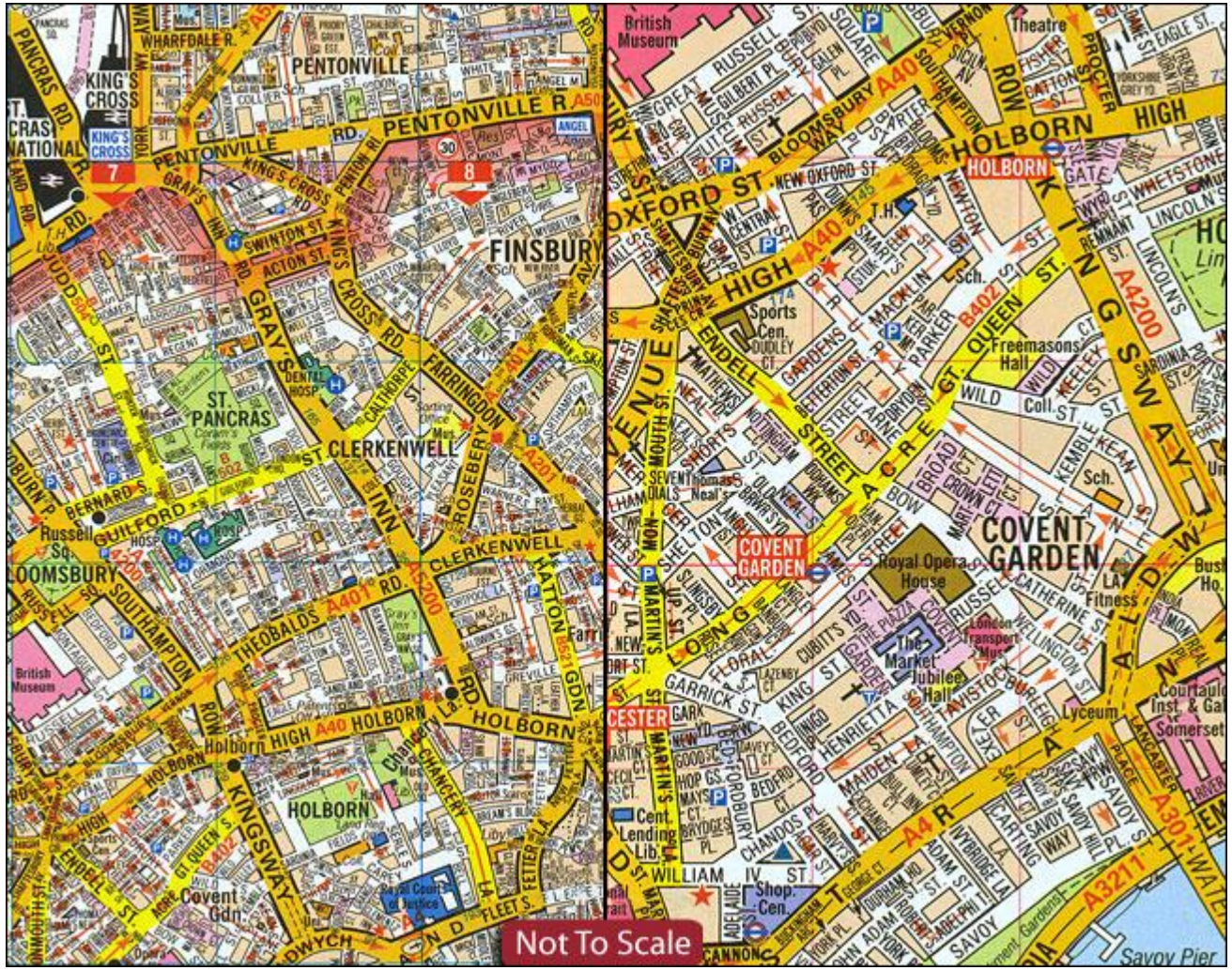
17 **»Tips + Tricks Bd. 1«** Für den Einsteiger und leicht fortgeschrittenen C 64-Benutzer unentbehrlich.

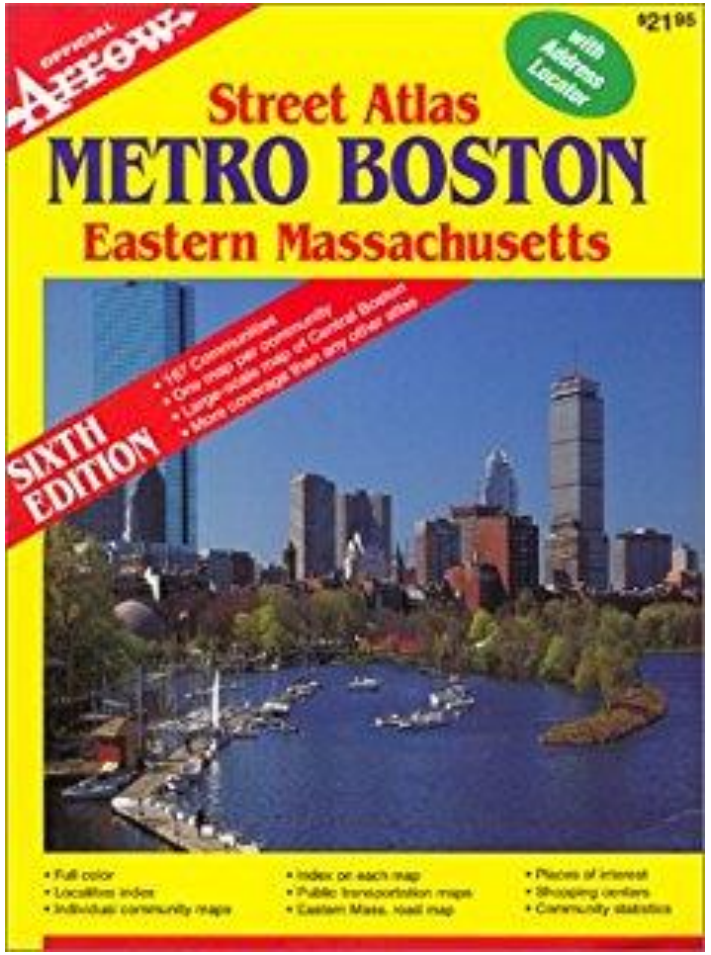
Bestell-Nummer	839 416

18 **Seikosha »SP 180«** – Matrixdrucker mit internationalem und deutschem Zeichensatz. Geschwindigkeit: 80 Z./Sek. bzw. 20 Z./Sek. bei NLQ. Druckmatrix 12×9 bzw. 24×18 (NLQ) Punkte. Vollgrafikfähig. Mit Traktor. Auch Einzelblattverarbeitung möglich. Direktanschluß für Commodore C 64/C 128. Schriftarten: Pica, Elite, Fettdruck, Hoch-Tief, automatisches Unterstreichen. DIP-Schalter vorhanden. Kabel für ATARI eingleich mitbestellen!

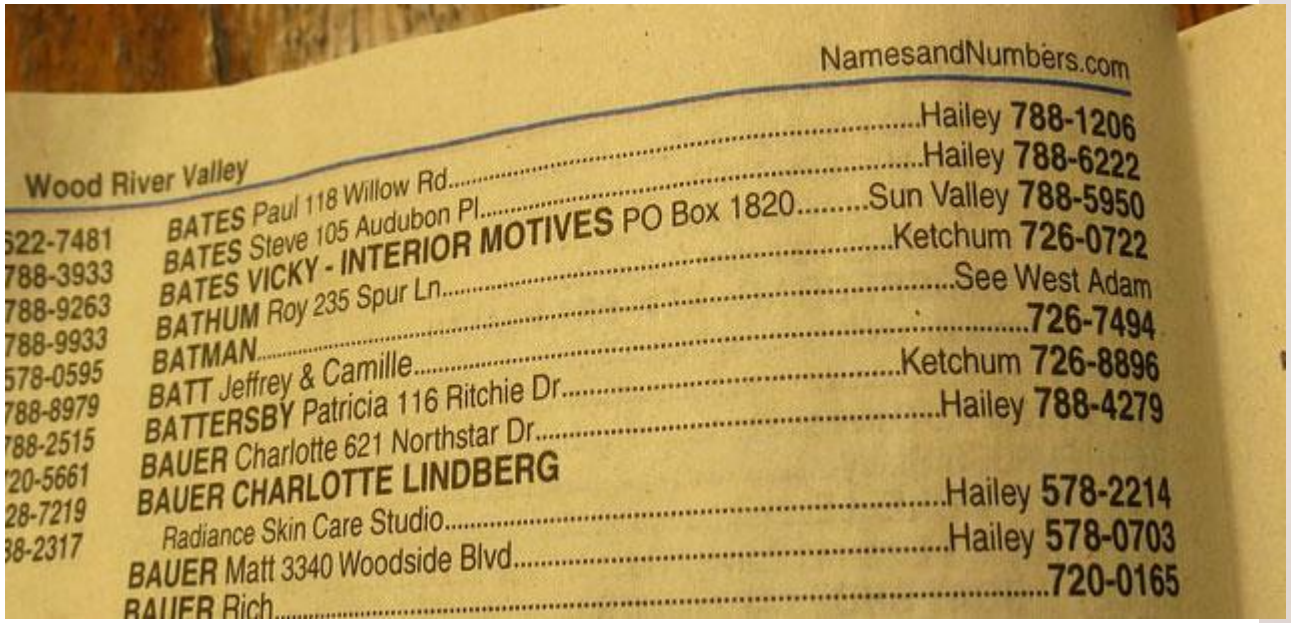
Für C64	839 241	598.–
Für ATARI-ST	649 192	598.–
Kabel für ATARI-ST	839 160	49,95
Farbband	649 839	24,90
2000 Blatt Papier	726 611	49,–

Road Maps


PENTONVILLE
KING'S CROSS
FINSBURY
ST PANCRAS
CLERKENWELL
HOLBORN
COVENT GARDEN
Royal Opera House
Freemasons Hall
Sports Cen.
British Museum
Not To Scale


Official Arrow
$21.95
with Address Locator
Street Atlas
METRO BOSTON
Eastern Massachusetts
SIXTH EDITION

Phone Books


NamesandNumbers.com
Wood River Valley
522-7481
788-3933
788-9263
788-9933
578-0595
788-8979
788-2515
720-5661
28-7219
38-2317
BATES Paul 118 Willow Rd.....Hailey 788-1206
BATES Steve 105 Audubon Pl.....Hailey 788-6222
BATES VICKY - INTERIOR MOTIVES PO Box 1820.....Sun Valley 788-5950
BATHUM Roy 235 Spur Ln.....Ketchum 726-0722
BATMAN.....See West Adam
BATT Jeffrey & Camille.....726-7494
BATTERSBY Patricia 116 Ritchie Dr.....Ketchum 726-8896
BAUER Charlotte 621 Northstar Dr.....Hailey 788-4279
BAUER CHARLOTTE LINDBERG
Radiance Skin Care Studio.....Hailey 578-2214
BAUER Matt 3340 Woodside Blvd.....Hailey 578-0703
BAUER Rich.....720-0165


verizon
PORTLAND & VANCOUVER METROPOLITAN AREA
DEX
503-PLUMBER
ACCIDENTS - MORGAN & ASSOCIATES
FREE
(888) INJ-ONLY

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

How does it work today?

Leibniz
Association

1-24 of over 3,000 results for **Cell Phones & Accessories** : **Cell Phones** : "smartphone"

Sort by Relevance

FREE Shipping

All customers get FREE Shipping on orders over $25 shipped by Amazon

Show results for

- Any Category
- Cell Phones & Accessories
  - **Cell Phones**
    - Unlocked Cell Phones
    - Carrier Cell Phones

Refine by

Amazon Prime
- prime

Eligible for Free Shipping
- Free Shipping by Amazon

Cell Phone Display Size
- 3.9 Inches & Under
- 4.0 to 4.4 Inches
- 4.5 to 4.9 Inches
- 5.0 to 5.4 Inches
- 5.5 Inches & Over

Avg. Customer Review
- & Up

Showing results in **Cell Phones & Accessories**. Show instead results in All Departments.

Sponsored
Nokia 6.1 (2018) - 32 GB - Unlocked Smartphone (AT&T/T-Mobile) - 5.5" Screen - Black
by Nokia mobile
$269.00 prime
Get it by **Monday, Jun 11**
FREE Shipping on eligible orders
60
- Operating System: **Android 8.0 (Oreo)**
- Display Size: **5.5 inches**
- Display Type: **LCD**

Sponsored
Samsung Galaxy S9 Unlocked Smartphone - Coral Blue - US Warranty
by Samsung
$719.90 prime
298

Smartphones (4.426)*

Sortieren nach: Beliebteste zuerst

Preis

- bis 95 € 895
- 95 € bis 150 € 1.214
- 150 € bis 200 € 797
- 200 € bis 300 € 722
- ab 300 € 798

Hersteller

Hersteller suchen...

- Acer 48
- Alcatel 164
- Allview 11
- Amazon Fire 2
- Amplicomms 2

Displaygröße

Displaygröße suchen...

- ab 6 Zoll 227
- 5,5 Zoll 1.134
- 5 Zoll 1.744
- 4,7 Zoll 242
- 4,5 Zoll 287

verfügbares Betriebssystem

Apple iPhone 8
LTE Smartphone, 4,7 Zoll, HD, 2 GB RAM, iOS 11, 12 Megapixel, 1821 mAh
Note Ø 1,9 ★★★★☆ 3
332 Angebote
623,00 – 1.319,95 €

Samsung Galaxy S7
LTE Smartphone, 5,1 Zoll, Quad HD, 4 GB RAM, 32 GB interner Speicher, Android 8.0 Oreo, 12 Megapixel, 3000 mAh, Speicher erweiterbar
Note Ø 1,4 ★★★★½ 11
259 Angebote
309,00 – 699,00 €

Apple iPhone X
LTE Smartphone, 5,8 Zoll, Full HD, 3 GB RAM, iOS 11, 12 Megapixel, 2700 mAh
Note Ø 1,5 ★★★★½ 2
138 Angebote
928,00 – 1.414,91 €

Samsung Galaxy S9
LTE Smartphone, 5,8 Zoll, Quad HD, 4 GB RAM, Android 8.0 Oreo, 12 Megapixel, 3000 mAh, Speicher erweiterbar
Note Ø 1,7
204 Angebote
548,00 – 1.199,99 €

Apple iPhone SE
LTE Smartphone, 4 Zoll, 2 GB RAM, iOS 11, 12 Megapixel, 1642 mAh
Note Ø 1,7 ★★★★☆ 2
365 Angebote
213,89 – 589,00 €

Samsung Galaxy S9 64GB midnight black
LTE Smartphone, 5,8 Zoll, Quad HD, 4 GB RAM, 64 GB interner Speicher, Android 8.0 Oreo, 12 Mega ...mehr
59 Angebote
548,00 – 1.014,44 €

Apple iPhone 7
LTE Smartphone, 4,7 Zoll, HD, 2 GB RAM, iOS 11, 12 Megapixel, 1960 mAh
Note Ø 1,8 ★★★★☆ 7
604 Angebote
450,00 – 1.090,55 €

Samsung Galaxy S8
LTE Smartphone, 5,8 Zoll, Quad HD, 4 GB RAM, 64 GB interner Speicher, Android 8.0 Oreo, 12 Megapixel, 3000 mAh, Speicher erweiterbar
Note Ø 1,4 ★★★★☆ 28
195 Angebote
459,90 – 872,03 €

Feedback

atlanta

Atlanta

Georgia
USA

Routenplaner

Überwiegend bewölkt · 26 °C
10:54

SPEICHERN | IN DER NÄHE | AN MEIN SMARTPHONE SENDEN | TEILEN

Kurzinfo

Atlanta ist die Hauptstadt des US-Bundesstaates Georgia. Die Stadt spielte sowohl während des Amerikanischen Bürgerkriegs als auch während der Bürgerrechtsbewegung in den 1960er-Jahren eine wichtige Rolle. Das Atlanta History Center erzählt von der Geschichte der Stadt, während die Martin Luther King Jr. National Historic Site dem Leben und Wirken des afroamerikanischen Bürgerrechtlers gewidmet ist. Im Centennial Olympic Park, der für die Olympischen Spiele 1996 in Downtown gebaut wurde, befindet sich das

The World of Publishing & Communication has profundely changed

New means adapted to the new posibilities were developed, e.g. „zooming“, dynamics

Business models changed completely

More focus on data, interlinking of **data/services and search** in the data

Integration, **crowdsourcing** play an important role

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

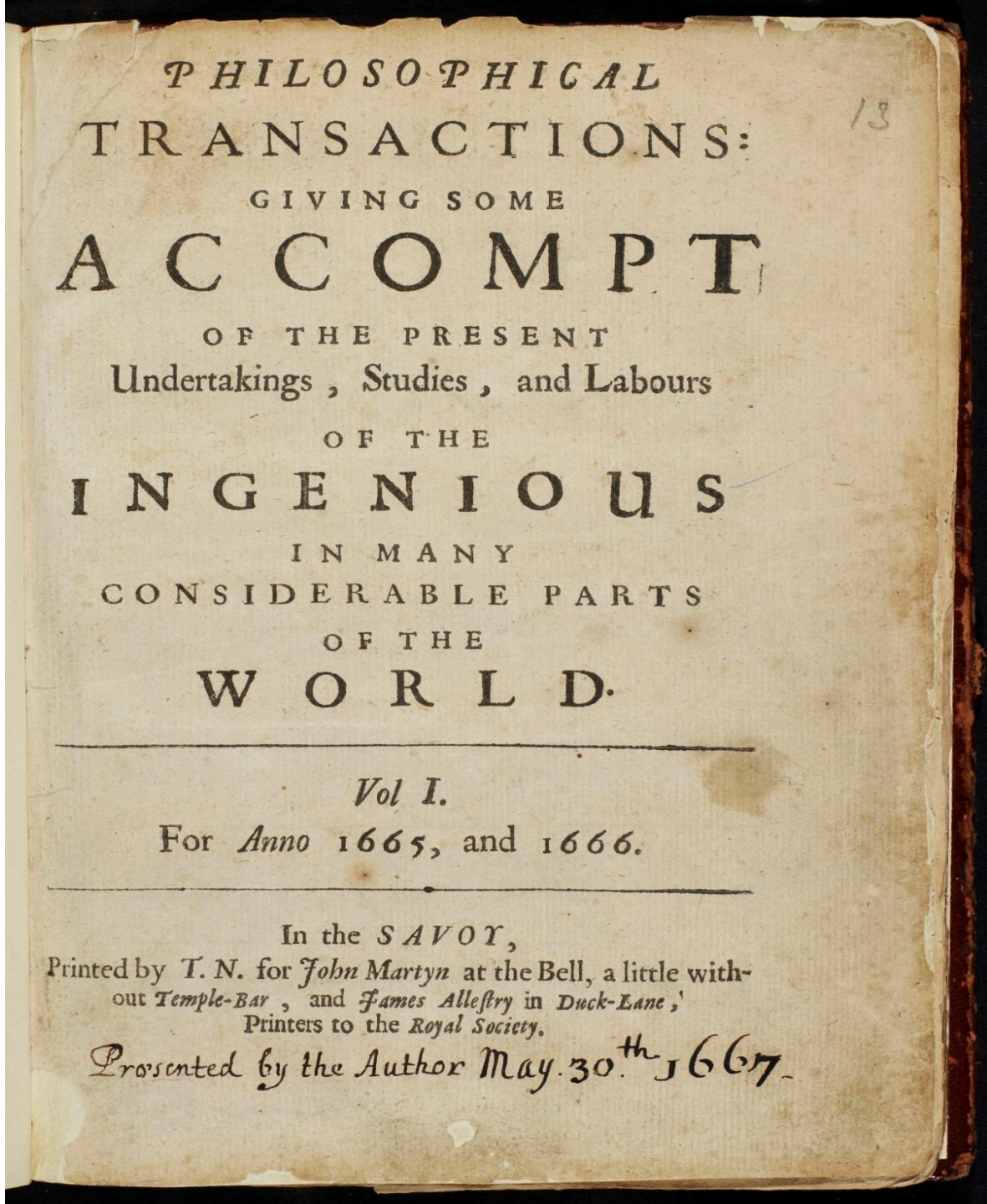
TIB

What about Scholarly Communication?

Leibniz Association

Scientific publishing in the 17th century

One of the earliest research journals: *Philosophical Transactions of the Royal Society*



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.

Scholarly communication in 1865

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

Publishing in 1970s

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

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 prop-

Scientific publishing today

We have:

BUT

- Mainly based on PDF
- Is only partially machine-readable
- Does not preserve structure
- Does not allow embedding of semantics
- Does not facilitate interactivity/dynamicity/ repurposing
- …

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.

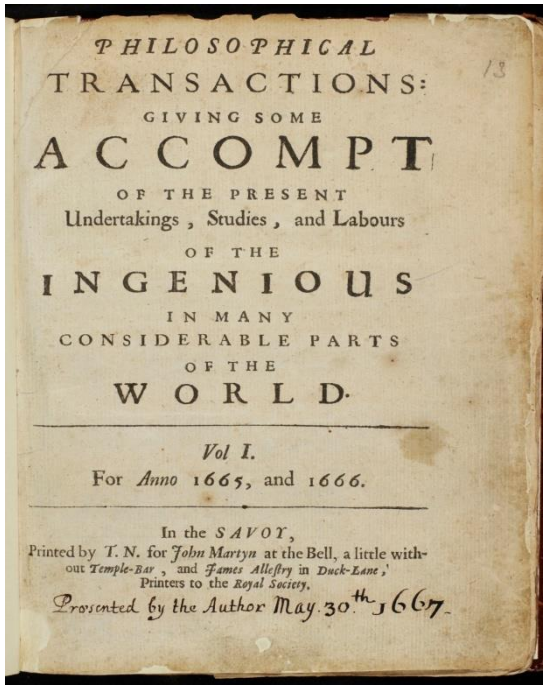
1 Introduction

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.

Scholarly Communication has not changed (much)

17th century

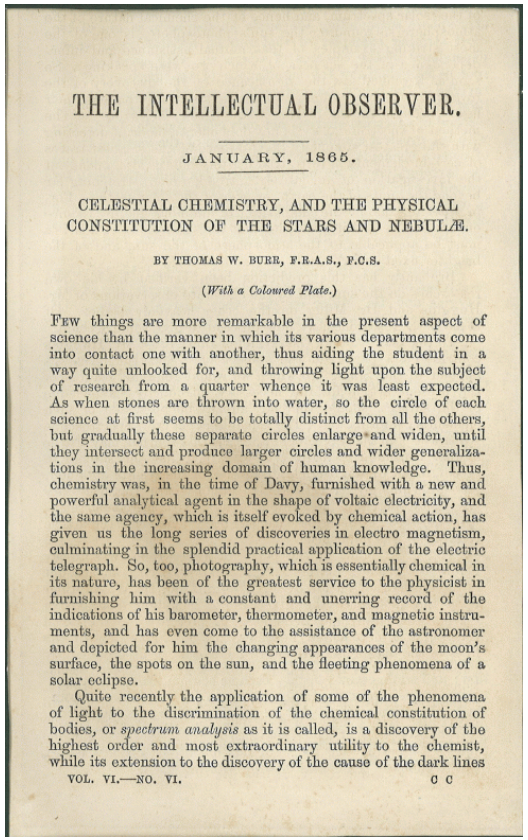


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

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.

Proliferation of scientific literature

Science and engineering articles by region, country: 2004 and 2014

Rank	Region, country, or economy	2004	2014	Average annual growth rate (%)	2014 world total (%)	2014 cumulative world total (%)
na	World	1,272,362	2,290,294	6.1	100.0	na
1	United States	336,194	431,623	2.5	18.8	18.8
2	China	110,388	395,588	13.6	17.3	36.1
3	Germany	72,177	107,747	4.1	4.7	40.8
4	India	28,752	106,574	14.0	4.7	45.5
5	Japan	95,999	103,793	0.8	4.5	50.0
6	United Kingdom	75,119	101,536	3.1	4.4	54.4
7	France	53,375	74,269	3.4	3.2	57.7
8	Italy	42,647	70,453	5.1	3.1	60.8
9	South Korea	27,029	63,748	9.0	2.8	63.5
10	Canada	40,624	60,916	4.1	2.7	66.2
11	Spain	30,977	56,604	6.2	2.5	68.7
12	Brazil	18,814	53,152	10.9	2.3	71.0
13	Australia	26,277	52,269	7.1	2.3	73.3
14	Russia	26,869	43,487	4.9	1.9	75.2
15	Iran	4,952	36,539	22.1	1.6	76.8

National Science Foundation: Science and Engineering Publication Output Trends: https://www.nsf.gov/statistics/2018/nsf18300/nsf18300.pdf

Reproducibility Crisis

1,500 scientists lift the lid on reproducibility

Monya Baker in *Nature,* 2016. **533** (7604): 452–454. doi:10.1038/533452a:

- 70% failed to reproduce at least one other scientist's experiment
- 50% failed to reproduce one of their own experiments

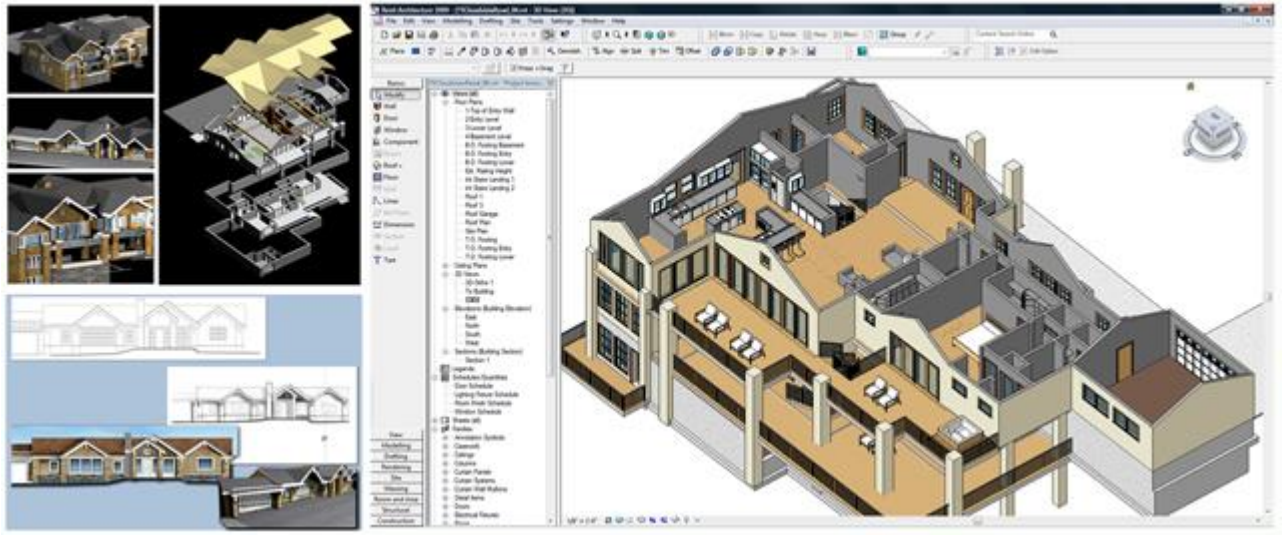
Failure to reproduce results among disciplines (in brackets own results):

- chemistry: 87% (64%),
- biology: 77% (60%),
- physics and engineering: 69% (51%),
- Earth sciences: 64% (41%).

Duplication and Inefficiency

How can we avoid duplication if the terminology, research problems, approaches, methods, characteristics, evaluations, … are not properly defined and identified?

How would you build an engine/building without properly defining their parts, relationships, materials, characteristics … ?



Root Cause - Deficiency of Scholarly Communication?

Lack of:

- **Transparency** – information is hidden in text
- **Integratability** – fitting different research results together
- **Machine assistance** – unstructured content is hard to process
- **Identifyability** of concepts beyond metadata
- **Collaboration** – one brain barrier
- **Overview** – scientists look for the needle in the haystack

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

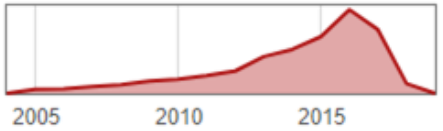
Seite 1 von 4.373 Ergebnissen

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**

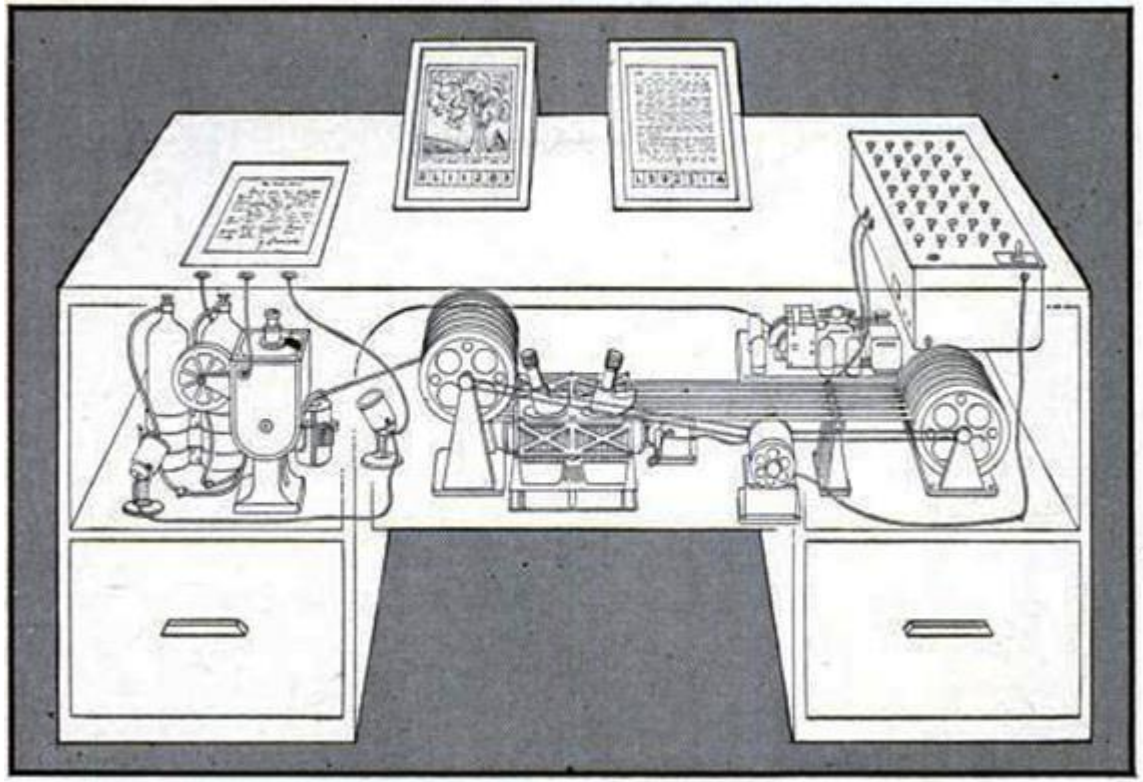
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

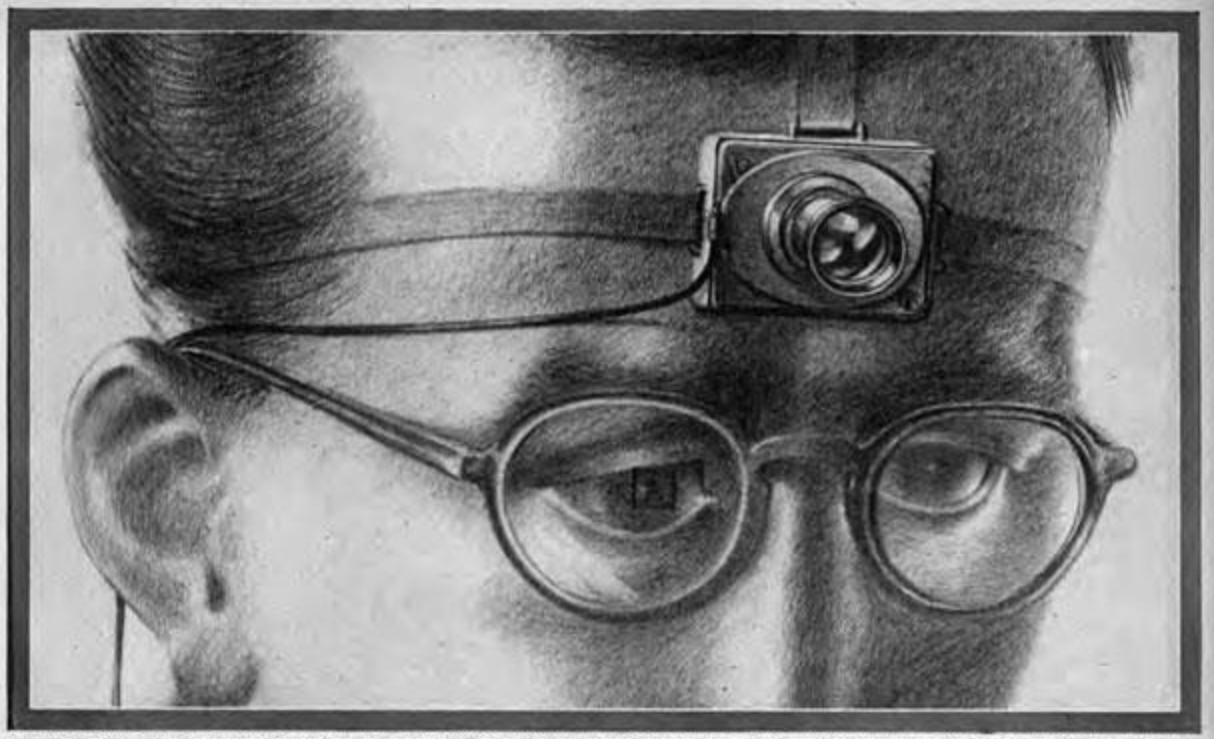
TIB

How can we fix it?

Leibniz Gemeinschaft

Realizing Vannevar Bush‘s vision of Memex





A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJECT

AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK

by VANNEVAR BUSH

DIRECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT

Condensed from the *Atlantic Monthly*, July 1945

This has not been a scientists' war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indecision, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who have had to devise new methods for their unanticipated assignments. They have done their part on the devices that made it possible to turn back the enemy. They have worked in combined effort with the physicists of our allies. They have felt within themselves the stir of achievement. They have been part of a great team. Now one asks where they will find objectives

ress, and the effort to bridge between disciplines is correspondingly superficial.

Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. If the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time might well be startling. Those who conscientiously attempt to keep abreast of current thought, even in restricted fields, by close and continuous reading might well shy away from an examination calculated to show how much of the previous month's efforts could be produced on call.

Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it. This sort of catastrophe is undoubtedly being repeated all about us as truly significant attainments become lost in the mass of the inconsequential.

Publication has been extended far beyond our present ability to make real

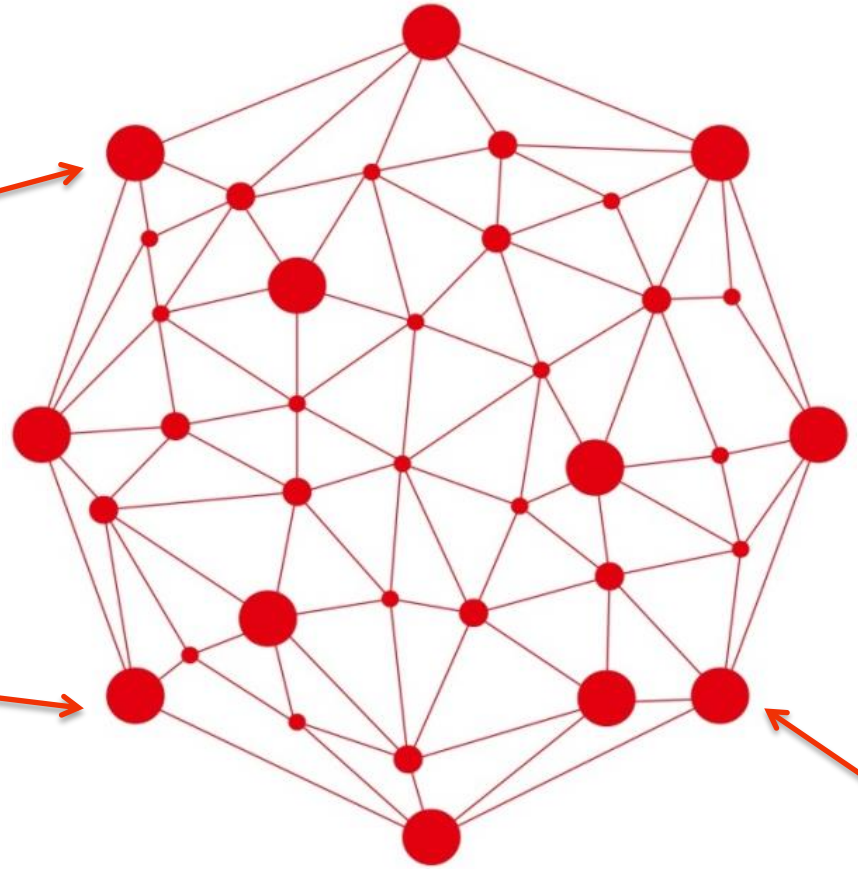
Concepts

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 • …

Linked Data Principles

1. Use **URIs** to **identify** the “things” in your data
2. Use **http:// URIs** so people (and machines) can **look** them **up** on the web
3. When a URI is looked up, **return** a **description** of the thing **in** the W3C **Resource Description Format** (RDF)
4. Include **links to related things**

http://www.w3.org/DesignIssues/LinkedData.html

[1] Auer, Lehmann, Ngomo, Zaveri: **Introduction to Linked Data and Its Lifecycle on the Web.** Reasoning Web 2013

RDF & Linked Data in a Nutshell

1. Graph based RDF data model consisting of S-P-O statements (facts)

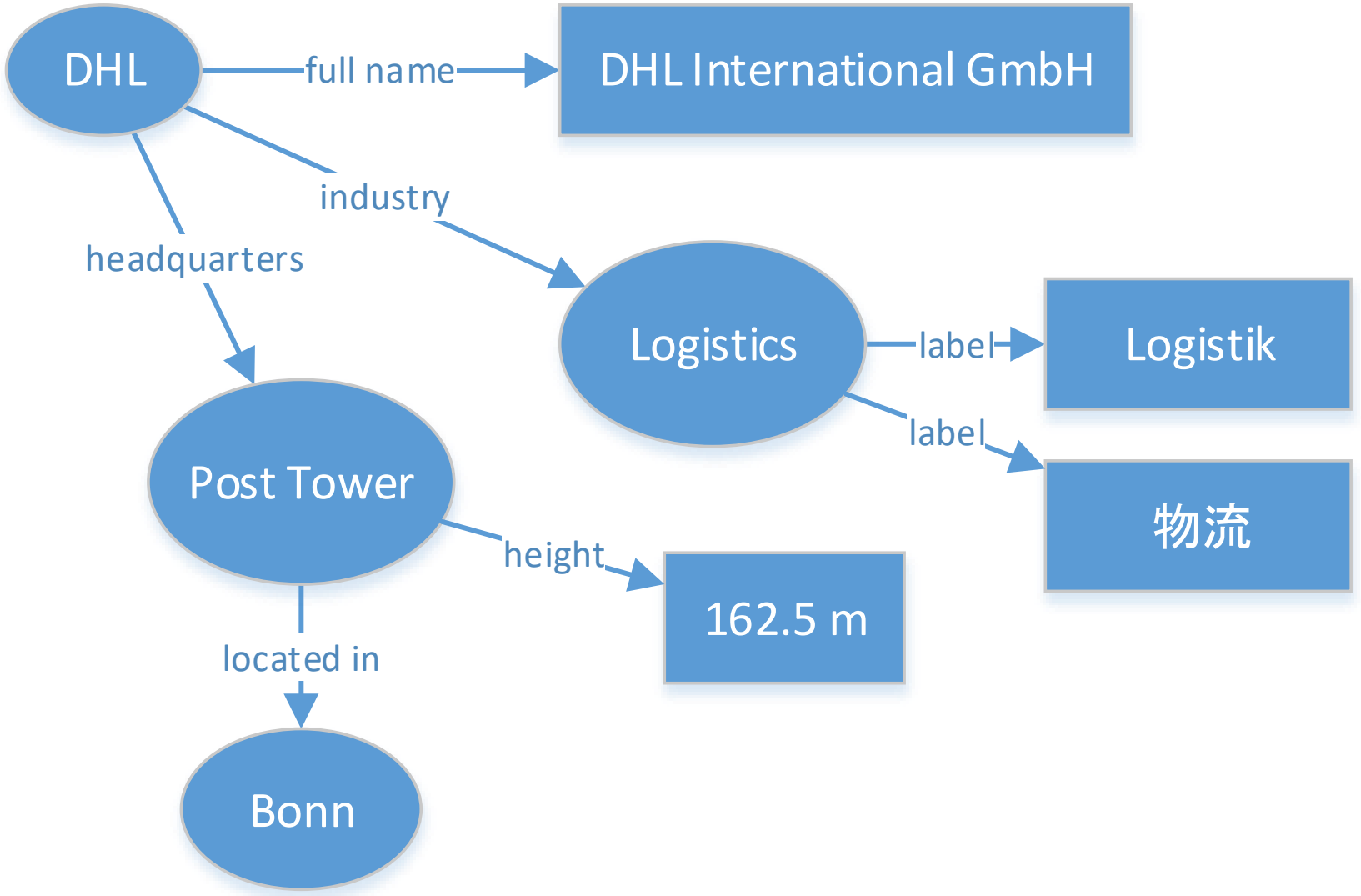
2. Serialised as RDF Triples:

```
ULB                 conf:organizes      DigitalScience .
DigitalScience      conf:starts         "2018-12-17"^^xsd:date .
DigitalScience      conf:takesPlaceAt   dbpedia:Bonn .
```

3. Publication under URL in Web, Intranet, Extranet

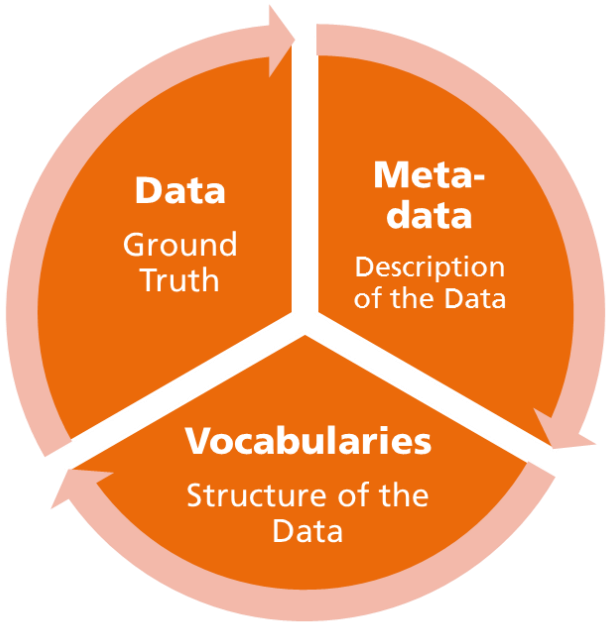
Linked Data

Creating Knowledge Graphs with RDF



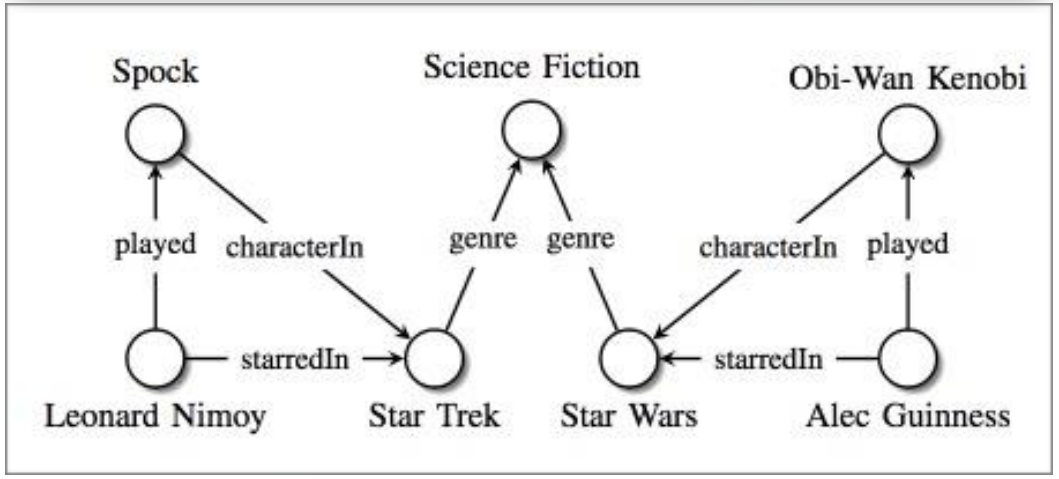
Knowledge Graphs – A definition

- Fabric of concept, class, property, relationships, entity descriptions
- Uses a knowledge representation formalism (typically RDF, RDF-Schema, OWL)
- Holistic knowledge (multi-domain, source, granularity):
  - **instance data** (ground truth),
    - open (e.g. DBpedia, WikiData), private (e.g. supply chain data), closed data (product models),
  - derived, aggregated data,
  - **schema data** (vocabularies, ontologies)
  - **meta-data** (e.g. provenance, versioning, documentation licensing)
  - comprehensive **taxonomies** to categorize entities
  - **links** between internal and external data
  - **mappings** to data stored in other systems and databases



Smart Data for Machine Learning

GND	
Link zu diesem Datensatz	http://d-nb.info/gnd/1021356255
Typ	Person (piz)
Person	Appelbaum, Anne
Geschlecht	weiblich
Zeit	Lebensdaten: 1908-1998
Land	Deutschland (XA-DE); USA (XD-US)
Geografischer Bezug	Wirkungsort: New York, NY
Beruf(e)	Psychologin
Beziehungen zu Personen	Cassirer, Ernst (Vater)
Beteiligt an	1 Publikation 1. *Nachgelassene Manuskripte und Texte / Bd. 1. Zur Metaphysik der symbolischen Formen* *1995*



Emerging Knowledge Graphs

Search Engine Optimization & Web-Commerce

- Schema.org used by >20% of Web sites
- Major search engines exploit semantic descriptions

Pharma, Lifesciences

- Mature, comprehensive vocabularies and ontologies
- Billions of disease, drug, clinical trial descriptions

Digital Libraries

- Many established vocabularies (DublinCore, FRBR, EDM)
- Millions of aggregated from thousands of memory institutions in Europeana, German Digital Library

Chemistry Example: CRISPR Genome Editing

New Results

A practical guide to CRISPR/Cas9 genome editing in Lepidoptera

Linlin Zhang, Robert Reed

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

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

Posted June 22, 2017.

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.

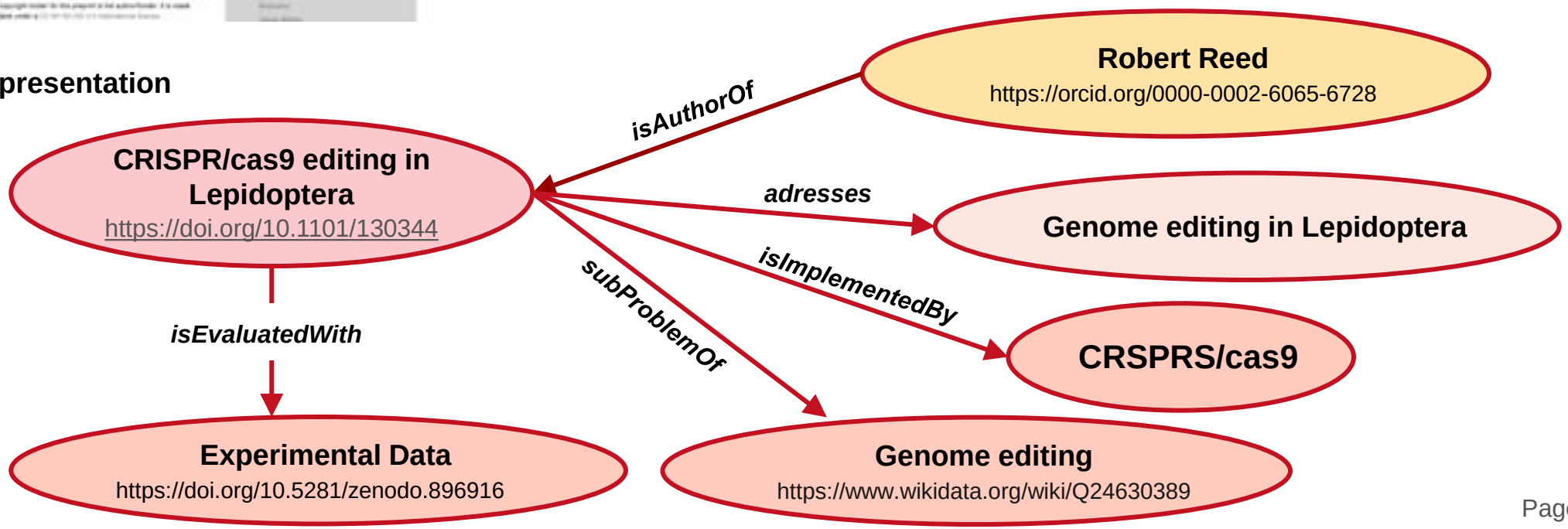
Chemistry Example: Populating the Graph

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

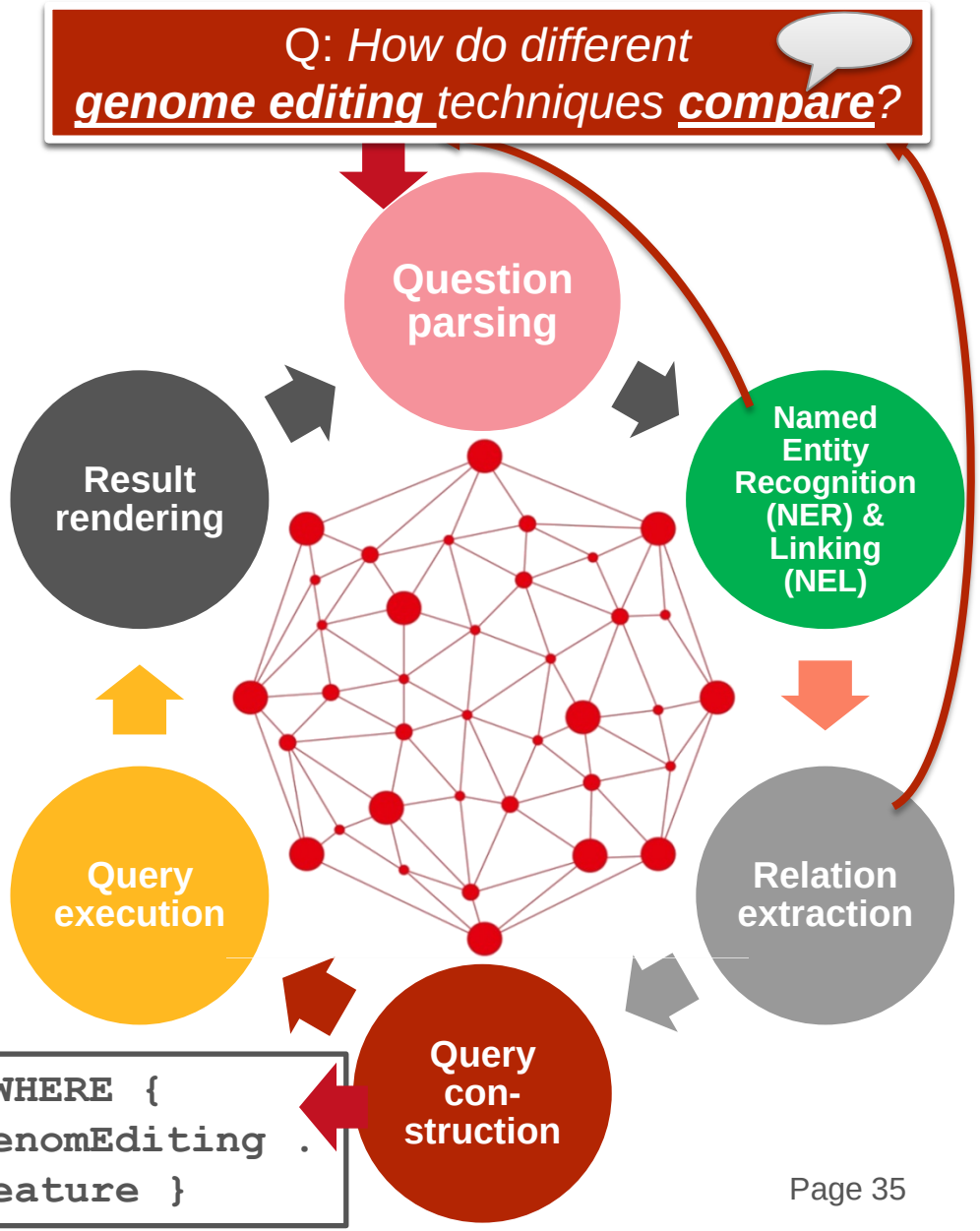
3. Graph representation



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

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

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, ...

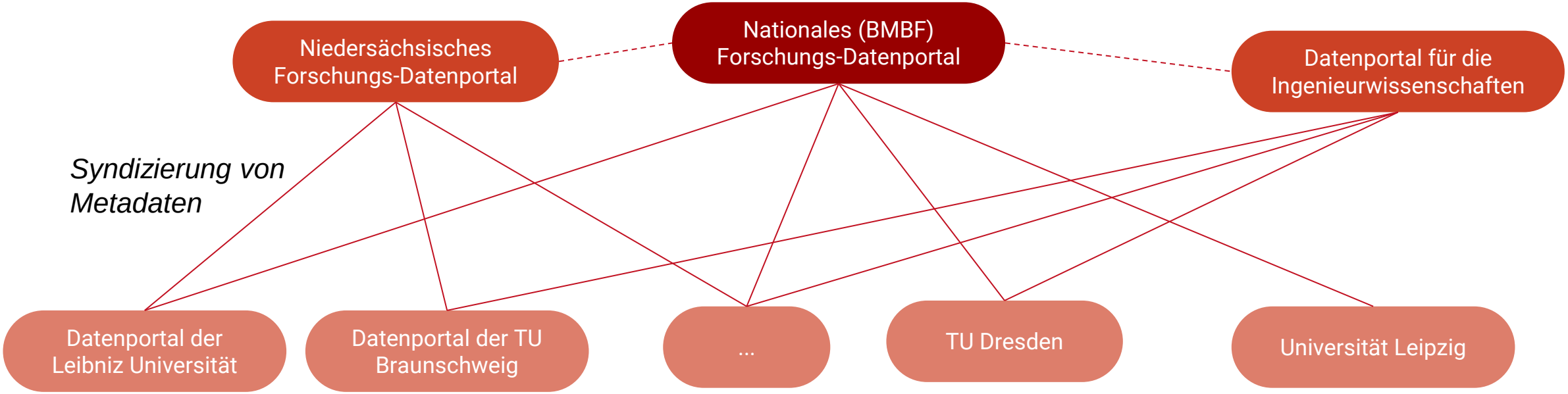
Netzwerk von Forschungsdateninfrastrukturen

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) 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

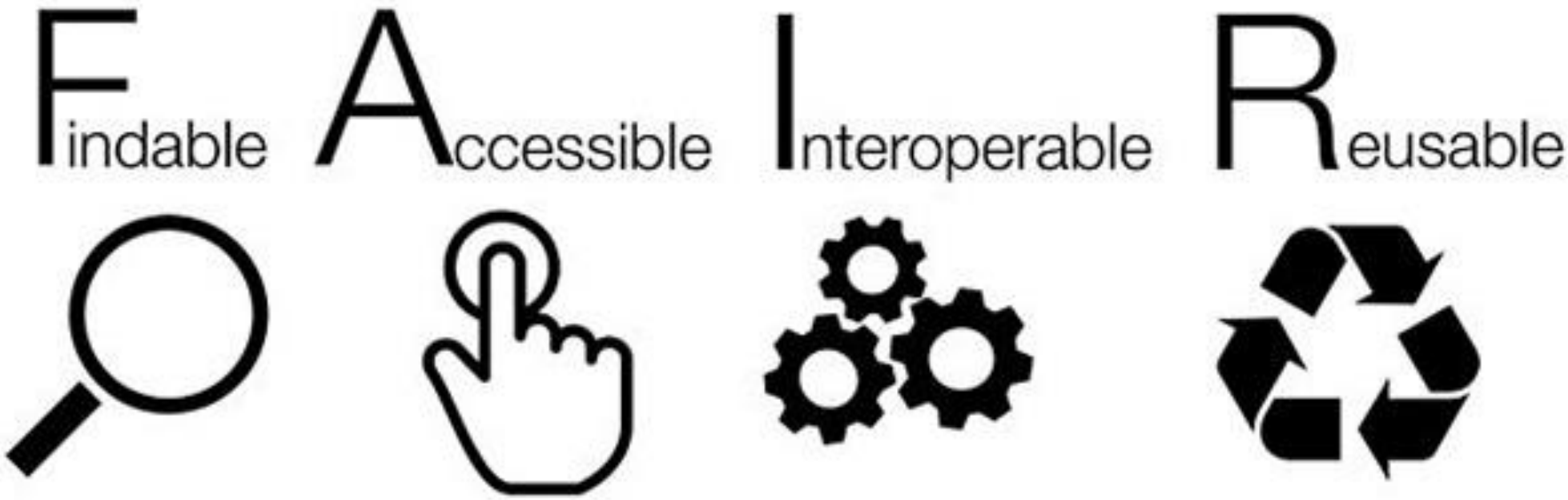
Interoperabilität zwischen Dateninfrastrukturen sollte realisiert werden über:

- **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, ...

FAIR Data Prinzipien



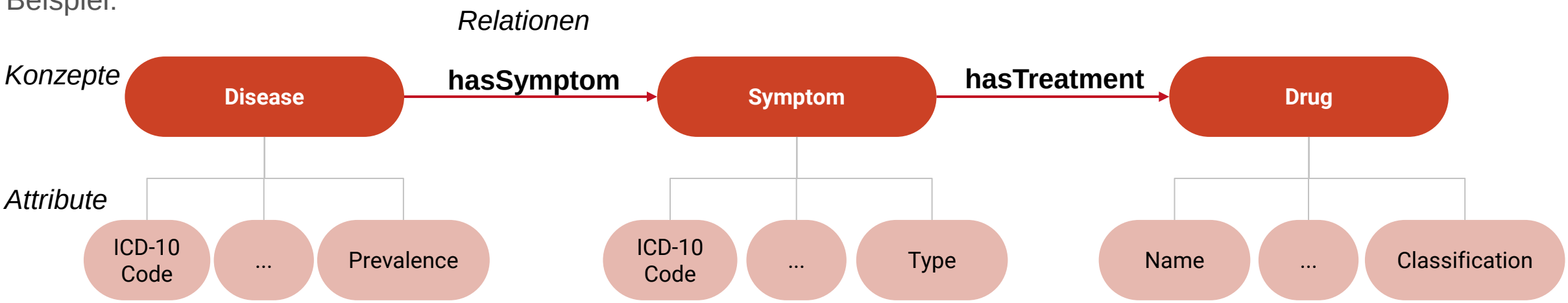
Initiativen zur dezentralen, semantischen Datenvernetzung

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

Erschließung und Integration mit Metadaten und Vokabularen

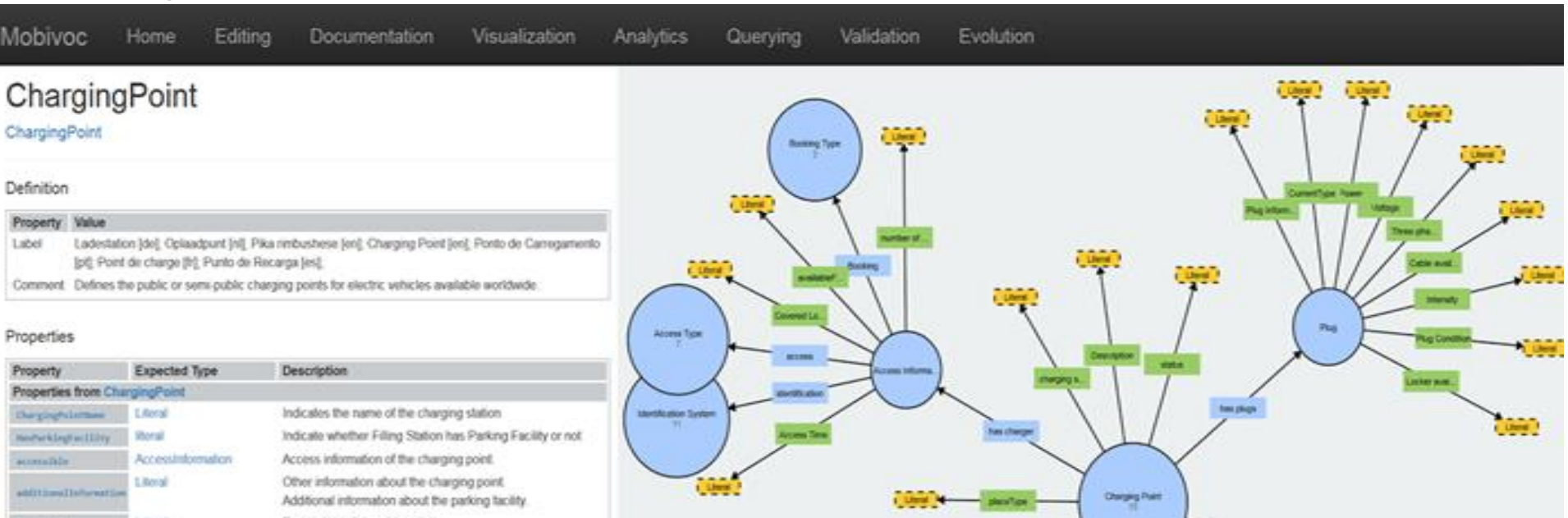
- Basismetadaten (Urheber, Lizenz, Fachgebiet, PID usw.) mit DCAT und W3C Data on the Web Best Practices machen Daten entsprechend den **FAIR Data Prinzipien** zugänglich
- Inhaltliche Erschließung der Daten erfolgt über **Mappings auf domänenspezifische Vokabulare**
- **Vokabulare etablieren gemeinsames Verständnis der Daten** und erfassen die domänenspezifische Semantik durch die Definition von Konzepten, zugeordneten Attributen und Relationen
- Daten-Mappings auf Vokabulare **ermöglichen Datenintegration** (z.B. Datenvernetzung, föderierten Zugriff) und **neue Explorationen** (semantische Suche, Visualisierung)

Beispiel:



Beispiel einer agilen, gemeinsamen Kuratierung von Vokabularen mit VoCol

- Folgt schema.org Ansatz; Basis: semantische Technologien (RDF, Linked Data)
- Methode und integrierte Entwicklungsumgebung zur Vokabular-Entwicklung
- Wissens-Ingenieure modellieren; Fachexperten prüfen
- Nutzt Git-Versionskontrolle zur Orchestrierung der Zusammenarbeit mit Branching/Merging, Push-Pull Requests
- Integriert Vielzahl von Diensten zur Kuratierung, Visualisierung, Publikation, Dokumentation, Issue Tracking, Validierung, ...
- http://vocol.iais.fraunhofer.de

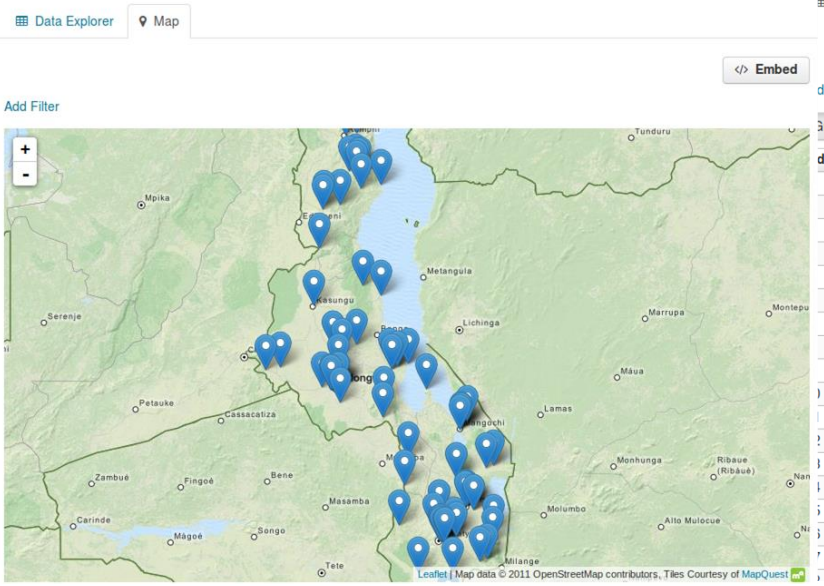


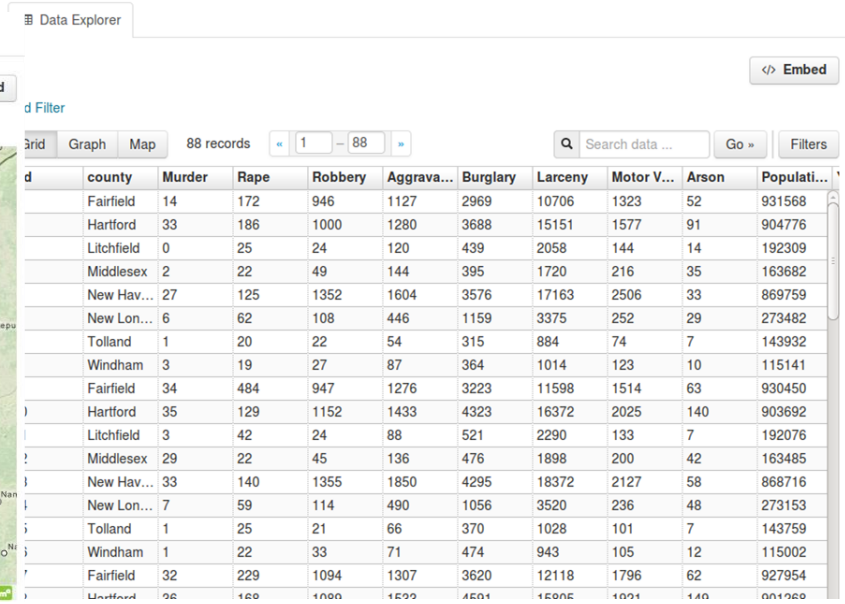
Vorteile:

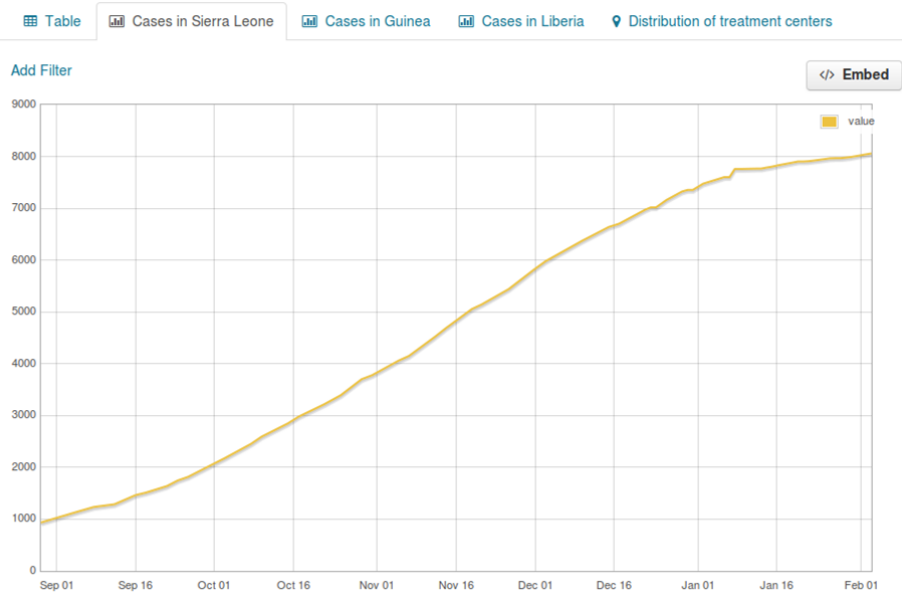
z.B. neue Möglichkeiten zur Suche, Exploration, Visualisierung

Die semantische Beschreibung ermöglicht vielfältige neue Such-, Explorations- und Visualisierungen

- **Facetenbasierte Suche** nach Datensätzen über Metadaten-Attribute
- **Direkte Visualisierungen** z.B. Kartendarstellung, Diagramme, Datenvorschau, 2D/3D, …
- **Semantische Suche** durch Harvesting/Integration von Daten in einen Research-Datawarehouse
- **Föderierte Suche** über verschiedene Datenquellen mittels W3C SPARQL Anfragesprache
- **Domänenspezifische Visualisierungsapps**, z.B. Chemische Reaktionen, CAD, …
- **Open Research Knowledge Graph:** Vernetzung von Daten, Forschungsinformationen, Publikationen







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 Forschungs

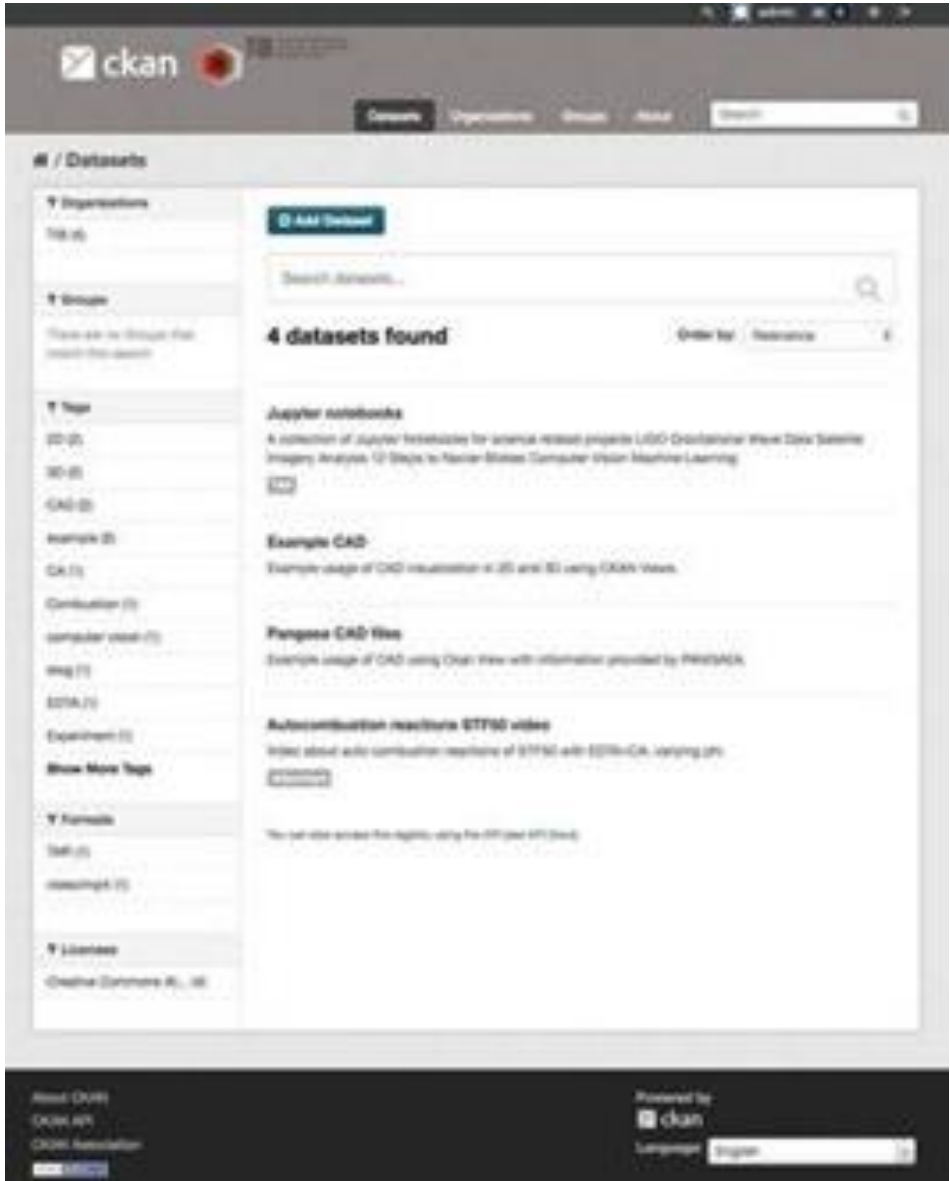
CKAN Data-Repository - wird seit über 10 Jahren für Open Government Data Portale eingesetzt (data.gov, data.gov.uk, govdata.de)

- 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

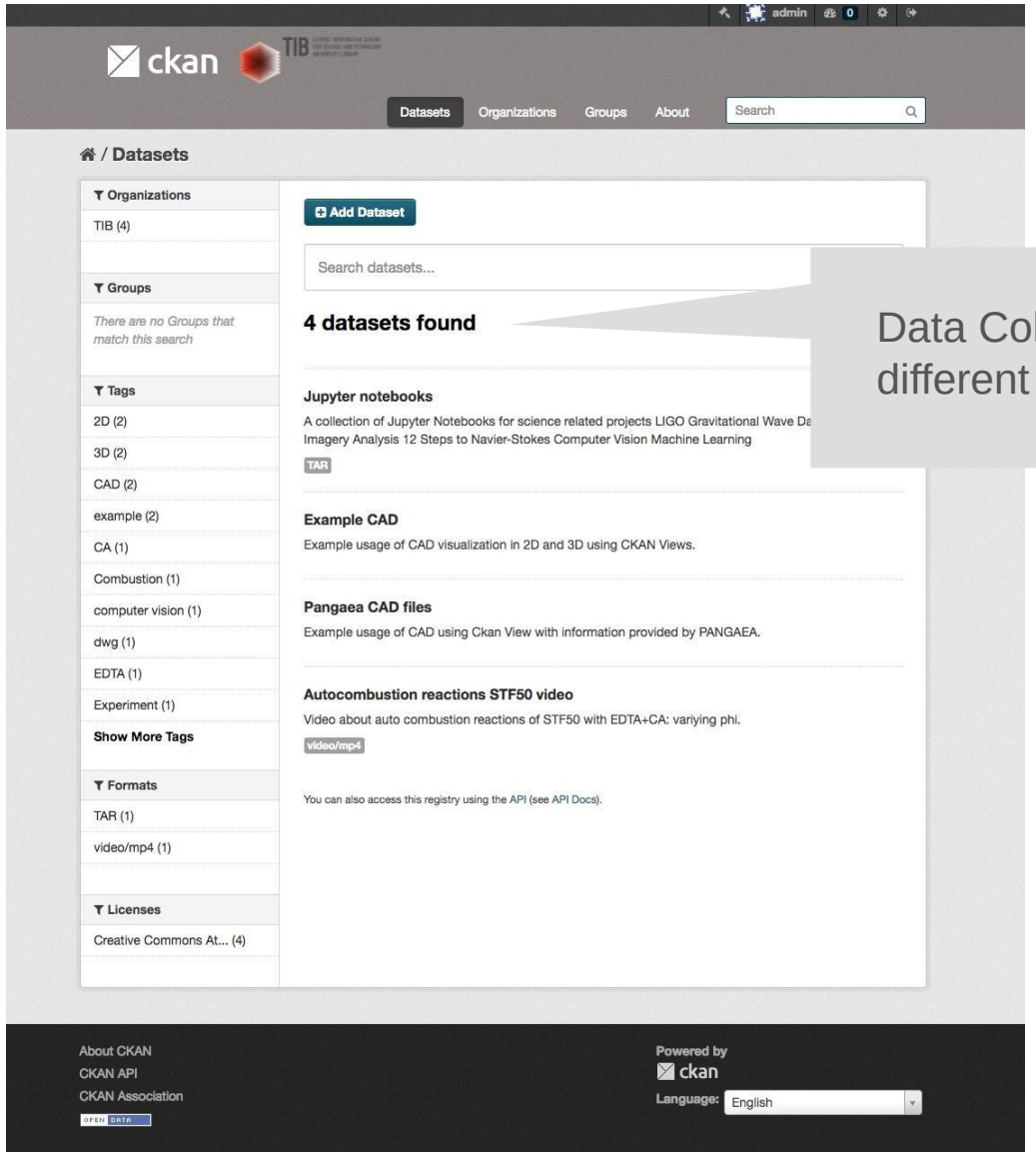
Leibniz Datamanager (http://datamanager.tib.eu) - spezifisch angepasste CKAN Distribution für Forschungsdatenmanagement

- 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

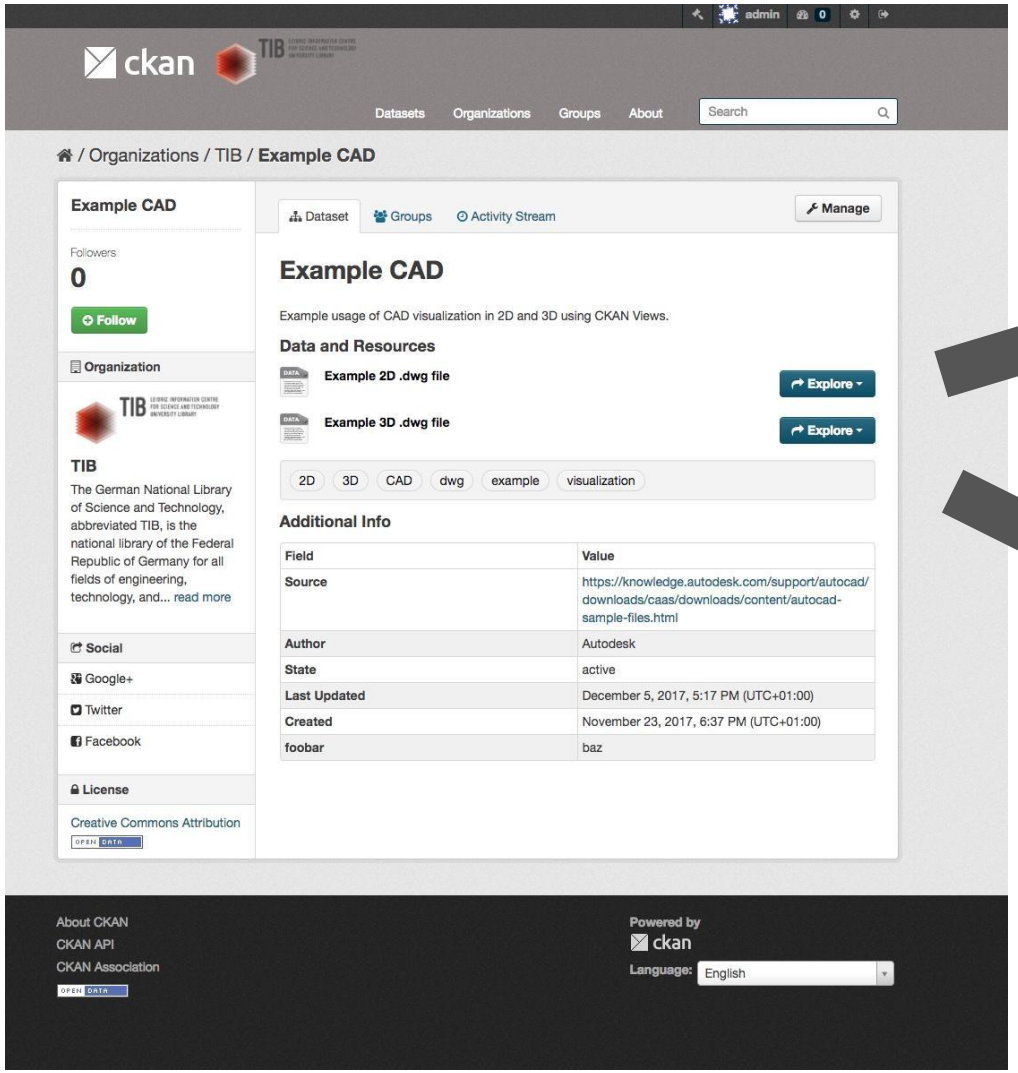


CKAN: A Repository for Heterogeneous Data Collections



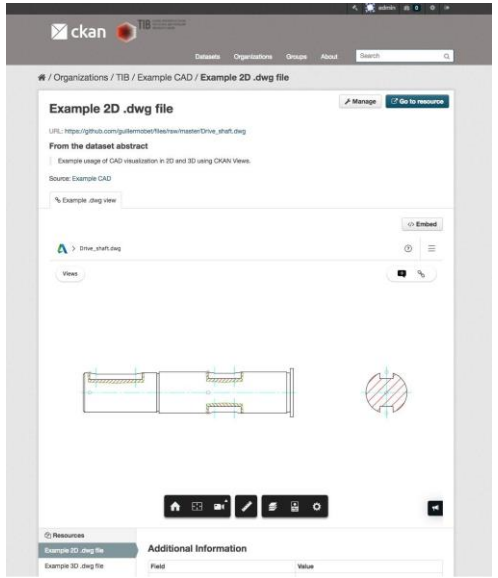
Data Collections with different formats

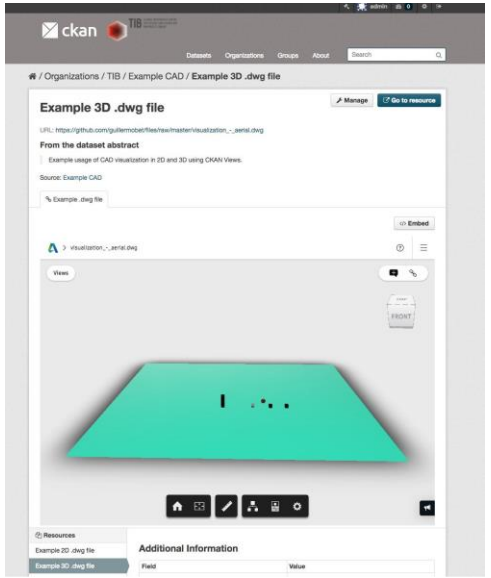
CKAN: Different Views of the Same Data Collections



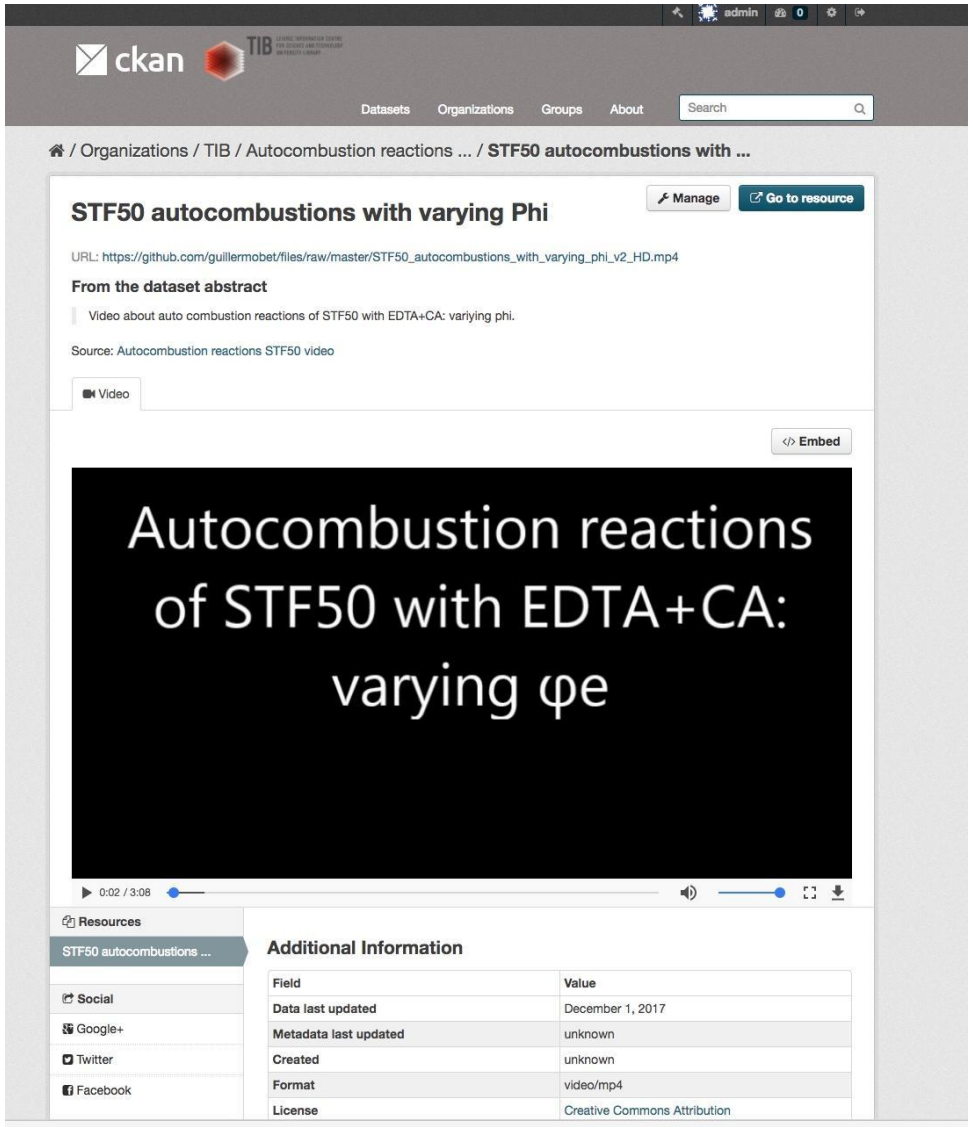
2D View

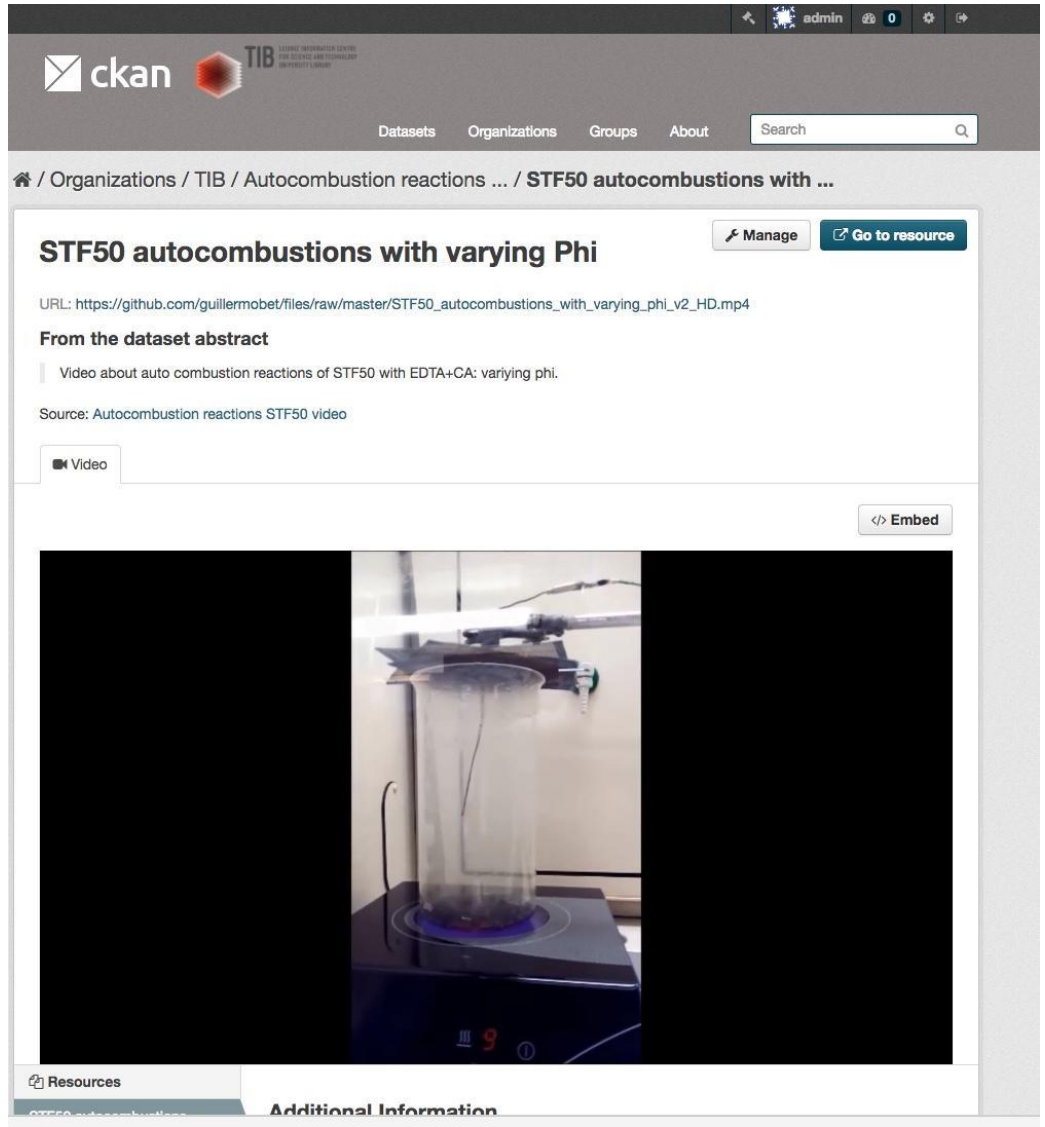
3D View



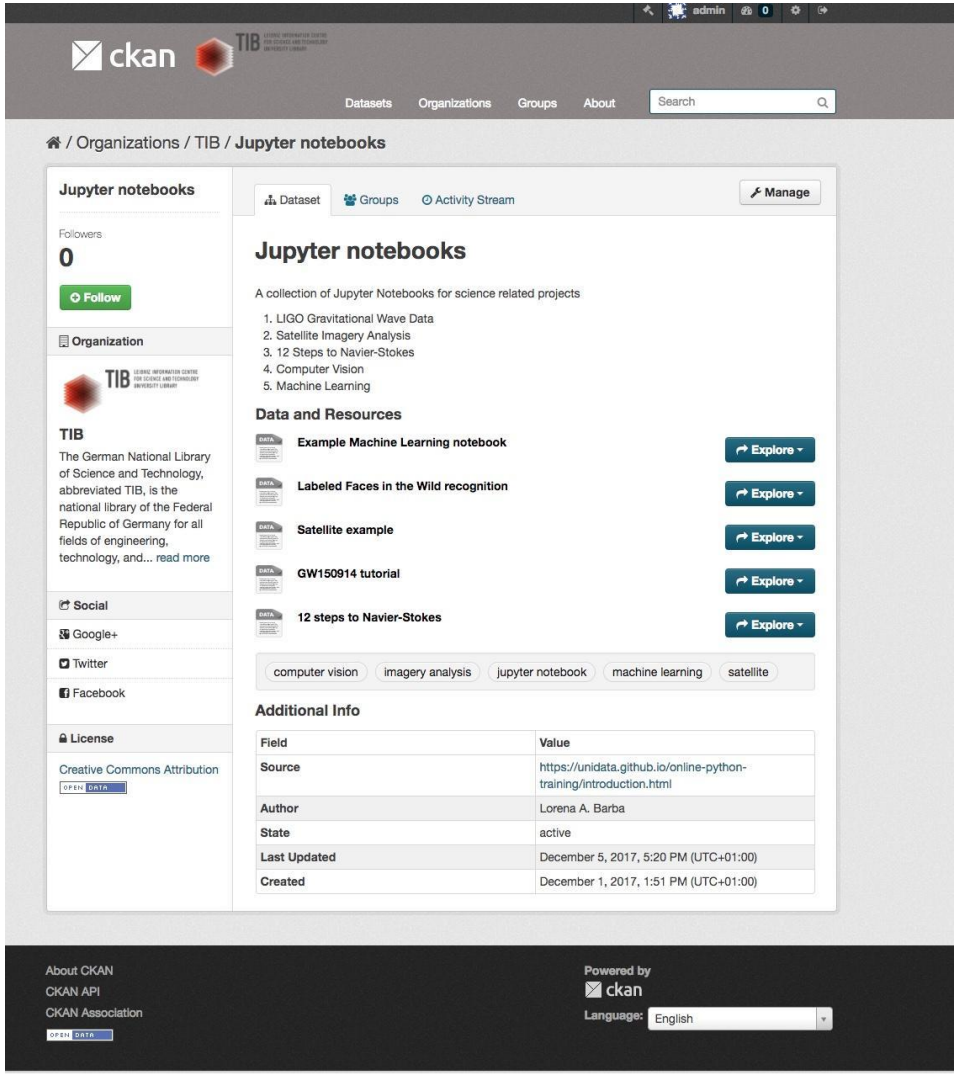


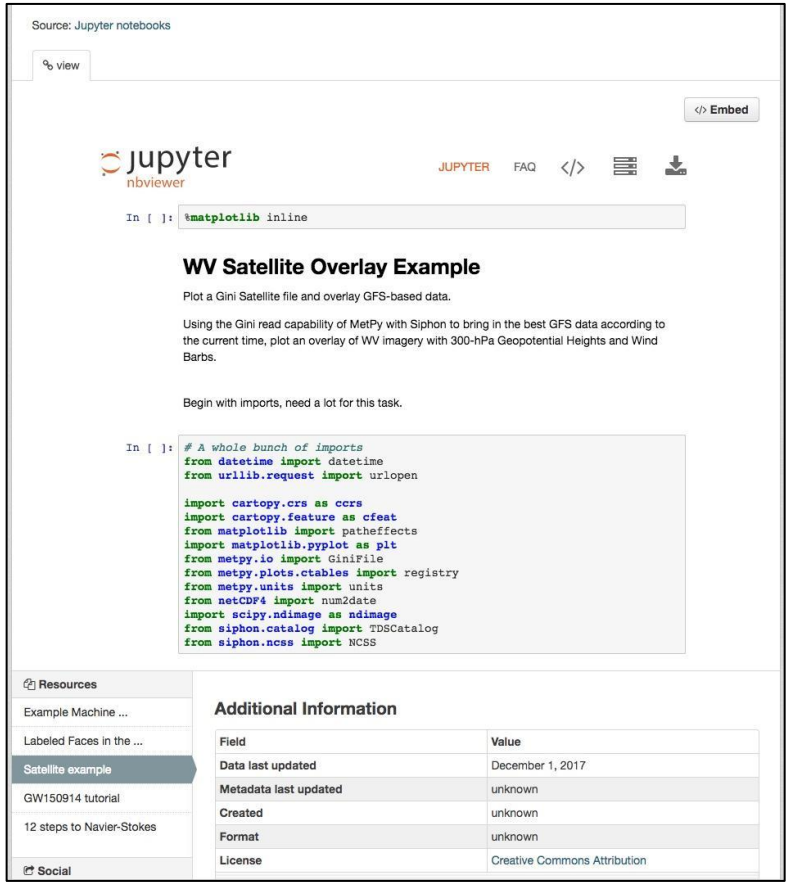
CKAN: Playing a Video



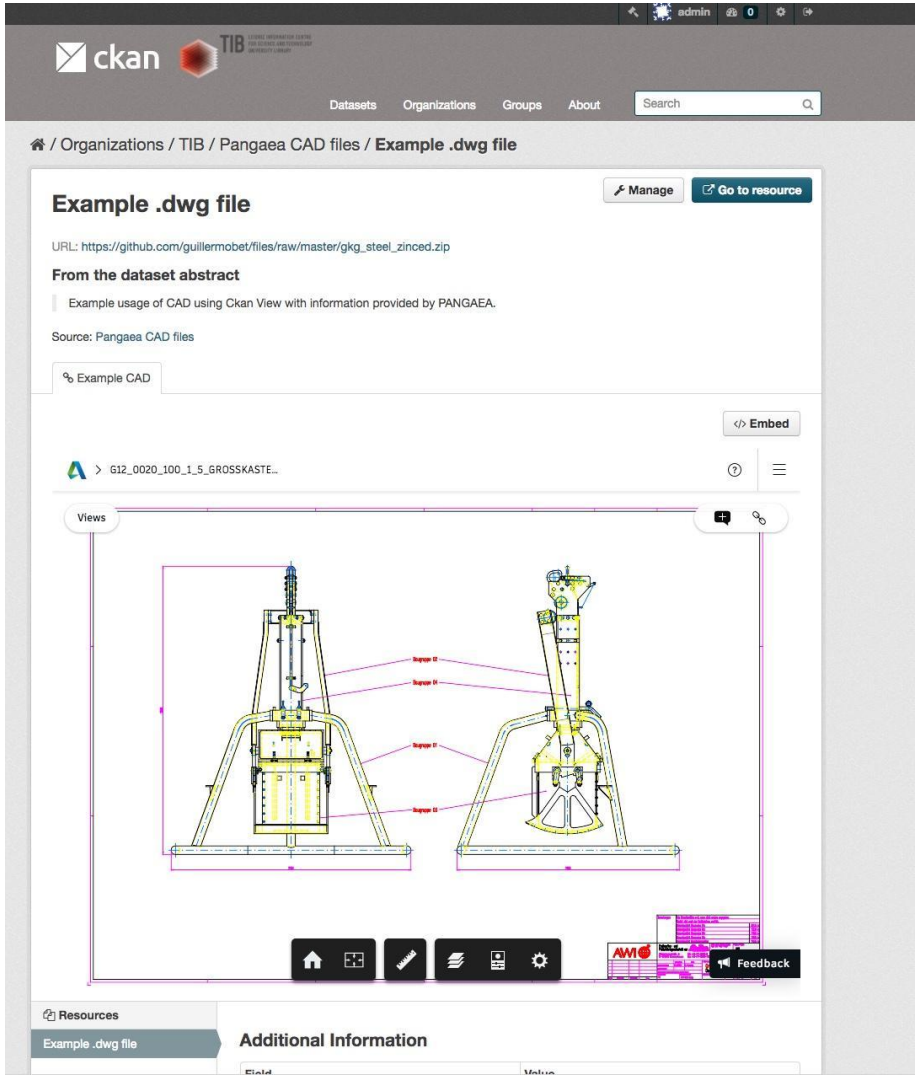


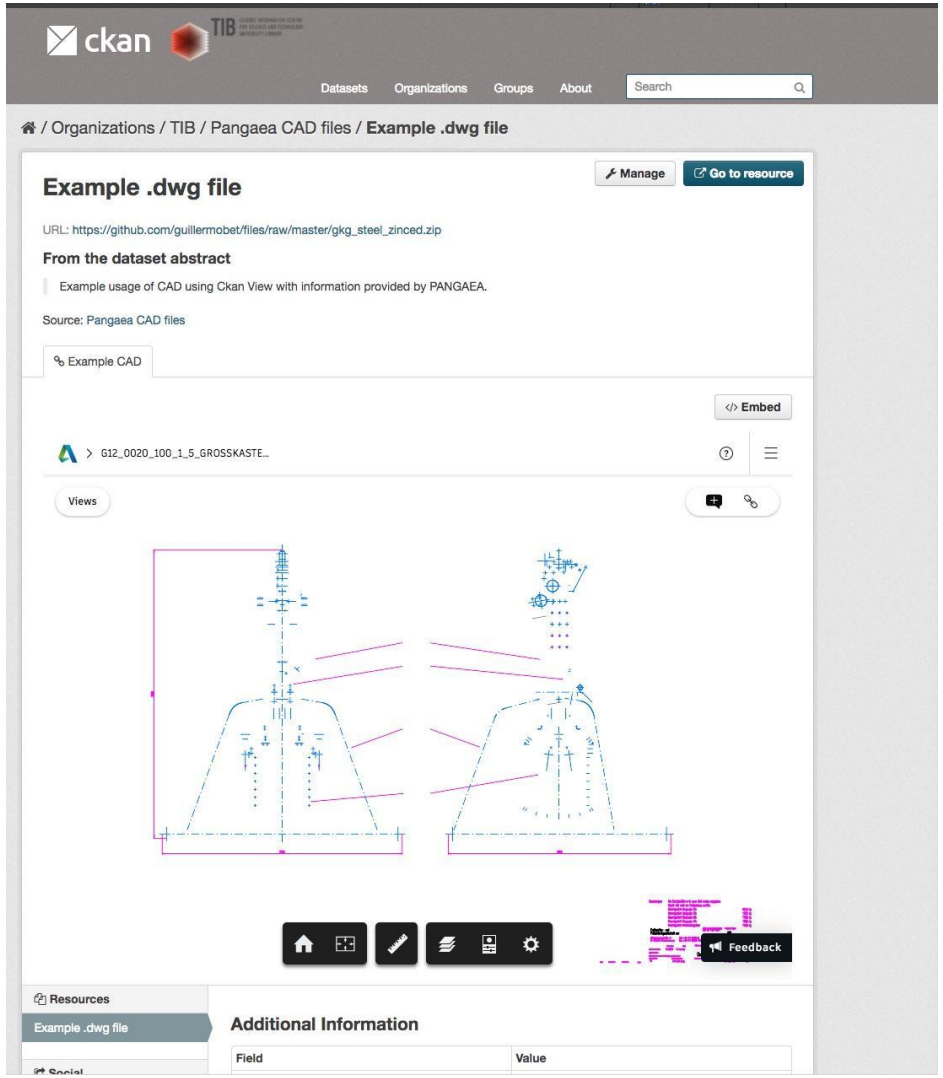
CKAN: Jupyter Notebooks for Demonstrating Live Code



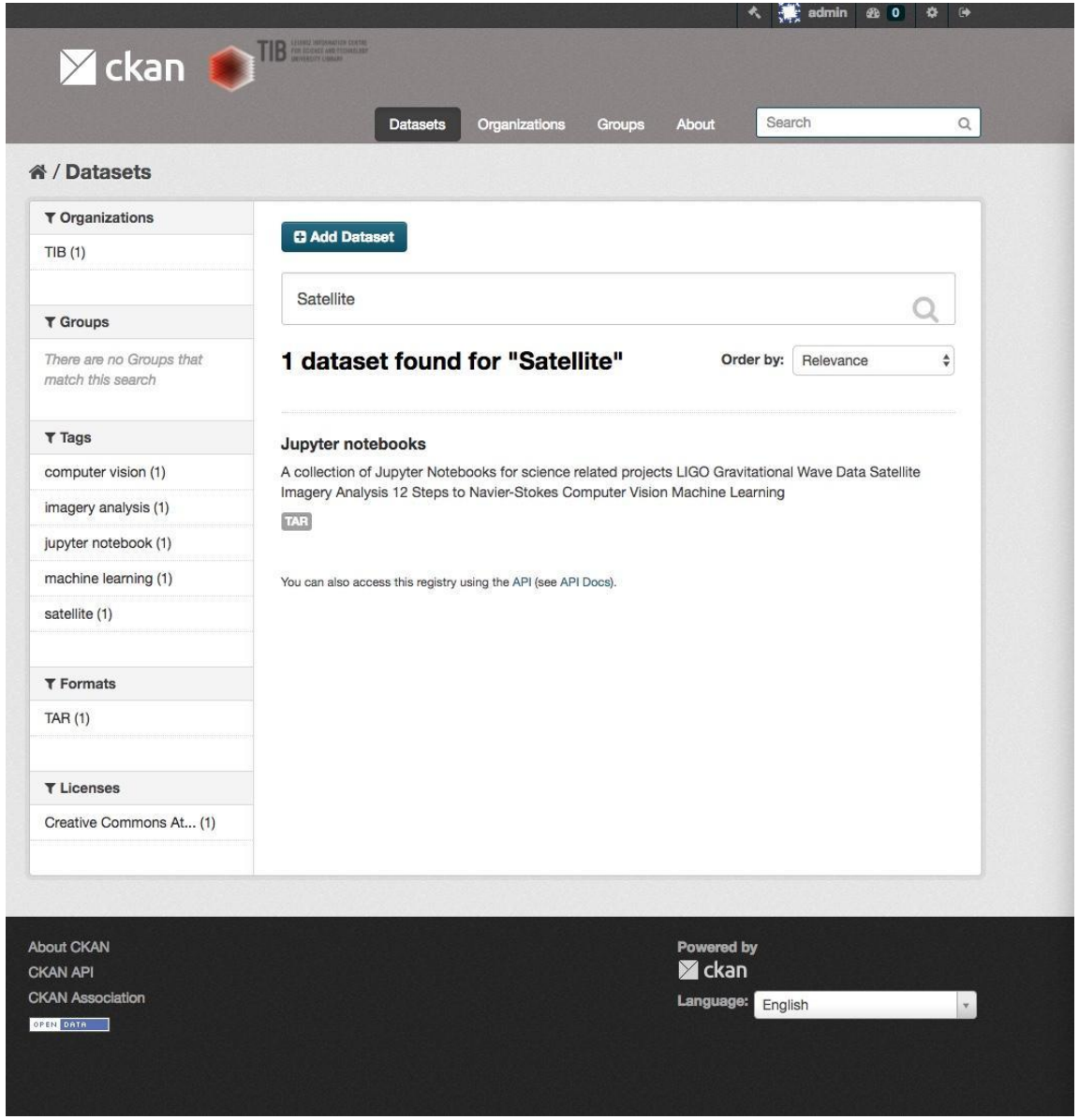


CKAN: Visualizations of Data Collections using Auto CAD

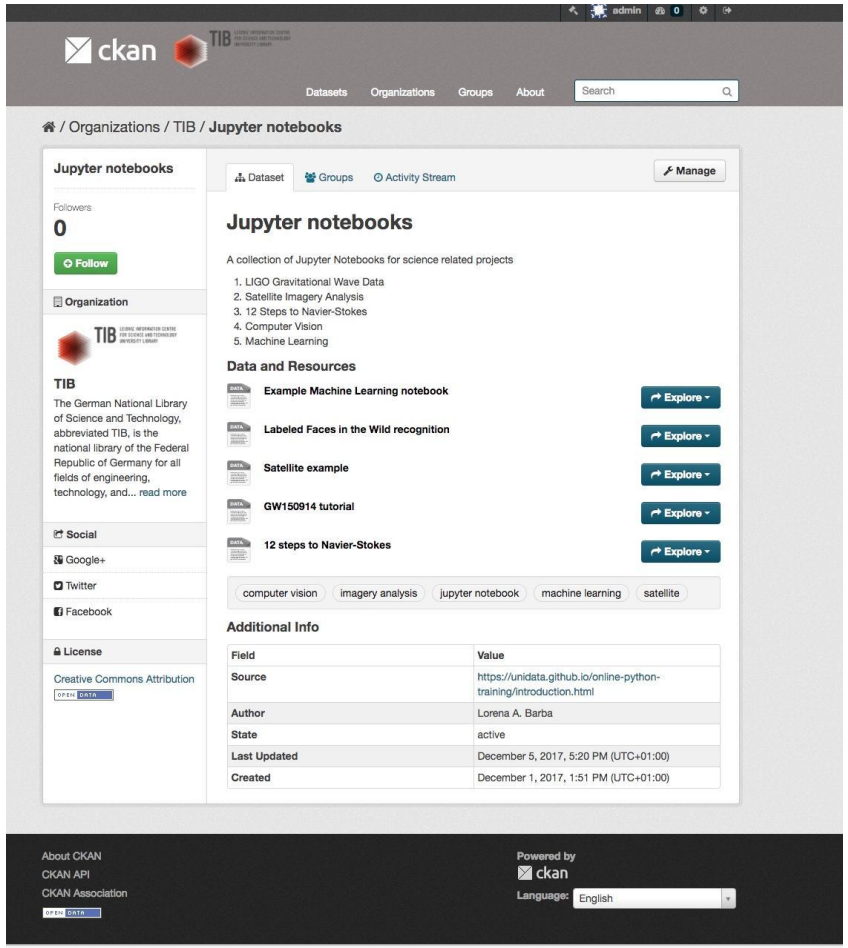




CKAN: Searching Data Collections



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

Zusammenfassung und Ausblick

Data Science erfordert mehr Zusammenarbeit –

- Fair/Open (Data/Knowledge/Source/Education) unterstützt dies

Wir brauchen

- mehr Werkzeuge zur offenen, gemeinsamen, kollaborativen Arbeit
- Mehr semantische Beschreibung und Vernetzung von Daten

Von daten-basierten zu daten-getriebenen Wissenschaft, z.B. automatische Hypothesengenerierung durch Datenanalyse

Stay tuned

- https://tib.eu
- Mailinglist/group: https://groups.google.com/forum/#!forum/orkg
- Open Research Knowledge Graph: https://orkg.org
- ERC Consolidator Grant ScienceGRAPH starting soon

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