# > Semantic Web Use Cases and Case Studies

## Case Study: A Digital Music Archive (DMA) for the Norwegian National Broadcaster (NRK) using Semantic Web techniques

*Dr. Robert H.P. Engels, ESIS, and Jon Roar Tønnesen, NRK*

*September 2007*

### General Description

**Introduction**

Digitizing the complete radio and television broadcasting production process is a major undertaking in many public and commercial broadcasters. Many public broadcasters possess enormous archives often ranging back 60+ years to include pre-WWII sound assets on bakelite, wax and even(!) chocolate. Whereas the older assets show a remarkable resistance against the tooth of time, more modern storage formats like digital video tape, certain CD's, tapes etc. are not that robust. At NRK (Norwegian National Broadcaster) it is expected that many tapes recorded in the late 80's and early 90's cannot be recovered within 5 years if no immediate action is taken and digitzing the assets is considered a correct way of action for preserving assets for the future.

Another effect of digitizing, besides the preservation argument, is that the assets become more easily available, with many manual or labor-intensive steps in a production process eliminated. At NRK it is estimated that during a year of broadcasting a maximum of 5% of all in-house available assets are really used in broadcasting.

Semantic Web technology is primarily used for enclosing the enormous amounts of metadata on music tracks available within the archives so that a larger amount of the "hidden treasures" will be used in broadcasting, potentially providing the broadcaster with an advantage over the competition, being better informed and more interesting.

**System objectives and components**

Metadata for all registered music in NRK has been handled by a group of librarians from the archiving department. From 1962 to 1982 all registrations for incoming records where made on paper, from 1981 until 2007 all registrations where made in a simple, file-based and non-relational database. Objectives of the system are to:

- significantly increase the efficiency and effectiveness of the production process in public broadcasting
- increase competitive advantage by make visible the "hidden assets" that are possessed by the broadcaster (you do not own what you cannot find)
- implement a solution capable of integration and alignment with several other production archives for radio, TV, movies, stills, etc. available within the organization.
- digitize all music tracks in a high-resolution format for usage in production allowing for multi-channel output
- digitalization of all qualified metadata available within the organization
- representation of all metadata using Semantic Web priniciples (graph-like, semantic network structure)
- make unexpected and potentially interesting relations visible to the users (journalists and program makers)
- allowing easy access to the archives and one-click ordering of relevant material streamed directly into TV and Radio production.

The complete system has been taken into production during summer 2007.

**Modeling the repository**

An important principle for design was the wish to use Semantic Web technology for the solution as to bring a Semantic Web scenario to end users in a commercial environment. The designed solution is business critical as well as a real-world production environment. During initial tests, some drawbacks were identified, as well as potential opportunities for expanding the solution:

- **Available metadata:** despite of mainly being available in analogue format, the data kept a high-quality and was rather well-structured. This does not make "translation" an easy task, but was a good help. Paper archives where scanned and post-processed with linguistic technologies to enhance results as much as possible. Authority lists on authors, composers and groups were generated from several sources for quality enhancement. The available electronic cards where augmented with relations and object types using manually defined business rules which could be automatically run on the data set available.
- **Semantics**: adding semantics was done during the same translation phase. All cards were first translated in an XML based semantic representation, on which business rules were used in order to get a SW representation.
- **Populating the repository**: after interpretation and cleansing, the repository was filled with all available data including, objects, properties and relations.
- **Administration & Maintenance**: Since a specialized tool for registration and maintenance of metadata in a broadcaster's music archive was not available, a tool was implemented for this very goal. The Archivist Module (as it is called) has web extraction capabilities and automated (linguistic-based) support for internet based information extraction (see Figure 1.

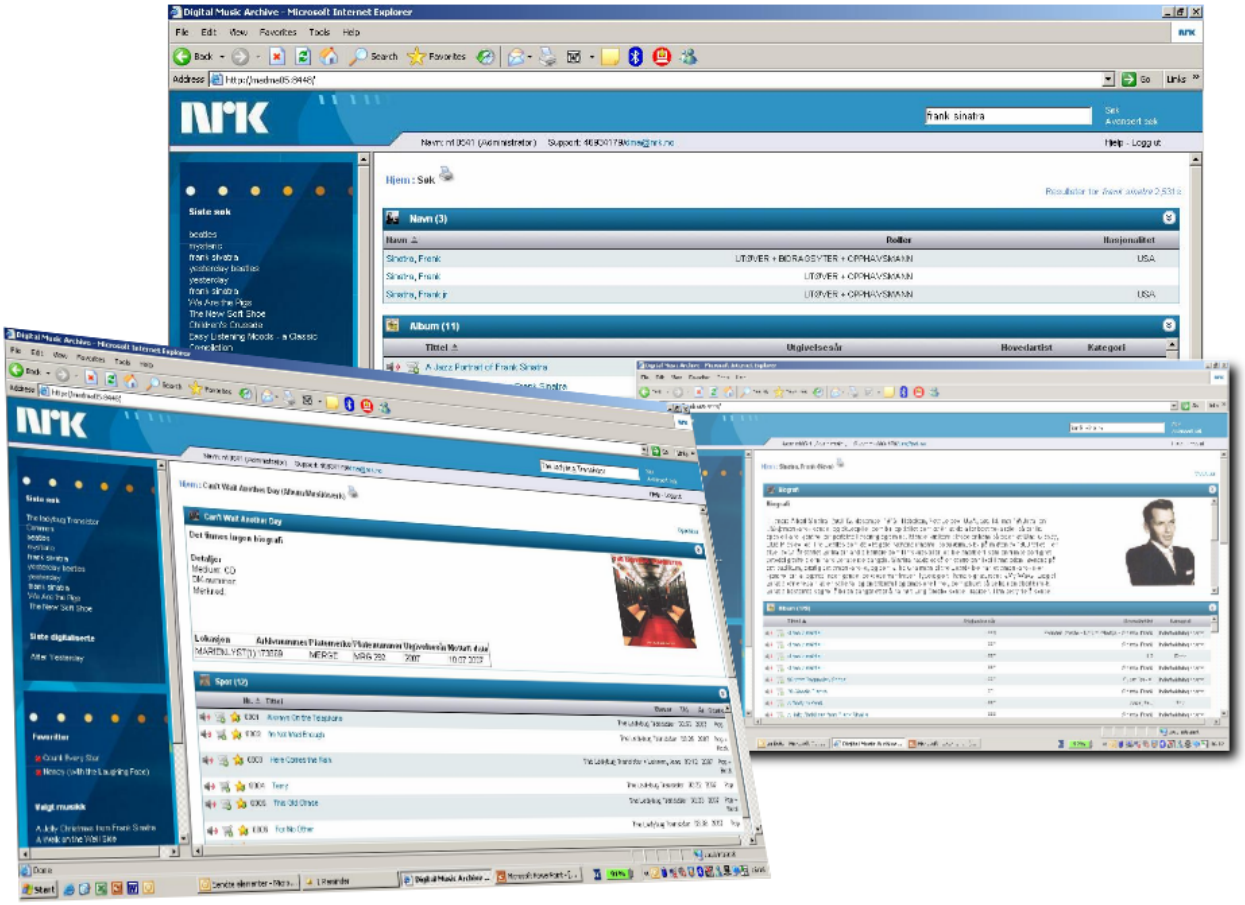


Figure 1: Some screenshots of the actual Digital Music Archive User Interface

**Conclusion and future work**

During design time, RDF stores and repositories seemed not to be able yet to store and reason with the sheer amount of triples needed for a proper representation of the archive assets (estimation in 2006: 150+ Million triples for the complete database). The evaluation showed that at that time not the whole stack could be served by proper Semantic Web technology. Therefore it was decided to do a production ready in-house development where objects, properties and relations are stored in a scalable RDBM mapped up to a Semantic Web based publication layer. This approach allowed for a production system while being able to show the benefits of using Semantic Web technology in Search & Navigation scenarios. Part of the solution is an export layer, where all metadata can be exported to a variety of formats, including XML/OWL.

Tests are currently conducted with parts of the archive and currently available technology in order to evaluate scalability of available systems to date.

As soon as SPARQL end-points for internet resources with metadata in the field of music become available, such connectors will be added to the administration module of the Digital Music Archive.

## Key Benefits of Using Semantic Web Technology

- Significantly enhance and facilitate archive access
- Navigation and Discovery of new, potentially interesting facts «hidden» in the repository
- Highly efficient music archive, combining multi-channel access with a fully automized ordering and production flow
- enhanced metadata representation, including multiple file formats (not only music files with flat metadata, but including pictures, links, interviews and many other resources) helping journalists to faster produce better trailers and talk-throughs
- Ease of integration across multiple archives and resources in the nearby future.