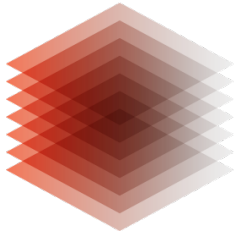
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



TIB

Future plans for TIB AV-Portal – modularization and replacement of software components

Felix Saurbier
Hamburg, November 7
Fedora User Group Meeting

Agenda

1. TIB and TIB AV-Portal
2. Technology
3. Systems Architecture
4. Future

Agenda

1. TIB and TIB AV-Portal
2. Technology
3. Systems Architecture
4. Future

Technische Informationsbibliothek (TIB)

- German National Library for Science and Technology
- University Library at Hannover
- The world's largest science and technology library
- An infrastructure provider for the whole scientific work process
- TIB strategy: **Move beyond text**
- **Competence Centre for Non-Textual Materials**

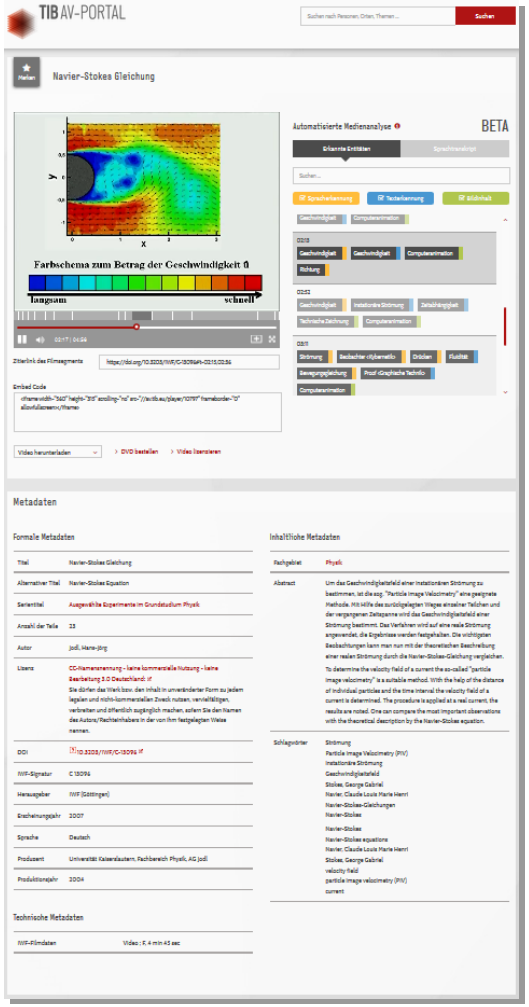
Leibniz Gemeinschaft

TIB AV-Portal (av.tib.eu)

- Platform for **quality-tested scientific videos**
- Online since April 2014
- Developed by TIB and Hasso Plattner Institute

- **Automatic metadata enrichment,** DOI/MFID, long-term preservation, semantic search

- More than 11,500 Videos (December 2017)
- Conference recordings, lectures, experiments, video abstracts, simulations, animations
- Videos predominantly under open access licences

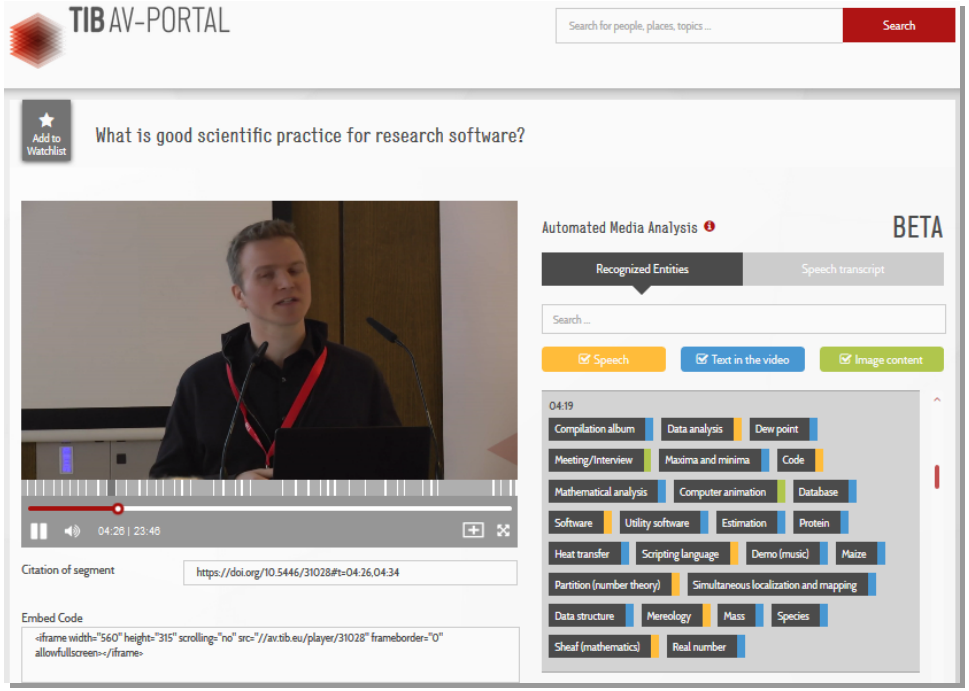


Agenda

1. TIB and TIB AV-Portal
2. Technology
3. Systems Architecture
4. Future

Semantic Video Analysis – Process

- Scene Recognition (SBD)
- Speech Recognition (ASR)
- Text Recognition (OCR)
- Image Recognition (VCD)
- Named Entity Linking (NEL)



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 botflies 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)

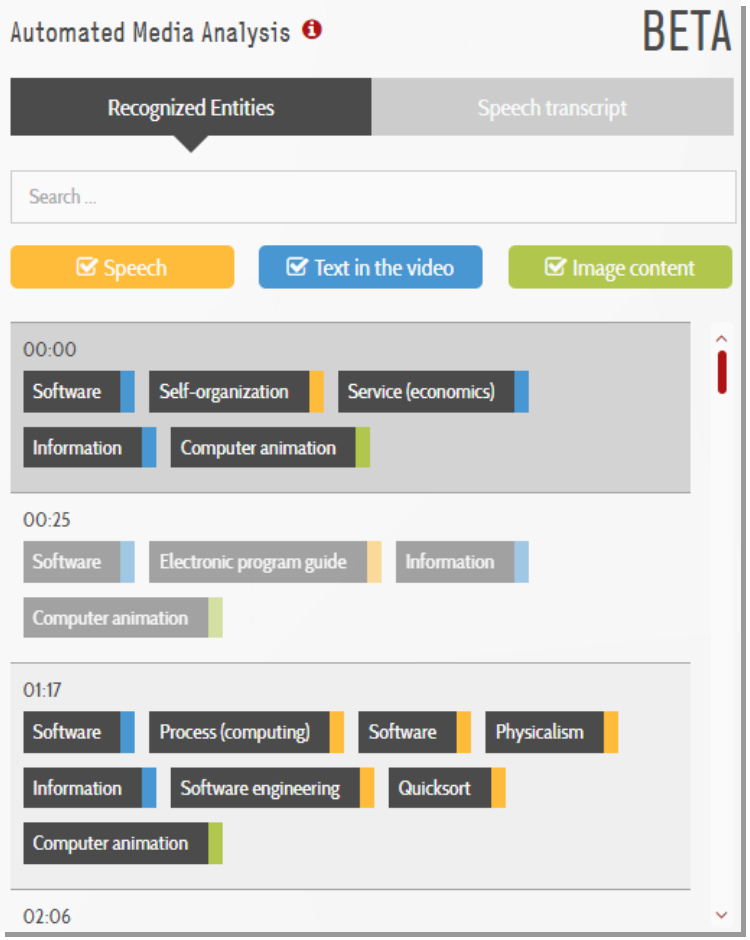
Semantic Analysis and Indexing

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



Live Demo

https://av.tib.eu/

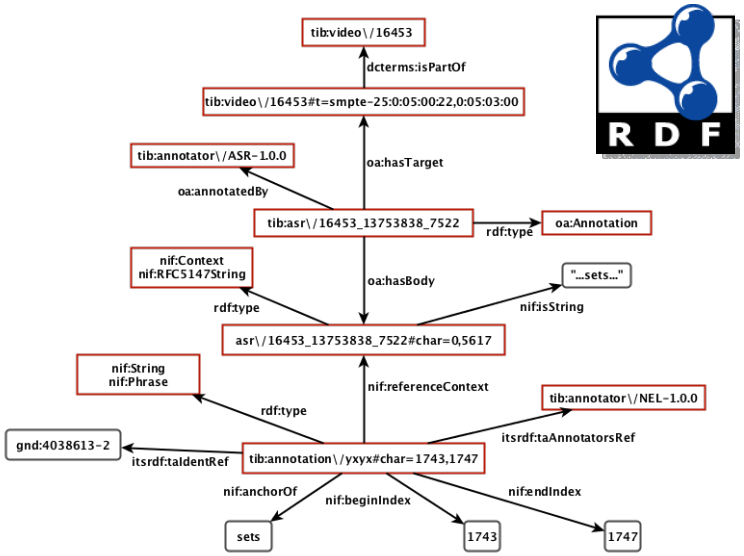
Data model

Resource Description Framework (RDF)

```
tib:asr/31031_26463533_23007#char=483,494 itsrdf:taAnnotatorsRef tib:annotator/NEL-1.0.0 ;
      itsrdf:taIdentRef <http://d-nb.info/gnd/4488209-9> ;
      nif:anchorOf "source code" .
```

Vocabularies

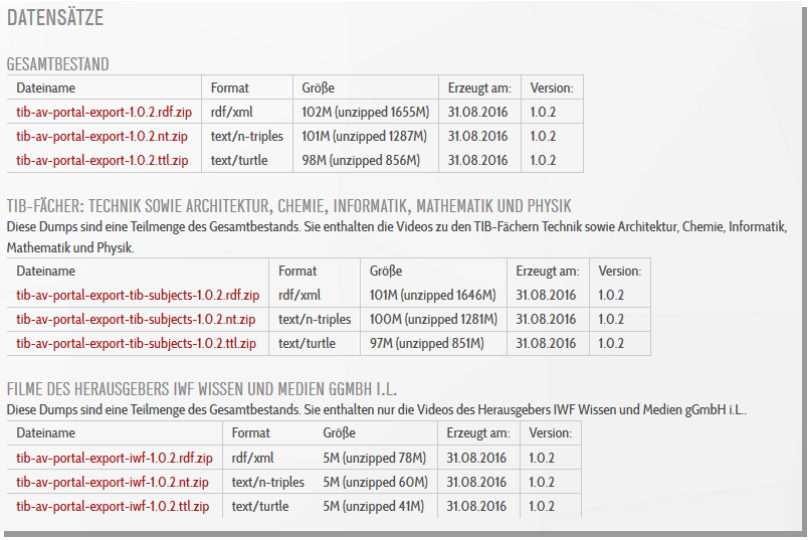
- Bibframe Vocabulary
- DCMI Metada Terms
- DCMI Type Vocabulary
- Friend of a Friend Vocabulary
- **Open Annotation Data Model**
- **NLP Interchange Format**
- **Internationalization Tag Set (ITS) Ontology**



https://av.tib.eu/opendata

Metadata Publication & Linked Open Data

- CC0 RDF-Dumps (since 2016)
- Dereferencable URIs & 303 (end of 2017)
- LDF-Server (since 2017)
- Public SPARQL-Endpoint (2018+)



DATENSÄTZE

GESAMTBESTAND

Dateiname	Format	Größe	Erzeugt am:	Version:
tib-av-portal-export-1.0.2.rdf.zip	rdf/xml	102M (unzipped 1655M)	31.08.2016	1.0.2
tib-av-portal-export-1.0.2.nt.zip	text/n-triples	101M (unzipped 1287M)	31.08.2016	1.0.2
tib-av-portal-export-1.0.2.ttl.zip	text/turtle	98M (unzipped 856M)	31.08.2016	1.0.2

TIB-FÄCHER: TECHNIK SOWIE ARCHITEKTUR, CHEMIE, INFORMATIK, MATHEMATIK UND PHYSIK

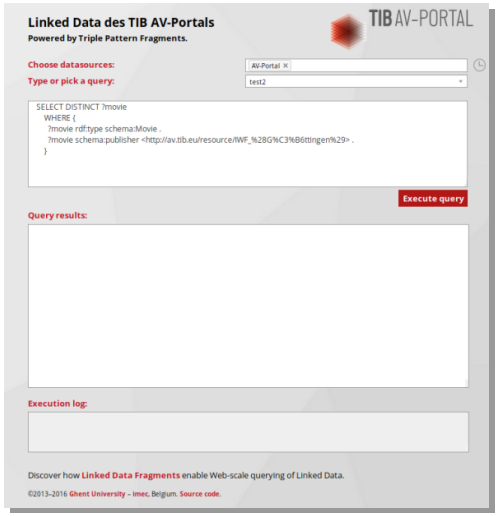
Diese Dumps sind eine Teilmenge des Gesamtbestands. Sie enthalten die Videos zu den TIB-Fächern Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik.

Dateiname	Format	Größe	Erzeugt am:	Version:
tib-av-portal-export-tib-subjects-1.0.2.rdf.zip	rdf/xml	101M (unzipped 1646M)	31.08.2016	1.0.2
tib-av-portal-export-tib-subjects-1.0.2.nt.zip	text/n-triples	100M (unzipped 1281M)	31.08.2016	1.0.2
tib-av-portal-export-tib-subjects-1.0.2.ttl.zip	text/turtle	97M (unzipped 851M)	31.08.2016	1.0.2

FILME DES HERAUSGEBERS IWF WISSEN UND MEDIEN GGMBH I.L.

Diese Dumps sind eine Teilmenge des Gesamtbestands. Sie enthalten nur die Videos des Herausgebers IWF Wissen und Medien gGmbH i.L..

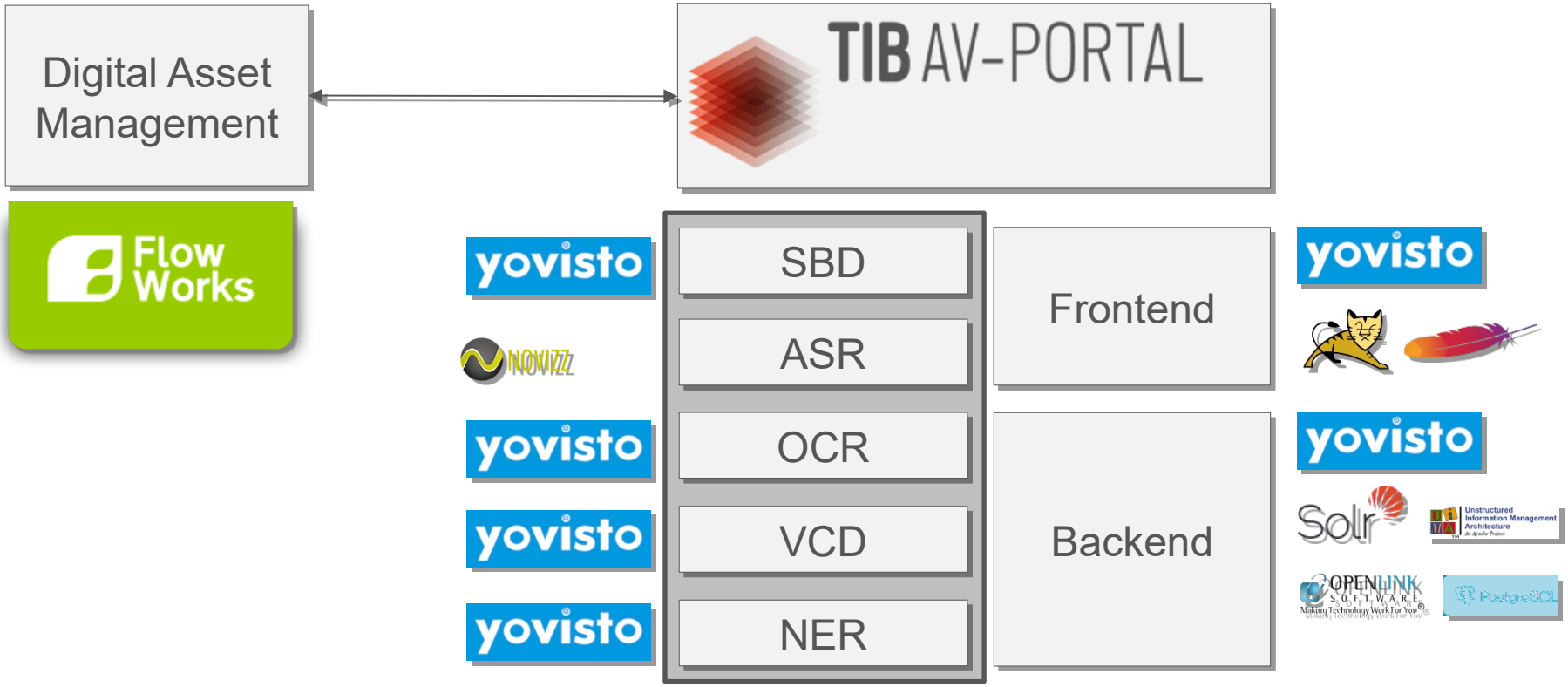
Dateiname	Format	Größe	Erzeugt am:	Version:
tib-av-portal-export-iwf-1.0.2.rdf.zip	rdf/xml	5M (unzipped 78M)	31.08.2016	1.0.2
tib-av-portal-export-iwf-1.0.2.nt.zip	text/n-triples	5M (unzipped 60M)	31.08.2016	1.0.2
tib-av-portal-export-iwf-1.0.2.ttl.zip	text/turtle	5M (unzipped 41M)	31.08.2016	1.0.2



Agenda

1. TIB and TIB AV-Portal
2. Technology
3. **Systems Architecture**
4. Future

Systems Architecture 2017



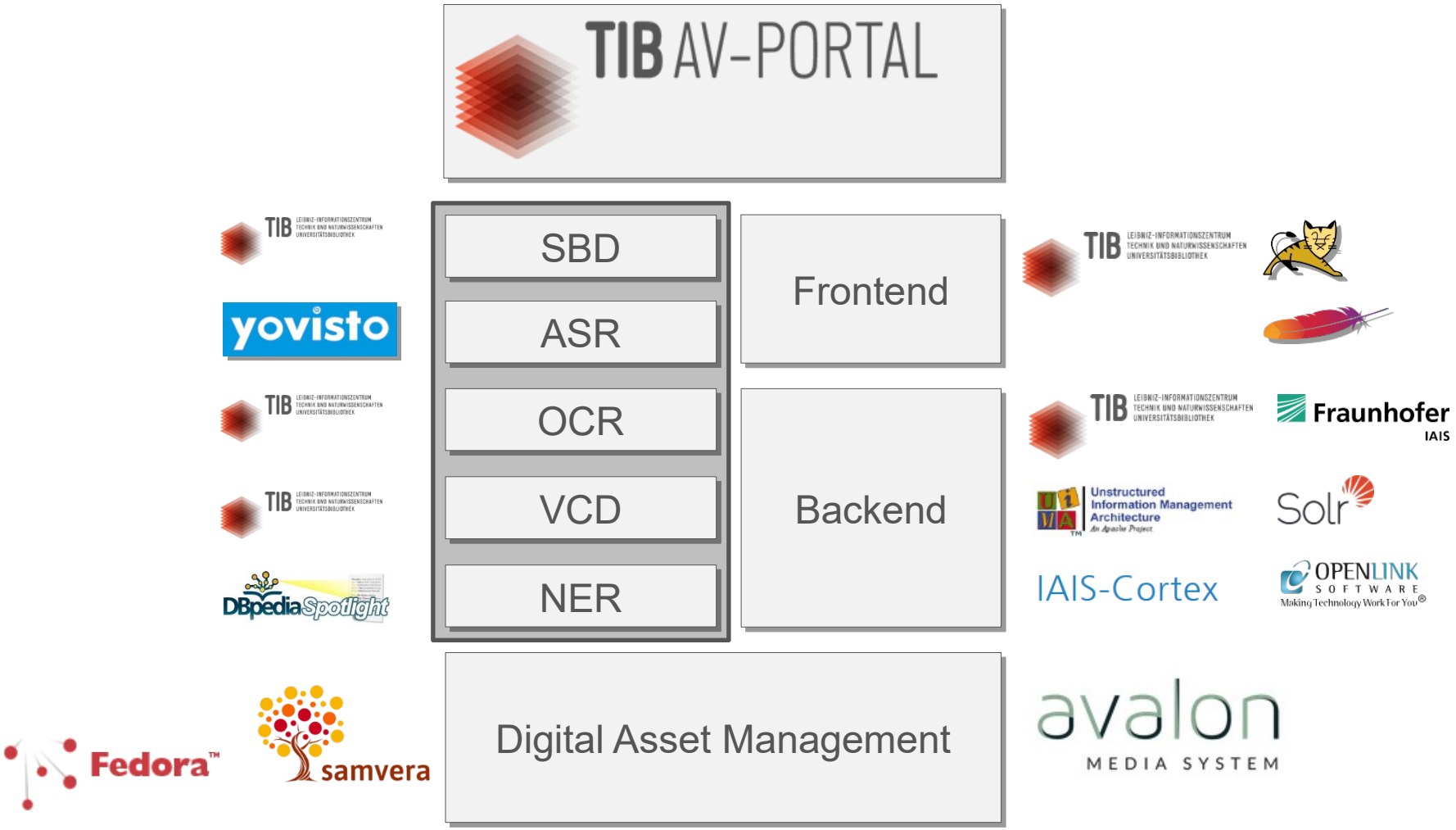
Agenda

1. TIB and TIB AV-Portal
2. Technology
3. Systems Architecture
4. **Future**

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



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

More Infos

KNM@tib.eu

av.tib.eu

Contact
Felix Saurbier
T +49 511 762-14645, felix.saurbier@tib.eu