DEALING WITH AV MEDIA AND DIGITAL FILES IN THE SWEDISH MEDIA DATABASE (SMDB) AT THE NATIONAL LIBRARY OF SWEDEN
Olle Johansson, Bibliographic Expert, Audiovisual Media Department, National Library of Sweden

Introduction

The Audiovisual Department within the National Library of Sweden receives a constant stream of material under the Legal Deposit Act: phonograms, videos, interactive multimedia products, radio and TV. The holdings total approximately 9 million hours of recordings.

The Swedish Media Database (SMDB) is a complex system with interacting parts:
- Supplier management system
- Ingesting system (FTP)
- Metadata funnel
- Mass digitisation system
- Digital archive
- Lending management system
- Media player
- File location system
- Browse copy production

The Supplier management system contains information about 6200 suppliers (3800 active). It is used for demanding Legal Deposit copies through email and mail (PDF). Here basic metadata is recorded and transferred to the client/server. It has a built-in FTP system for receiving digital files and it manages surveillance of incoming files and metadata.

The on-going migration/digitization of our material – with a capacity of 2800 hours of recordings per day – gave us further copies to deal with. Totally, the archive holds more than 4 million hours of digitized recordings.
Most of this material demands extensive description of the content: tracks, parts, programs etc. Not only do we get a copy of each work, we get all the different editions, versions, kits, boxed sets, etc. of phonograms, videos and interactive multimedia products published in Sweden. An important difference for the description of our material is that phonograms, films/videos and multimedia are catalogued manually, but for radio and TV we use program schedules to describe the broadcasts.

The publication of kits, e.g. a concert on CD and DVD published together, and boxed sets consisting of separate works requires special treatment in order to make the catalogue information clear and easy to understand for the end user.

The publication of digital files on the web raises new questions of how to record metadata, both regarding descriptive and technical metadata.

Technical metadata is an especially difficult chapter for AV Media, and no existing cataloguing standard addresses the question sufficiently.

In order to deal with all these complex questions we have created the Swedish Media Database (Svensk Mediedatabas), a catalogue system that, in a simplified form, is based on Functional Requirements for Bibliographic Records (FRBR). It is also compatible with the MARC21 format, to make it possible to exchange catalogue information with others.

The result is a system where we separate the catalogue information into two different types of catalogue records: content records and carrier records. This enables us to link carrier records to existing content records, e.g. for new editions, versions etc. A carrier record for a digital copy made in-house is also linked to the content record.

By doing this we gather different editions etc. together, so that instead of getting several hits in a hit list when searching for a title in the database, you only get one hit, thus making it easier for the end user to navigate.

It is also possible to describe kits and boxed sets consisting of separate works separately and link the records together. This makes the catalogue information about each work easier to understand, and you can still show the end user that the works have been published together.

For the optimal description of parts, e.g. tracks on a CD, short films on a DVD compilation etc., we create records parts within the content record. Every part can thus be catalogued on a deep level, with composers, lyricists, performers, recording data, classification etc. A content record may include an optional number of part records.

**Catalogue records (content and carrier records) and linking**

Normally, a catalogue record consists of a content record with an integrated carrier record. The resource first catalogued is described in the content record with the integrated carrier record.

Other resources may then be catalogued in carrier records that are linked to the content record, when the content is the same – or almost the same – as in the content record.
A carrier record for a digital browse copy made in-house is also linked to the content record. By doing this we gather different editions, versions, etc. together, so that instead of getting several hits/results in a hit list when searching for a title in the database, you only get one hit. This makes it easier for the end user to navigate.

In the first view, the carrier records are shown in short form, to improve the overview for the users. You expand the carrier records by clicking on them.
Different works published together

For different works published together, as in certain boxed sets, it is possible to describe each work (here meaning albums, films etc. published on separate carriers) individually, in separate catalogue records, and link the records together. The records are linked together by the title of the boxed set recorded in the Box field in each content record. The catalogue information about each work is described in separate catalogue records, which makes the information easier to understand.

Boxed set containing three different albums on CD

Boxed set containing three different albums

Kit containing the same concert on CD and DVD

Kits, which contain works published on different media types, are described separately and linked together with the title of the kit recorded in the Box field, in the same way as boxed sets.
The Box field shows the boxed set edition with title, year, extent, and technical format. The overview of the carrier records shows the different editions of the album *Kiss and tell*. If you click on the box title you get an overview of the phonograms published in that particular boxed set. If you click on the series title you get all phonograms published in the ‘3 original albums classics’ series.
Different works recorded on one carrier

Different works recorded on one carrier, e.g. short films transferred to video, are described in separate content records and linked to a separate carrier record.

Thus the catalogue information about each work is described in separate content records. This makes the information easier to understand. The end user can still get an overview of the works recorded on the same carrier by clicking on the carrier record.

This model is sometimes used for radio/TV program schedules, which can be linked to a carrier record containing broadcast from a longer period, for example a month.

Part records (for tracks, short films etc)

For the optimal description of parts (for instance tracks on a CD or short films on a DVD compilation) we create part records within the content record. Every part can be catalogued on a deep level, with composers, lyricists, performers, recording data, classification, etc., if you have the time and the resources. Part records make the content record easy to read and understand. A content record may include an optional number of part records.
In the first view, the parts records are shown in short form, to improve the overview for the users. You expand the part records by clicking on the arrow before them.

<table>
<thead>
<tr>
<th>Title</th>
<th>Artists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covers / Steve Dobrogosz, Anna Christoffersson</td>
<td></td>
</tr>
<tr>
<td>[Stockholm : Amigo, p 2010]</td>
<td></td>
</tr>
</tbody>
</table>

- Ave Maria
- Jealous Guy
- Cohn Child
- Like a hurricane
- Just the way you are
- I have nothing
- Through the morning, through the night
- Somewhere
- A song for you
- Dig me a grave
- All around
- Akaake ni

### Composer
- Christoffersson, Anna
- Dobrogosz, Steve (producer)

### Urförelse
- År utgiven: p 2010
- Utgivare: Stockholm: Amigo
- Utgivningsland: Sverige

In this case, information on composers and lyricists are given when the part records are expanded. (Compare with a content note created according to AACR2.)
Cataloguing rules for AV media

The Audiovisual Media Department within the National Library of Sweden uses in-house cataloguing rules for AV Media. They are based on The IASA Cataloguing Rules – a manual for the description of sound recordings and related audiovisual media (http://www.iasa-web.org/iasa-cataloguing-rules). The IASA Rules are in their turn based on AACR2.

Rules for cataloguing different works published together

The rules for when to catalogue an item in separate catalogue records are, in short, that all kits, and phonograms and videos where each carrier holds a separate work/expression should be catalogued separately. Boxed sets containing a compilation, TV-series etc, are catalogued in one catalogue record.

Rules for linking

The following elements may differ when linking:

- Format
- Date of publication, distribution etc.
- Place and publisher, distributor etc.
- Edition including if published in boxed set or separately
- Duration
- Language
- Console for computer games
- Minor differences in content

In this example, describing two editions with a minor difference in content – with and without bonus tracks – all tracks including the bonus tracks are catalogued as part records in the content record. In the content record there is a note saying: “Tracks 12-13 only on CD06-
Differences in content are always described in the content record. (If the difference in content is considerable, as is the case with an extended version of a film, you have to create a new content record.) Other differences than in content are described in the carrier records. Hence, differences in place, publisher and date of publication, label and catalogue number are described in the carrier record.

Physical description: technical formats etc.

To solve the problems with technical metadata, we have created a hierarchical list with all technical formats and their characteristics that are used in (or, in some cases, known by) the archive.

<table>
<thead>
<tr>
<th>Format</th>
<th>Codec</th>
<th>Display</th>
<th>File format</th>
<th>Disc format</th>
<th>Data format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analog</td>
<td>LP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analog</td>
<td>EP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analog</td>
<td>78 rpm disc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analog</td>
<td>single</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analog</td>
<td>multi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>MO-disc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>MD-disc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>floppy disk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>hard drive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>server (SIP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic optical</td>
<td>DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital</td>
<td>CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital</td>
<td>DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital</td>
<td>R-DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Page from SMDB – Format specification
This specification is the structure behind the drop-down boxes that the cataloguer has to use when choosing technical format.

**Radio and TV**

For radio and TV program schedules from the nationwide radio and TV broadcasters are imported (often in the XML format) and converted into content records in the SMDB. Either they are imported directly from the broadcasters, which is the case with Swedish Public Radio/TV and the commercial channel TV4, or from the news agency TT Spektra (TV channels in the digital terrestrial net or satellite channels). For some 70 selected programs from Swedish Public Radio/TV and from TV4, the metadata is enhanced with additions made manually by cataloguers.

The actual broadcasts are delivered either as digital files (MPEG-4, MPEG-1 or MP3) or as physical carriers (DVD-R or CD-R).

For the broadcasts delivered as digital files, carrier records are created automatically and connected to the content records, as integrated carrier records. The carrier records get automated archive numbers, which link them to the digital files. The files are linked to the catalogue records via the archive number. This number is automatically constructed from broadcasting company, channel and broadcast date, e.g. XA_tv4_tv4_2012-07-10.

The digital files are instantly accessible via the SMDB on the premises in Stockholm.

We get broadcasts from:

- Swedish Public Radio (19 channels)
- Swedish Public TV (5 regular, 19 regional channels)
- TV4 (11 regular, 60 regional channels)
- Axess, Kanal Global, NTM and Canal plus
- MTG (42 channels) and SBS (48 channels) – 4 selected weeks/year
In this view only the title for each program is shown. Next example shows expanded part records with full information on each program.

The information on each program in the program schedule can be quite extensive (here we even get the names of the actors in name fields). This information is delivered to us from the broadcasting company. The arrows within orange circles are “Play buttons”, which indicate that the programs are already digitized and instantly playable for the researchers.
For the broadcasts delivered on physical carriers, carrier records must be created and numbered manually and connected to the content records as integrated carrier records. This particular carrier record contains one month of broadcasts, which means that it is linked to thirty program schedules, each containing one day of broadcasts. The shopping basket before the title indicates that you can order a digitization of the program.

**Swedish Public Radio**

In 2007, a special project concerning the program schedules for Swedish Public Radio broadcasts from 1978 to 1984 was completed. These program schedules were scanned, and by using Optical Character Recognition (OCR), and manual editing, we were able to put them into the Swedish Media Database in a searchable form. And since the nationwide broadcasts from the Swedish Public Radio have been migrated into digital files, all programs from 1978 and onwards are now searchable and accessible via SMDB on the premises in Stockholm.
The open reels – six hours duration with four channels in mono recorded simultaneously – have been digitized into files with six hours duration.

**Swedish Public TV**

All Swedish Public TV nationwide broadcasts from 1978 and onwards are searchable in the SMDB. In the 1990s, the broadcasts from 1978-1982 were transferred from VCR tapes to Digital Betacam. In this on-going migration project, the years 1983-2008 have so far been migrated into digital files. In the near future, all nationwide broadcasts from the Swedish Public TV will be accessible as digital files.

The technical metadata for the digital file is described in the carrier record for the file. The prefix Y stands for mass migrated files. YA is an archive file, YS is a browse file.

**Mass digitization project – CDs from 1984 to the present**

We have an on-going mass digitization project for CDs. So far CDs from 1984 to 1997 have been digitized – around 30 000 objects. The files have been described and linked to the database, and are playable for our researchers on the premises.

**Digitization on demand for researchers**

Phonograms, films, video, radio and TV programs are digitized on demand, when researchers want to listen to or view a production. Most of our collections are available only to researchers but on the premises in Stockholm certain special collections are available to the public. On request from researchers, we can also order and digitize older material that we don’t have in our archive: films (from the Swedish Film Institute), and public radio/TV broadcasts (from the Swedish Broadcasting Corporation).

**Downloading digital files on the web**

A selection of Swedish music, Swedish TV programs and Swedish video productions has been downloaded from the Internet and catalogued in SMDB (productions published on the web only). This has been done in preparation for the new Legal Deposit Act, which will include web
publications. While waiting for the new law, we learn how to work with the material and also save some productions for the afterwards.

The digital files are described in the carrier record and the actual sound files are linked to the part records. The prefix X stands for files made from external material, XA is an archive file, and XS is a browse file. In cases like this, the archive files are also browse files since the files imported to our system are already in a browse format.

SMDB – the Swedish Media Database

SMDB, the Swedish Media Database is available at the Swedish National Library home page: www.kb.se > Sök i SMDB KB:s Audiovisuella medier or directly at: smdb.kb.se

Please note that no digital files are available on the web, for copyright reasons.