TIB | Technische Informationsbibliothek

New developments in digital library services: textual and non-textual information in a science portal

Uwe Rosemann
Tokyo
March 8, 2011

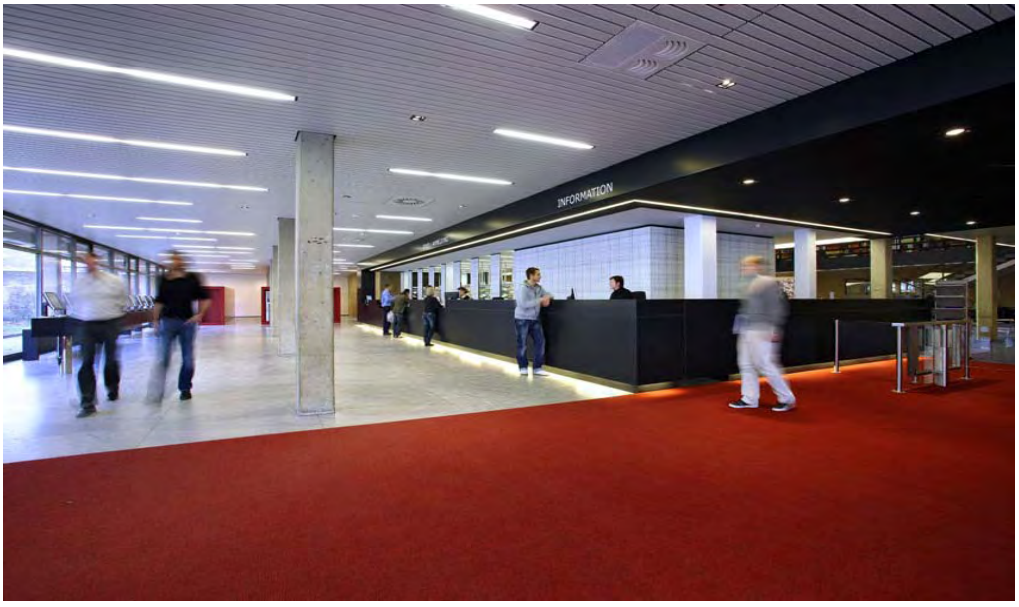
TIB Hannover – Some of the facts

= **German National Library of Science and Technology**

- engineering, architecture, chemistry, information technology, mathematics and physics
- Founded in 1959
- Financed by Federal Government and all Federal States

Main Building

Entrance Hall



Reading Room

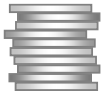
Marstall Building

Marstall Building – a former horse stable

Castle

Main Stacks

TIB Hannover – Some of the facts



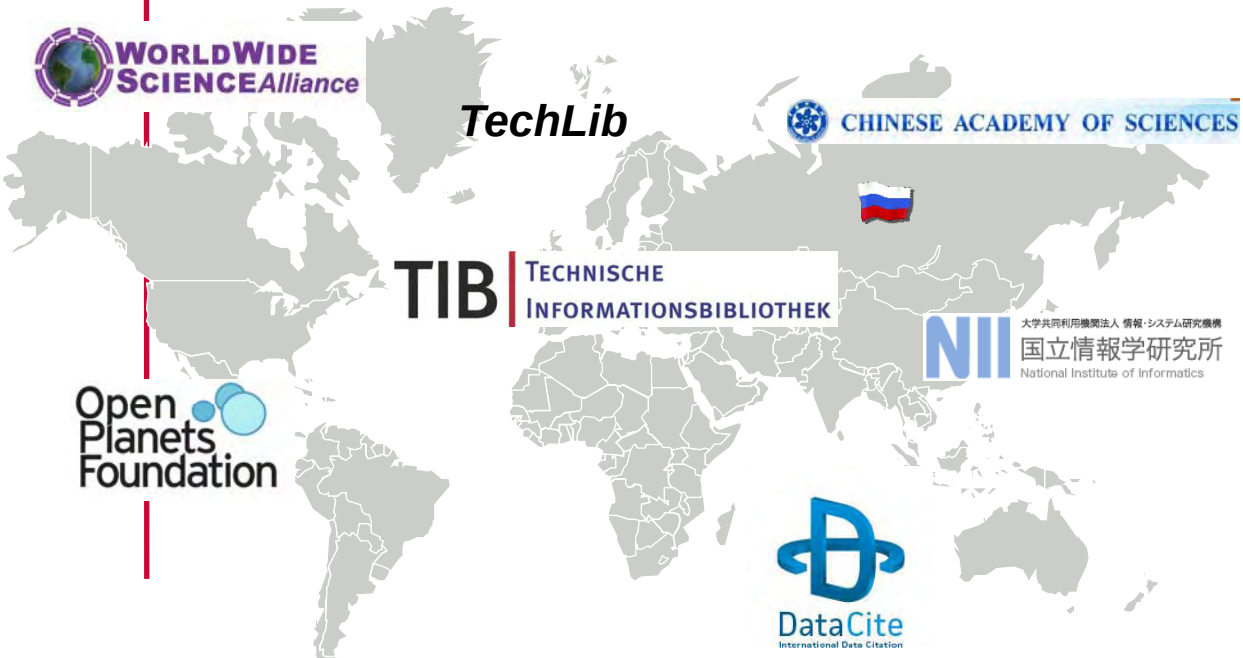
- € 8m annual acquisition budget

- 18,500 journal subscriptions
- 7m items

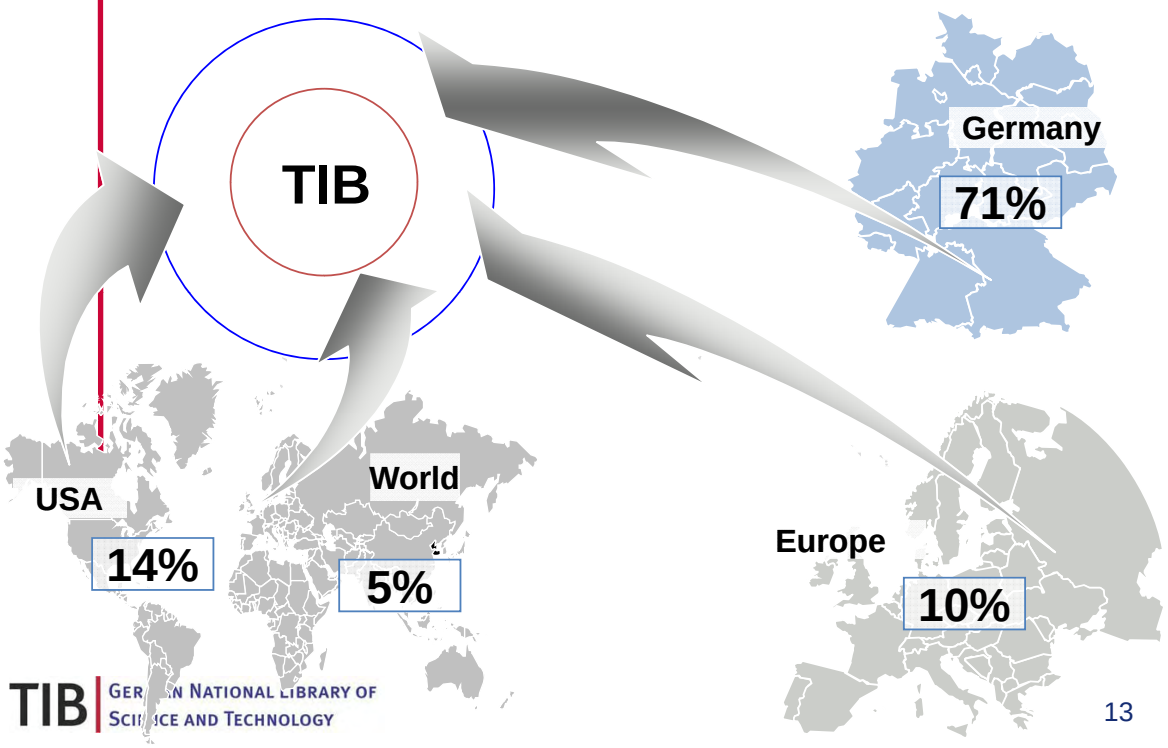
- Staff: ca. 175 FTE

Networks in Germany

International Networks

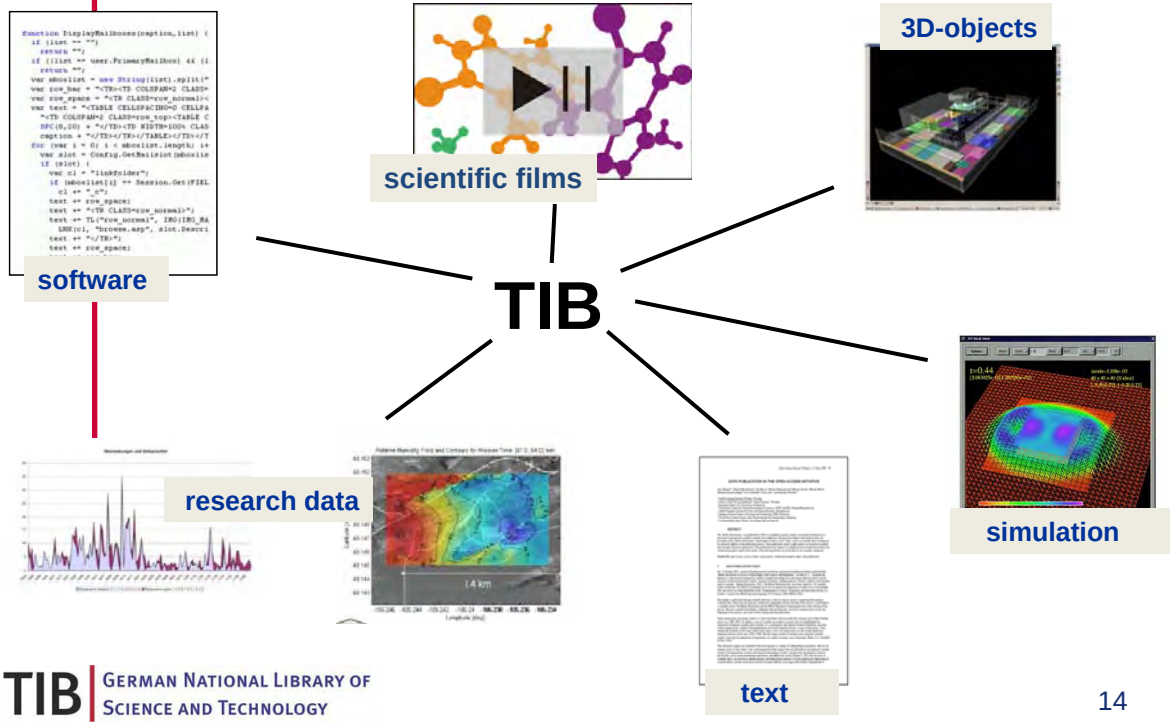


Customer worldwide



TIB | German National Library of Science and Technology

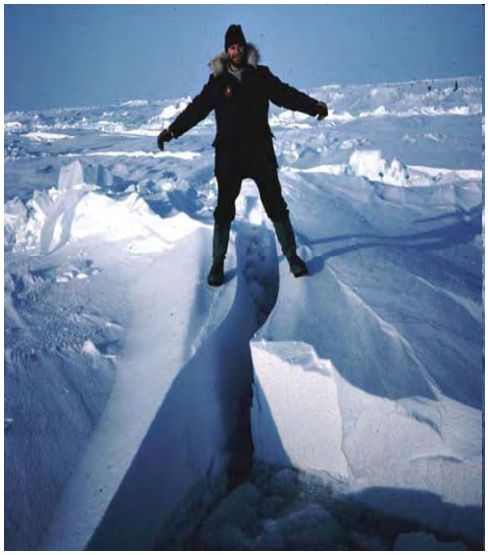
Vision and Strategy



TIB | German National Library of Science and Technology

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**



Competence Centre for non-textual materials (CNM)

The vision of the Competence Centre is to make efficient tools and technologies available to researchers, teachers and students, to enable them to search in a multimedia database through a central portal, which is characterised by extensive, cross-media content, simple and media-specific access and high hit rates.

CNM
Goal I

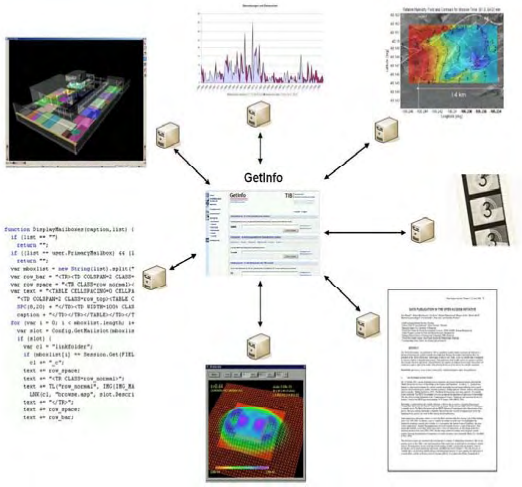
1. Establishing a media centre for multimedia
    - Collection, indexing, provision
    - Automatic content analysis
    - Digitalisation and archiving
    - Citability by allocating DOIs
    - Integration of suitable search and presentation processes
    - Advice from knowledge providers on questions of standardisation, archiving and provision

CNM
Goal II

2. Establishing an infrastructure for referencing research data
    - Collection and provision of metadata
    - Citability by allocating DOIs
    - Coordination of infrastructure for data storage
    - Development of a visual search process
    - Scientific advice

An Access System: GetInfo

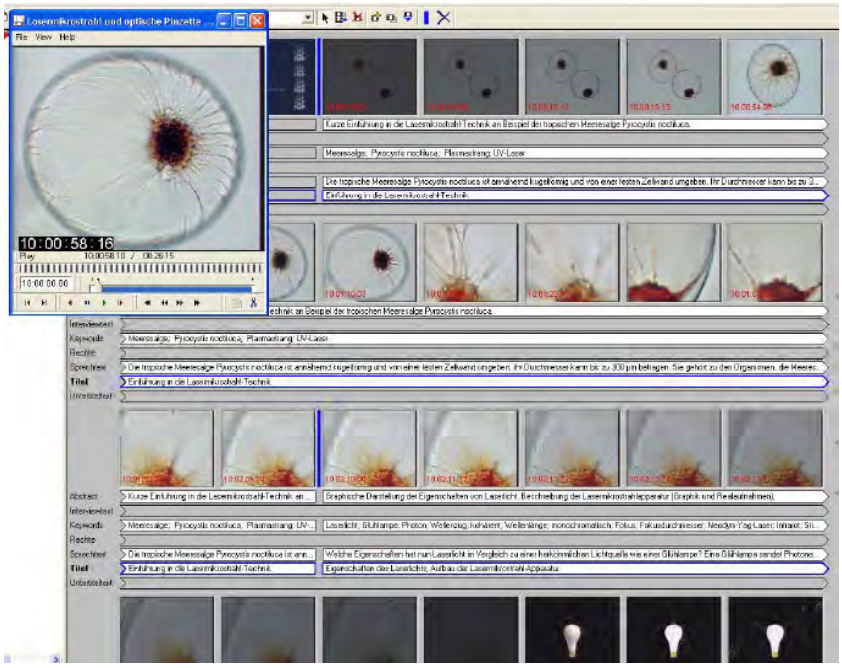
GetInfo



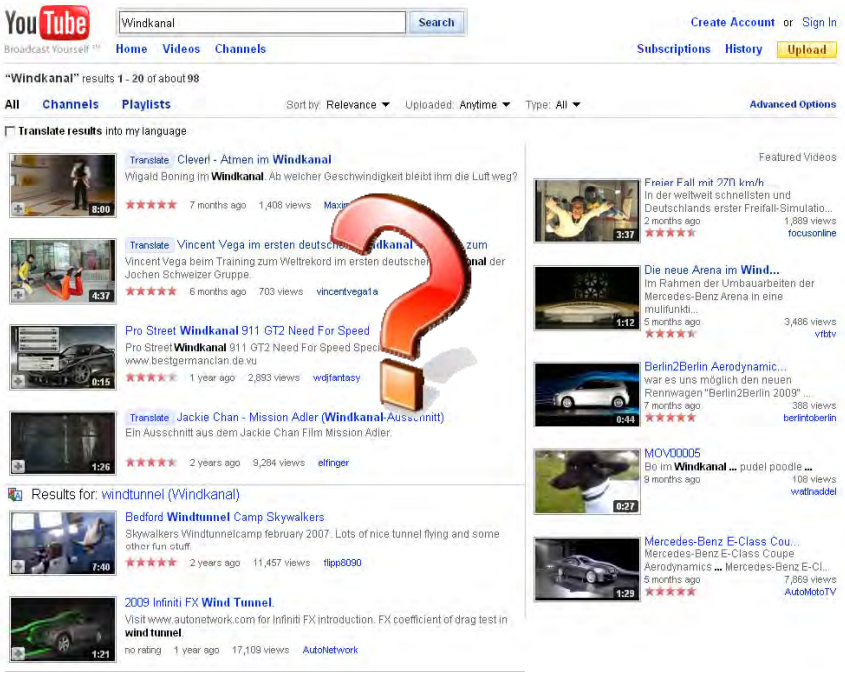
Starting Point

- Increase in data volumes
- Heterogeneity of media types and formats
- Interoperability of metadata
- Automatic indexing of metadata
- Search and presentation process for multimedia retrieval

Media Indexing



Media Provision



Archiving of Analogue Films

Analogue	
Source material	16 mm / 35 mm
Digitalisation	Digibeta (in the future JPEG 2000, file based)
Encoding	MPEG1, MPEG2, wma, flash …
Storage	Digibeta Master and Digibeta DUP remain in the media archive Original is in the Federal Archive

Archiving Video Material

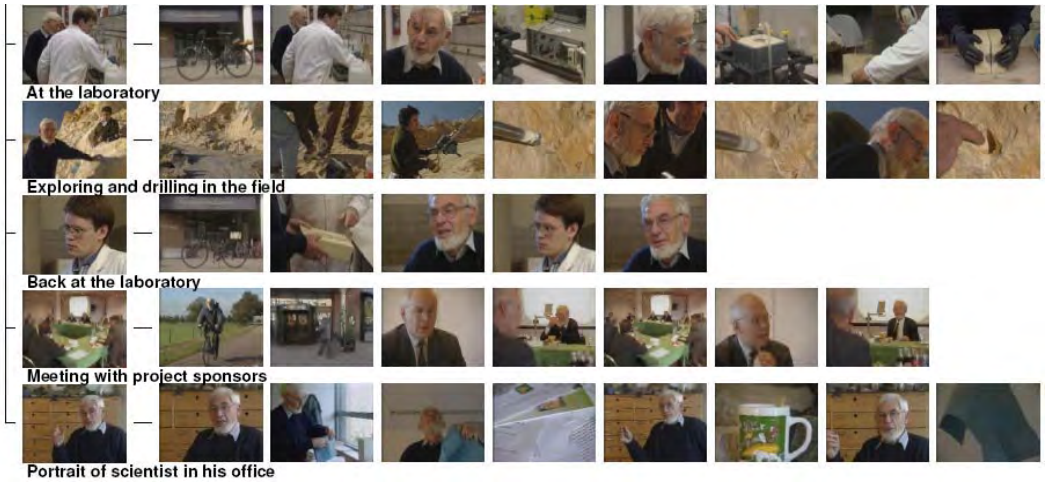
Digital	
Source material	SD > Digibeta, D1, D2, DV (Consumer), DVCam HD > Hdcam, HDCAM SR, HDV
Digitalisation	SD > Digibeta (in the future JPEG 2000) HD > HDCAM SR
Encoding	MPEG1, MPEG2, wma, flash …
Storage	Digibeta Master and Digibeta DUP remain in the media archive Regular copying onto the best possible quality carriers

Automatic Content Analysis

- Shot and scene detection
- Camera movement detection (tilt and zoom)
- Face detection / identification
- Object detection
- Context detection (indoor/outdoor, countryside, buildings)
- Genre detection (news, sport, advertising)
- Event analysis (tennis, football, etc.)
- Text recognition
- Speech recognition

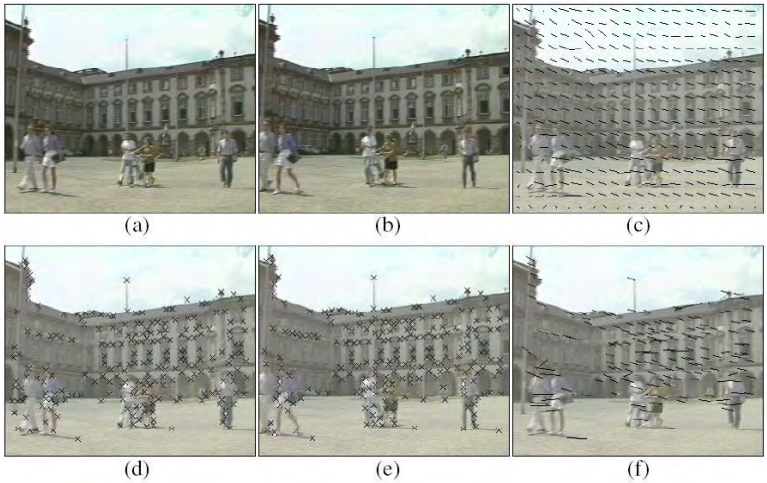
Automatic Content Analysis

- Scene recognition: A *scene* is a timed video segment with 3 consistent properties: Event, Location and Time period



Dr Patrick Ndjiki-Nya: Automatic video analysis for creation of metadata, http://www.cinearchiv.de

Automatic Content Analysis



Estimation of the movement vectors for two pictures (a) and (b) from a video sequence.

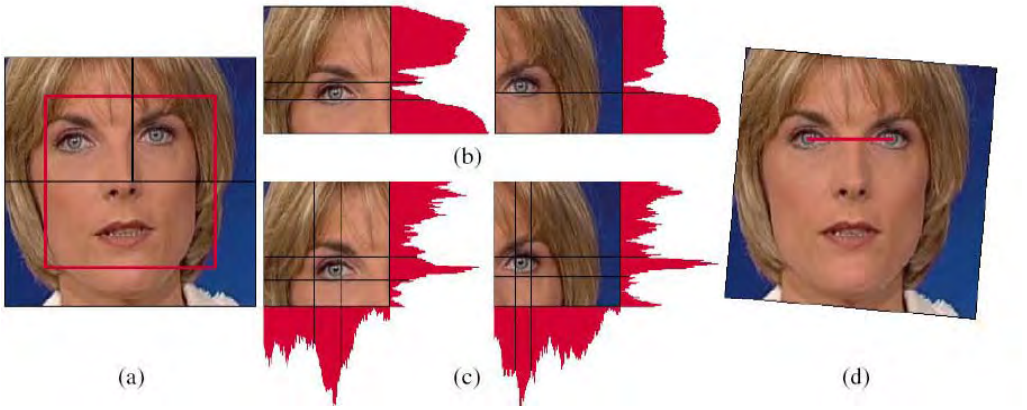
Kopf, S. Computer-aided content analysis of digital video archives, Mannheim. 2006

Automatic Content Analysis

Object recognition through motion analysis

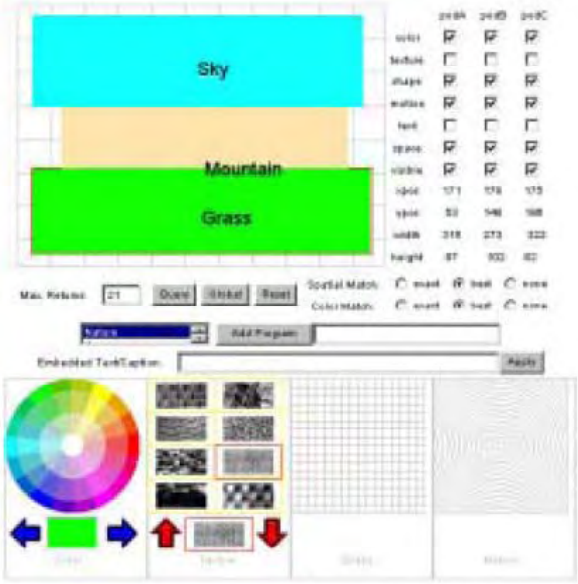
Kopf, S. Computer-aided content analysis of digital video archives, Mannheim. 2006

Automatic Content Analysis

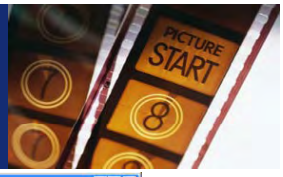


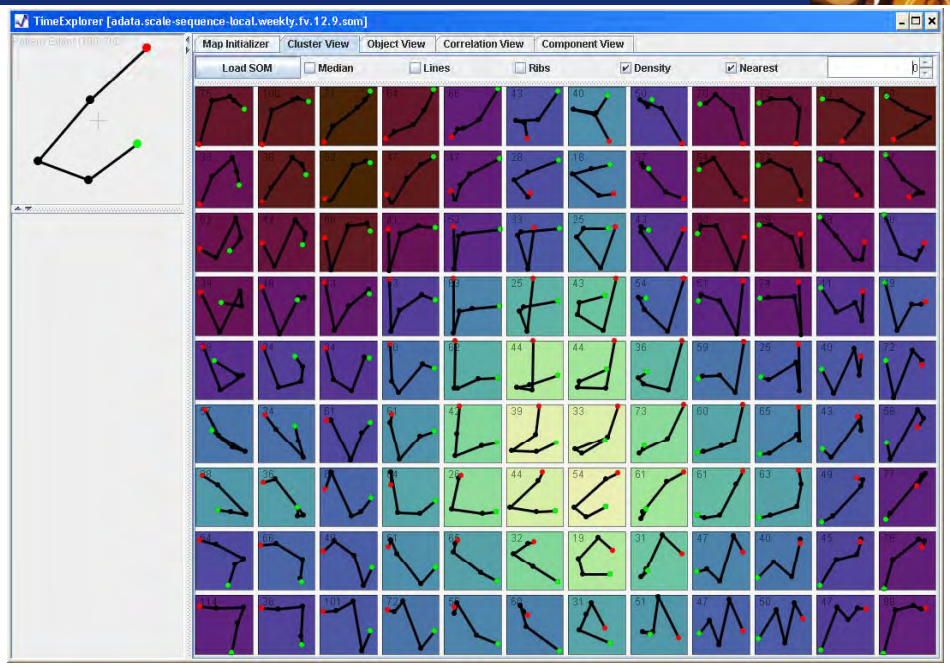
Kopf, S. Computer-aided content analysis of digital video archives, Mannheim. 2006

Research Topic: Search Process



Research Topic: Visualisation



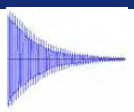


Research data

What is the general problem with data?

The Research Trajectory

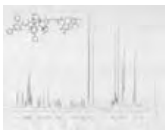
Data



... is lost!

analysed
interpreted

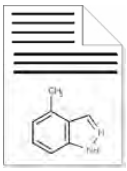
Information



... is traceable

published

Knowledge



... is accessible

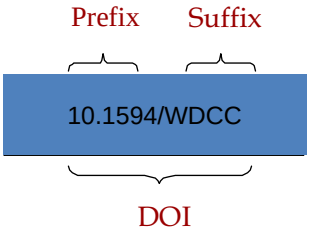
Publication

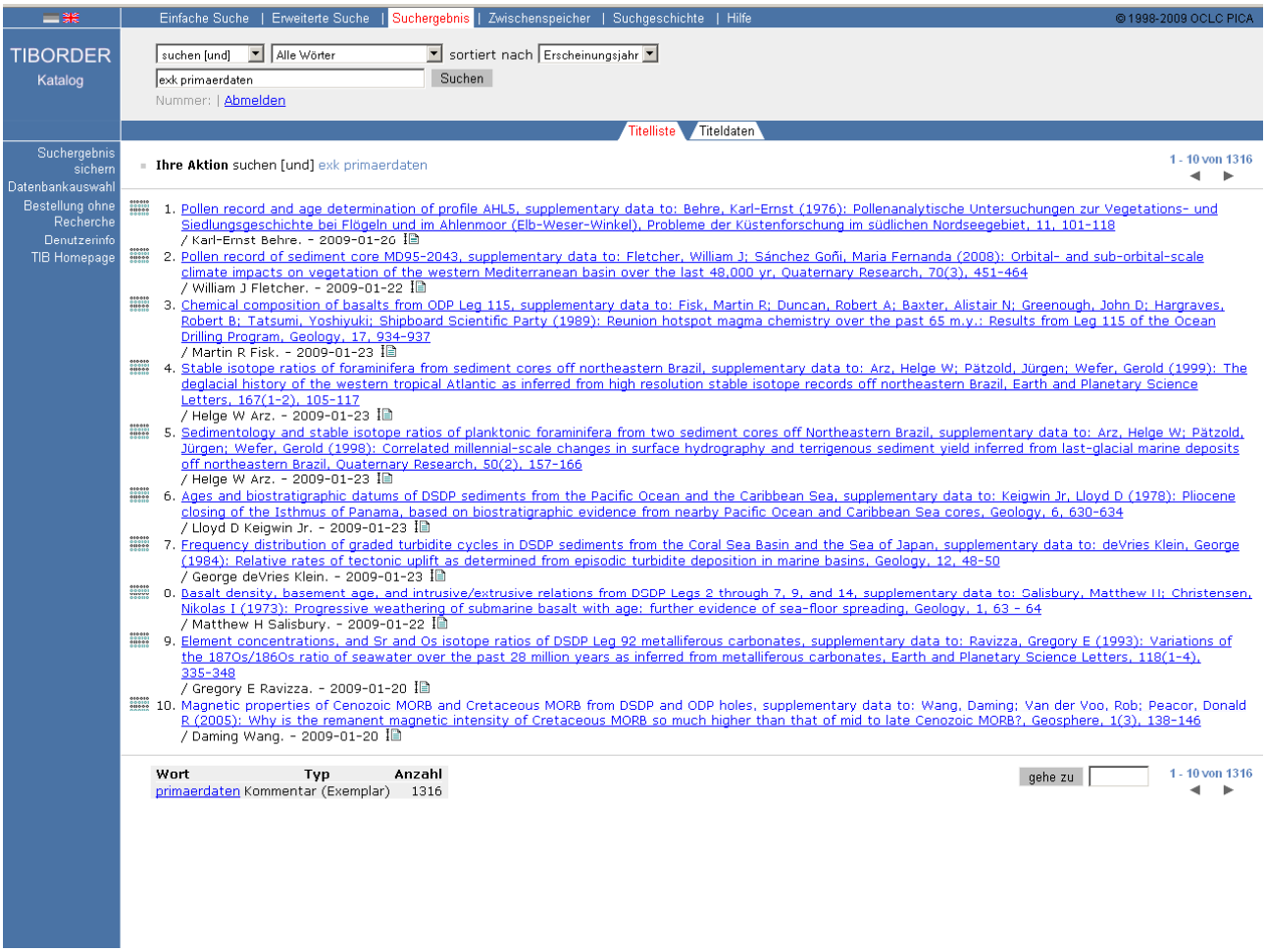
Solution

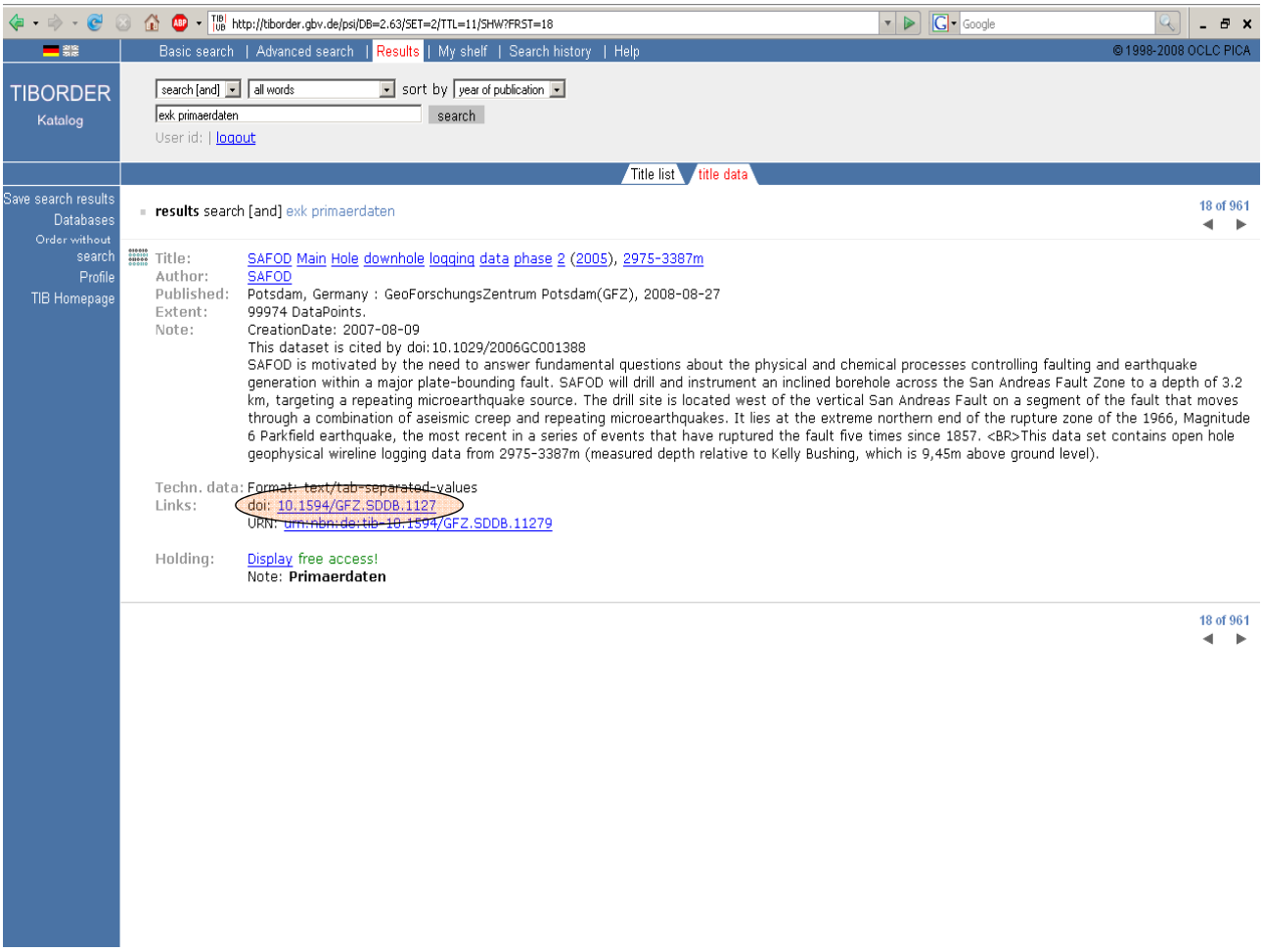
- Creation of new and strenghtening of existing data centres
- Global access to data sets and their metadata through existing catalogues
- By the use of persistent identifiers

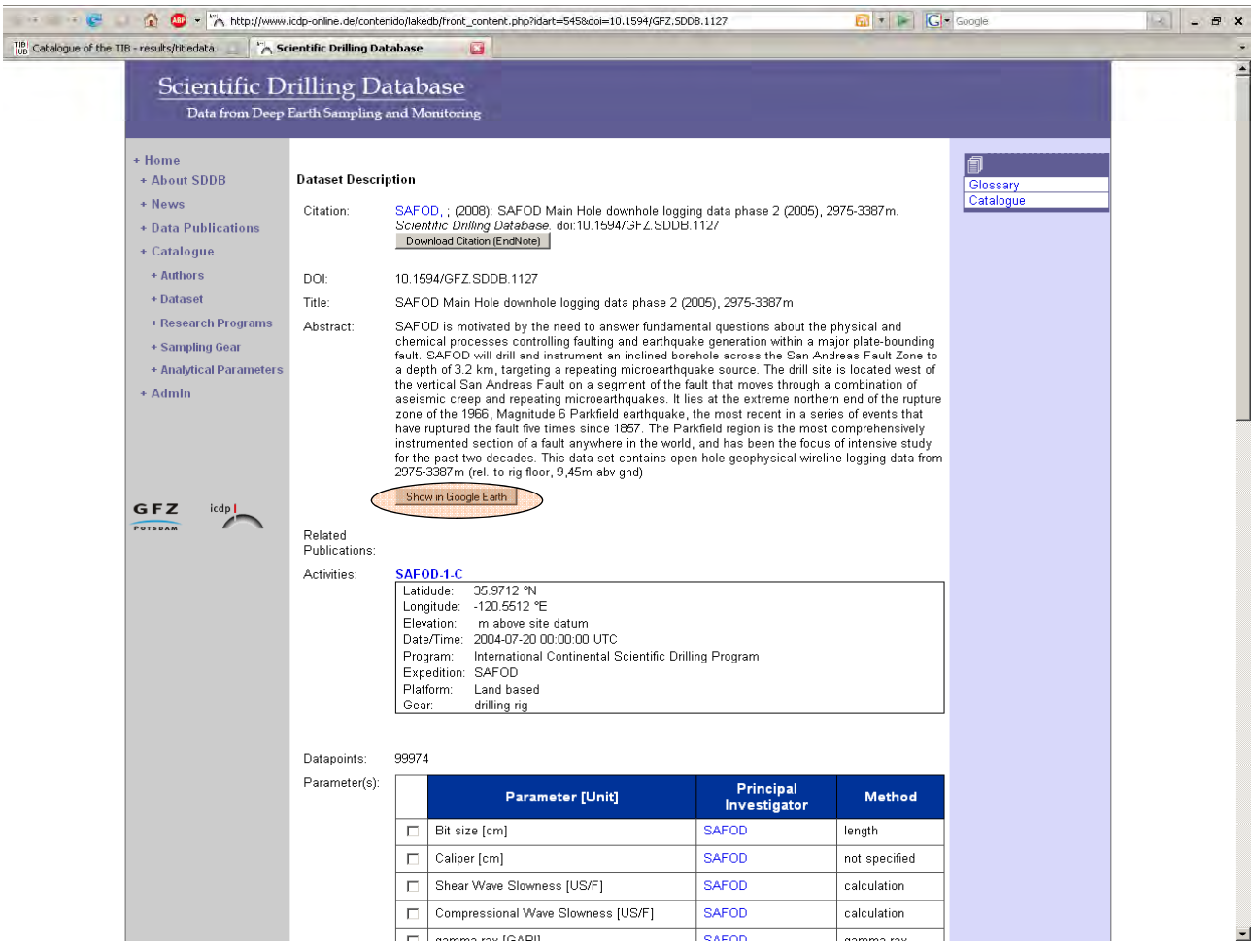
Digital Object Identifier (DOI)

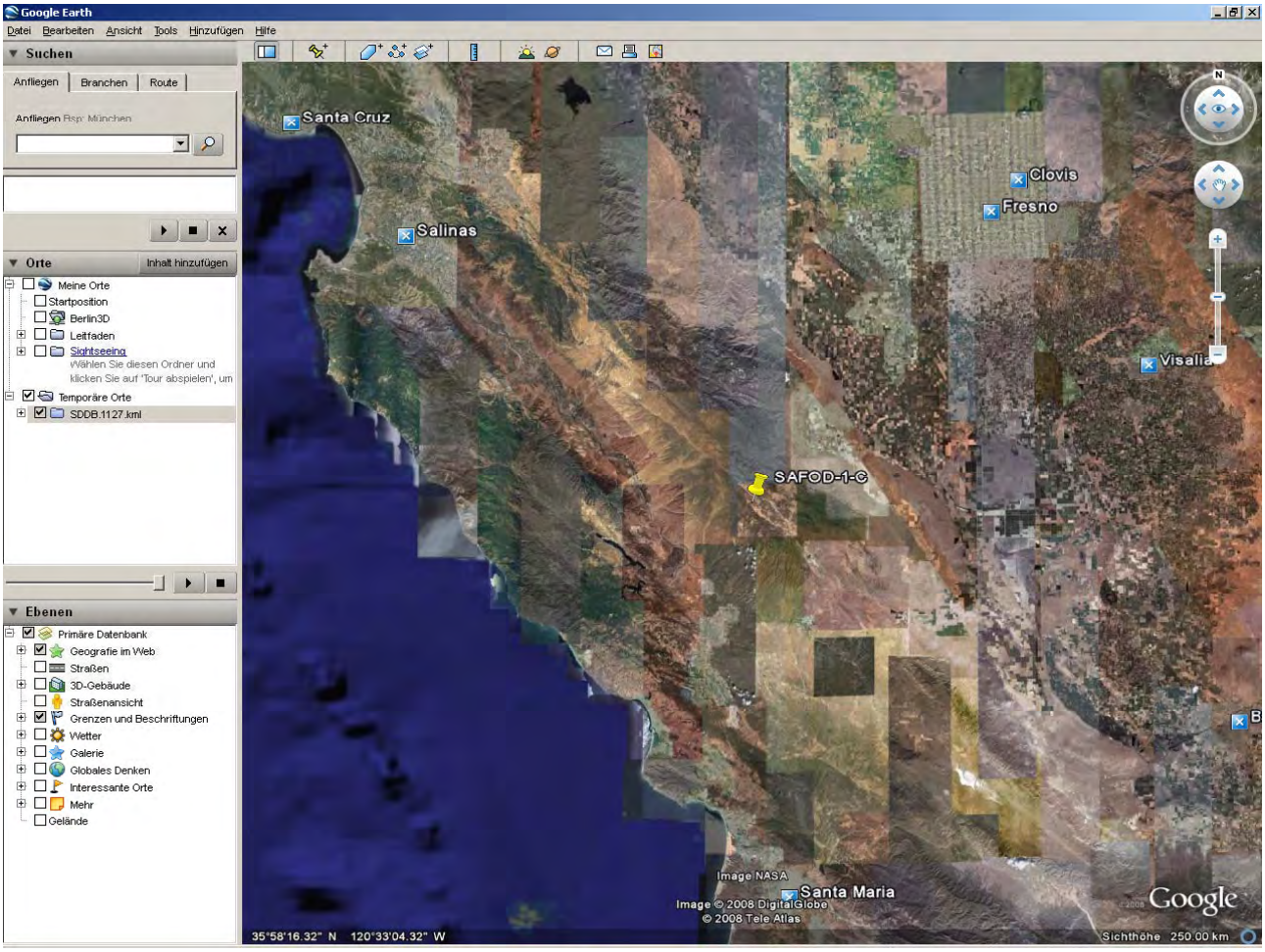
- The DOI system is a worldwide system for persistent and actionable identification and interoperable exchange of intellectual property on digital networks
- A DOI name is made up of two components, the prefix and the suffix






http://tiborder.gbv.de/psi/DB=2.63/SET=2/TTL=11/SHW?FRST=18
Basic search | Advanced search | Results | My shelf | Search history | Help
© 1998-2008 OCLC PICA
TIBORDER
Katalog
search [and]
all words
sort by
year of publication
exk primaerdaten
search
User id: | logout
Title list
title data
Save search results
Databases
Order without search
Profile
TIB Homepage
results search [and] exk primaerdaten
18 of 961
Title: SAFOD Main Hole downhole logging data phase 2 (2005), 2975-3387m
Author: SAFOD
Published: Potsdam, Germany : GeoForschungsZentrum Potsdam(GFZ), 2008-08-27
Extent: 99974 DataPoints.
Note: CreationDate: 2007-08-09
This dataset is cited by doi:10.1029/2006GC001388
SAFOD is motivated by the need to answer fundamental questions about the physical and chemical processes controlling faulting and earthquake generation within a major plate-bounding fault. SAFOD will drill and instrument an inclined borehole across the San Andreas Fault Zone to a depth of 3.2 km, targeting a repeating microearthquake source. The drill site is located west of the vertical San Andreas Fault on a segment of the fault that moves through a combination of aseismic creep and repeating microearthquakes. It lies at the extreme northern end of the rupture zone of the 1966, Magnitude 6 Parkfield earthquake, the most recent in a series of events that have ruptured the fault five times since 1857.
This data set contains open hole geophysical wireline logging data from 2975-3387m (measured depth relative to Kelly Bushing, which is 9,45m above ground level).
Techn. data: Format: text/tab-separated-values
Links: doi: 10.1594/GFZ.SDDB.1127
URN: urn:nbn:de:tib-10.1594/GFZ.SDDB.11279
Holding: Display free access!
Note: Primaerdaten
18 of 961


http://www.icdp-online.de/contenido/lakedb/front_content.php?idart=545&doi=10.1594/GFZ.SDDB.1127
Catalogue of the TIB - results/titledata
Scientific Drilling Database
Scientific Drilling Database
Data from Deep Earth Sampling and Monitoring
+ Home
+ About SDDB
+ News
+ Data Publications
+ Catalogue
+ Authors
+ Dataset
+ Research Programs
+ Sampling Gear
+ Analytical Parameters
+ Admin
Glossary
Catalogue
Dataset Description
Citation: SAFOD, ; (2008): SAFOD Main Hole downhole logging data phase 2 (2005), 2975-3387m. Scientific Drilling Database. doi:10.1594/GFZ.SDDB.1127
Download Citation (EndNote)
DOI: 10.1594/GFZ.SDDB.1127
Title: SAFOD Main Hole downhole logging data phase 2 (2005), 2975-3387m
Abstract: SAFOD is motivated by the need to answer fundamental questions about the physical and chemical processes controlling faulting and earthquake generation within a major plate-bounding fault. SAFOD will drill and instrument an inclined borehole across the San Andreas Fault Zone to a depth of 3.2 km, targeting a repeating microearthquake source. The drill site is located west of the vertical San Andreas Fault on a segment of the fault that moves through a combination of aseismic creep and repeating microearthquakes. It lies at the extreme northern end of the rupture zone of the 1966, Magnitude 6 Parkfield earthquake, the most recent in a series of events that have ruptured the fault five times since 1857. The Parkfield region is the most comprehensively instrumented section of a fault anywhere in the world, and has been the focus of intensive study for the past two decades. This data set contains open hole geophysical wireline logging data from 2975-3387m (rel. to rig floor, 9,45m abv gnd)
Show in Google Earth
Related Publications:
Activities: SAFOD-1-C
Latitude: 35.9712 °N
Longitude: -120.5512 °E
Elevation: m above site datum
Date/Time: 2004-07-20 00:00:00 UTC
Program: International Continental Scientific Drilling Program
Expedition: SAFOD
Platform: Land based
Gear: drilling rig
Datapoints: 99974
Parameter(s):
Parameter [Unit]
Principal Investigator
Method
Bit size [cm] SAFOD length
Caliper [cm] SAFOD not specified
Shear Wave Slowness [US/F] SAFOD calculation
Compressional Wave Slowness [US/F] SAFOD calculation



☐	gamma ray [GAPI]	SAFOD	gamma ray logging
☐	Relative Bearing [°]	SAFOD	not specified
☐	Caliper from x axis of x-y Caliper(s) [cm]	SAFOD	not specified
☐	Cablehead Tension [N]	SAFOD	not specified
☐	Field Normalized Compensated Neutron Porosity [PU]	SAFOD	calculation
☐	gamma ray [GAPI]	SAFOD	gamma ray logging
☐	Potassium [%]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 10 inch [Ohmm]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 20 inch [Ohmm]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 30 inch [Ohmm]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 60 inch [Ohmm]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 90 inch [Ohmm]	SAFOD	calculation
☐	Vert resolution matched (2 ft) res - DOI 120 inch [Ohmm]	SAFOD	calculation
☐	Porosity from ZDEN or ZDNC [PU]	SAFOD	not specified
☐	Spontaneous Potential Shifted [mV]	SAFOD	not specified
☐	Speed [m/s]	SAFOD	calculation
☐	Thorium [ppm]	SAFOD	calculation
☐	Uranium [ppm]	SAFOD	calculation
☐	ZDL correction [g/cm3]	SAFOD	calculation
☐	ZDL bulk density [g/cm3]	SAFOD	not specified

Separator: Comma

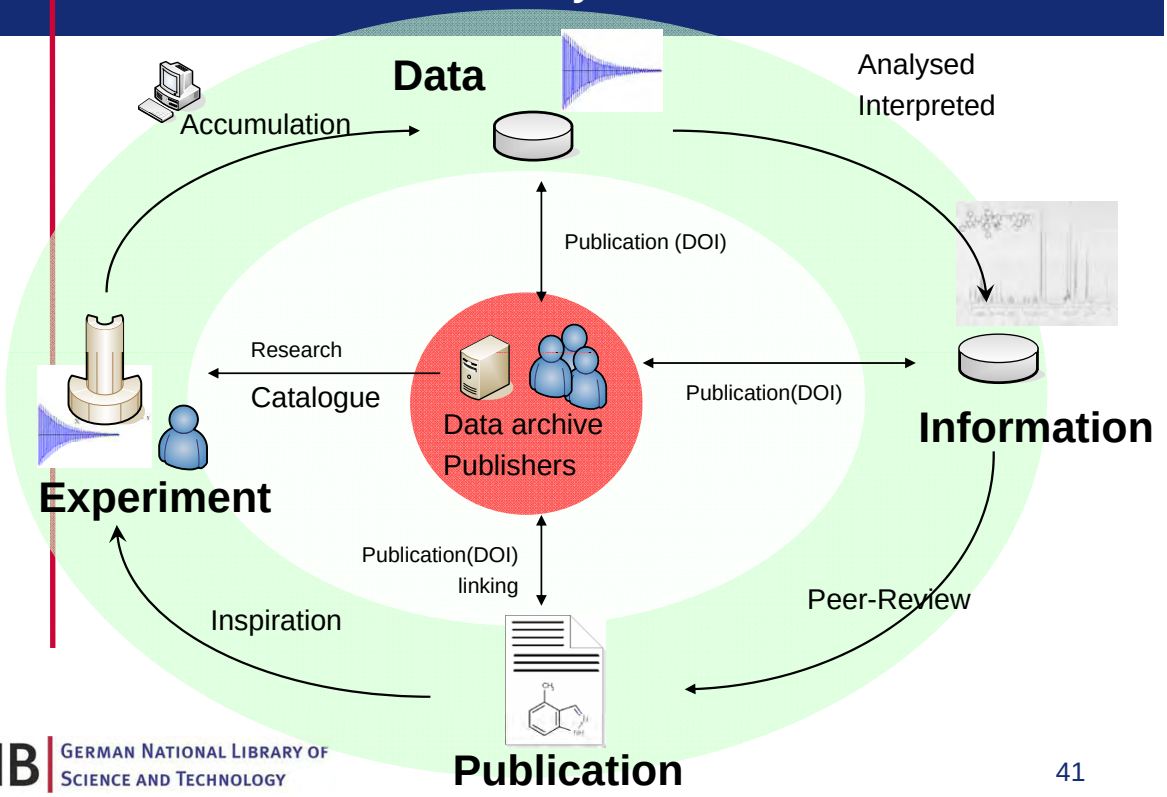
This dataset is under the licence of Attribution Non-commercial Share Alike

☐ I accept the Terms and Conditions.

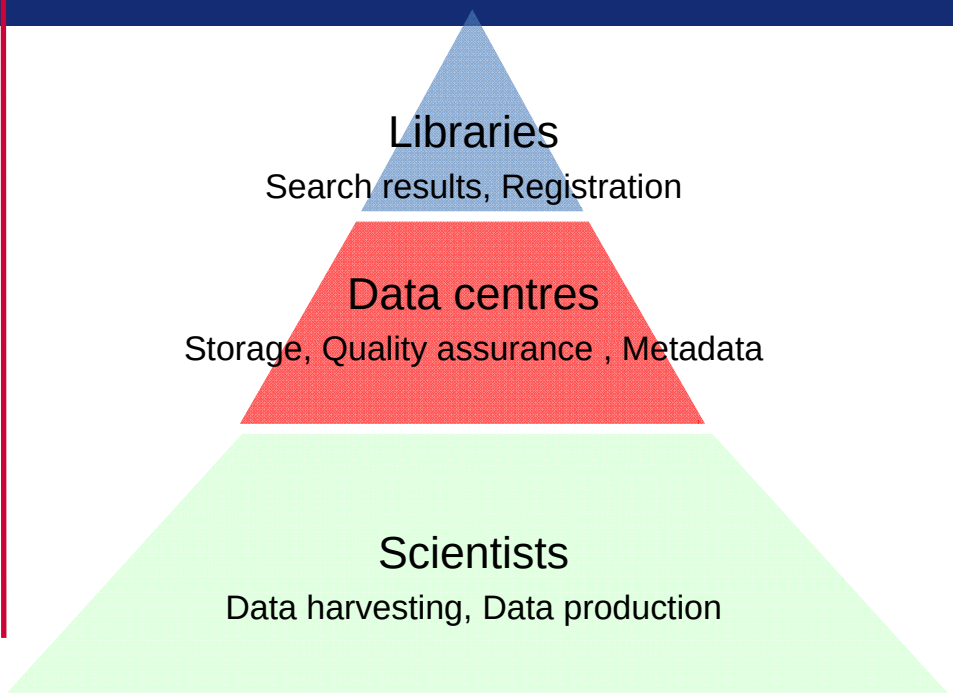
Download selected | Download all

SOME RIGHTS RESERVED
The contents of this site are licensed under a Creative Commons License (cc by).
Last Modification: 2006-07-21 | Some rights reserved by Daten- und Rechenzentrum, GeoForschungsZentrum Potsdam

An Ideal Research Cycle



3 layered Data Infrastructure



State of Developments

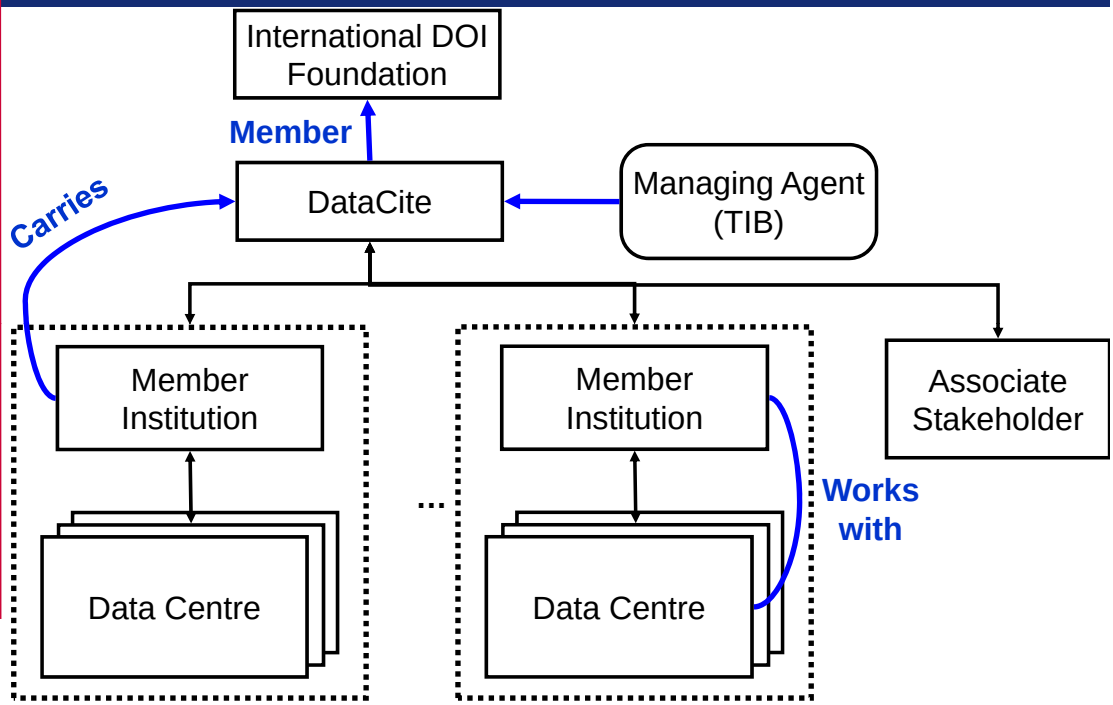
- So far the TIB is a DOI registration agency for primary data (and other non-commercial scientific information), with other local institutions as its customers
- In 1999 the publishers funded their independent DOI agency CrossRef
- In 2009 TIB has been transited the DOI registration to a new worldwide agency, named DataCite.

DataCite

- DataCite supports researchers by enabling them to locate, identify, and cite research datasets with confidence
- DataCite supports data centres by providing workflows and standards for data publication
- DataCite supports publisher by enabling linking from articles to the underlying data

http://www.datacite.org

DataCite Structure



Members

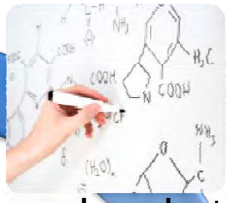
- Technische Informationsbibliothek (TIB)
- Canada Institute for Scientific and Technical Information (CISTI),
- California Digital Library, USA
- Purdue University, USA
- Office of Scientific and Technical Information (OSTI), USA
- Library of TU Delft, The Netherlands
- Technical Information Center of Denmark
- The British Library
- ZB Med, Deutschland
- ZBW, Deutschland
- Gesis, Deutschland
- Library of ETH Zürich
- L'Institut de l'Information Scientifique et Technique (INIST), Frankreich
- Swedish National Data Service (SND)
- Australian National Data Service (ANDS)

Affiliated members:

- Digital Curation Center (UK)
- Microsoft Research
- Interuniversity Consortium for Political and Social Research (ICPSR)
- Korea Institute of Science and Technology Information (KISTI)

Virtual Research Environment

planning



query-by-sketch

Databases connected with GetInfo

download

scientific environment

TIB | German National Library of Science and Technology

TIB | Technische Informationsbibliothek

Thank you for your attention