CONTENTUS
Next Generation Multimedia Libraries

THESEUS
New Technologies for the Internet of Services

CONTENTUS

Technologies for Next Generation Multimedia Libraries

AIEMPro'10, Firenze
Andreas Heß, German National Library
Jan Nandzik, Acosta Consult

DEUTSCHE NATIONAL BIBLIOTHEK
technicolor
Fraunhofer Heinrich-Hertz-Institut
HPI Hasso Plattner Institut
Fraunhofer IAIS
Institut für Rundfunktechnik IRT
mufin
Supported by
Federal Ministry of Economics and Technology

Motivation

Supported by
Federal Ministry of Economics and Technology

Motivation

More than 30 millions hours of audio-visual content are stored in European archives

Supported by
Federal Ministry of Economics and Technology

Motivation

National Libraries contain millions of printed media

More than 30 millions hours of audio-visual content are stored in European archives

Supported by
Federal Ministry of Economics and Technology

Motivation

National Libraries contain millions of printed media

World film production hit 5,039 feature films in 2007 and represented 60% of the audio-visual revenues.

More than 30 millions hours of audio-visual content are stored in European archives

Supported by
Federal Ministry of Economics and Technology

Next Generation Multimedia Archives (ca. 2015)

- Everything is digital
  - Mass processing, administration and provision of multimedia content is day-to-day business
  - All media are available in high quality
- Always accessible
  - Access from anywhere, at any time
  - Resources are not added to knowledge networks – they are created within them
- Always up-to-date
  - Desired information finds the user
- Knowledge Journeys
  - An interactive exploration of cultural and scientific collections
  - Largely replace traditional access (e.g. search engines)

Supported by
Federal Ministry of Economics and Technology

Present Generation Multimedia Archives (ca. 2010)

- Still large amounts of non-digitised material
  - Mass digitisation is still expensive, if quality is important
  - Problem: Media deterioration
- Restriced Access
  - Only from reading room, even if digital (legal reasons...)
- Not necessarily up-to-date
  - Indexing requires manual and intellectual effort
- Search is the paradigm
  - User must know what he/she is looking for
  - Different search engines for different collections
  - Media discontinuity
  - Situation in present libraries/archives: search engines often slow and outdated

Supported by
Federal Ministry of Economics and Technology

CONTENTUS
Next Generation Multimedia Libraries

THESEUS
New Technologies for the Internet of Services

Supported by

Federal Ministry of Economics and Technology

CONTENTUS
Next Generation Multimedia Libraries

THESEUS
New Technologies for the Internet of Services

Supported by
Federal Ministry of Economics and Technology

Problems on the Way

- **Deterioration**
- **Digitisation**
- **Metadata**
- **Accessibility**

Supported by
Federal Ministry of Economics and Technology

Deterioration

- Improper storage and handling
- Magnetic tape: drop-outs, wow and flutter, ...
- Film: dirt, scratches, blotches, ...
- Paper: bleaching, acid, mice, ...
- Optical discs: coating decay, ...

Supported by
Federal Ministry of Economics and Technology

Digitisation

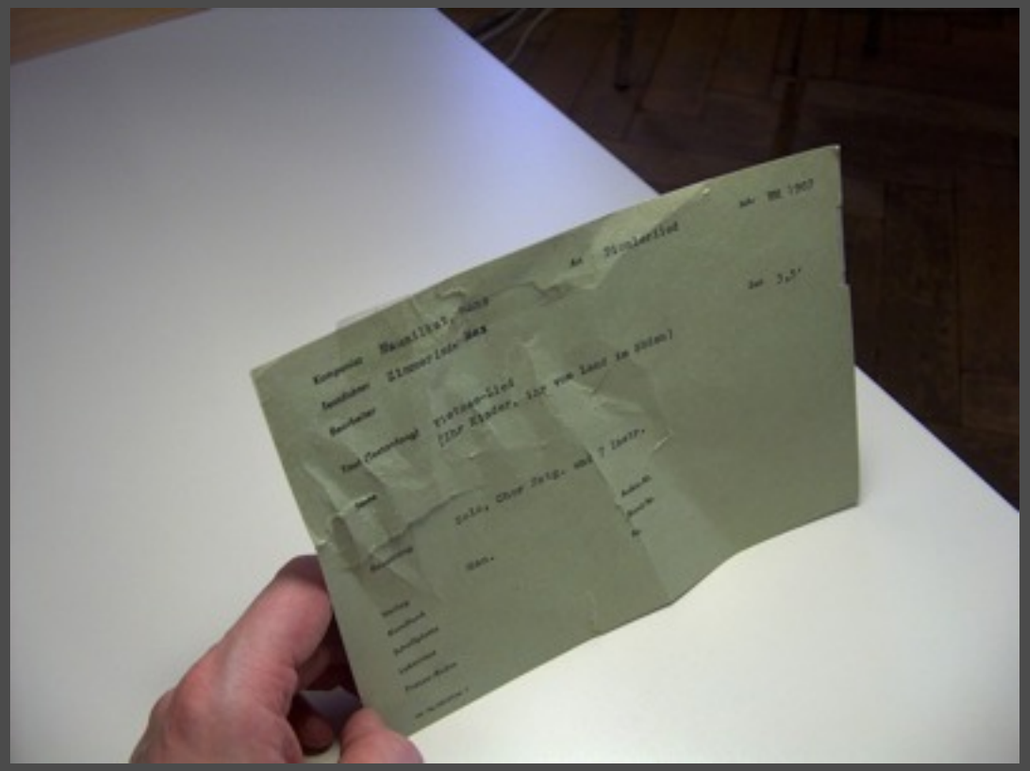
- **Often lack of quality awareness**
- **Quality is crucial for preserving cultural heritage**

Supported by
Federal Ministry of Economics and Technology

Digitisation – Problems

- **Causes for quality issues:**
  - **Unsuitable hardware**
  - **Unsuitable configuration**
  - **Errors during digitisation**

- **Goals:**
  - **Automatisation and efficiency**
  - **Continuous checks while job is being processed**



Supported by
Federal Ministry of Economics and Technology

Metadata – Problems

- **Not always present**
- **Indexing and annotation are time-consuming and error-prone**
- **Incompatibility of different sources of metadata**

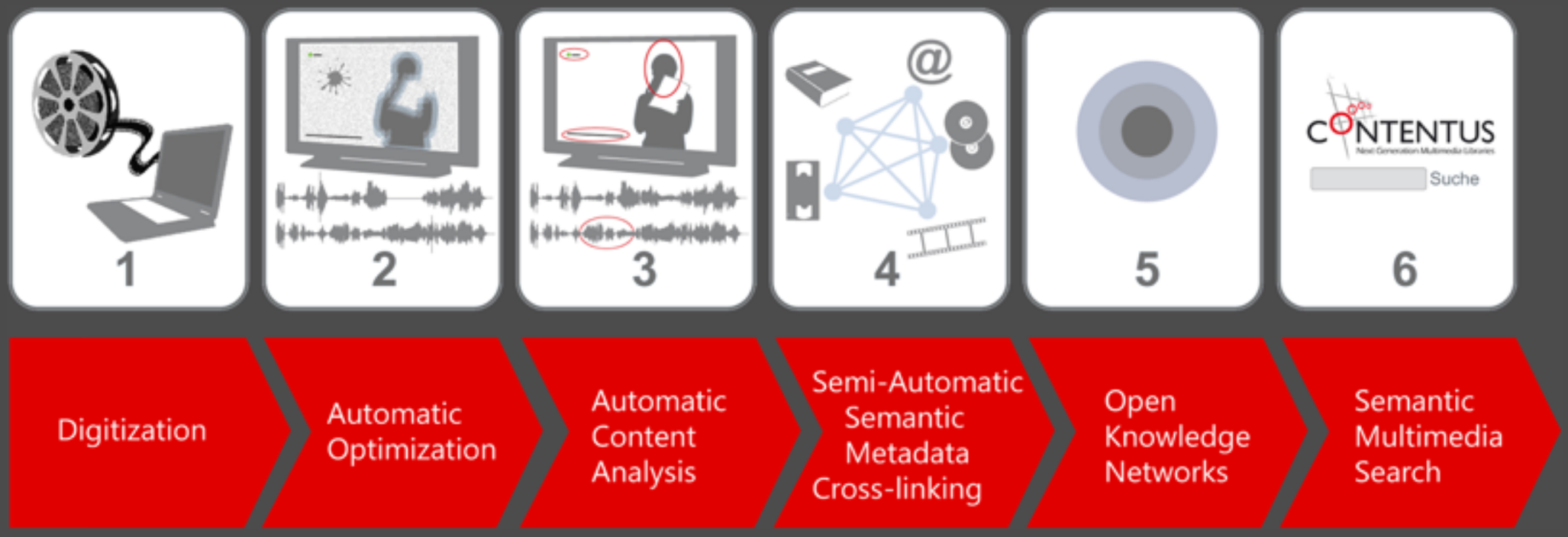
Supported by
Federal Ministry of Economics and Technology

Accessibility – Problems

- **Current search approaches not really suitable for multi-media content**
- **Search and consumption is separated**
- **Physical presence of media**
- **Data is nothing without meta-data!**

Supported by
Federal Ministry of Economics and Technology

Supported by
Federal Ministry of Economics and Technology



Supported by
Federal Ministry of Economics and Technology

Florenz ×

Search

Results 100

Article	77
Person	8
Video	8
Book	5
Audio	1
Location	1

History

Florenz	100
Florenz	100
Florenz	100
Florenz	100
Firenze	5
Armani	8
hallo andreas	1
hallo andreas	1
hallo andreas	1
Florenz	100

Relevance Popularity Quality Date↑ Date↓

collect similar

Florenz 0 views

Florenz, Firenze, Florentia

Location Florenz

Also known as Florentia; Florenz; Firenze

collect similar

Florence Nightingale 0 views

*Geboren: 1820 in **Florenz**; Gestorben: 1910 in London; Krankenschwester, Krankenpflegerin; Brit. Krankenpflegerin.*

collect similar

Luigi Dallapiccola 0 views

*Gestorben: 19.02.1975 in **Florenz**; Komponist, Musiker; Ital. Komponist.*

Professions Komponist Musiker **PreferredName** Dallapiccola, Luigi
PlaceOfBirth Pazin **DateOfDeath** 19.02.1975
BiographicalInformation Ital. Komponist **DateOfBirth** 03.02.1904

collect similar

Arnold Böcklin 0 views

Bildhauer; Basel; Weimar;

collect similar

Cesar Bresgen 0 views

*Geboren: 1913 in **Florenz**;*

1313 2008

filter / reset

Location 5 of 50

Florenz	73
Italien	47
Rom	32
Deutschland	28
Europa	23

Contributor 5 of 5

Heinz Krause-Graumnitz	1
Walther Siegmund-Schultze	1
Eugen Schmitz	1
Hellmuth Christian Wolff	1
Kurt Schwaen	1

Supported by
Federal Ministry of Economics and Technology

Our Project

- **THESEUS**
  - **Research intiative in the area of Internet-based technologies, focus: semantic technologies**
  - **funded by German Federal Ministry of Economics and Technology**
  - **consortium of approx. 60 partners from academia and industry**
  - **„application scenario"- and „basic technology"-subprojects**

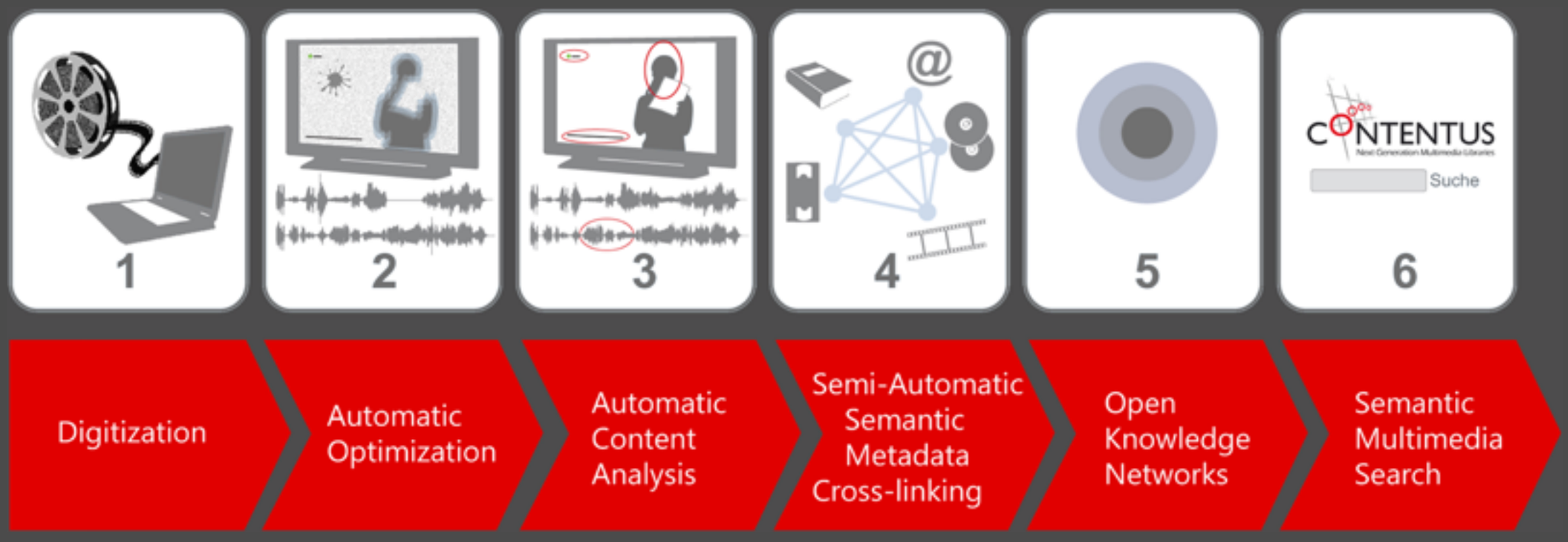
- **CONTENTUS**
  - **application scenario-subproject of THESEUS**
  - **concepts and technologies for multimedia-libraries and archives**

Supported by Federal Ministry of Economics and Technology

CONTENTUS
Next Generation Multimedia Libraries

THESEUS
New Technologies for the Internet of Services

The CONTENTUS Processing Chain



Media-specific

Media-independent

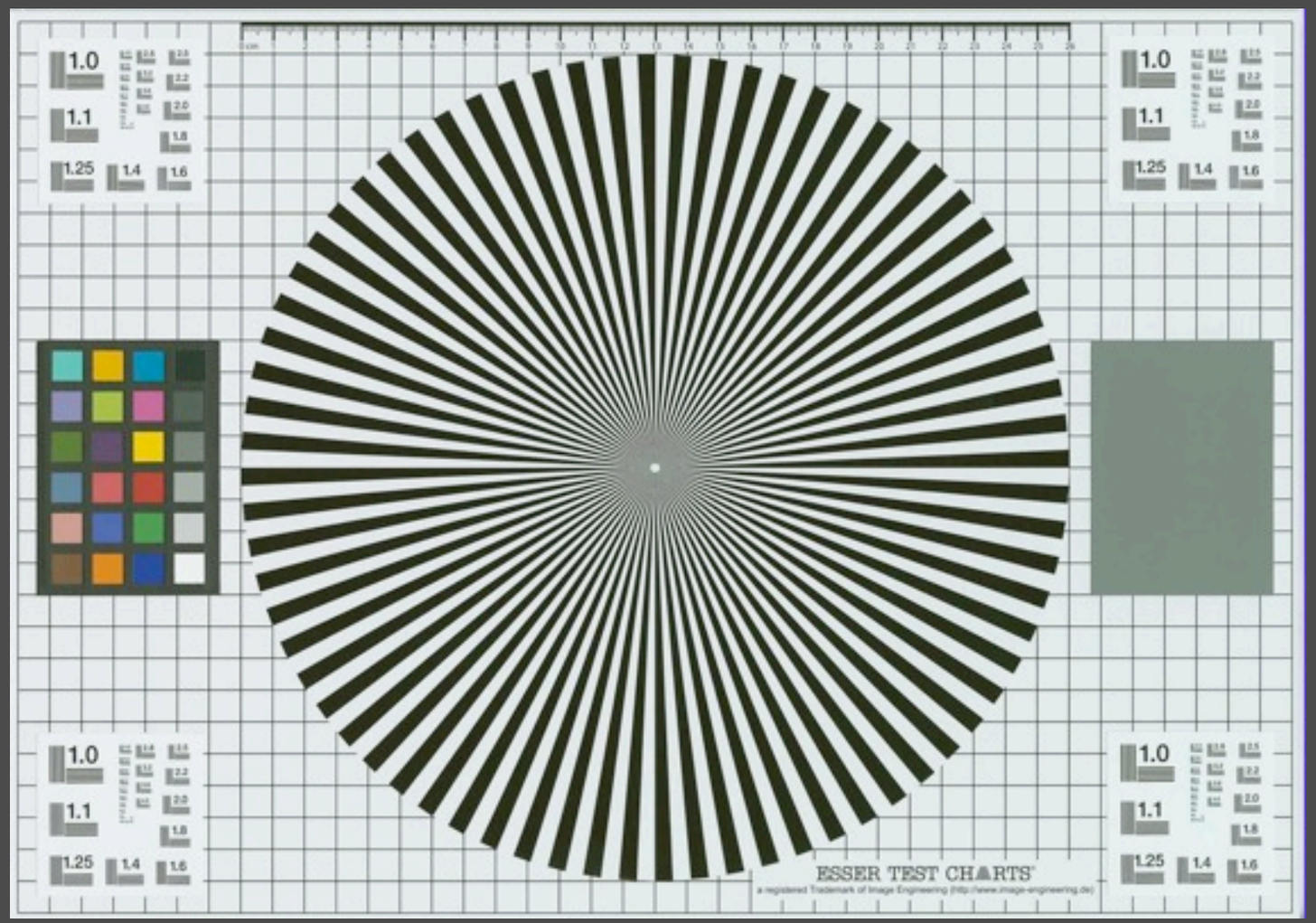
Supported by
Federal Ministry of Economics and Technology

Outline

- **Media-specific processing**
  - **Print**
  - **Video**
  - **Audio**
- **Media-independent processing**
  - **Entity recognition and disambiguation**
  - **Semantic Linking**
  - **Semantic Multi-Media Search**

Supported by
Federal Ministry of Economics and Technology

Print – Quality Assessment



Das Unternehmen erstarrt in dem Stil, den der alte Renz prägte. Aber das Publikum hat sich gewandelt; eine neue Generation ist herangewachsen. In den Wohlstand hineingeboren, frönen die jungen Bourgeois der neunziger Jahre nicht mehr der naiven Pferdeliebe ihrer reich gewordenen Väter. Das Protzentum der Emporkömmlinge aus der Gründerzeit ist in Blasiertheit und Snobismus umgeschlagen. Von der Manege werden nervenkitzelnde Sensationen und Effekte, von Podium und Bühne des Vergnügungsetablissements mondäne Eleganz und ein Hauch pseudo-romantischer Verruchtheit verlangt. Die altpreußischen Haudegen schwärmen für die sportlich-flotte, strammschenkelige Kunstreiterin, die den Logenbesuchern kokett – und dennoch bieder – ein Kaffhändchen zuwarf. Die Gardeoffiziere von 1895 finden eindeutig zweideutige Chansons »superb« und flirten im Chambre séparée mit einer offenherzig dekolletierten Varietétänzerin oder Tingeltangeldame.

Franz Renz versucht sein abflauendes, von der Konkurrenz hart bedrängtes Zirkusgeschäft zu beleben, indem er den Aufwand der Programme und den Luxus des Hauses steigert. Bei Premieren steckt er das Personal in Ballkleidern. Frisch erworbene Pferde werden, prächtig aufgezäumt, einzeln um die Piste geführt. In den Stallgängen liegen kostbare Teppiche, neben den Boxen stehen Blumenarrangements, und im Foyer wird ein Sektrestaurant eröffnet. Aber nichts vermag Franz Renz zu retten, nicht einmal die kurzpatriotische Monsterschau »1870/71«, für die er 150 Tänzerinnen, 400 Statisten und mehrere Männergesangvereine aufbietet. 1896 werden in der Revue »Lustige Blätter« die letzten zirkusspezifischen Elemente aufgegeben, zum finanziellen gesellt sich der künstlerische Bankrott.

Franz Renz resigniert. Am 31. Juli 1897 wird sein Zirkus geschlossen. Das Ensemble löst sich auf; Tiere und Requisiten kommen unter den Hammer des Auktionators.

Ernst Renz, der Sohn der Drahtseilkünstlerin Ottaxa, erwirbt bei der Versteigerung ein paar Dutzend Pferde und geht 1897 mit einem neuen, kleineren Zirkus Renz auf Gastspiel-

41

Supported by
Federal Ministry of Economics and Technology

Content-aware optimisation

Original

Supported by
Federal Ministry of Economics and Technology

Content-aware optimisation

Original

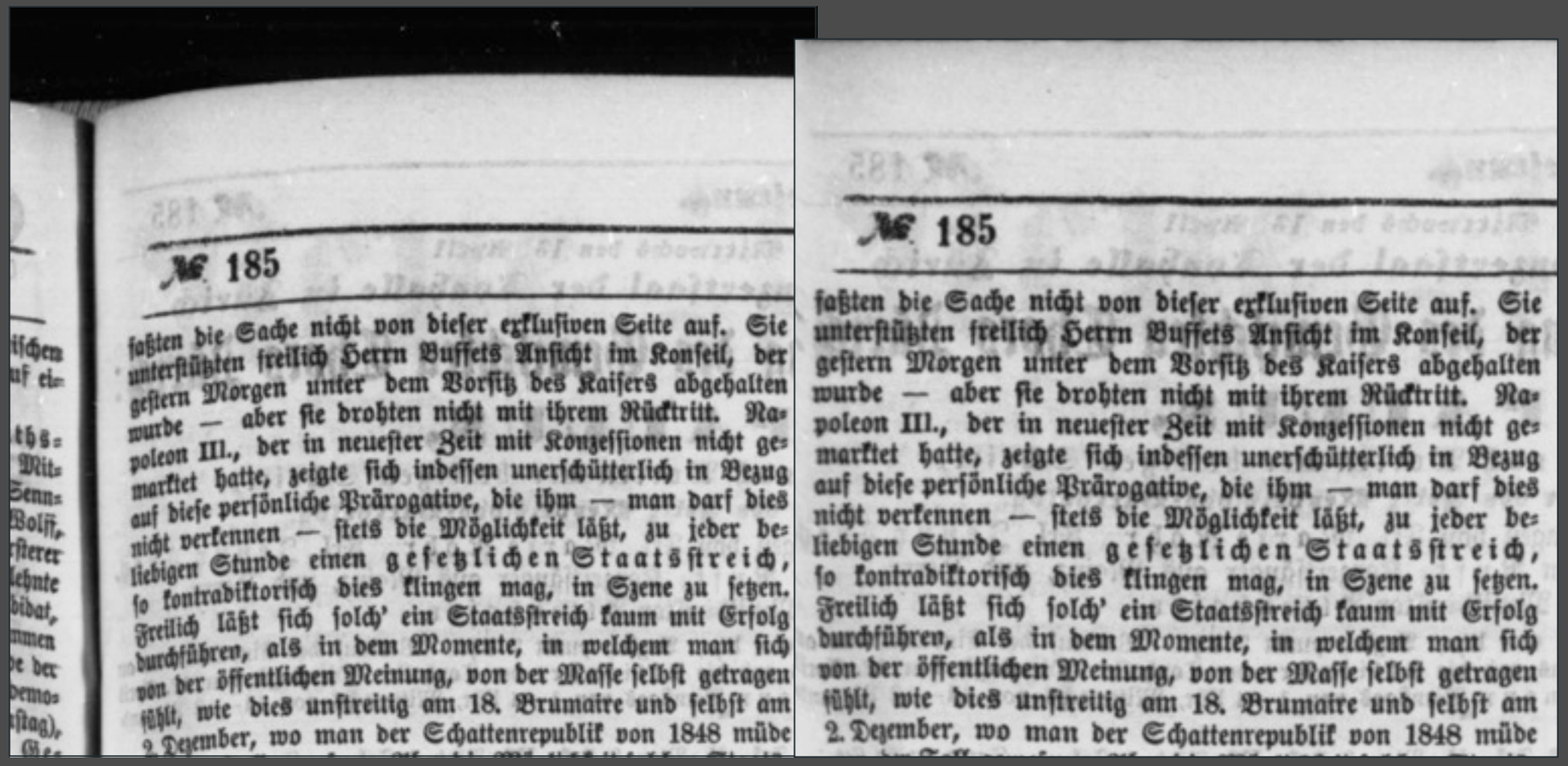
Otsu
Binarisation

Sauvola
Binarisation

CONTENTUS Approach +
Content-specific optimisation

Supported by
Federal Ministry
of Economics
and Technology

De-Warping



№ 185

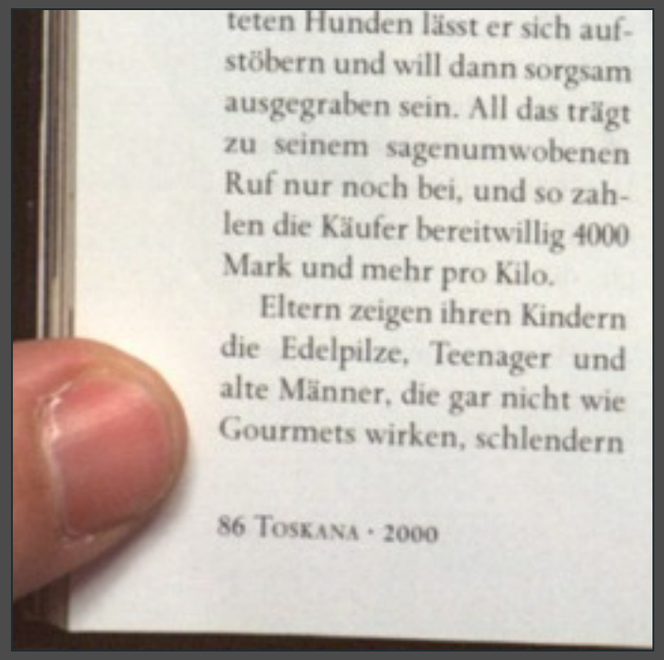
faßten die Sache nicht von dieser exklusiven Seite auf. Sie unterstützten freilich Herrn Buffets Ansicht im Konseil, der gestern Morgen unter dem Vorsitz des Kaisers abgehalten wurde — aber sie drohten nicht mit ihrem Rücktritt. Napoleon III., der in neuester Zeit mit Konzessionen nicht gemarktet hatte, zeigte sich indessen unerschütterlich in Bezug auf diese persönliche Prärogative, die ihm — man darf dies nicht verkennen — stets die Möglichkeit läßt, zu jeder beliebigen Stunde einen gesetzlichen Staatsstreich, so kontradiktorisch dies klingen mag, in Szene zu setzen. Freilich läßt sich solch' ein Staatsstreich kaum mit Erfolg durchführen, als in dem Momente, in welchem man sich von der öffentlichen Meinung, von der Masse selbst getragen fühlt, wie dies unstreitig am 18. Brumaire und selbst am 2. Dezember, wo man der Schattenrepublik von 1848 müde

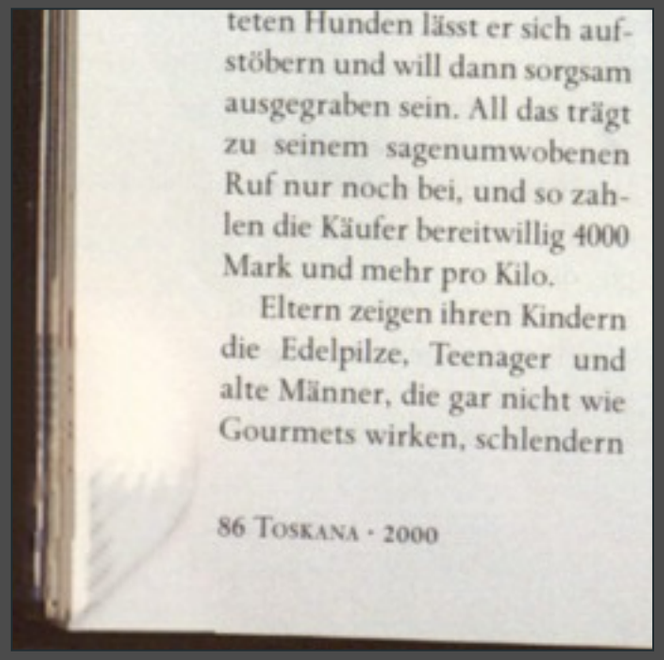
№ 185

faßten die Sache nicht von dieser exklusiven Seite auf. Sie unterstützten freilich Herrn Buffets Ansicht im Konseil, der gestern Morgen unter dem Vorsitz des Kaisers abgehalten wurde — aber sie drohten nicht mit ihrem Rücktritt. Napoleon III., der in neuester Zeit mit Konzessionen nicht gemarktet hatte, zeigte sich indessen unerschütterlich in Bezug auf diese persönliche Prärogative, die ihm — man darf dies nicht verkennen — stets die Möglichkeit läßt, zu jeder beliebigen Stunde einen gesetzlichen Staatsstreich, so kontradiktorisch dies klingen mag, in Szene zu setzen. Freilich läßt sich solch' ein Staatsstreich kaum mit Erfolg durchführen, als in dem Momente, in welchem man sich von der öffentlichen Meinung, von der Masse selbst getragen fühlt, wie dies unstreitig am 18. Brumaire und selbst am 2. Dezember, wo man der Schattenrepublik von 1848 müde

Supported by
Federal Ministry of Economics and Technology

Removal of Unwanted Objects


teten Hunden lässt er sich auf-
stöbern und will dann sorgsam
ausgegraben sein. All das trägt
zu seinem sagenumwobenen
Ruf nur noch bei, und so zah-
len die Käufer bereitwillig 4000
Mark und mehr pro Kilo.
Eltern zeigen ihren Kindern
die Edelpilze, Teenager und
alte Männer, die gar nicht wie
Gourmets wirken, schlendern
86 Toskana · 2000


teten Hunden lässt er sich auf-
stöbern und will dann sorgsam
ausgegraben sein. All das trägt
zu seinem sagenumwobenen
Ruf nur noch bei, und so zah-
len die Käufer bereitwillig 4000
Mark und mehr pro Kilo.
Eltern zeigen ihren Kindern
die Edelpilze, Teenager und
alte Männer, die gar nicht wie
Gourmets wirken, schlendern
86 Toskana · 2000

Supported by
Federal Ministry of Economics and Technology

Page Segmentation

- **Automatic identification of:**
  - **Articles**
  - **Headings**
  - **Tables of content**
  - **Figures / pictures and captions**
  - **Bullet Points**

Supported by
Federal Ministry of Economics and Technology

Video – Quality Assessment

- Quality survey essential to determine value of content
- Use perceptual no-reference quality metrics
- Check for specific image artifacts
- Based on human visual models
- Use specific restoration modules

Restoration

- Current solutions require at least 4 hours of work per hour of video
- Automation necessary

Supported by
Federal Ministry of Economics and Technology

Video – Restoration

Supported by
Federal Ministry of Economics and Technology

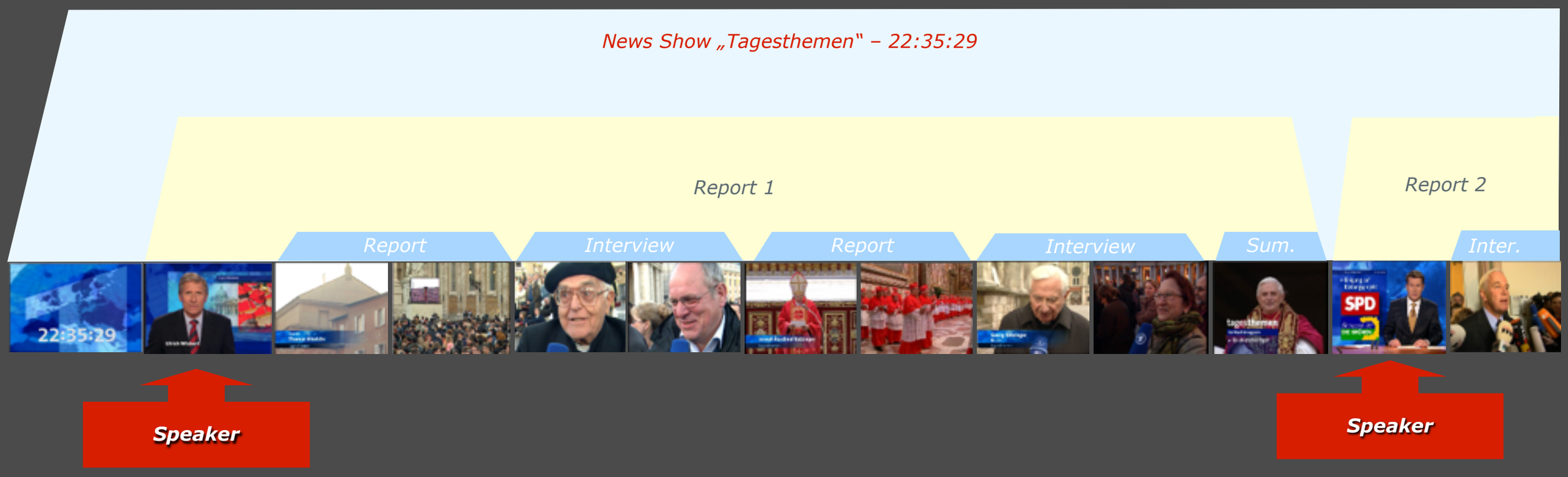
Video – Restoration

Scratches are automatically identified

Scratches are automatically removed

Supported by
Federal Ministry of Economics and Technology

Video – Segmentation



Supported by
Federal Ministry of Economics and Technology

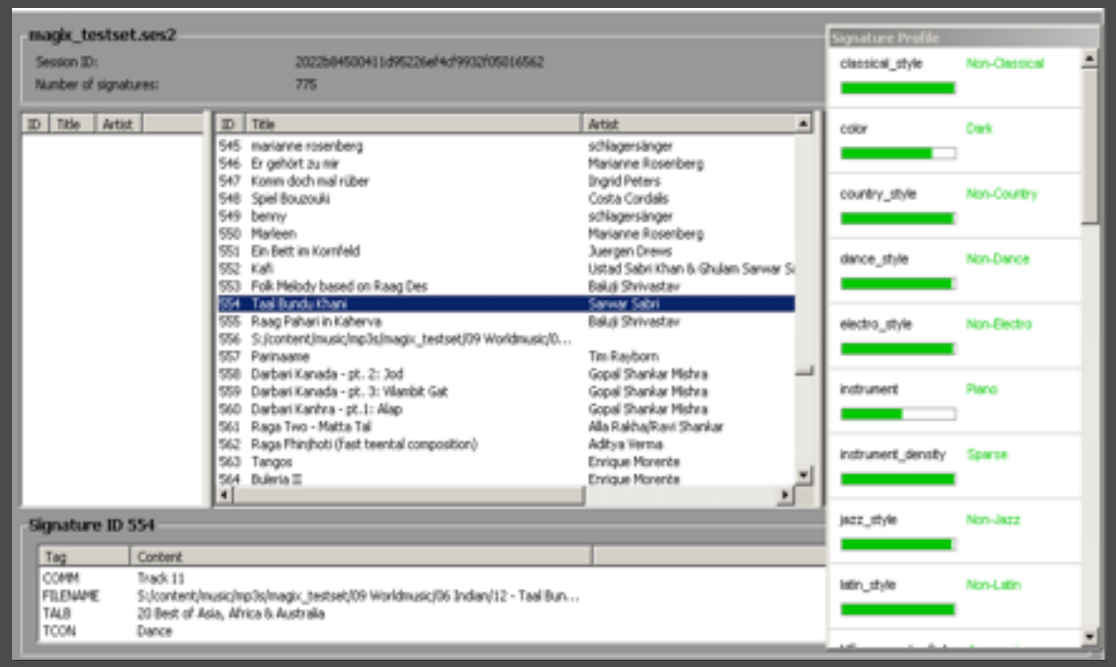
Video – Annotation

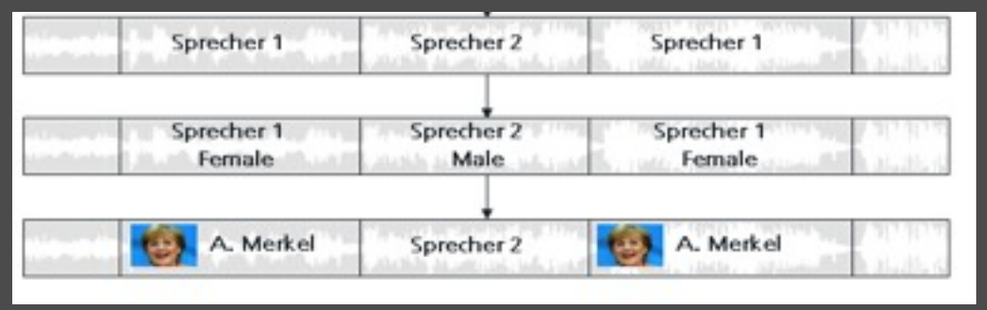
- Face detection and annotation
- Prerequisite for further indexing
- Text and logo detection

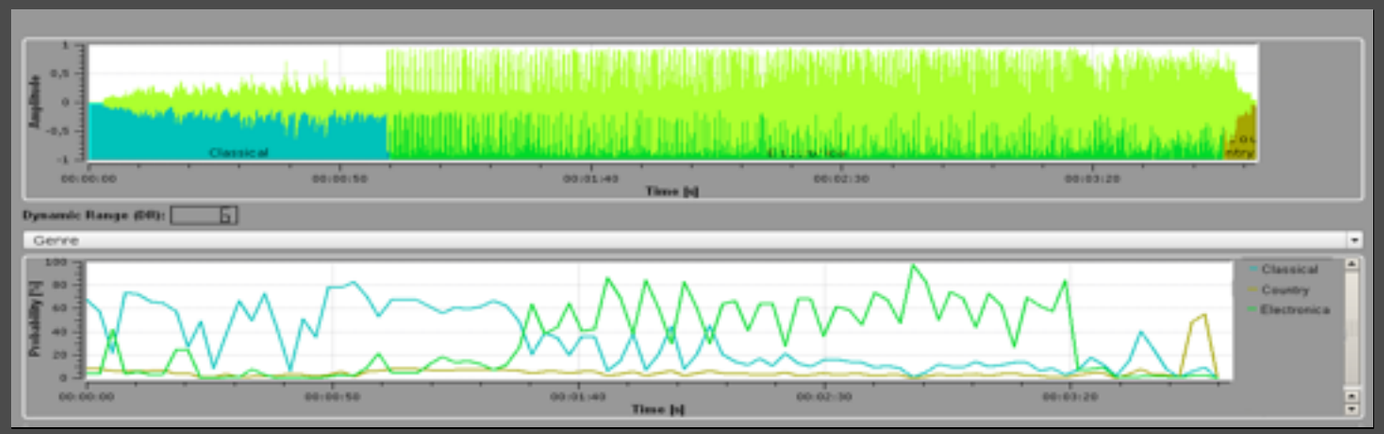
Supported by
Federal Ministry of Economics and Technology

Audio

- Segmentation speech / non-speech
- Extraction of musical features
- Speech transcription
- Speaker recognition
- Similarity search







Supported by
Federal Ministry of Economics and Technology

Media-Independent Processing

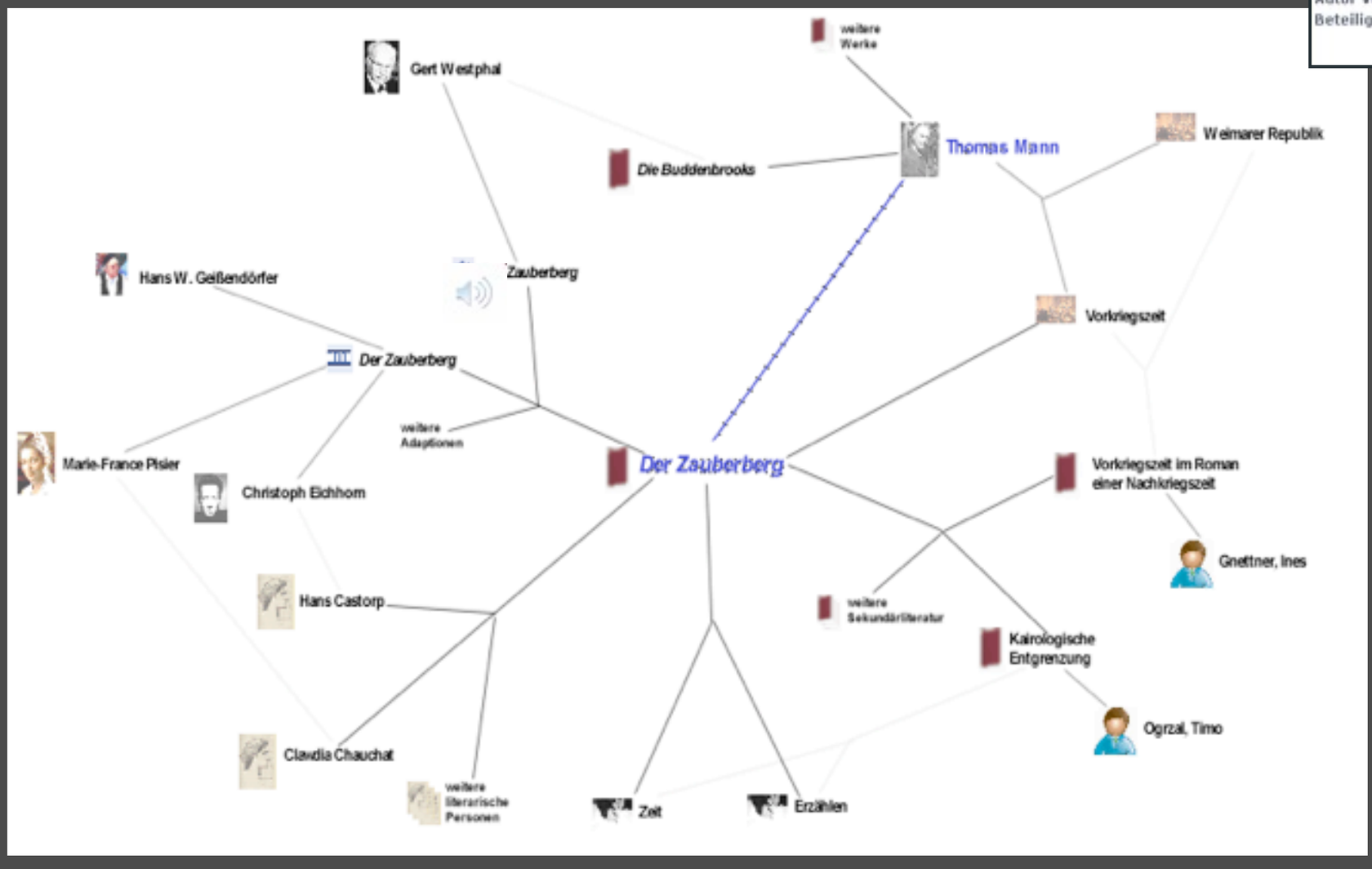
Supported by
Federal Ministry of Economics and Technology

Named Entity Disambiguation

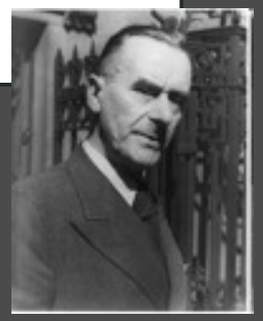
- **Named entities are extracted from text**
- **Context is used for disambiguation**
- **Compare to reference text, e.g. from Wikipedia**

Supported by
Federal Ministry of Economics and Technology

Semantic Linking



Person	Mann, Thomas (männlich)
Weitere Angaben	Nobelpreis für Literatur 1929; emigrierte 1933 in die Schweiz, 1938 in die USA, 1952 in die Schweiz; lebte ab 1933 in der Schweiz und den USA
Beziehungen zu anderen Personen	Mann, Katia (Ehefrau); Mann, Erika (Tochter); Mann, Frido (Enkel); Mann, Klaus (Sohn); Mann, Golo (Sohn); Mann, Monika (Tochter); Mann-Borgese, Elisabeth (Tochter); Mann, Michael (Sohn); Mann, Stefan (Urenkel); Pringsheim, Peter (Schwager); Reach, Hilde (Sekretärin)
Autor von / Beteiligt an	3102 Publikationen 1. *Der Tod in Venedig [Tonträger]* *München : Der Hörverl., 2009*



Authority Files

Wikipedia

MusicBrainz

Supported by
Federal Ministry of Economics and Technology

Semantic Multimedia Search

provides

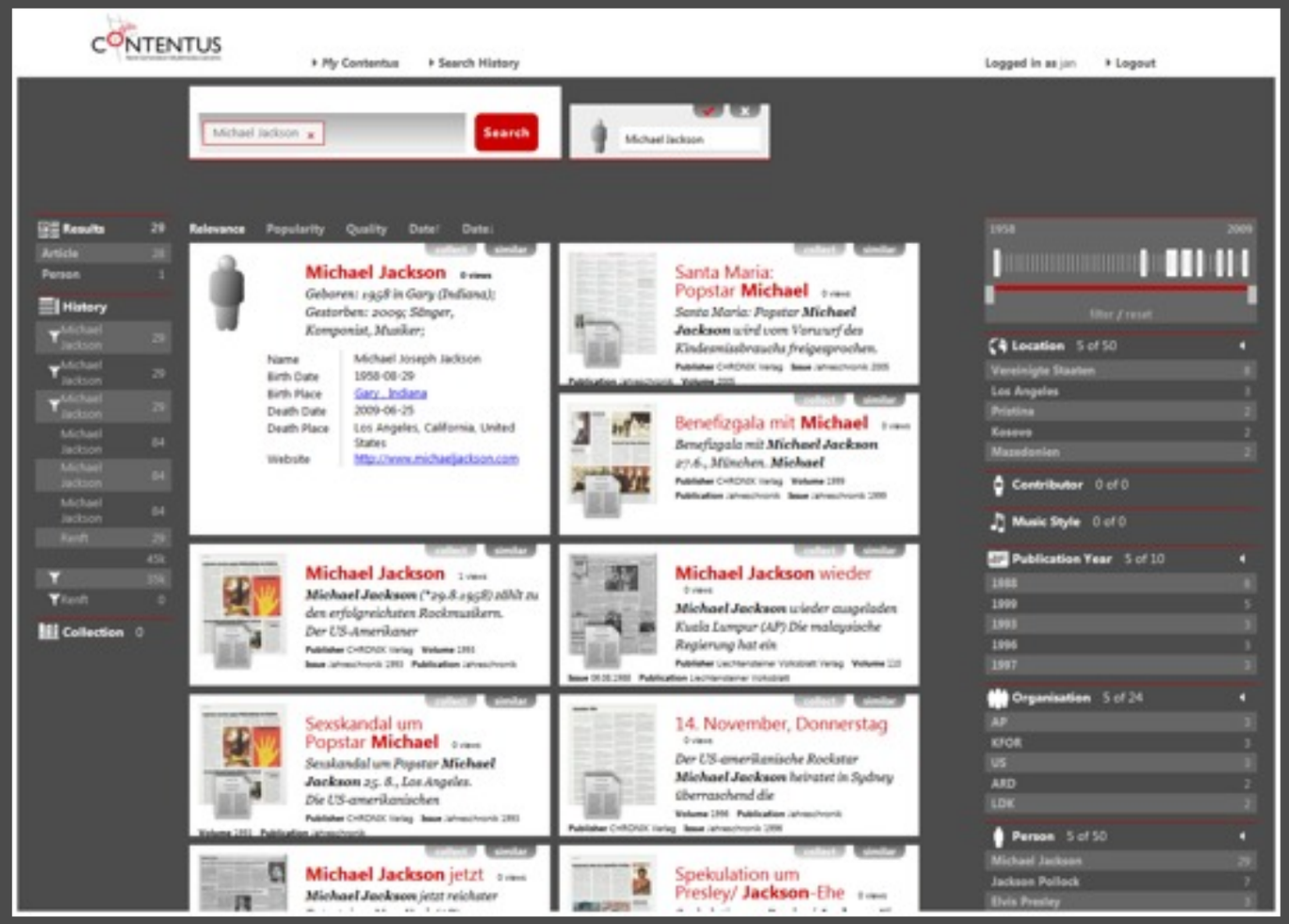
- **Seamless searching in multimedia**
- **Query expansion / narrowing**

combines

- **Full-text search and Semantic Web Stack (RDF based ontology)**

integrates

- **Multiple media (Video, Audio, Text)**
- **Clickable filter facets**
- **Text classification**
- **Similarity search**



Supported by
Federal Ministry of Economics and Technology

The CONTENTUS-Collection

- Digitisation of Music Information Center of the GDR
- Now a special collection of the German Music Archive
  - 1600 books
  - 200.000 press clippings
  - 4000 audio records
  - 10.000 photos

- Other media
  - Newspapers (Neues Deutschland)
  - News broadcasts
  - Historical film material

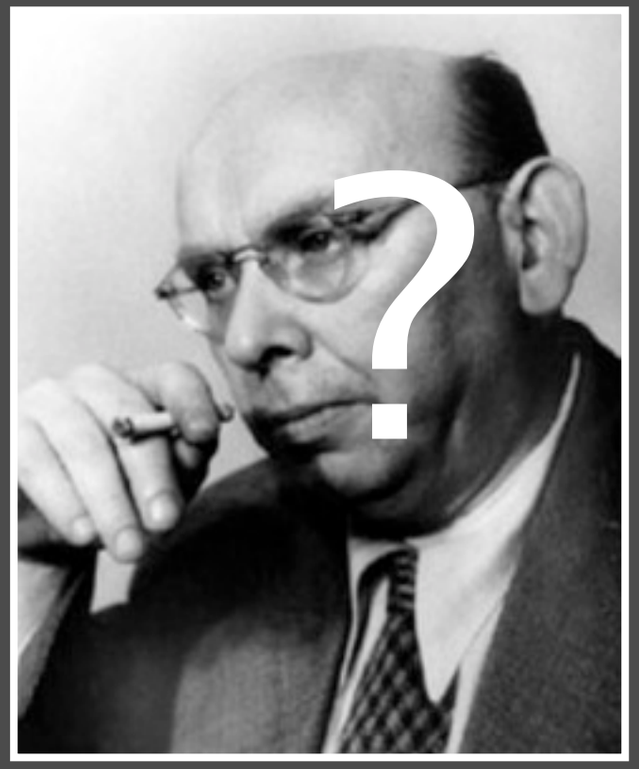
Supported by
Federal Ministry of Economics and Technology

Demonstration

Supported by
Federal Ministry of Economics and Technology

SMMS Demo 1 – Who was Hanns Eisler?

- Hanns Eisler, GDR-Composer
- Composed the music of the GDR national anthem



SMMS-Features:

- Multimedia
- One unified index
- Facetted search
- Roles of persons
- Timeline

Supported by
Federal Ministry of Economics and Technology

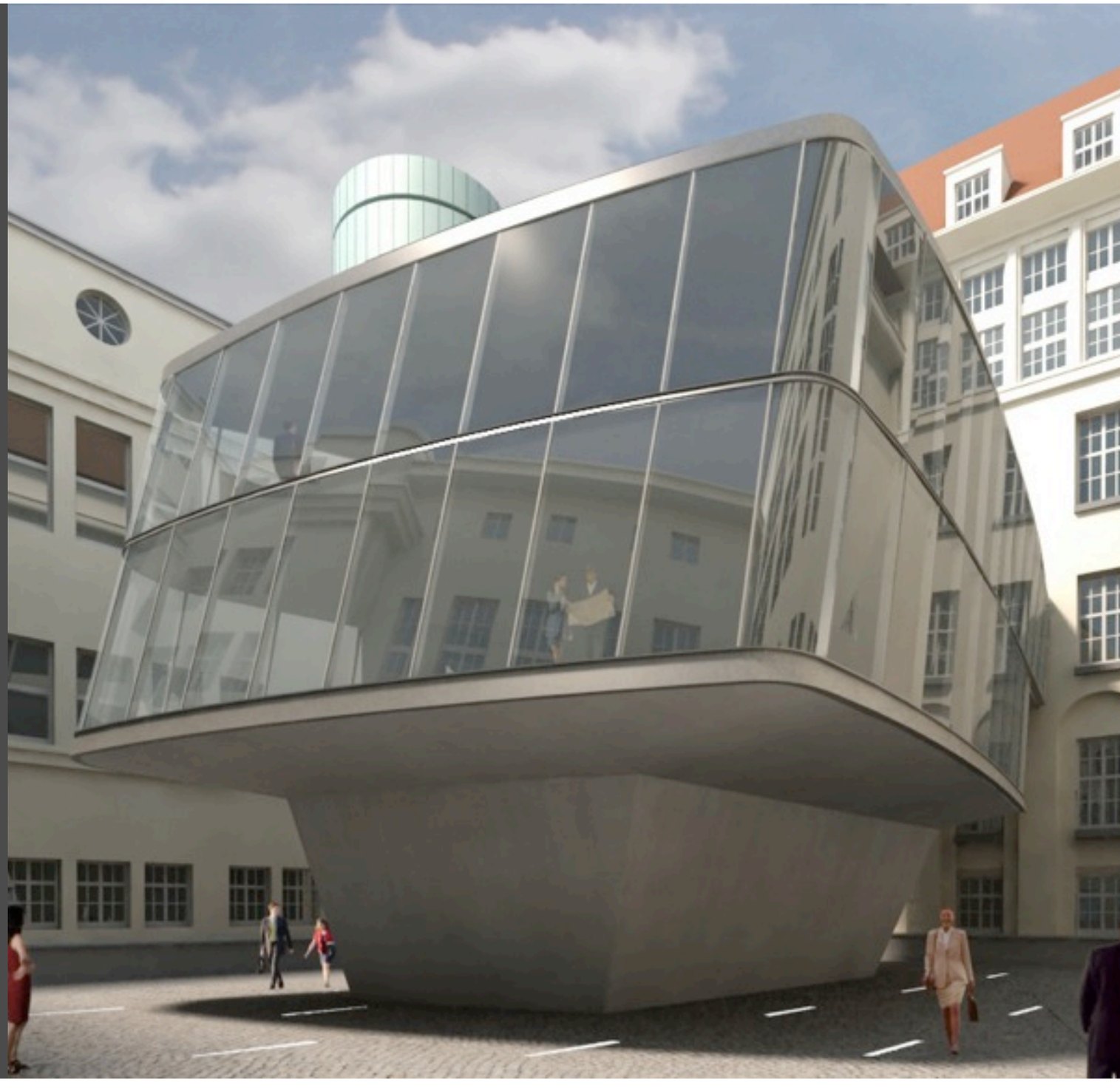
Summary

- Challenges and obstacles on the way to the Next Generation Multimedia Library
- Our project
  - High degree of automation during complete processing chain
  - Semantic multimedia search engine

Supported by
Federal Ministry of Economics and Technology

Conclusion

- The next generation library is not here yet
- We‘re on the way…
- We need your! help



Supported by
Federal Ministry of Economics and Technology

Thank you for your attention

- **Visit THESEUS:**
  http://www.theseus-programm.de/

- **Mail us:**
  a.hess@d-nb.de
  jn@acosta-consult.de

Supported by
Federal Ministry of Economics and Technology