

IAML presentation

Eric van Balkum (Dutch Radio and Television)

Sheet music archive Muziekschatten: pioneering with digital sheet music, linked open data and more...

Abstract:

In a 2-year project the Dutch Radio and Television Sheet Music Archive discloses its collection of about 450,000 items of sheet music (all genres), digitizing and making available 60,000 items of it (for its members, at €20 per annum), publishing the metadata as Linked Open Data and seeking connections with other Dutch music collections. In this presentation Van Balkum will elaborate on the experiences in this projects, on its results and on its future.

DIA 1 + DIA 2

In this presentation I will tell you about the sheet music digitizing project that we at the Dutch Radio Sheet Music Archive were able to realize in the past two years.

First I will take you for a short trip through radio music history in the Netherlands.

Then I will briefly describe the collection and the way it has been catalogued.

The part about the project I will divide into five sections. Consectively the processes of **selecting** the music to digitize, of **publishing** our metadata as **linked data** and of **digitization**. Then I'll dive into the matter of **availability** and finally I will depict our efforts in adding **context** from external sources to the catalogue items

- HISTORIC TIMELINE
- THE COLLECTION AND CATALOGUING
- THE PROJECT
 - SELECTING
 - PUBLISHING LINKED DATA
 - DIGITIZING
 - AVAILABILITY
 - CONTEXTUALIZATION

DIA 3

DIA 4

In the 1920s the first Public Broadcasting Corporations emerged having various religious and social denominations (VARA, AVRO, NCRV, KRO, VPRO). Each had their own ensembles, band leaders and (sheet) music collection

These sheet music collections were brought together in media city Hilversum during WW II by order of the German occupier.

After the war more broadcasting corporations emerged. Every one of them should represent a social or religious group. Ensembles and orchestras played an important role in recruiting and maintaining members and giving every corporation its special look and feel.

Up to 1980 about 1000 different ensembles had played live music of all genres in radio and later also television programs: jazz, Latin, light and salon music, classical music, band music, pop, socialist and national-socialist propaganda music, radio plays etc. Some of these ensembles existed only for a few weeks, others for over 30 years.

Dozens of arrangers and copyists adapted the popular tunes of those days for all the specific instrumentations and vocal configurations of these ensembles. There were, for example, accordion ensembles, groups consisting of different kinds mandolins, Dixieland bands, ensembles with Hawaiian guitars. You name it.

To date only two radio ensembles remain, both performing lesser known classical music or newly composed contemporary music: the Radio Philharmonic Orchestra and the Dutch Radio Choir, each with their own orchestra librarian. Together with the sheet music archive they are part of the Stichting Omroep Muziek (SOM), an organization that provides services to the various public broadcasting corporations.

DIA 5

Since 2013 the Dutch Radio Music Library is closed due to budget cuts and turned into an archive. No longer new items are added to its collection.

The collection consists of about 450.0000 sheet music items (published and handwritten). In our online catalogue 300.000 of them can be found.

It is divided into three sections:

- classical music (scores, parts, piano reductions, mostly published)
- popular music (song sheets and song books, mostly published)
- arrangements (repertoire of the Metropole Orkest, the world's leading pop and jazz orchestra, practically all handwritten)

The classical and popular music have been catalogued in accordance with the International Standard Bibliographic Description for Printed Music ISBD(PM) and international rules for authority control.

The metadata are moulded in the UNIMARC bibliographic format, with some local fields added to it. The light music arrangements are catalogued according to house rules.

DIA 6

The project

At the end of 2017 the Dutch Ministry of Culture acknowledged a request for a digitization project for sheet music manuscripts. The main goal was to make this music online available, both for performance and research. The ministry granted 1.000.000 Euro's, provided that the future of the collection would be ensured by other funders. The city of Hilversum will guarantee this from the beginning of 2020. Moreover we should adhere to the principles of the National Strategy for Digital Heritage, of which I come to speak later in this presentation. So after nearly 4 years of *status quo* the sheet music collection was revived. And for me and my colleague Jan Jaap Kassies there's paid employment again. We share 1 fulltime equivalent of 36 hours.

- **Selecting**

As I said the sheet music collection of the music archive comprises of some 450.000 items. A representative selection of 5.000 items were already digitized and made available in 2011, when the first version of the Muziekschatten-website was created. The grant permitted us to scan 800.000 pages. In this new project we focused on the repertoire that the light music ensembles played live on radio and television until their demise in 1980.

DIA 7

This part of the collection comprises of more than 150.000 titles, so we had to make a selection. Making choices out of such a large body of titles cannot be quickly done by looking in envelopes to see if the material would be relevant for the project. We had to decide to select the titles from our automated library catalogue.

In order to be able to do that, we had the paper inventory cards entered in Excel by data typists. These cards merely state an archive number, the title of the piece, sometimes the composer's surname and initials, in general the arranger's surnames and initials and a code indicating the ensemble that performed it.

DIA 8

In a crowdsourcing project we ask the public to help us enrich the metadata by entering data from the first page of each scan.

From the datatypists I received every week, for seven months, a file with 5.000 titles which I loaded into our database. From this my colleague and I selected the pieces for digitization, based on the ensemble, the arrangers and the titles. This without knowing in what form the chosen sheet music would present itself. The process of selecting was further complicated by the fact that we did not know of how many pages each sheet music item was composed.

We estimated that we should select about 45.000 pieces.

The focus in the project was on digitizing scores or piano/violin/accordion directions mainly, but when no score was available we did send parts if the repertoire was considered important enough.

DIA 9

- **Publishing Linked Data**

The Dutch Digital Heritage Network is a partnership that focuses on developing a system of national facilities and services for improving the visibility, usability, and sustainability of digital heritage. It encourages heritage institutions to adhere to the principles of the National Strategy for Digital Heritage. Muziekschatten has signed this covenant.

Two of its principles are: “to ensure that heritage information refers” (principle 5) and “to ensure that heritage information is referable” (principle 6).

That is why we decided – being the first Dutch music collection to do so – to publish our metadata set of over 600.000 records as linked data.

Linked Data is one of the core pillars of the Semantic Web, also known as the Web of Data. The Semantic Web is about making links between datasets that are understandable not only to humans, but also to machines, and Linked Data provides the best practices for making these links possible. In other words, Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web.

In this short presentation will not elaborate on the technical process, but with a few screenshots I will picture it and show some resulting applications.

DIA 10

As there are still few parties that are capable of creating linked data and we being inexperienced and in some kind of a hurry, we took little time in finding suitable music ontologies. With that I mean that we did not find a data structure fittingly describing and defining all aspects of our sheet music metadata.

Our collaborative partner suggested we create our own ontology by mapping the bibliographic format to the vocabulary of schema.org. And create our own properties for the data fields that had no match. This is what we did.

We have not had time yet to properly describe the structure of this SOM-ontology, so other, related organisations will have difficulty using it at this moment.

DIA 11

We've publishes our dataset on this web address: <https://data.muziekschatten.nl/som/>.

In three subsets – Music Compositions, Persons and Music Groups – we strive to link with data and web thesauruses other organizations. In the part about contextualization I will tell you more about this.

Now I will take you through the process of digitization.

DIA 12

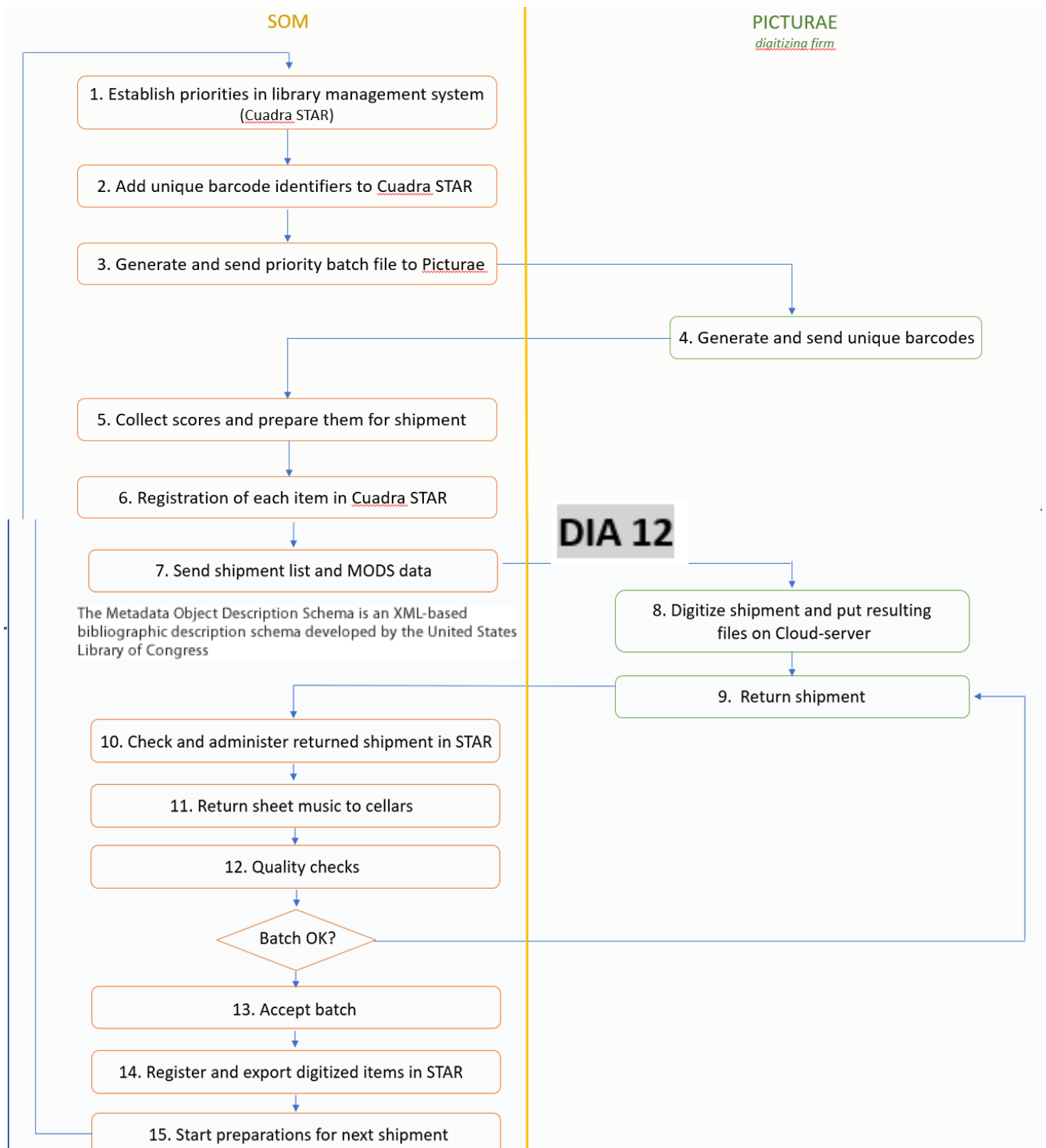
- **Digitizing**

In the process of digitizing the sheet music we also had to comply with national standards concerning the quality and the file format of the scans. Automated checksums had to be programmed to judge the quality of the resulting .jp2-pictures. METS-files with technical and bibliographical data should accompany them. A valid xml-structure was a requirement for these files.

With the help of the municipality of radio city Hilversum we formed a team of volunteers that helped us registering, sending and receiving the crates with sheet music for the digitizing firm.

First we did a small pilot in which we established the following procedure.

Flowchart digitalizeringsproces



- **Availability**

The matter of availability of the digital sheet music I will divide into two subjects. First I will discuss the copyright issue, then I will tell you about how the public have access to the digitized material.

DIA 14

- **Copyright**

As I have said earlier, most of the content concerns handwritten, unpublished sheet music from the period between 1930 and 1980. This music was created in commission for the different broadcasting corporations by arrangers and copyists. Regrettably these people used to have freelance contracts and we are not able to find more information about the type of contract they signed. If they would have been regular employees the rights of their works would rest with their employers (i.e. the broadcasting corporations). This would have made our lives easier.

During the project we were contacted by a representant of the organization of Dutch music publishers FEMU, with whom we reached an agreement.

In addition to that we made top 10 of important arrangers, whom we contacted (or sometimes their heirs). Most of them agreed to making their handwritten music available, from others we got no permission. During the process we stopped selecting works from these musicians for digitization.

For the large group of arrangers we did not contact I can say we just put the music online with the agreement with the FEMU as back-up. Up to now we have had no claims or complaints. Occasionally a relative sends us an e-mail telling us how great it is that the work of their father or granddad – yes, they were almost all men - can be studied and played again.

DIA 15

-

We hope that the recent European Directive on Copyright in the Digital Single Market will increase our possibilities. This includes a copyright exception that allows cultural heritage institutions to make Out of Commerce Works in their collection available online.

DIA 16

○ Website

From the start of the digitization project – in June 2017 - we looked for a website developer that is experienced in building a site that uses linked data and would fit within the project's budget.

DIA 17

Our ambition was to present a website with the following functionality:

Looking at the clock I will refrain from reading them all to you

- with a catalogue that reflects the full collection in an accessible manner (our still existing online catalogue does not contain the metadata produced within the project)
- with the possibility to search in one of five preselected catalogue sections per genre
- with a news section to provide updates on new functionalities and articles on radio music history
- with a search engine that has features like finding works by their instrumentation or selecting only digitized music
- with a Twitter-feed showing up-to-date information and occasionally showcasing items from our musical treasury
- with a means of monitoring visits and producing statistics
- with a membership at low fee for downloading the digitized music in .pdf-format (including a means to see who downloaded what, for copyright reasons). Of course we have no commercial intentions, but the € 20 are meant as a small compensation for our hosting costs.
- with the possibility to present context from other cultural institutions
- savouring the other benefits from linked data (like interactivity or automated suggestions to the visitor for related content)

DIA 18

On Thursday the first of March 2018 our new website went online. Since then some of the above mentioned features were added. We are still working on improvements and looking additional functionality.

Up to now the website has no English version, because we still are somewhat careful with the copyright situation.

Finally I will tell you something about our attempts to provide the website's visitors with additional service and context from other sources.

DIA 19

- **Contextualization**

I mentioned earlier that we strive to link with data and web thesauruses of other organizations.

Taking three subsets from our linked data set I will show you some promising – though not yet perfect - results.

Music Compositions

We are striving to provide the service of presenting a sounding fragment with the sheet music titles on our website. Copyright law permits us to play a fragment of 30 seconds. For this service we are collaborating with Muziekweb, the national music library which has a very large collection of records and cd's and which provides a streaming service to Dutch public libraries. As Muziekweb does not yet produce linked data, we have used database exports for matching. For classical music this matching has been rather successful, as Muziekweb applies similar cataloguing rules, including uniform titles. I've been able to add 20.000 links to our bibliographic database, one percent of which links digitized sheet music to digital sound. The other 99% links titles of non-digitized sheet music to the sound recording.

DIA 20

From their web catalogue Muziekweb presents links to classical sheet music titles on the Muziekschatten website.

For popular music matching is more complex as we at SOM focus on who wrote the music whereas Muziekweb lies more weight on its performers. But we are in the process of improving the matching results.

DIA 21

As the non-digitized part of our collection is not open to the public we make another effort to serve the visitors to our site. For this we try to match music editions in our collection with those in public libraries. The Dutch National Library has published a linked open dataset which – in theory - includes all sheet music by Dutch composers or published in the Netherlands. We have succeeded in linking our datasets, so we can refer the public to the equivalent sheet music title in public libraries. Due to differences in cataloguing, these links do not yet have the quality we think they should have.

DIA 22

- **Persons**

Apart from providing context to music compositions we try to do the same on the level of the names of persons in our catalogue. Many of them are listed in our controlled authority file, except for those figuring in the metadata generated in the project. The authority file records are published as linked data so that we can match the names with those in the [Nederlandse Thesaurus van Auteursnamen \(NTA\)](#) (the National Thesaurus of Author's Names).

Through these we can also access sources like Wikidata and VIAF which adds many new possibilities.

DIA 23

By publishing our own [Radio Music Wikipedia](#) as linked data we are able to present biographies of radio and television musicians and ensembles from this source.

DIA 24

- **Music Groups**

And with that we have come to the third category for which we try to find standardisation and extra context; that of the music groups.

In this field we work together with the Dutch Institute of Sound and Vision. When we succeed in matching our database with ensemble names to those in the Thesaurus for Audiovisual Archives we have made a first step in being able to show digitized radio or television music content side by side with the sheet music the musicians played from.

DIA 25 + 26

Eric van Balkum

**application manager/metadata specialist
at Muziekschatten/SOM**

E-mail : muziekschatten@omroepmuziek.nl

Twitter: @wibv + @muziekschatten

Web : www.muziekschatten.nl