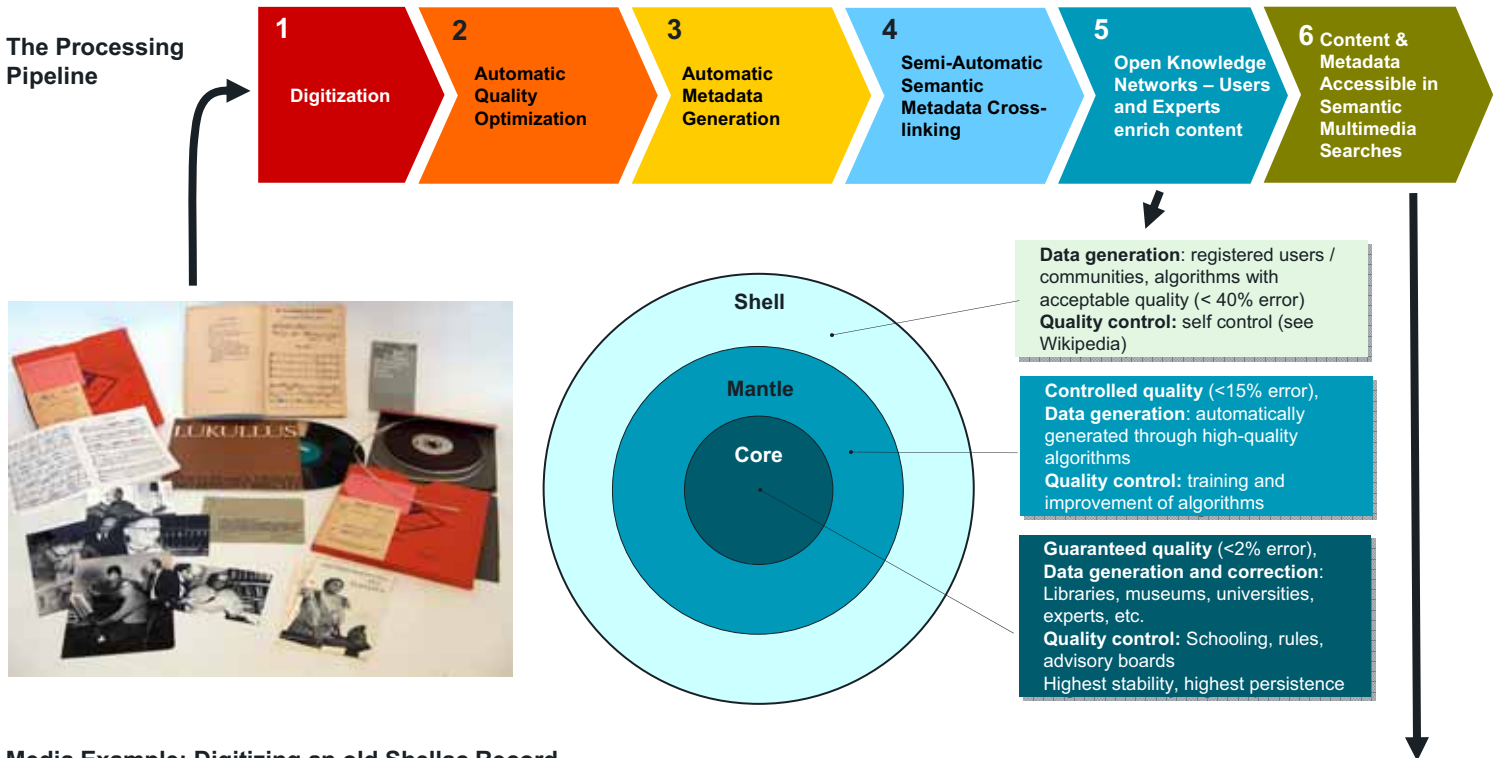


# CONTENTUS – From Information Asset to Knowledge Network

Berениke Loos, Jan Hannemann, Jürgen Kett, Reinhard Althenhöner, Nicolas Flores-Herr  
German National Library - Information Technology - Adickesallee 1, D-60322 Frankfurt am Main - b.loos@d-nb.de

## Motivation: Making our Cultural Heritage Accessible Through Semantic Multimedia Searches

- **Problem: Deterioration of Media and Accessibility** – Ever since knowledge has been transferred onto books, film, tapes etc., multimedia collections have comprised a unique and invaluable cultural heritage. Due to obsolete analog and digital formats and the physical deterioration of traditional storage media (books, documents, film, tapes etc.) this unique knowledge is in jeopardy. To preserve our cultural heritage for future generations, the ability to access knowledge is as important as its adequate archiving.
- **Incorporating Social Networks** – To access knowledge and digitized cultural heritage, social networks and communities will become fundamental platforms representing academic and cultural institutions as well as private and commercial content providers. These platforms will facilitate the visibility of content owners and, simultaneously, draw attention to their content and the institution itself.
- **Community-Based Sharing and Creation of Resources** – A challenge faced by national libraries is a new publishing paradigm fostered by Web 2.0 and Semantic Web technologies. With the advent of Web 2.0 and Semantic Web technologies the object-centric (e.g. book, CD, DVD etc.) archetype of art and knowledge publishing on one side and consumption on the other side has evolved into a community-based sharing and creation of resources and relations between them.
- **Solution: The CONTENTUS Project** – The German National Library cooperates with several research institutions and industry partners to develop technologies that facilitate the preservation of our cultural heritage. The core element of this project is an efficient processing pipeline "from information asset to knowledge network", that describes the workflow for processing our cultural heritage from traditional storage media to community-enriched content accessible through semantic multimedia searches.



## Media Example: Digitizing an old Shellac Record

- 1 The source is digitized, which involves sampling the record and scanning the record covers. A basic set of technical metadata (audio format, sampling rate, recording length, source type, type of digitization device etc.) is also generated during the digitization process.
- 2 The quality of the resulting audio file can be optimized by removing background noises and compensating for scratches, if desired. Similarly, the quality of cover scans is automatically optimized by adjusting image parameters (e.g. tonal curve, sharpness, hue, saturation).
- 3 Further information about the source is generated. This includes descriptive metadata (e.g. a textual version of the cover text obtained via OCR, and the subsequent identification of the composer etc. from the text).
- 4 The metadata is connected with existing content, e.g., the composer is linked with the appropriate entry in the "Personennamendatei" (name authority file). Similarly, an analysis of the audio file can determine other audio sources in the database with similar properties and link those as well. This step is a semi-automatic extension of the knowledge base.
- 5 In a community-based approach, the existing content is corrected and extended by (expert) users. For example, a user could add the fact that this recording was performed by the Staatskapelle at the Staatstheater Dresden in March 1962.
- 6 Finally, the content is made available via semantic multimedia searches, where users can access it.



## Project Partners:

