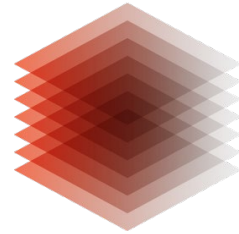

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

It's all about the viewers: How user-centered design helps to develop and operate an open repository for audiovisual media

Margret Plank, Bastian Drees, Abiodun Ogunyemi
Open Repositories Conference 2019
11. June 2019

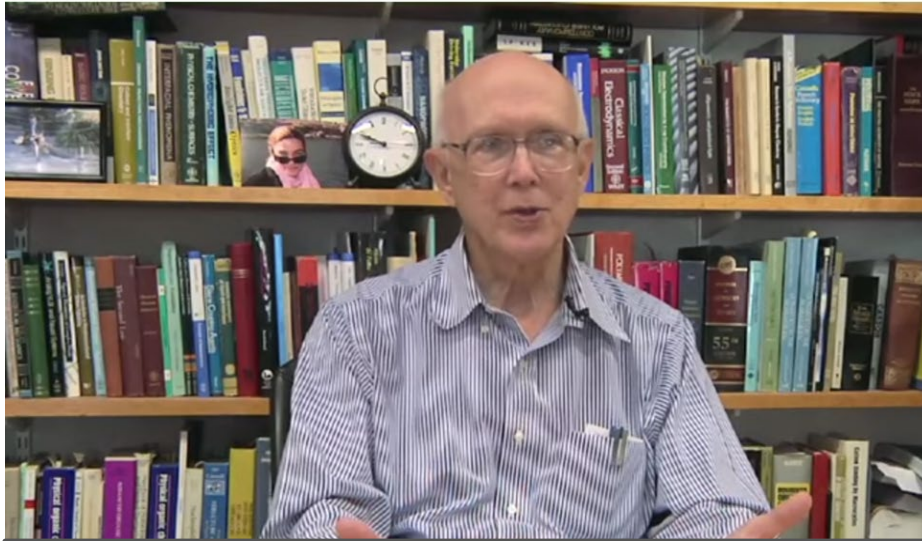
German National Library of Science and Technology



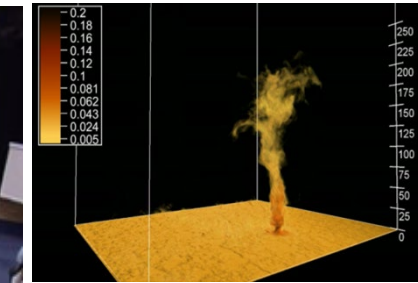
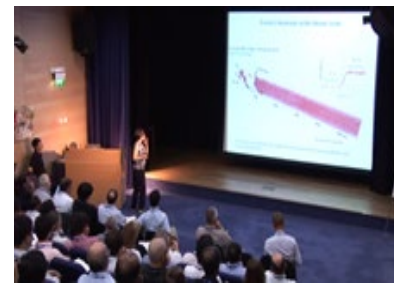
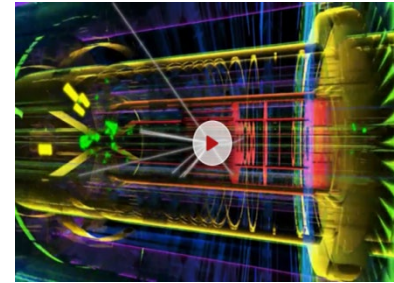
- The Leibniz Information Centre for Science and Technology and University Library
- The worlds largest library for science and technology
- Joint funding of federal government (30 %) and federal states (70 %)
- Global supplier for scientific and technical information
- An infrastructure provider for the scientific work process
- TIB-Strategy: **Move beyond text**
- **Competence Center for non-textual materials**
- **Research Group Visual Analytics**

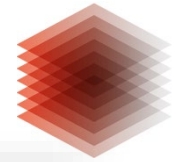


Videos in Higher Education



“I increasingly find that the stuff that we do that has **probably the greatest impact** in some metrics is **done in video format** not in paper format.”
 (<https://www.youtube.com/watch?v=NHuC5yZeHYQ>)





Videos are an endangered species



The 'Bit List' of Digitally Endangered Species

The DPC's 'Bit List' of Digitally Endangered Species is a crowd-sourcing exercise to discover which digital materials our community thinks are most at risk, as well as those which are relatively safe thanks to digital preservation. By compiling and maintaining this list over the coming years, the DPC aims to celebrate great digital preservation endeavors as entries become less of a 'concern,' whilst still highlighting the need for efforts to safeguard those still considered 'critically endangered.'

Click on each of the Risk Classifications to see the entries for each category.

LOWER RISK



Digital materials are listed as *Lower Risk* when it does not meet the requirements for other categories but where there is a distinct preservation requirement. Failure or removal of the preservation function would result in re-classification to one of the threatened categories.

CRITICALLY ENDANGERED



Digital materials are listed *Critically Endangered* when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs. This classification includes *Endangered* materials in the

Endangered

Digital materials are listed *Endangered* when they face material technical challenges to preservation or responsibility for care is poorly understood, or where the responsible agencies are poorly equipped to meet preservation needs. This classification includes *Vulnerable* materials in the presence of aggravating conditions.

In alphabetical order, Digital Species found to be *Endangered* are:

Born Digital Photos and Video Shared on Social Media or Uploaded to Cloud Services

Digital images or video with no analogue equivalent and where the only copy is online with a social media platform or cloud image hosting service.

Action:

Assessment is Most Urgent

(Action in 12 months in presence of aggravating conditions)

Assessment Completed:

November 2017

Examples:

Flickr

Vimeo

YouTube

Instagram

Periscope

DropBox

Facebook?

Aggravating conditions:

lack of preservation capacity in provider;

lack of explicit preservation commitment or incentive from provider to preserve;

lack of storage replication by provider

dependence on proprietary products or formats;

poor management of data protection;

inaccessibility to web archiving services;

political or commercial interference;

lack of offline equivalent;

over-abundance;

poorly managed intellectual property rights

lossy compression applied in upload scripts

Critically Endangered in the presence of Aggravating Conditions.

Vulnerable where good practice can be demonstrated.

few known examples are inaccessible by most practical means and methods. This classification includes *Critically Endangered* materials in the presence of aggravating conditions.

member of the digital preservation community has expressed a legitimate concern but the concern has not yet been assessed by the BITList jury. They will be assessed for inclusion in the subsequent year.

Trusted Home for Videos from higher education: TIB AV-Portal (av.tib.eu)

Profile

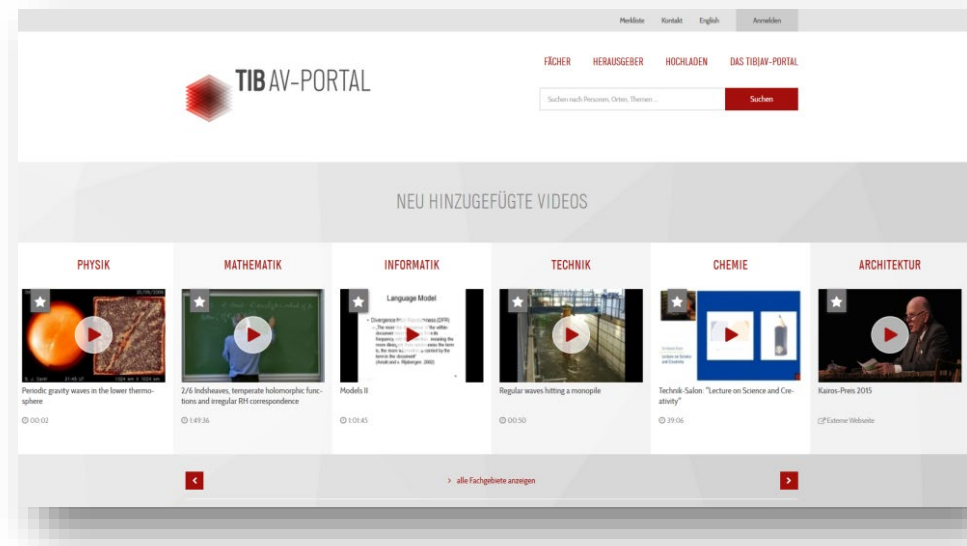
- The TIB AV-Portal is a web-based platform for curated scientific videos

Development

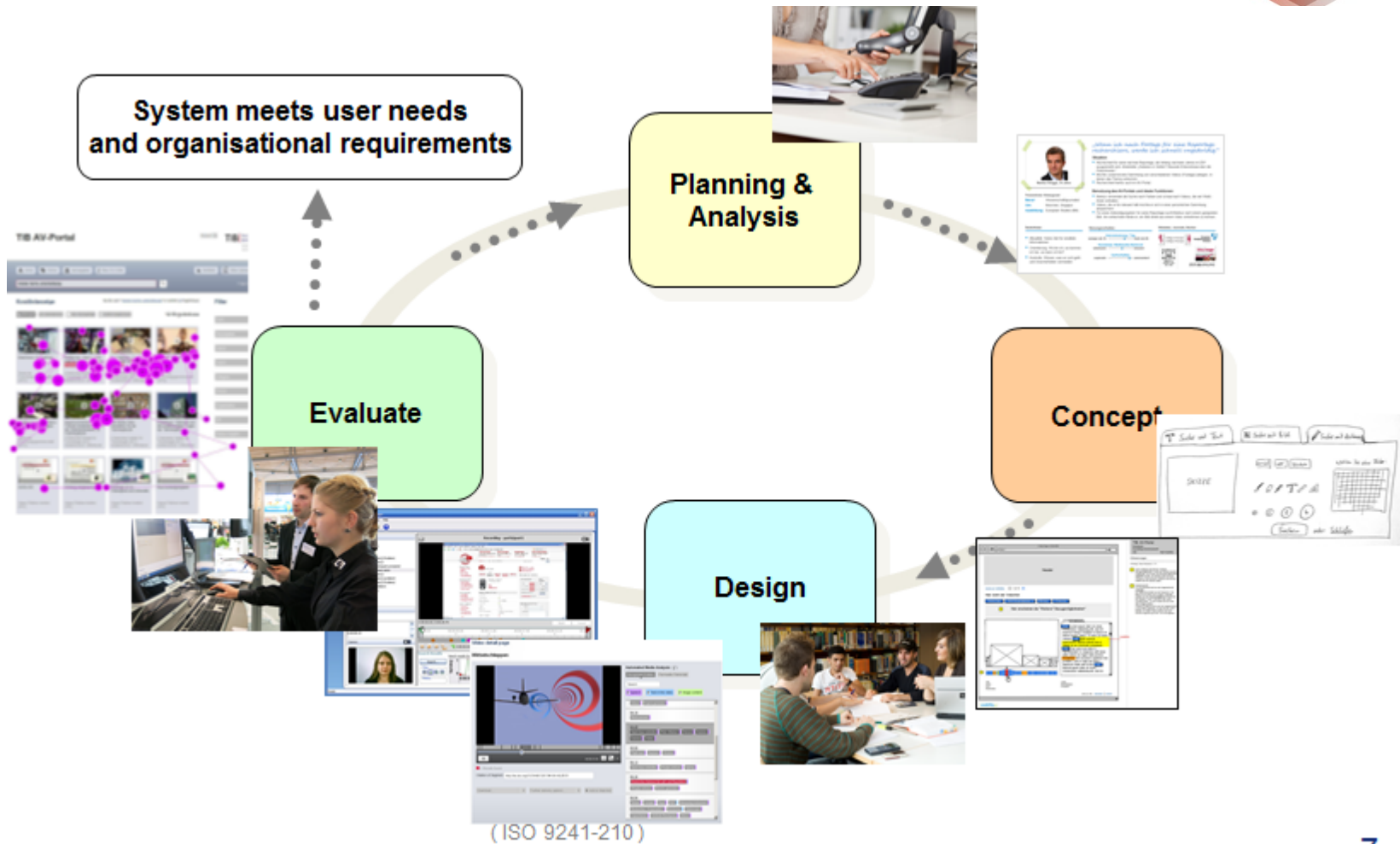
- TIB and Hasso Plattner Institute
- Online since: 04/2014

Content

- More than 20.000 videos (05/2019)
- English / German
- Predominantly under open access licence
- Conference recordings, lectures, experiments, video abstracts, simulations, animations....



UCD (2012-2014)



Enrichment: Semantic Video Analysis

Scene Recognition (SBD)



Speech Recognition (ASR)



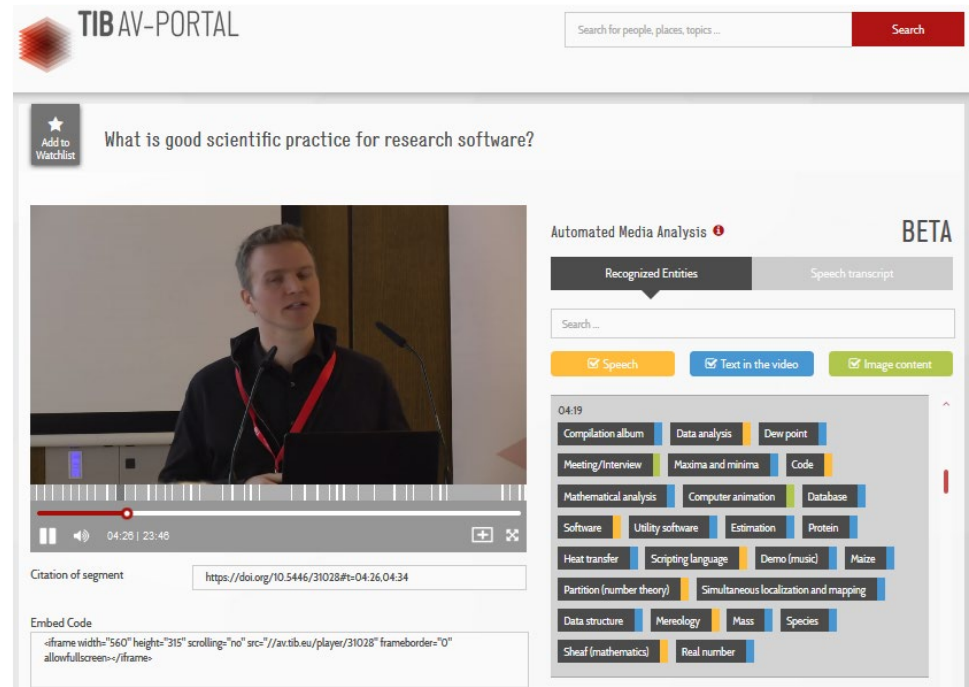
Text Recognition (OCR)



Image Recognition (VCD)



Named Entity Linking (NEL)



The screenshot displays the TIB AV-Portal interface. At the top, there is a search bar with the text "Search for people, places, topics ..." and a "Search" button. Below the search bar, the page title is "What is good scientific practice for research software?". A video player is embedded in the center, showing a man speaking at a podium. To the right of the video player is the "Automated Media Analysis" sidebar, which is currently in "Recognized Entities" mode. The sidebar shows a search bar and three filter buttons: "Speech", "Text in the video", and "Image content". Below these buttons, a grid of tags is displayed, including "Compilation album", "Data analysis", "Dew point", "Meeting/Interview", "Maxima and minima", "Code", "Mathematical analysis", "Computer animation", "Database", "Software", "Utility software", "Estimation", "Protein", "Heat transfer", "Scripting language", "Demo (music)", "Maize", "Partition (number theory)", "Simultaneous localization and mapping", "Data structure", "Mereology", "Mass", "Species", "Sheaf (mathematics)", and "Real number".

Semantic Video Analysis – Results

Video Segments



Audio Transcript

02:46

hitting science well about what people images that in software to crimes this data right you need these tools there is no signs without software today and minus catching maybe bottflies that at some point you would use this to a generator a database of the butterflies again and researchers need maybe to classify all this so software is essential for science today the it can also

Named Entities

02:44

Resultant

Point (geometry)

Database

Service (economics)

Information

Whiteboard

Computer animation

Software

Software

Right angle

Sign (mathematics)

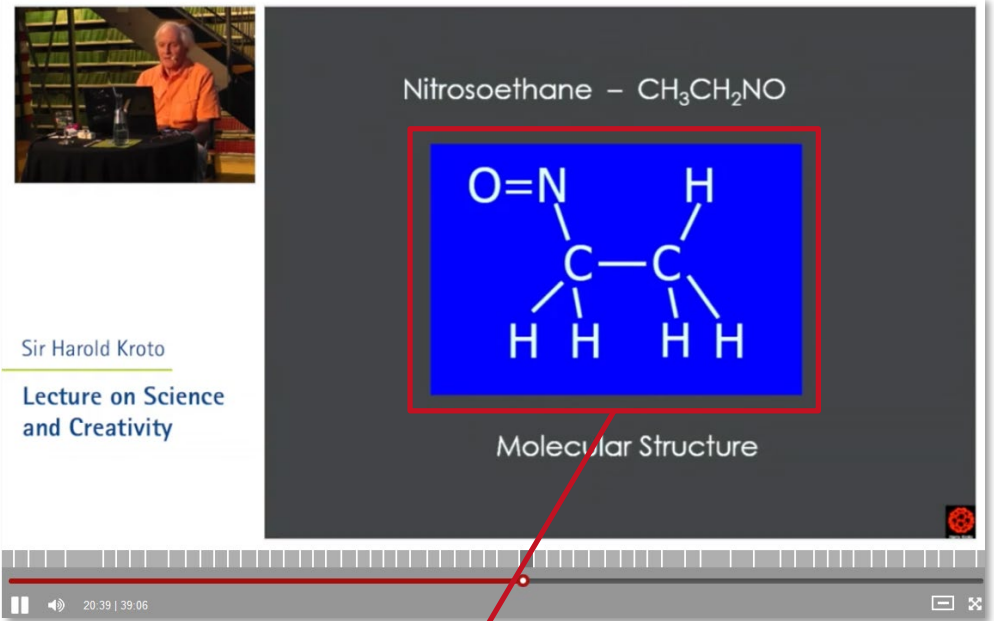
Game controller

Medical imaging

Function (mathematics)

TIB AV-Portal – Video Concept Detection

Video Keyframes

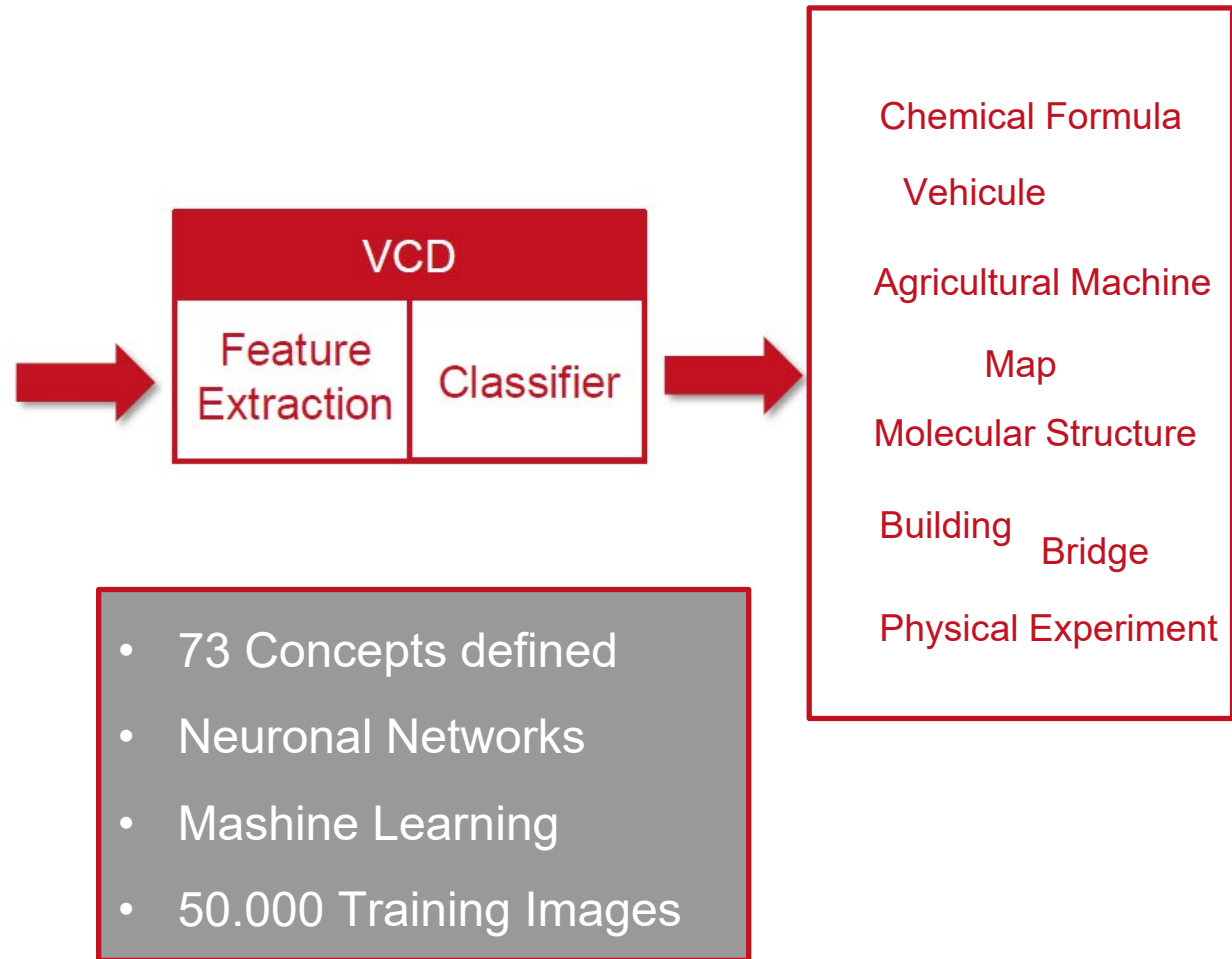
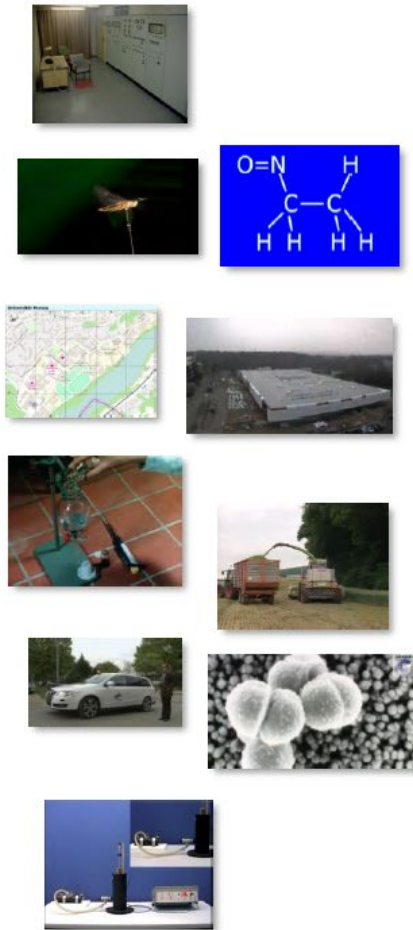


Visual Concepts

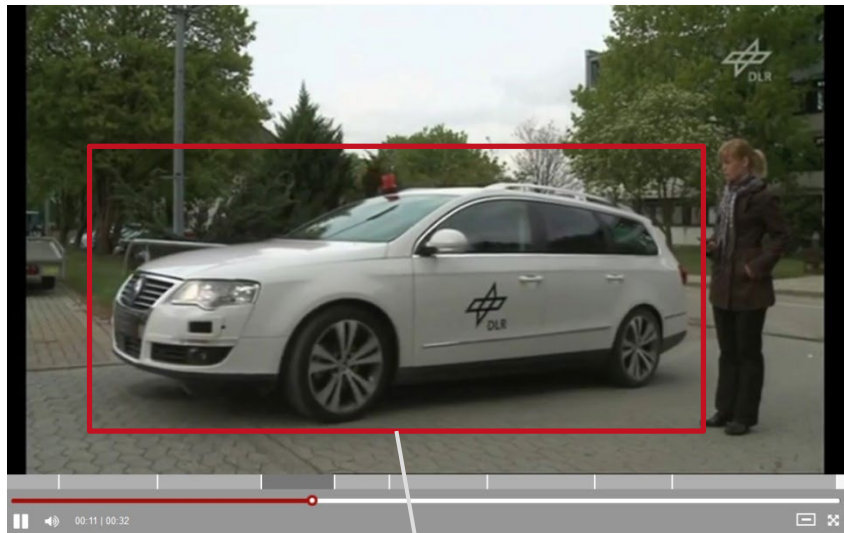
20:33

Nitrosoverbindungen	Molecular geometry	Hydrierung	Cyclooxygenase
Emission spectrum	Carbon (fiber)	Chemical structure	River source
Molecule	Molecule		

Visual Concept Detection

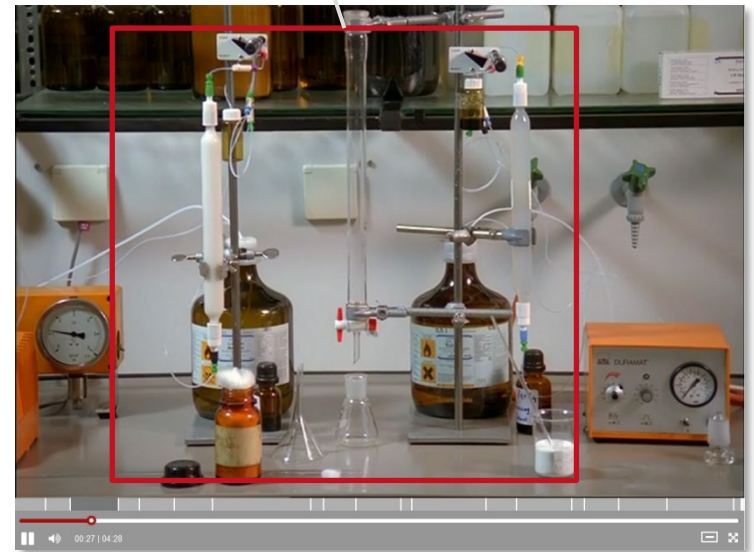


Visual Concept Detection



Kraftfahrzeug

Chemisches Experiment



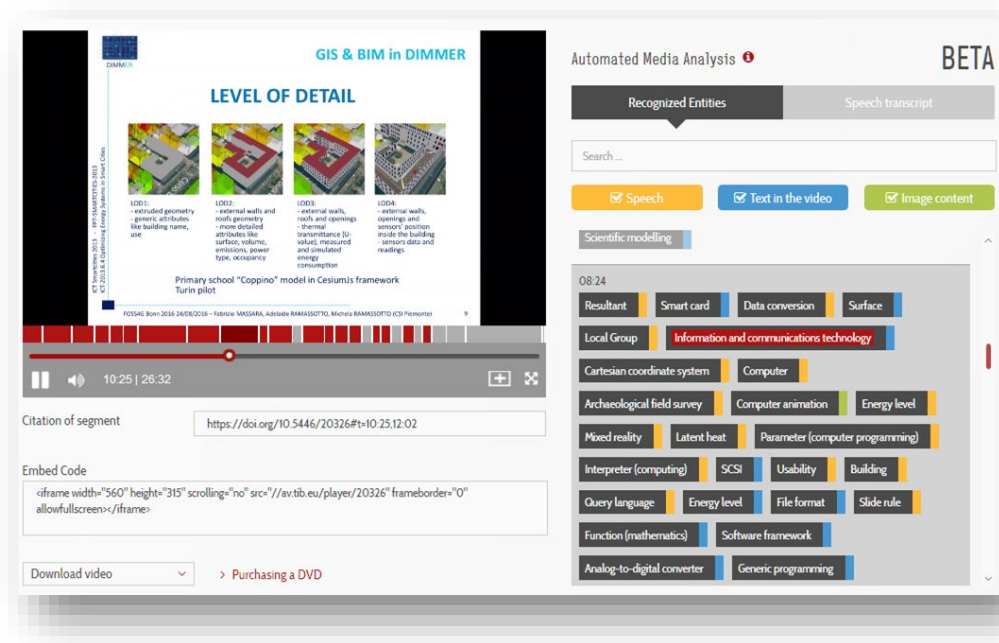
Semantic Analysis and Indexing

Named Entity Linking

- Based on Audio-/OCR-Transcripts
- Knowledge Base:
 - 63.356 GND Entities
 - TIB core subjects

Cross-lingual Mapping

- Heuristic Mapping approach
- English labels from
 - DBpedia
 - LCSH, MACS, TEMA



The screenshot displays a video player interface. The main video area shows a presentation slide titled "GIS & BIM in DIMMER" with a sub-heading "LEVEL OF DETAIL". The slide contains four columns of information:

LOD1	LOD2	LOD3	LOD4
- extruded geometry - generic attributes like building name, use	- external walls and roof geometry - more detailed attributes like surface, volume, emissions, power type, occupancy	- external walls, roofs and openings - thermal transmittance (U values), measured and simulated energy consumption	- external walls, openings and sensors' position inside the building - sensors data and readings

Below the slide, it mentions "Primary school 'Coppino' model in Cesium's framework Turin pilot".

The video player interface includes a progress bar at 10:25 / 26:32, a citation field with the URL <https://doi.org/10.5446/20326#t=10:25:12:02>, and an embed code field containing: `<iframe width="560" height="315" scrolling="no" src="//av.tib.eu/player/20326" frameborder="0" allowfullscreen></iframe>`. There is also a "Download video" button and a note "> Purchasing a DVD".

On the right side, there is a sidebar for "Automated Media Analysis" (BETA). It features a search bar and three active filters: "Speech", "Text in the video", and "Image content". Below these, a "Scientific modelling" section displays a list of tags such as "Resultant", "Smart card", "Data conversion", "Surface", "Local Group", "Information and communications technology", "Cartesian coordinate system", "Computer", "Archaeological field survey", "Computer animation", "Energy level", "Mixed reality", "Latent heat", "Parameter (computer programming)", "Interpreter (computing)", "SCSI", "Usability", "Building", "Query language", "Energy level", "File format", "Slide rule", "Function (mathematics)", "Software framework", "Analog-to-digital converter", and "Generic programming".

Cite videos to the second

DOI + MFID:

- Complete Video:

<https://doi.org/10.3203/IWF/C-13096>

- Particle Image Velocimetry:

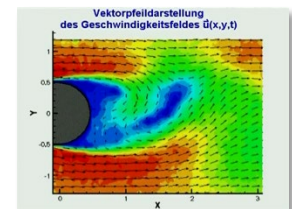
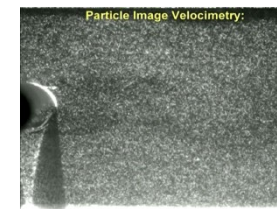
<https://doi.org/10.3203/IWF/C-13096#t=00:11,01:58>

- Navier-Stokes-Equation

<https://doi.org/10.3203/IWF/C-13096#t=03:19,04:32>

- Quote: „the basic idea of this approach is [...].“

<https://doi.org/10.3203/IWF/C-13096#t=00:18,00:27>



Formulierung der Bewegungsgleichung
 für ein Volumenelement dV
 einer realen viskosen strömenden Flüssigkeit

Ausgangspunkt: Newtonsches Axiom

$$\vec{F} = m \cdot \frac{d\vec{u}}{dt} \quad \text{mit } m = \rho \cdot dV$$

Navier-Stokes-Gleichung:

$$-\text{grad } p + \rho \cdot \vec{g} + \eta \Delta \vec{u} = \rho \cdot \left[\frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \nabla) \vec{u} \right]$$


Results from User Interviews

What: In-depth Interviews, Questionnaire

When: October/ November 2018

Who:

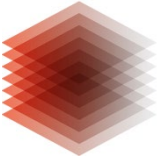
- General public: 2
- Educators: 9
- Researchers: 6

From: Hungarian, Germany, France, Belgium, Ireland, Italy

Frequently used video platforms:

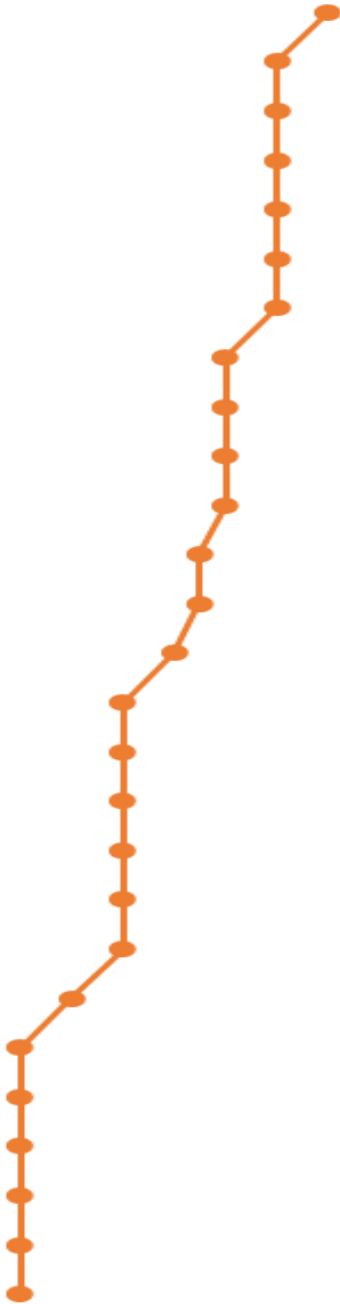
YouTube, Vimeo, VIAA, TIB AV Portal, archive.org, Facebook, Kahn Academy, ksnapda, moovly, Netflix, SchoolTV, Teleblik, Visus, VRTnws

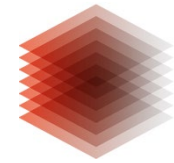
60% 80% 100%



TIB

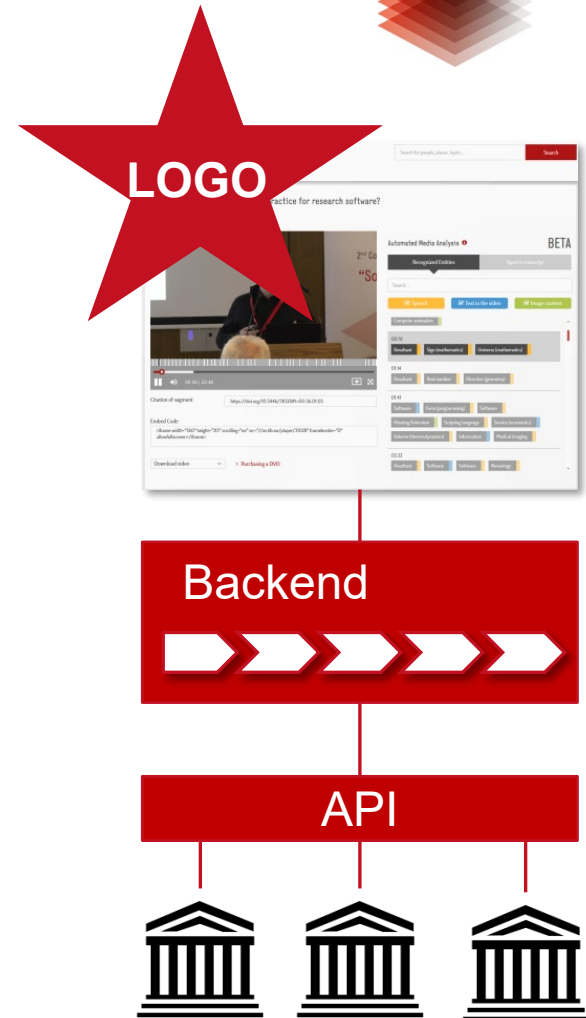
In text search
Add annotations (video)
Add comments (video)
Add annotations (segment)
Add comments (segment)
Thumbnail images: snippet navigation
Download still images
Download video
Embed video
Display/view (existing) subtitling
Create collection/playlist
Export bibliographic (meta)data (segment) (*)
Display/view DOI/MFID (*)
Display/view copyright information
Export bibliographic (meta)data (video)
Embed segment
Translate and generate subtitling
Increase/decrease speed
Store and recall personal information (search strings, results etc.)
Display/view number of views
Add keywords (video)
Store keywords (video)
Edit keywords (video)
Share keywords (video)
Download segments
Enlarge view (zoom in on video)
Jump x seconds forward/backward



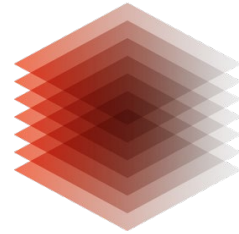


NEW!

- **Co-branding / Whitelabeling Frontend**
 - Provision of an instance of the frontend with its own CI
- **Future Work: Videoanalyse as a service**
 - Provision of video analysis such as speech-, text-, image recognition via API



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Thank you for your attention!

Contact

Margret Plank(**ORCID:** 0000-0001-8941-7563)

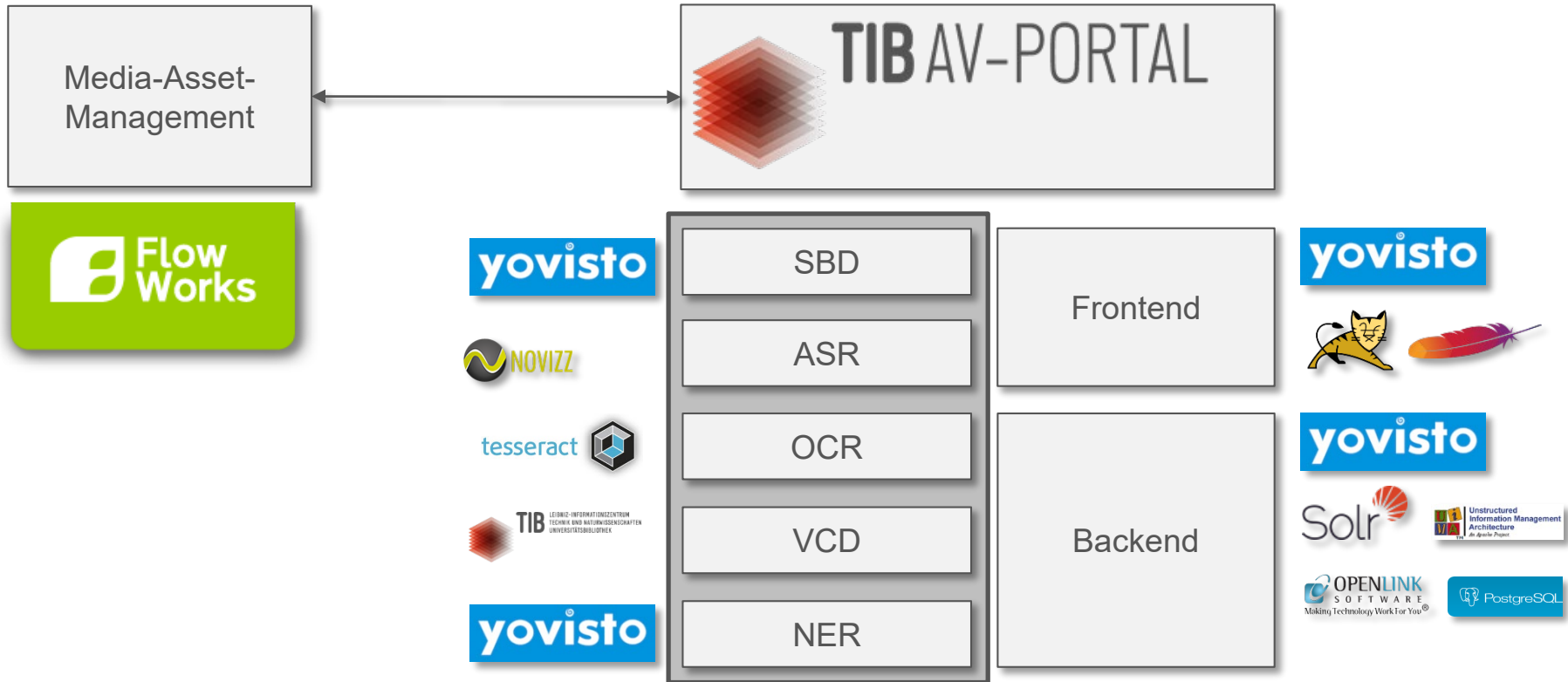
Phone: +49 511 762-4884

Email: Margret.plank@tib.eu



[@TIB_AVPortal](https://twitter.com/TIB_AVPortal)

Systems Architecture 2017



Goals 2018 – 2020

- Inhousing and Hosting at TIB
- Modernization of all System Components
- Modularization of Systems Architecture
- Technical Development at TIB
- Basis: Open Source Software

Systems Architecture 2018 – 2020

