Archive Migration Solutions for the German National Library

Autor und Fotos: Peter Kaminski

The German National Library "Deutsche Nationalbibliothek" has the statutory duty to collect, preserve, archive and provide public access to cultural assets.

The German Music Archive "Deutsches Musikarchiv" takes on this task regarding musical supplies and sound recordings, and works as the audio-bibliographical center of information in Germany. The National Library receives two copies of every book, magazine and audio carrier published in Germany for its records.



The Music Archive was founded in 1970 in Berlin. Since 2010 it is now based in Leipzig, at the site of the National Library. With its fourth expansion, new depot rooms were built. This expansion consists of three floors below ground and four floors overground. Also new are the Music Reading Room and a newly equipped recording studio.

Audio Monitoring

The archive is used by many students and musicologists, but also by labels which are looking for material that is no longer available on the market. For monitoring the audio, various possibilities exist in the archive. First, there is an auditorium for a larger group of listeners. Also historical playout systems are ready to use here.



The audio monitoring rooms are situated on two floors and consist of 18 user seats, of which four have pull-out piano keyboards. All seats use top quality audio-technical components: PCs with RME audio cards type HDSP9632, Lake People HPA V100 headset amplifier, and AKG K271 MKII and Sennheiser HD-650 headsets.

Phoca PDF



In the listening room it is possible to play audio material in 5.1-surround.



Archive

To give an idea of the size of the archive, here are a few numbers. The reported assets of the Music Archive in 2010 were stated as follows: Optical media (CDs, DVDs, SACDs): more than 462,000 Vinyls and Compact Cassettes: 335,000, Historical sound carriers like shellac records, piano rolls and audio cylinders: 163,000.



We talked on premises in Leipzig with Hinnerk Gehrckens from the IT system development about strategies of archiving the collected media.

Hinnerk Gehrckens: "Two years ago, we started a major project on audio migration. First, the audio CDs are ingested, and then we will continue with the ~30,000 audio compact cassettes and vinyl and shellac records. Additionally, we will migrate other historically valuable audio carriers in other formats. The migration project will run until about 2018, however, the funding is not secured, yet."

Phoca PDF



The central digital long time archive is located in Frankfurt am Main. So work is not only done locally for the Music Archive in Leipzig, but for the long time digital archive of the entire National Library with a storage capacity of hundreds of terabytes. Analog media is digitized in resolutions up to 96kHz/24bit. But not all is fixed yet.

Phoca PDF



The German Music Archive as a department of the German National Library watches with great concern the increasing deterioration of their CD collection, especially of the early items from the 80s. The statutory mission of the German National Library, to collect all media creations published in Germany, to preserve them and make them accessible to the public, is no longer possible in the form of audio CDs, at least not with their original audio carriers.

As a practical resolution, only preservation of their content on mass storage is feasible. Additionally, this enables a significantly improved deployment and usage scenario at the site of the National Library in Frankfurt and the new location of the Music Archive in Leipzig.

CD-Audio Migration

The German Music archive decided against outsourcing and for the purchase of own CD-ingest systems after testing different options . This was economically the most sensible option and also offers the logistic advantage of only having to transport the CDs in-house. For ingestion of the CDs, the German Music Archive chose the product CD-Inspector, a robotic solution by the German archive specialists Cube-Tec International.

Phoca PDF



proaudio.de: What has led to the decision on this system?

Hinnerk Gehrckens: In the short list at the end of the decision process, only two systems were in question. All standard CD-ripping software available on the market is out quickly, if one takes into account what types of errors get detected, and how. An important point for us was to gain knowledge of the exact types of errors the CDs to be migrated have. For further testing, we prepared CDs with physical defects and looked closely at the reported errors. The result was that there simply was no alternative to Cube-Tec's CD-Inspector. In these comprehensive tests, it has also been discovered that there is a correlation between the type of detected errors and their audibility. Based on these findings Cube-Tec has subsequently implemented an algorithm that sorts the material according the "audible" and "non-audible" defects. As the migration goes on it is planned to optimized the setting further.

proaudio.de: Who does the CD day-by-day digitizing work?

Hinnerk Gehrckens: This work is done by student assistants, while the project is supervised by me and another colleague of the IT department. Since 2012, after the move to Leipzig, the project will be continued here. As for the performance of the system, it is able to import 500 CDs overnight, so the required amount can be

Phoca PDF

processed easily. This allows us to import about 100,000 CDs per year. We planned for four to five years of work for the complete migration of our whole CD collection. One should not underestimate the extra effort, because this project does an inventory of the CD Audio portfolio, as well.

proaudio.de: How high is your error quota on the migrated CDs?

Hinnerk Gehrckens: For the most critical years 1983 to 1985 about five percent of audio CDs that have been marked as flawed by the mass transfer process (read errors and other errors). In the majority of these CDs, these errors were correctable during the migration process. Other CDs can be read without errors after cleaning or with special CD drives supplied by Cube-Tec for problematic Audio CDs. This work is not yet complete for all 17,000 CDs transferred until now, but the prediction of CDs unable to be migrated stays below one percent.

Studios

As part of the sound studio installation in Leipzig, accomplish by Salzbrenner/Stagetec, Cube-Tec installed a QUADRIGA archive ingest workstation. This installation includes, additional to the QUADRIGA with 4 import modules, a restoration workstation and a customization for the input of special metadata. Up to four playback machines can be used simultaneously to digitize the archival material. The playback devices are distributed in a studio with Stagetec AURUS mixer with NEXUS router and high-quality 5.1-Speaker set-up.

Phoca PDF



[left to right: Torsten Ahl and Hinnerk Gehrckens of Deutschen Musikarchiv and Tom Lorenz of Cube-Tec International]

Before moving to Leipzig the German National Library Music Archive had his main sound studio in Berlin. Only some of the "good old" analog equipment like EMT turntables and the Studer tape machines have been moved from this Berlin studio to Leipzig (see picture below).



We talked to Torsten Ahl, responsible for overdubs and restoration in the studio. "Mostly we get requests from the users of the Music Archive. We try to do the signal extraction in such a high a quality that no additional processing is needed, but of course there are sound carriers that require some digital restoration before we can hand over the Compact Discs copy to the user. As for the real-time playout, until the new staging system is set-up, we use another system which creates CD-Images which are able to be played back directly, without been copied to an optical carrier again.

Phoca PDF



On our AudioCube Restoration Workstation from Cube-Tec which we use in addition to the QUADRIGA ingest workstation we have a fine suite of restoration tools in use. The computer for the QUADRIGA ingest workstation is setup with much drive space in a RAID configuration to hold the latest digitizations locally, so that they can be reworked, if necessary. To increase our flexibility further, we hope to be able add more restoration tools in future. Particularly desirable would be if we could get more digitization tools for the video area. At some point, the Blu-ray discs will need a migration as well, as they are just as vulnerable as CD Audio."

Hinnerk Gehrckens added: "In comparison, vinyl records are a relatively robust sound carrier. Shellacs are different, because they also are aging more quickly."

Phoca PDF



proaudio.de: What about the networking?

Torsten Ahl: "We use a NEXUS matrix router with very flexible possibilities. While I use the re-recording studio to digitize and restore, I can supply audio data to the QUADRIGA in our studio with 96 kHz in parallel, for example. Thanks to the NEXUS there are many possibilities. The control room is networked, as well. In the future, we plan to make the auditorium accessible, too. This step already in preparation."

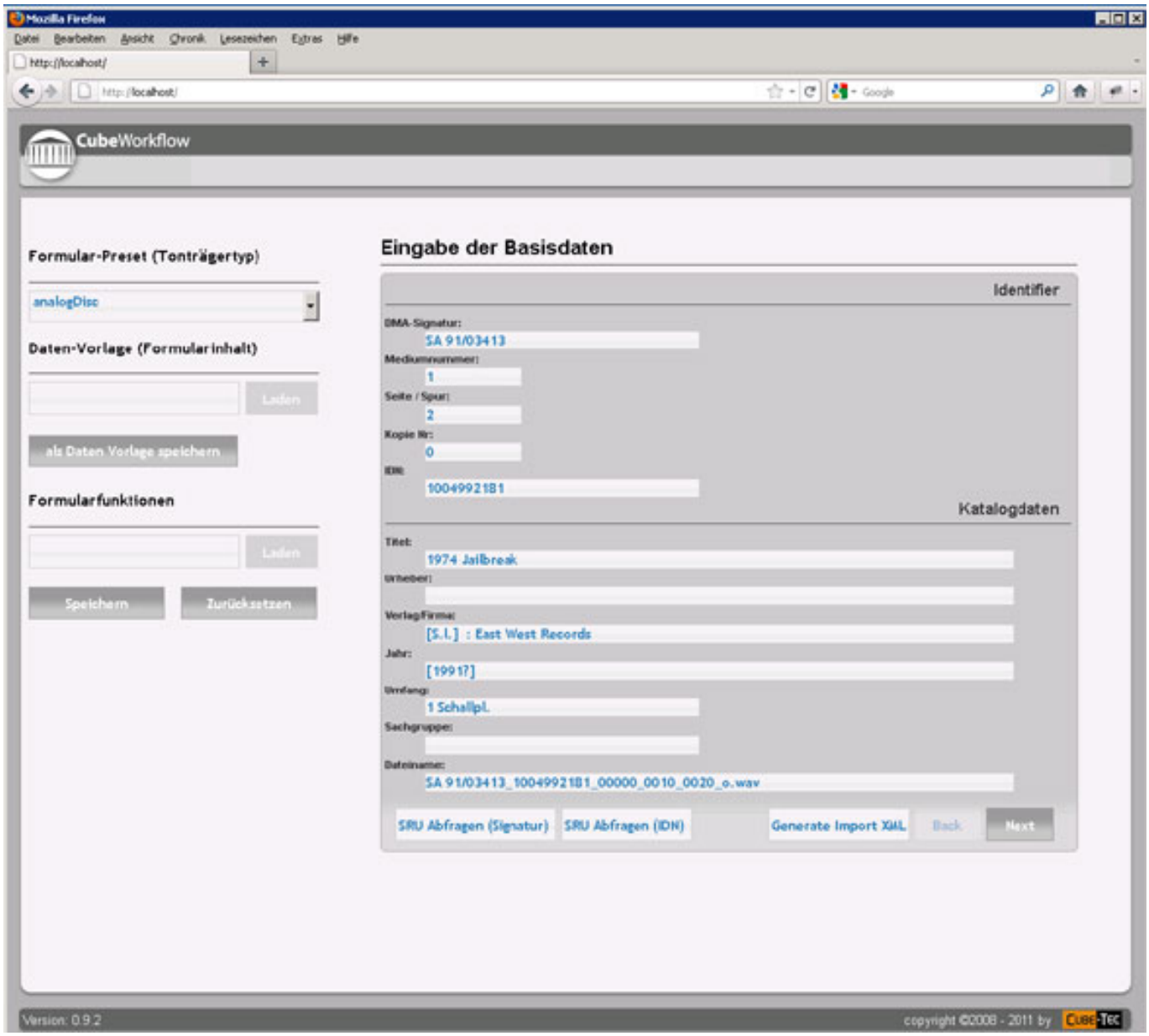
Phoca PDF



Gathering The Metadata

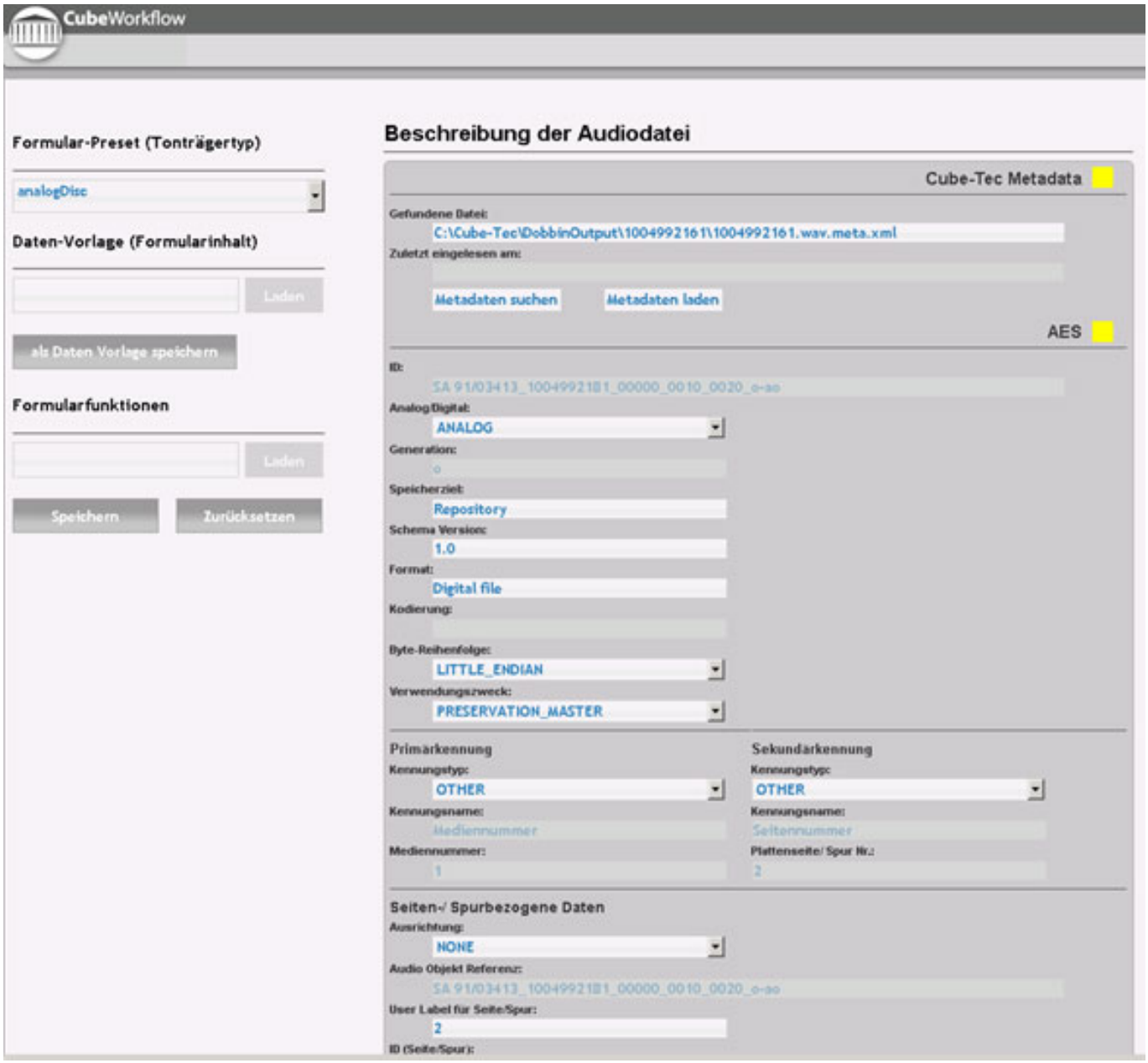
Regarding metadata capturing Hinnerk Gehrckens told us: "A big improvement compared to the technical infrastructure in Berlin is, that - through the software from Cube-Tec - we are able to take over the files that Thorsten Ahl digitized here directly into our long-term storage, because all necessary metadata is automatically captured as well. After our earlier digitizations in Berlin, this metadata had to be captured separately after the audio digitization. Now everything is done in one step. So when we get a user request, the material is digitized and automatically added to the long-term storage and can be accessed in the reading room with a mouse click, without a lengthy ordering process.

Phoca PDF



A special input form, which extends over three pages us used to capture all relevant descriptive metadata. For historical reasons, the Music Archive had its own legacy identifier, which has to be translated to the current unique identifier now used by the German National Library. Through the input form, this identifier first is searched for in a database. For later use, the track positions set by the QUADRIGA software are important. Additionally, metadata on the original sound carrier can be entered in detail, such as media type, size, material, ribbon type, playback speed etc.

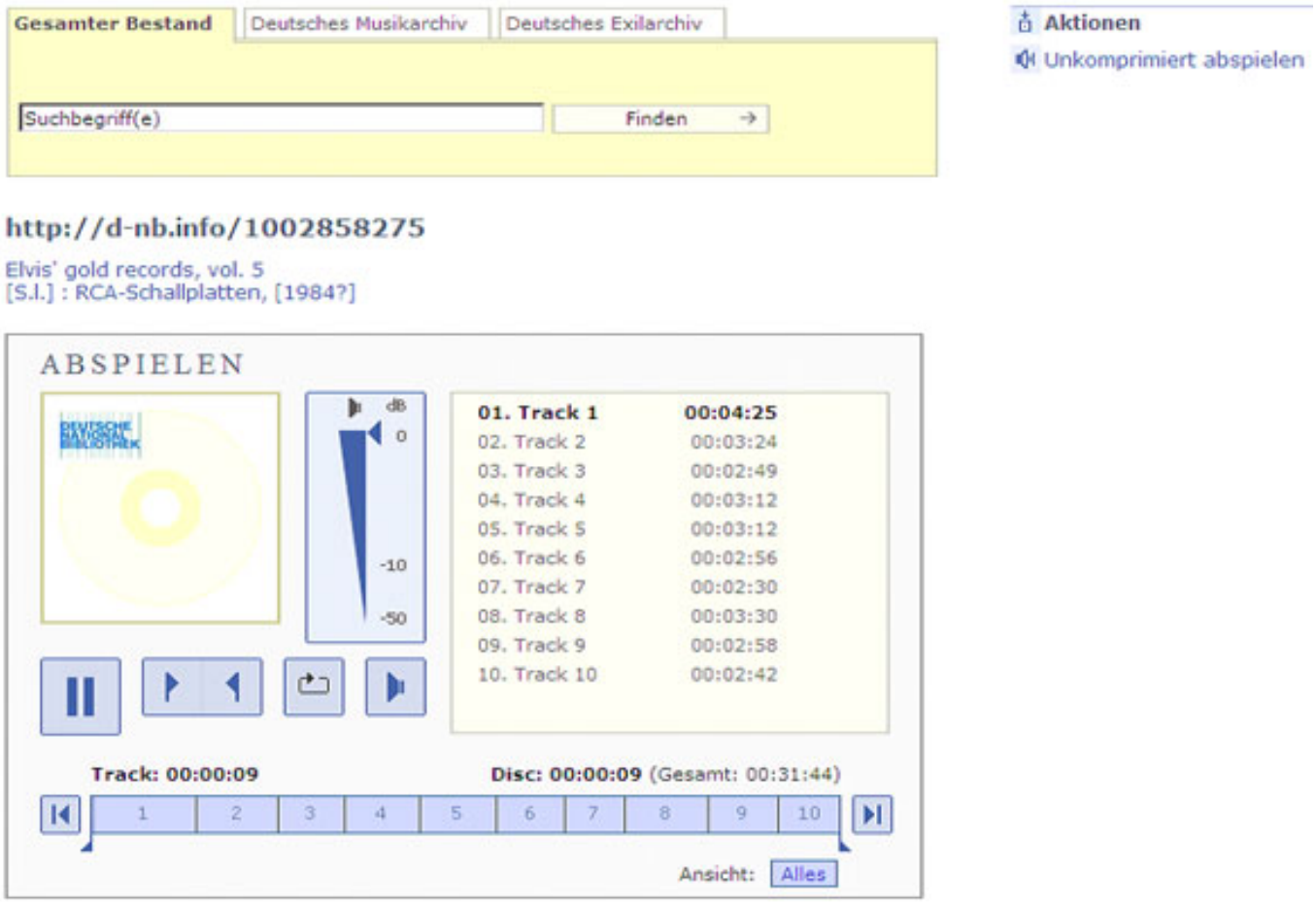
Phoca PDF



The input parameters displayed in the browser-based input mask are adapted to the particular media type. We have many different sound carriers up to exotic formats like audio cylinders and magnetic wire recordings. Furthermore, metadata about the digitization process itself is included, like type of playback device, needle, playback speed, analog to digital converter, sample rate and more.

Through presets, typical settings can be called back directly into the form. The input format follows the new metadata standard AES57 introduced by the Audio Engineering Society in 2011.

Phoca PDF



This data is stored into an XML file format and is preserved together with the audio essence. Currently, our catalog system is not able to work with this additional metadata fields, but it is one of the next steps to make this information accessible to our users.

www.dnb.de
www.cube-tec.com

Phoca PDF