

Textual and non-textual objects: Seamless access for scientists

Uwe Rosemann
ICIC 2013
Vienna



German National Library of Science and Technology (TIB)

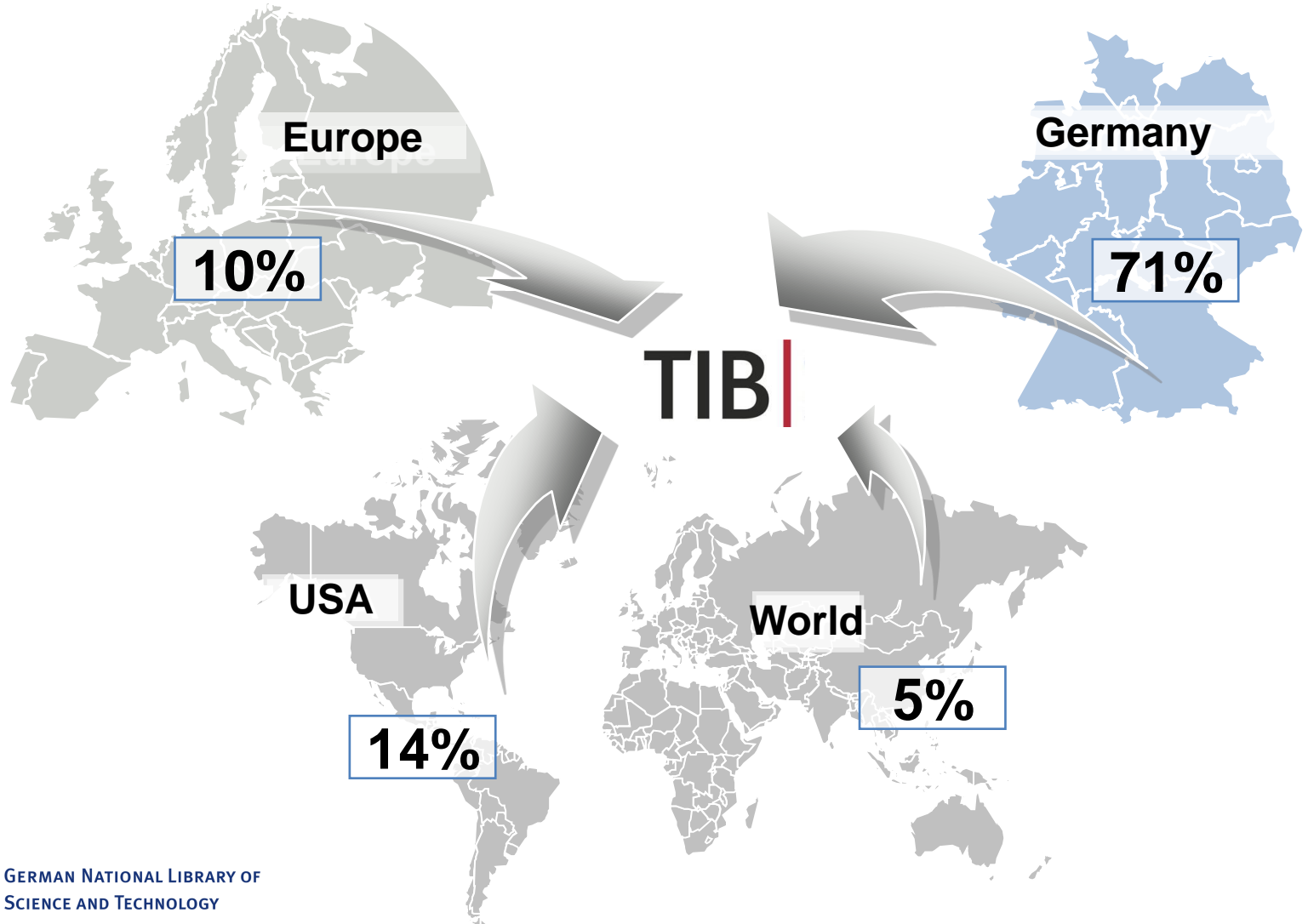
- Specialized Library for Architecture, Chemistry, Computer Science, Mathematics, Physics, Engineering Technology
- Financed by Federal Government and all Federal States
- Member of the Leibniz Association
- Global supplier for scientific and technical information



Global Network



Customers



Main Services

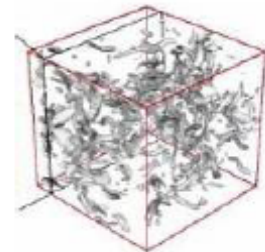
- Provision of scientific content
 - full texts, document delivery, interlibrary loan
- Scientific retrieval
 - portal GetInfo
- Long-term preservation
- DOI-Service for research data
- Research and development

Changes in the scientific process

Science Paradigms

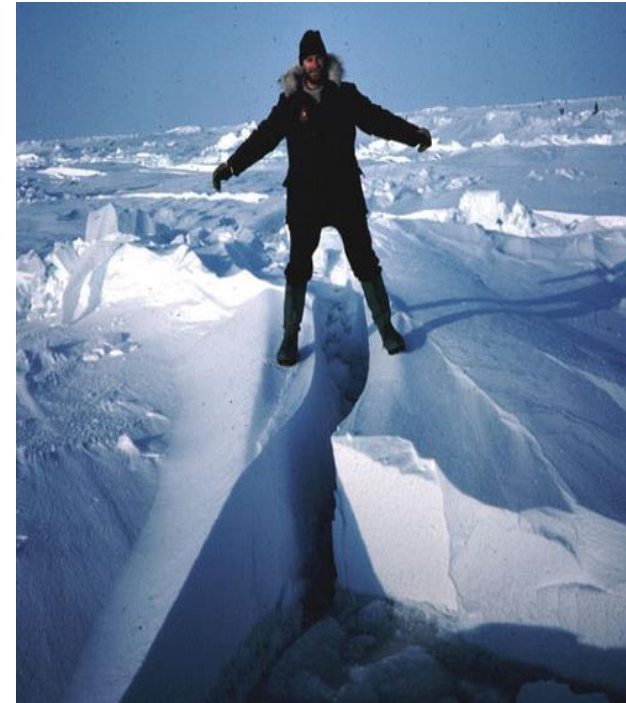
- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a **computational** branch
simulating complex phenomena
- Today:
data exploration (eScience)
unify theory, experiment, and simulation
using data management and statistics
 - Data captured by instruments
 - Or generated by simulator
 - Processed by software
 - Scientist analyzes database / files

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



A gap

- A widening gap in the scientific record between published research in a text document and the data that underlies it
- As a result, datasets are
 - difficult to discover
 - difficult to access
- Scientific information gets lost

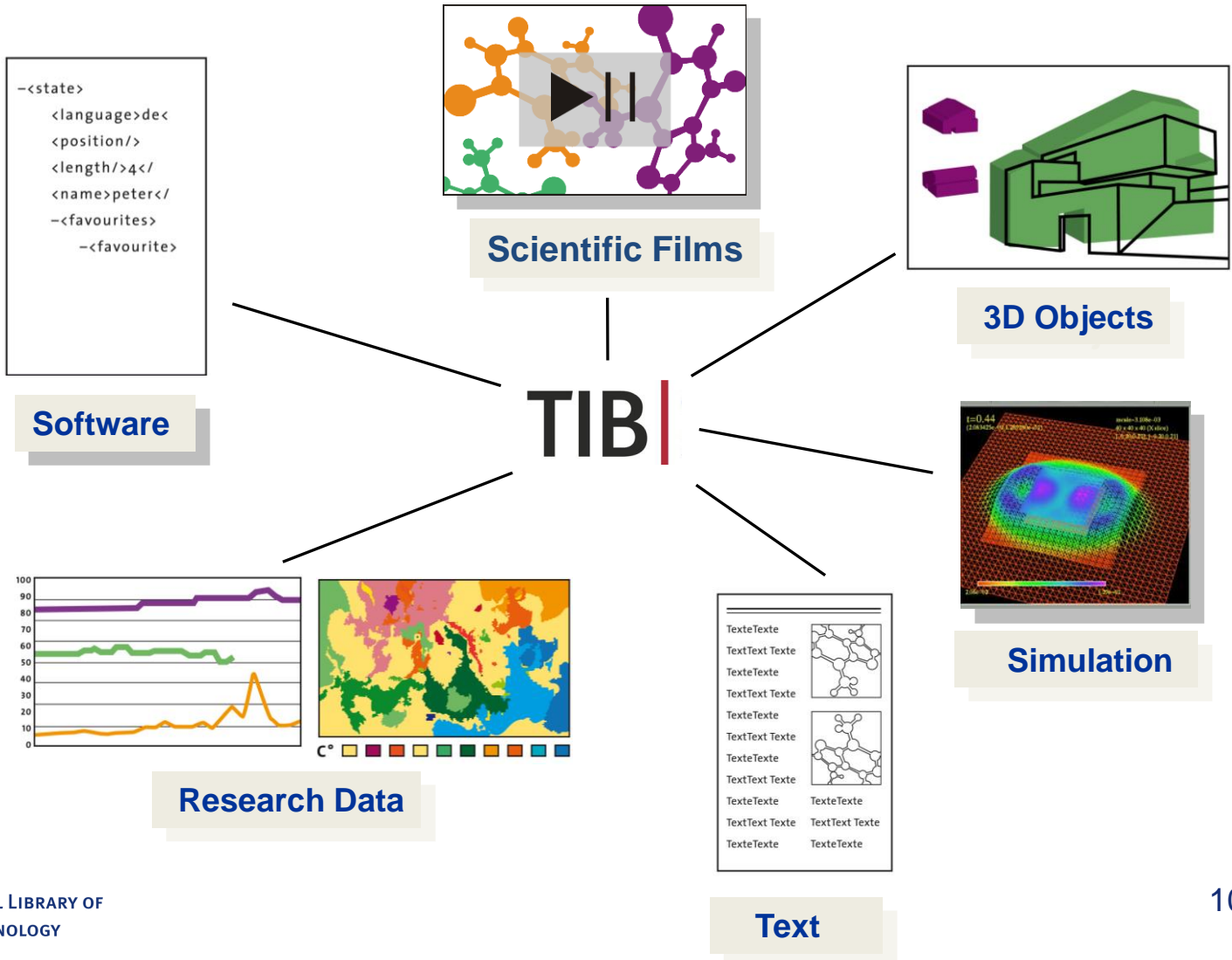


„Riding the wave“ – How Europe can gain access from the rising tide of scientific data

Final report of the High Level Expert Group on Scientific Data.



Strategy – Move beyond text



Move beyond text – Consequences for TIB

- Research communities produce many types of scientific and technical information
- Each has its own unique characteristics and life cycle
- Must become capable of accepting and managing new media formats

Competence Center for Non-textual Materials I

- Develop a clear strategy for the use and integration of non-textual materials at the TIB
- Systematically collect non-textual materials from research and teaching
- Define, integrate and establish technical infrastructure
- Define and establish workflows for indexing, cataloguing, digital preservation, DOI names, licencing

Competence Center for Non-textual Materials II

- Develop innovative media-specific portals enabled by e.g. an automated video analysis with scene, speech, text and image recognition
 - Linking non-textual materials to other research information such as full texts and research data via the specialist portal GetInfo
 - Engage in communities, provide support and advice to media providers
- TIB will establish its own research capacity**

How have we been preparing ?

- ***Infrastructure for research data***
- Visual search tools for AV-media
- 3D-Objects
- chemOCR

Collaboration – Research Data

- In 2005, the TIB became a non-commercial DOI registration agency for research data
- In 2010, the TIB became co-founder of the international DataCite consortium to establish easier access to scientific research data on the Internet

Mission

- Citability of research data
- High visibility of the data
- Easy re-use and verification of the data sets
- Increasing quality of published papers

DataCite Members



Example: EHEC virus

Metadata Search beta

[Options](#) | [Advanced Search](#) | [About Us](#) | [Contact](#) | [Help](#)

DataCite

Filter

allocator

datacentre

prefix

resourceType

contributor

creator

publicationYear

publisher

language

refQuality

has_metadata

No active filters. Use the sidebar to filter search results.

20 documents found in 54ms

Page 1 of 2 

Data from: Genetic variation for antibiotic persistence in Escherichia coli

1

 [doi:10.5061/DRYAD.H7D843J6](#)

Stewart, Balint • Rozen, Daniel E

Data from: Genetic variation for antibiotic persistence in **Escherichia coli**

Leitfaden Labordiagnostik von shigatoxinbildenden und anderen darmpathogenen Escherichia coli-Stämmen # 2

 [doi:10.4126/38M-003183531](#) Text : Book

Heißenhuber, Annette

Leitfaden Labordiagnostik von shigatoxinbildenden und anderen darmpathogenen **Escherichia coli**-Stämmen

Removal of bacterial indicators of fecal contamination in urban stormwater using a natural riparian buffer # 3

[version 1]

 [doi:10.4122/1.1000001379](#) Event : Conference presentation

Casteel, M. • Bartow, G. • Taylor, S. R. • Sweetland, P.

showed that levels of **Escherichia coli** and total coliforms increased significantly during storm events **Escherichia coli**

Removal of bacterial indicators of fecal contamination in urban stormwater using a natural riparian buffer # 4

[version 1]

 [doi:10.4122/1.1000001380](#) Text : Conference full text

Casteel, M. • Bartow, G. • Taylor, S. R. • Sweetland, P.

showed that levels of **Escherichia coli** and total coliforms increased significantly during storm events **Escherichia coli**

Genomic data from Escherichia coli O104:H4 isolate TY-2482

5

 [doi:10.5524/100001](#)

Li, D • Xi, F • Zhao, M • Liang, Y • Chen, W • (et. al.)

Genomic data from **Escherichia coli** O104:H4 isolate TY-2482 the **Escherichia coli** O104:H4 TY-2482 isolate genome sequencing consortium

Example: EHEC virus

E. coli

The May 2011 outbreak of an *E. coli* infection in Europe resulted in serious concerns about the potential appearance of a new deadly strain of bacteria, *Escherichia coli* O104:H4 TY-2482. In response to this situation, and immediately after the reports of deaths, the University Medical Centre Hamburg-Eppendorf and BGI-Shenzhen worked together to sequence the bacterium and assess its human health risk.

The bacterium's genome was first sequenced using Life Technologies' Ion Torrent sequencing platform. According to the results of the draft assembly, the estimated genome size of this new *E. coli* strain is about 5.2 Mb. Sequence analysis indicated this bacterium is an EHEC serotype O104 *E. coli* strain. Comparative analysis showed that this bacterium has 93% sequence similarity with the EAEC 55989 *E. coli* strain, which was isolated in the Central African Republic and known to cause serious diarrhea. This strain of *E. coli*, however, has also acquired specific sequences that appear to be similar to those involved in the pathogenicity of hemorrhagic colitis and hemolytic-uremic syndrome. The acquisition of these genes may have occurred through horizontal gene transfer.

To maximize its utility to the research community and aid those fighting the epidemic, this genomic data was released into the public domain under a [CC0 license](#).

To the extent possible under law, [BGI Shenzhen](#) has waived all copyright and related or neighboring rights to genomic data from the 2011 *E. coli* outbreak. This work is published from: China.

Workflow and further information can be found at:

http://climb.genomics.cn/Ecoli_TY-2482



Citation

In further accordance with our [terms of use](#), please cite this dataset as:

Li, D; Xi, F; Zhao, M; Chen, W; Cao, S; Xu, R; Wang, G; Wang, J; Zhang, Z; Li, Y; Cui, C; Chang, C; Cui, C; Luo, Y; Qin, J; Li, S; Li, J; Peng, Y; Pu, F; Sun, Y; Chen, Y; Zong, Y; Ma, X; Yang, X; Cen, Z; Song, Y; Zhao, X; Chen, F; Yin, X; Rohde, H; Liang, Y; Li, Y and the

DOI Services

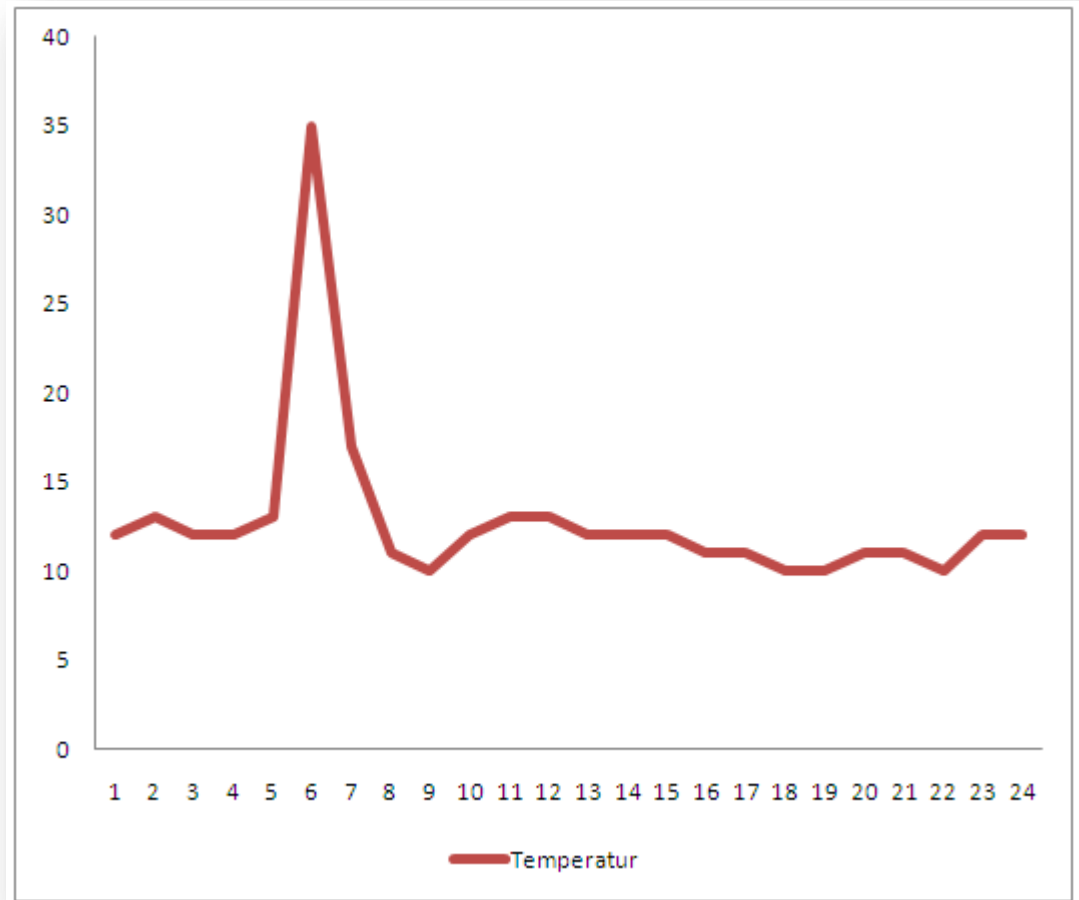
- Contracts with 60 data centres
 - Research Institutes
 - Universities
 - Libraries
 - Publisher
- 776.454 DOI registrations
 - 22.533 up to September 2013

Research data – Further developments

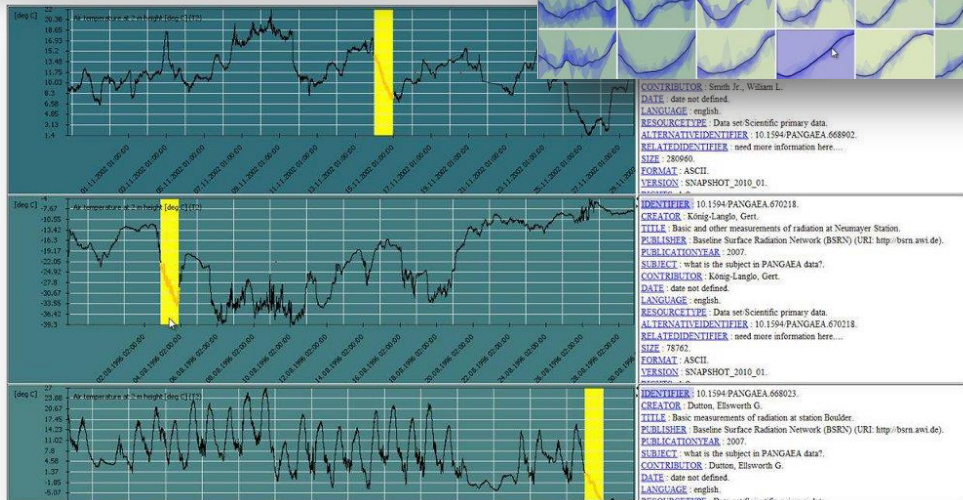
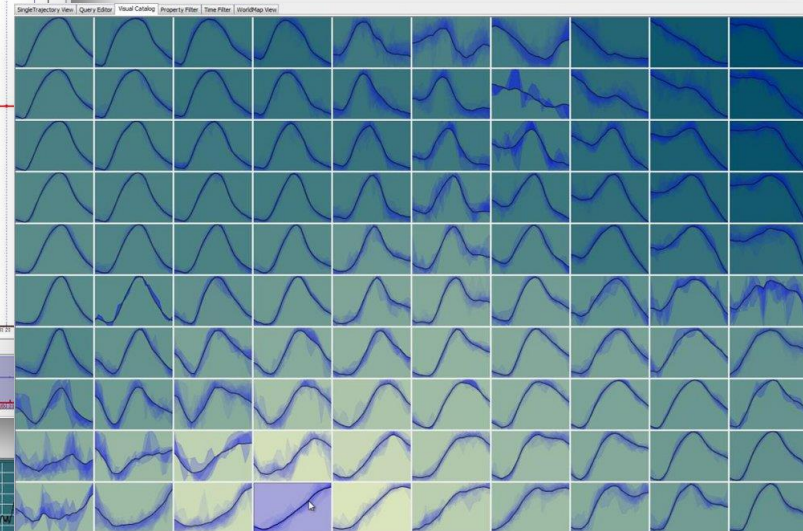
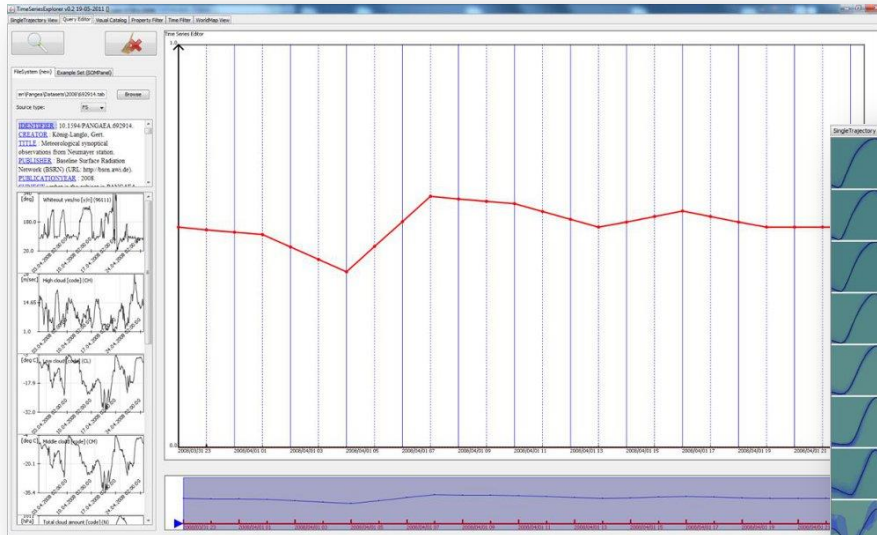
- KomFor
 - Centre of Expertise for Research Data from the „Earth and Environment“ project
- RADAR
 - RADAR - Research Data Repository
- Visual Analysis
 - VisInfo Methods

Numerical data

Zeit [h]	T [°C]
1	12
2	13
3	12
4	12
5	13
6	35
7	17
8	11
9	10
10	12
11	13
12	13
13	12
14	12
15	12
16	11
17	11
18	10
19	10
20	11
21	11
22	10
23	12
24	12



Visual access to research data



How have we been preparing ?

- Infrastructure for research data
- ***Visual search tools for AV-media***
- 3D-Objects
- chemOCR

TIB's portal for audiovisual media

Project	Development of a portal for audiovisual media
Aim	Improve access to AV-Media
Time	July 2011 – December 2013
Partner	Hasso-Plattner Institut for Softwaresystemtechnology GmbH

The screenshot shows the TIB AV-Portal website. At the top, there is a navigation bar with the title "TIB AV-Portal" on the left, a language dropdown set to "English", and the TIB logo "TIB TECHNISCHE INFORMATIONSBIBLIOTHEK" on the right. Below the navigation bar is a search bar with the placeholder text "Search for People, Places, Subjects ..." and a magnifying glass icon. To the right of the search bar are social media icons for Facebook and Twitter, and a "Follow us:" label. Below the search bar is a "Startpage" section with a "Like" button (0), a "Tweet" button (0), and an "Email" button (0). The main content area is divided into two columns. The left column is titled "About the AV-Portal" and contains a welcome message, a description of the portal's purpose, and a list of features: "Search in scientific films", "Simple publishing from copyrighted videos", "Professional quality control and indexing", and "Semantic search by use of multimedia analysis method". Below this list is a link for "More Information". The right column is titled "Subjects" and displays a grid of video thumbnails. Each thumbnail includes a "Recently added:" label, a video player icon, and a title. The subjects shown are: Mathematics (Subject) (12 videos) with "02A.1 Kehwert ableiten"; Physics (Subject) (12 videos) with "Erkenntnisse aus der Unendlichkeit"; Engineering (Subject) (16 videos) with "Technik, die versteht"; Chemistry (Subject) (4 videos) with "Katalyse für die Zukunft"; Information Technology (Subject) (21 videos) with "Ultra high-Speed Mobile Information and"; and Architecture (Subject) (10 videos) with "Mit den Menschen bauen".

TIB's portal for audiovisual media

How do I find what I'm looking for in videos?

Today: Manual annotation of the whole video

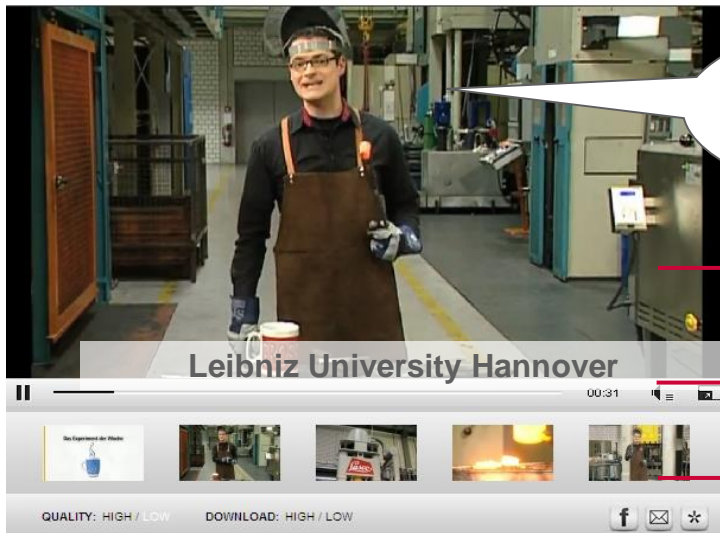


Metadata

- Titel
- Author
- Description
- Publisher
- Publication year
- Rightsholder
-

TIB's portal for audiovisual media

Future: Manual Annotation plus content-based information



source: Scorupka, Sascha, Experiment der Woche, 2011

1. Speech



2. Visual features

e.g. Indoor, Experiment, Technology

3. Textual information

4. Structural Information

Scenes, Shots, Segments

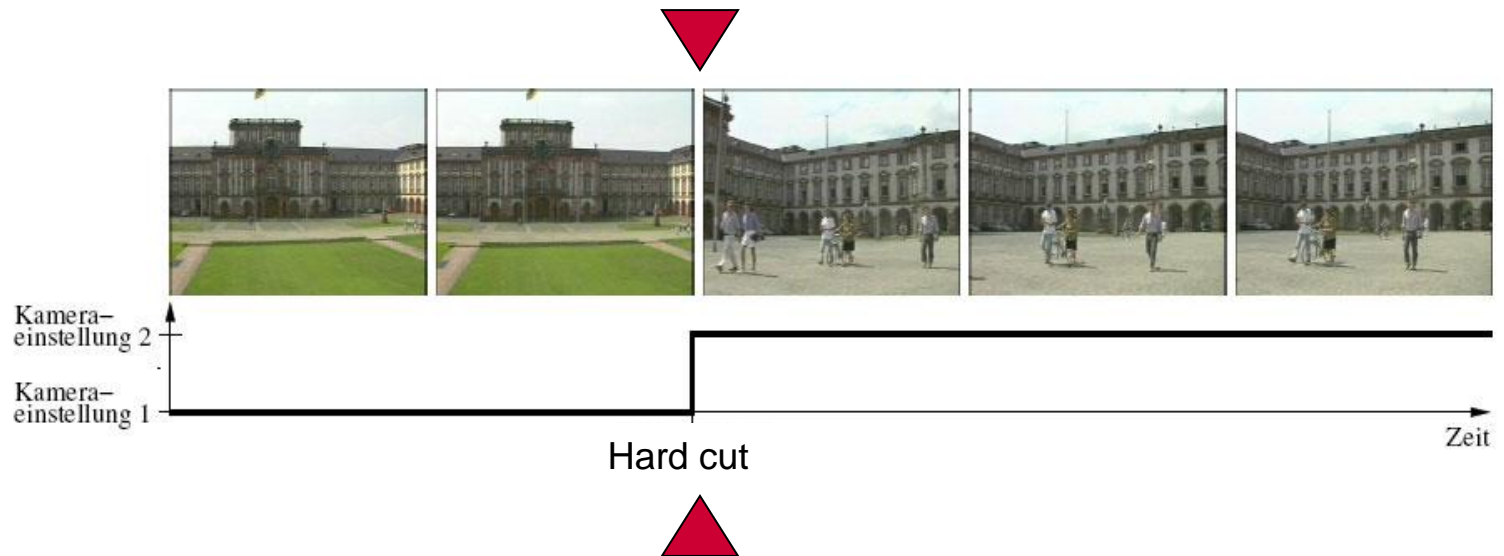
TIB's portal for audiovisual media

Media analysis process



TIB's portal for audiovisual media

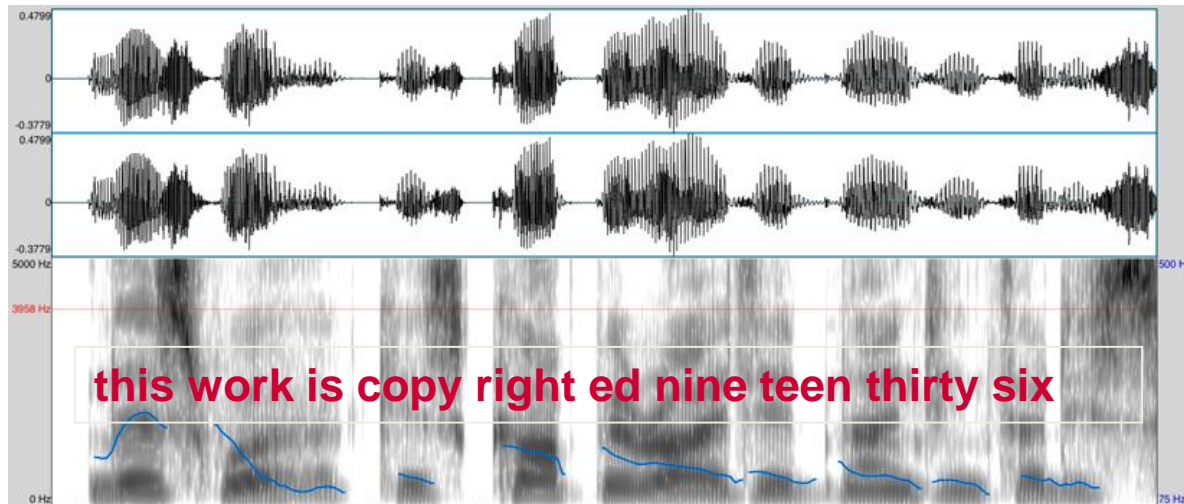
Scene recognition



- Automatic cut detection
- luminance / contrast
 - colour distribution / colour histogram
 - edges

TIB's portal for audiovisual media

Automatic speech recognition



Quality of results is dependent upon

- quality of the speaker
- dialects
- background noises
- voice overlaps

TIB's portal for audiovisual media

Intelligent Character Recognition

Intelligent Character Recognition (ICR)

- Character/Logo Detection
- Character Filtering
- Character Recognition

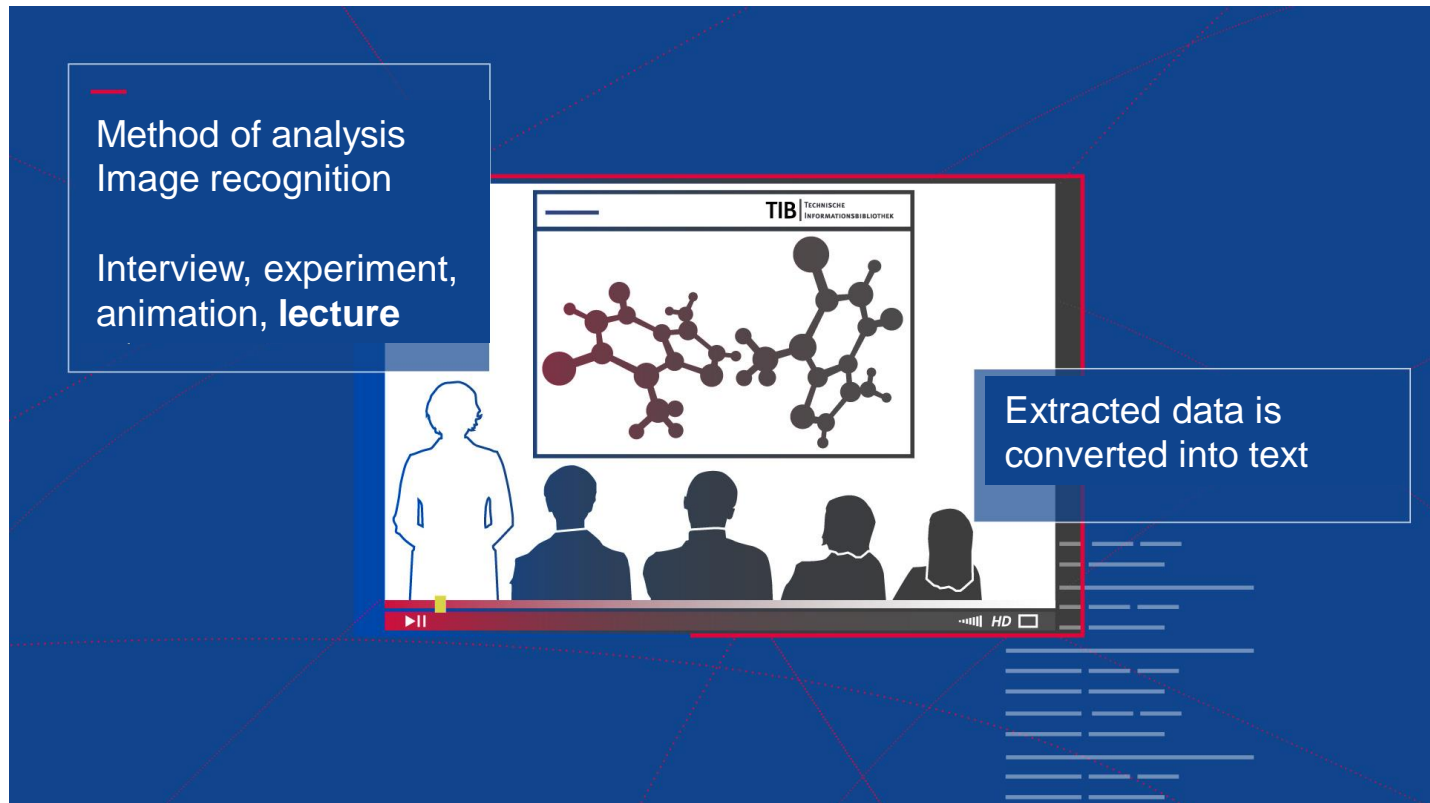
11.1 Video Similarity

- **Solution:** consider quantities of similar frames as fundamental units
 - Without regarding the temporal structure (representation as a set of feature vectors) we combine all visually similar frames to clusters
 - Two frames $x, y \in X$ belong to the same cluster if $d(x, y) \leq \epsilon$
 - **Problem:** consistent cutting is not always possible
 - if $d(x, y) \leq \epsilon$ and $d(y, z) \leq \epsilon$, then what is with $d(x, z)$?

Multimedia Databases – Wolf-Tilo Balke – Institut für Informationssysteme – TU Braunschweig 18

TIB's portal for audiovisual media

Automated analysis: Image recognition



TIB's portal for audiovisual media

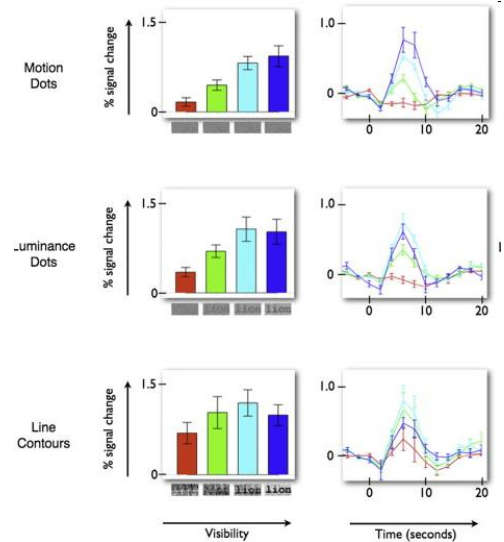
Keyframes

Annotation

Machine learning
using visual features

05:58 - 14:25 s / 8,68 s (28 Frames)	
14:25 - 20:02 s / 5,77 s (19 Frames)	
09:42 - 10:22 s / 40 s (14 Frames)	
15:12 - 17:21 s / 2,09 s (7 Frames)	
17:21 - 18:11 s / 90 s (30 Frames)	

Visual Concepts
 Graphical : Animation
 Graphical : Drawing
 Graphical : Diagram
 Real : Outdoor
 Real : Indoor
 Real : Lecture /
 Conference
 Real : Interview
 Real : Buildings
 ...



TIB's portal for audiovisual media

TIB AV-Portal

English

TIB TECHNISCHE
INFORMATIONSBIBLIOTHEK

Deutsche Zentrale Fachbibliothek für
Technik sowie Architektur, Chemie,
Informatik, Mathematik und Physik.

[Home](#) [Subjects](#) [Publisher](#) [About AV-Portal](#)

[Watchlist](#) [Video-Upload](#) [Login](#)

Search for People, Places, Subjects ...



Follow us: [f](#) [t](#)

Startpage

[Like](#) 0 [Tweet](#) 0 [+1](#) 0 [Email](#) 0

About the AV-Portal

Welcome to the TIB AV-Portal

Do you use, publish or cite scientific audiovisual material?
The AV portal of the German National Library of Science and Technology (TIB) provides new multimedia retrieval methods for the search in high grade scientific films from the fields of technology and the natural sciences.

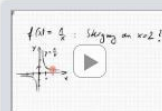
- ✓ Search in scientific films
- ✓ Simple publishing from copyrighted videos
- ✓ Professional quality control and indexing
- ✓ Semantic search by use of multimedia analysis method

[More Information](#)

Subjects

Mathematics (Subject) (12 videos)

Recently added:



02A.1 Kehrwert ableiten

Physics (Subject) (12 videos)

Recently added:



Erkenntnisse aus der Unendlichkeit

Engineering (Subject) (16 videos)

Recently added:



Technik, die versteht

Chemistry (Subject) (4 videos)

Recently added:



Katalyse für die Zukunft

Information Technology (Subject) (21 videos)

Recently added:



Ultra high-Speed Mobile Information and

Architecture (Subject) (10 videos)

Recently added:



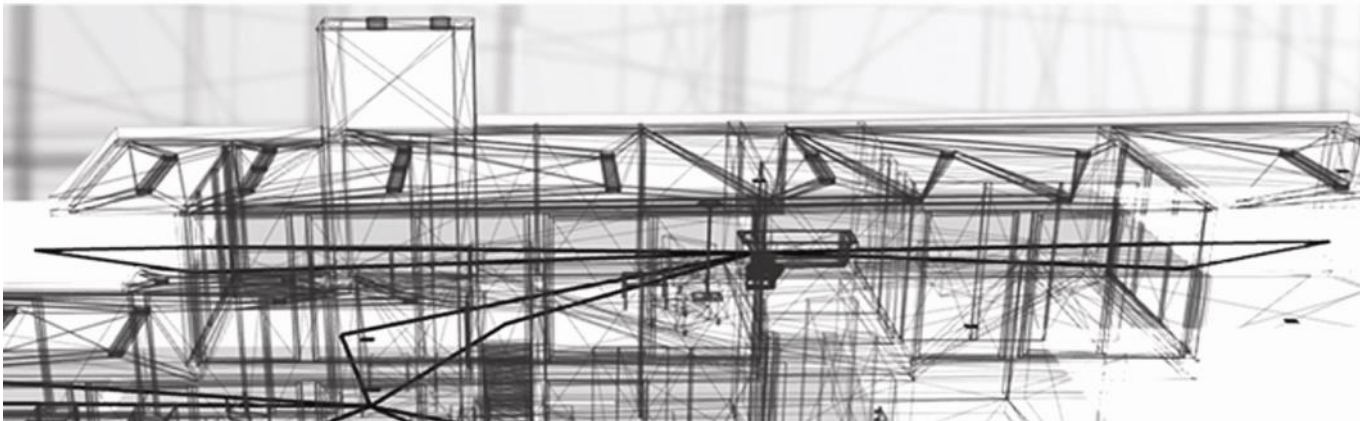
Mit den Menschen bauen

How have we been preparing?

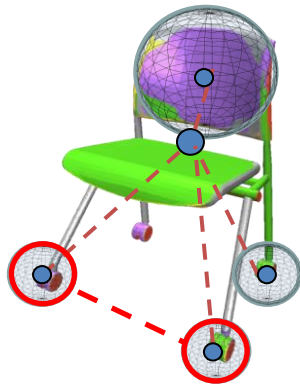
- Infrastructure for research data
- Visual search tools for AV-media
- ***3D Objects***
- chemOCR

3D Objects – an excursion to Architecture

PROBADO

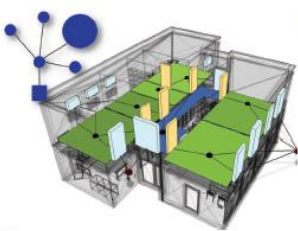


Content based indexing

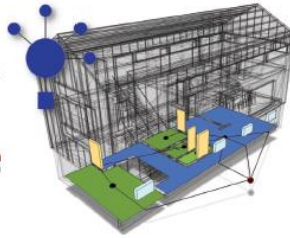


*segmentation with
form-primitives*

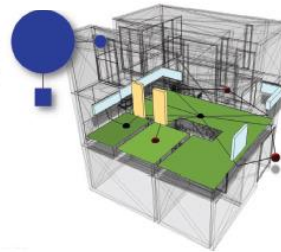
*extraction of
room connectivity
graphs*



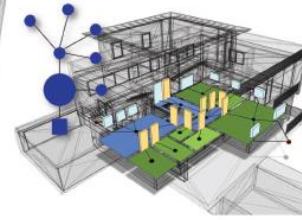
(a) corridor type



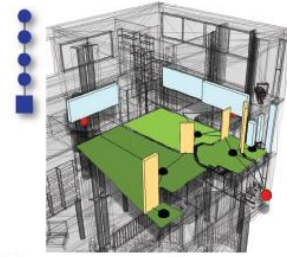
(b) distributor type



(c) loft type

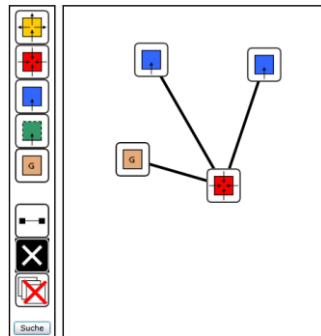


(d) hierarchical type

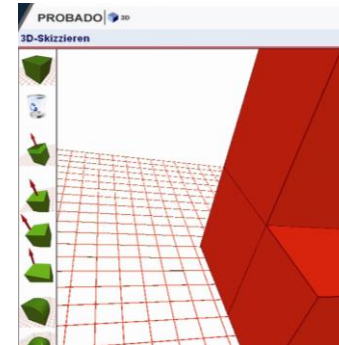


(e) chain type

Visual search









attributed graph



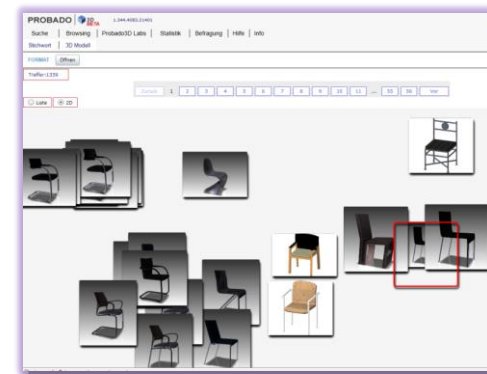
3D sketch



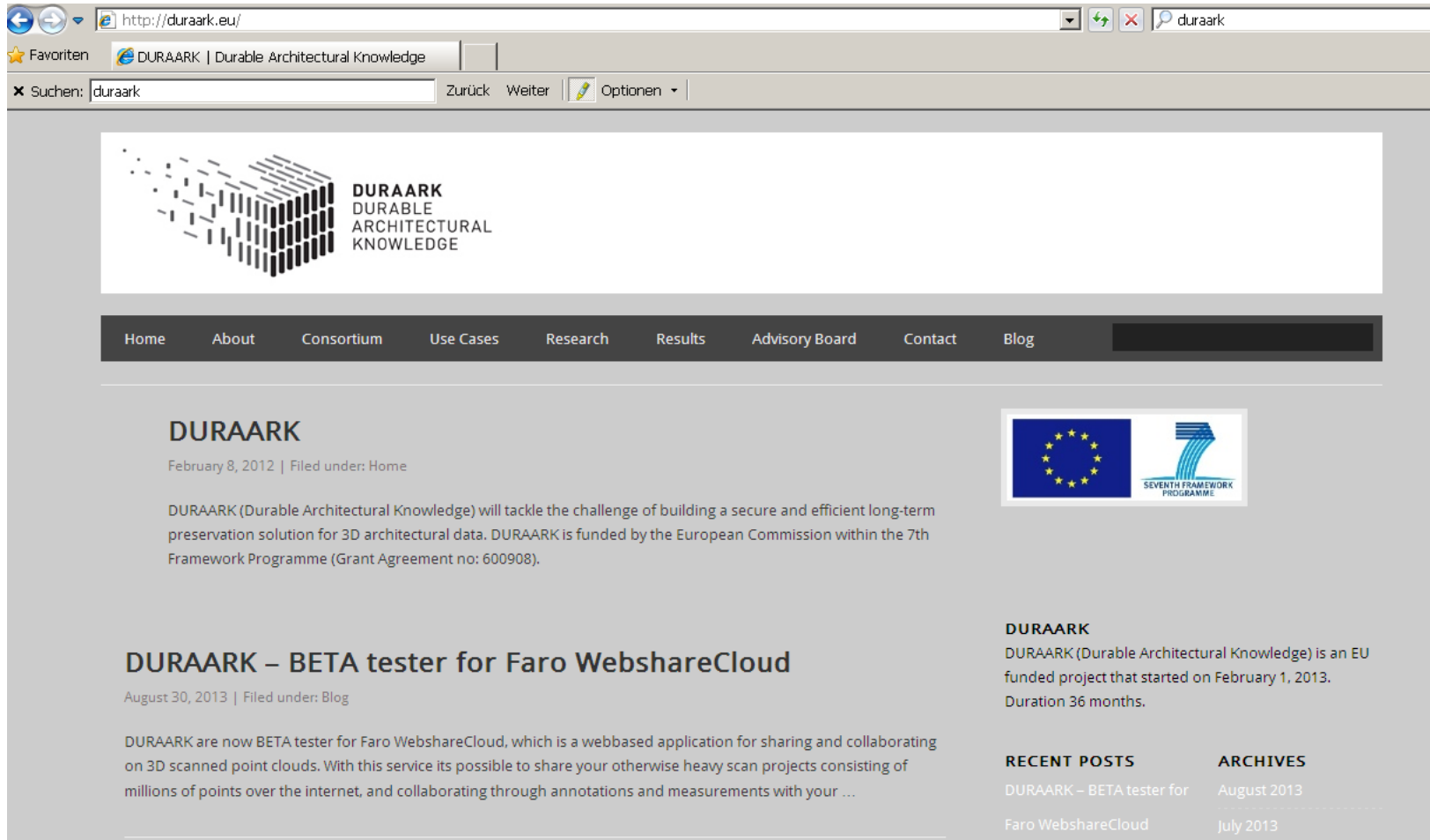
Treffer: 6 angezeigte Treffer/Seite: 24

	ModelInfoID: 238 Van Basshuysen House, West Pennant - Download: FULL
	ModelInfoID: 236 Modernes Atriumhaus in Trier - Download: FULL
	ModelInfoID: 242 Magnolia Row - Download: FULL
	ModelInfoID: 244 House in Minato, Tokyo - Download: FULL
	ModelInfoID: 239 Haus Mueller in Staufen - Download: FULL
	ModelInfoID: 231 Y House

result
visualization



Further developments



The screenshot shows a web browser window with the address bar displaying 'http://duraark.eu/'. The page features a header with the DURAARK logo (a 3D grid of vertical bars) and the text 'DURAARK DURABLE ARCHITECTURAL KNOWLEDGE'. Below the header is a navigation menu with links: Home, About, Consortium, Use Cases, Research, Results, Advisory Board, Contact, and Blog. The main content area contains three sections: 1) A post titled 'DURAARK' dated February 8, 2012, filed under Home, with a description of the project's goal and funding. 2) A post titled 'DURAARK – BETA tester for Faro WebshareCloud' dated August 30, 2013, filed under Blog, with a description of the service. 3) A sidebar on the right containing the European Union flag and the Seventh Framework Programme logo, followed by a section titled 'DURAARK' with a description of the project, and two columns: 'RECENT POSTS' and 'ARCHIVES'.

DURAARK
February 8, 2012 | Filed under: Home

DURAARK (Durable Architectural Knowledge) will tackle the challenge of building a secure and efficient long-term preservation solution for 3D architectural data. DURAARK is funded by the European Commission within the 7th Framework Programme (Grant Agreement no: 600908).

DURAARK – BETA tester for Faro WebshareCloud
August 30, 2013 | Filed under: Blog

DURAARK are now BETA tester for Faro WebshareCloud, which is a webbased application for sharing and collaborating on 3D scanned point clouds. With this service its possible to share your otherwise heavy scan projects consisting of millions of points over the internet, and collaborating through annotations and measurements with your ...

DURAARK
DURAARK (Durable Architectural Knowledge) is an EU funded project that started on February 1, 2013. Duration 36 months.

RECENT POSTS
DURAARK – BETA tester for Faro WebshareCloud

ARCHIVES
August 2013
July 2013

How have we been preparing ?

- Infrastructure for research data
- Visual search tools for AV-media
- 3D Objects
- ***chemOCR***

Information retrieval in Chemistry

Search for chemical structures – how?

Chemists are used to drawing

?

Google™
Deutschland

Suche: Das Web Seiten auf Deutsch Seiten aus Deutschland

[Erweiterte Suche](#)
[Einstellungen](#)
[Sprachtools](#)

Google-Suche Auf gut Glück!

[Werben mit Google](#) - [Unternehmensangebote](#) - [Über Google](#) - [Google.com in English](#)

[Machen Sie Google zu Ihrer Startseite!](#)

©2008 - [Datenschutz](#)

Textual and non-textual chemical information

Chemical Names

The copper(I)-catalyzed 1,3-dipolar cycloaddition [33-38] of organic azides and alkynes (also called “click chemistry”) resulting in the formation of 1,2,3-triazoles has become an increasingly attractive area [39]. According to the literature [33-38], the Cu(I) species can be used directly (e.g. CuI), or generated by oxidation of a Cu(0) or reduction of a Cu(II) species. Catalysis by the CuI is known to yield exclusively the 1,4-disubstituted regioisomer [33,34]. First, the *N*-(*p*-methoxyphenyl)-1-(trifluoromethyl)propargylamine was reacted with benzyl azide in the presence of CuI (10 mol%) and showed good selectivity with completion of the reaction within 24 h, whereas the use of CuSO₄/Na ascorbate afforded the cycloadduct in low yield. The reaction was then carried out with different propargylamines (*N*-(*p*-methoxyphenyl) and *N*-benzyl) and various azides at room temperature in acetonitrile within 24 h which afforded the compounds 2a-i with good yields (63-92%) after purification by column chromatography. The results are summarized in Table 1.

Linked entities from the table

As expected the new triazoles were formed in a fully regioselective manner affording the 1,4-regioisomer as highlighted from NOE experiments on compound 2c (Figure 1). A strong correlation was observed between the hydrogen H_a and H_b respectively. The structure of the other compounds 2a-i was assigned by analogy with 2c.

In our goal to study the influence of the CF₃ group on the conformation of peptidomimetics, we applied our strategy to the enantiopure trifluoromethyl-propargylamine **3** bearing the removable (*R*)-phenylglycyl chiral auxiliary (Scheme 2) [30-32].

The reaction was carried out under the same condition with azidoacetic acid methyl ester and afforded the cycloadduct **4** in

Reaction scheme

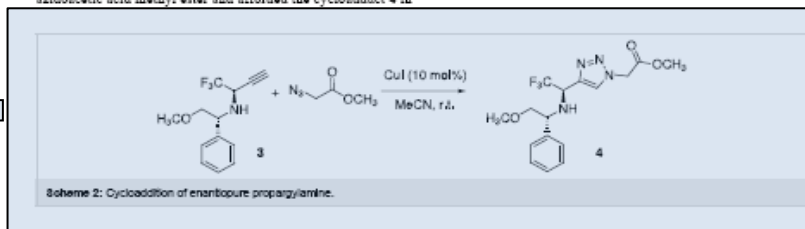


Table 1: Copper(I)-catalyzed synthesis of 1,4-disubstituted triazoles

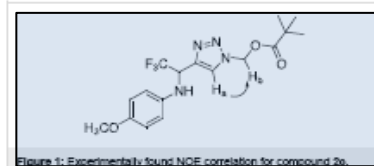
Entry	R ₁	R ₂	Product	Yield (%) ^a
1	-PMP ^a	-Bn	2a	82
2	-PMP	-CH ₂ COPh	2b	76
3	-PMP	-CH ₂ COCC(CH ₃) ₃	2c	73
4	-PMP	-CH ₂ CO ₂ CH ₃	2d	83
5	-PMP	-CH ₂ CH ₂ OH	2e	87
6	-Bn ^a	-Bn	2f	78
7	-Bn	-CH ₂ COPh	2g	63
8	-Bn	-CH ₂ CO ₂ CH ₃	2h	92
9	-Bn	-(CH ₂) ₂ OH	2i	73

^aPMP: *p*-methoxyphenyl, Bn: benzyl. ^bYield after flash purification.

Table with reaction scheme

2a-i: Derivates from the reaction

Chemical structure



good yield (79%) and as a single isomer without any racemization. This compound can easily afford the free amino ester which is a promising trifluoromethyl building block for the synthesis of new triazole-based trifluoromethyl oligomers.

Non-textual data processing – chemOCR

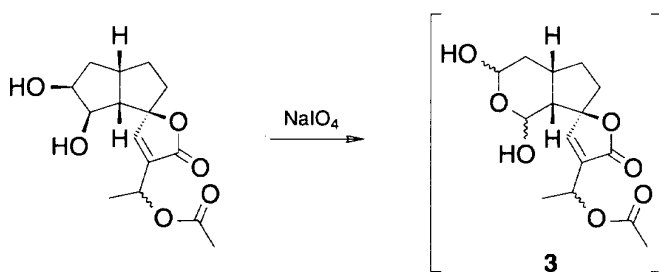
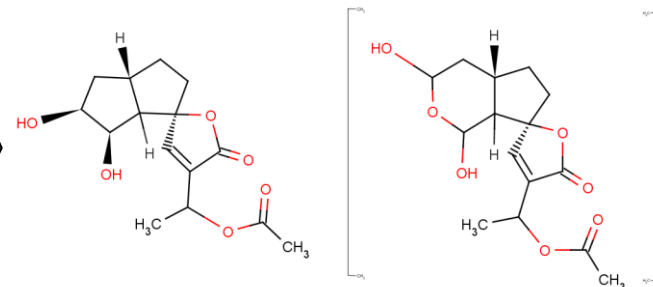


image data



chemical structure data

CLiDE



chemOCR



Fraunhofer
SCAI

Information retrieval in chemistry

Text AND formulas

Pattern formation on t

Matthias Bertram, Carsten E
Fritzsche

Experiments with
system can be supp
patterns, seen near t
include intermittent
method based on the
reconstructed from t

DOI: 10.1103/PhysR

L INTRODU

Control of chaos is one of the
ear dynamics. In contrast to ex
heuristic approach based on the i
backs does not require extensive
For extended systems, where sp
ficult, global delayed feedbacks
methods, information continuo
ments of the system is summed
signal which is applied back to
parameter. Global feedbacks ca
otherwise unstable trajectories. I
new spatiotemporal patterns. Ac
chaotic extended systems has b
retically investigated for lasers [;
semiconductors [5,6], populatio
tors [7], and surface chemical r
cussed in the more general cou
Landau equation [9,10]. Furth
global feedback have been suc
pattern formation in nonchaotic
latory [13-17] chemical system

Nonequilibrium systems on t
of generating a broad variety of
system to a boundary between
ics, its parameters may be app
forcing may be introduced. How
tion of such predefined contr
because a system at the edge o
small parameter variations. An
ing global delayed feedback. Th
that an acting force is generat
therefore automatically adjust
mental conditions.

In this paper, we apply a glo
vestigate spatiotemporal pattern
to chaos in an oscillatory surfac
onstrate that chemical turbule
means of such a feedback and t
and nonchaotic patterns can be induced. The system we consi
der is the catalytic oxidation of carbon monoxide on
Pt(110), representing the best studied example among oscil

Home

Order without search

Collective search

Subject search

- ▶ Architecture
- ▶ Chemistry
- ▶ Computer Science
- ▶ Mathematics
- ▶ Physics
- ▶ Engineering

TIB-Catalogue Search

Watch list

Pricing

MyGetInfo

Register

Guided Tour

About GetInfo

News

Newsletter

Customer Service

GetInfo

FIND THE WORLD OF
SCIENCE AND TECHNOLOGY

GetInfo - Get ahead and
Fast search engine technology

Collective search

Interdisciplinary search in selected

GetInfo

Example: (gear* OR Getriebe*) AND Hain

Subject search

Architecture
Chemistry
Computer Science
Mathematics
Physics
Engineering

Search in selected subject specific databases.

Engineering

Example: (gear* OR Getriebe*) AND Hain

TIB-Catalogue Search

Search in the TIB's stocks - the largest specialist subject library for science and technology in the world.

**GERMAN NATIONAL LIBRARY OF
SCIENCE AND TECHNOLOGY**

German National Library for all areas of
engineering as well as architecture,
chemistry, information technology,
mathematics and physics.

Full text delivery

Contact ▶ **Deutsch**

Advanced Search
Database selection

Advanced Search
Database selection

f heterogeneous catalysis : 8 volumes / ed. by Gerhard Ertl ...; Vol. 4

Further subjects

- Open Science Lab
- Ontology



HOME CONTACT IMPRINT

TIB | OPEN SCIENCE LAB

TIB's Open Science Lab

DEUTSCH | ENGLISH

OPEN SCIENCE IM TIB | BLOG

VivoCamp13, das Barcamp über Linked Open Data für Forschungsinformationen, findet am 25.11.2013 in Hamburg statt

TIB's Open Science Lab experiments with VIVO — to elucidate networks in expert scientific communities

Big data, big science, common infrastructures: e-IRG, think-tank of European eScience service providers, speaks out

Big Data, Big Science, gemeinsame Infrastrukturen: e-IRG, der Think Tank europäischer eScience-Dienstleister, meldet sich zu Wort

TIB Open Science Lab experimentiert mit VIVO — Ziel ist ein freier Blick auf die Netzwerke in wissenschaftlichen Fachcommunities

The Open Science Lab, established in 2013, is a team of experts at the [German National Library of Science and Technology](#) (TIB) who test and refine web applications for scientists. The internet influences the way in which researchers work. Today, e-mails, Google and PDF files are taken for granted, any yet developments will not stop there:

 **Social software:** Tools such as [blogs](#), RSS, Twitter and [social network services](#) enable us to communicate with one another quickly and more effectively. It is easier to receive feedback about provisional results from peers. Or, seen from the other perspective, it is easier to look over the shoulders of researchers working on their current projects. Traditional research publications are often [discussed](#) and [criticised](#) in blogs. ([Post publication peer review.](#))

 **Open Collaboration:** Collaborative platforms such as [Github](#), wikis (e.g. [OpenWetWare](#)) and Stack Exchange (e.g. [MathOverflow](#)) enable users to contribute quickly and easily to projects of third parties. Despite sometimes being only minor contributions, their authorship and significance for the final product can be presented with precision. As with software, projects “live” on such platforms: they can be refined to make new versions or copied at any time, e.g. for use in other projects.

Conclusion

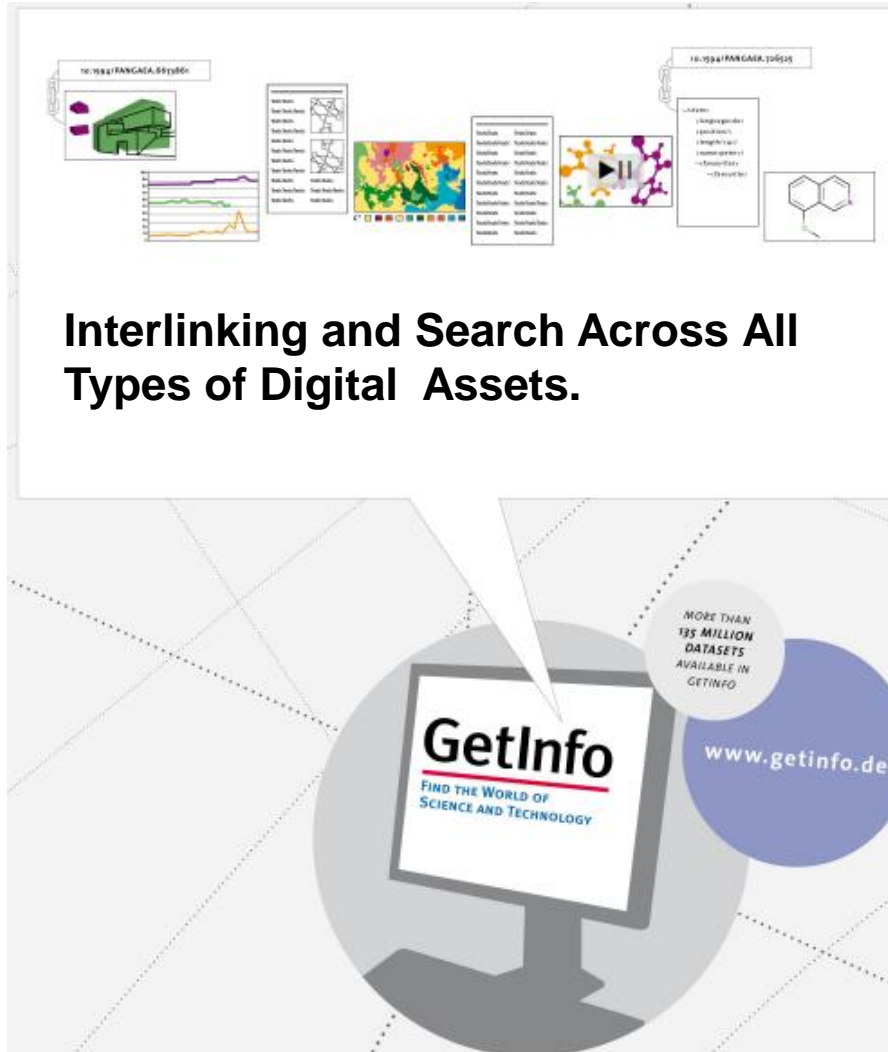
Dissemination of scientific and technical information has been a foundational mission.

The methods have completely changed, but the mission remains the same.

Conclusion

Ultimate Goal: →

Interlinking and Search Across All Types of Digital Assets.



GetInfo – Portal for Science and Technology

- 58 m metadata in internal index
- 390 m metadata in external sources
- 900.000 pdf fulltexts
- Data, AV-Media, 3D Objects

GetInfo

FIND THE WORLD OF
SCIENCE AND TECHNOLOGY

GetInfo

FIND THE WORLD OF
SCIENCE AND TECHNOLOGY

10.1 Information Gain

- Attribute selection by gain computation, example
 - Class P: buys_computer = "yes"
 - Class N: buys_computer = "no"

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

$$-I(p, n) = I(9, 5) = 0.94$$

price	quality	avail	credit	buys_computer
<=30	high	no	fair	no
<=30	high	no	fair	yes
>=40	medium	no	excellent	yes
>=40	low	yes	excellent	yes
>=40	low	yes	fair	no
31..40	low	yes	excellent	yes
<=20	medium	no	fair	no
<=20	low	yes	fair	yes
>=40	medium	yes	excellent	yes
>=20	medium	yes	excellent	yes
31..40	medium	no	excellent	yes
31..40	high	yes	fair	yes
>=40	medium	no	fair	no

QUALITY: HIGH / LOW DOWNLOAD: HIGH / LOW

Portal for audiovisual Media

BEREITSTELLE INHALTSBASIERTE ERSCHEISSUNG

VISUELLE SUCHE

Probado 3D

A photograph of a surfer riding a large, curling blue wave. The surfer is positioned in the lower center of the frame, leaning forward. The wave's crest is on the left, and the water is a vibrant blue. The sky is a pale, overcast blue.

A SURFBOARD FOR RIDING THE WAVE

TOWARDS A FOUR COUNTRY ACTION PROGRAMME
ON RESEARCH DATA

Questions?