Audio cataloguing in the British Library Sound Archive

Robert Smith, Assistant Librarian, NHS Greater Glasgow and Clyde

Introduction

From May 2016 until recently, the author had the opportunity to work as an Audio Project Cataloguer in the sound archive at the British Library (BL). This exciting and engaging experience informs the outline on audio cataloguing offered here. The Sound and Moving Image catalogue (SAMI) is a unique catalogue that has organically grown and weathered many changes. Working on the catalogue brought into focus various non-text based metadata challenges that will be elaborated upon. The sound archive itself is rich with collections, some of which are still waiting to be revealed. How similar collections are catalogued and how the cataloguing can respond to various user/curatorial concerns will also be considered.

Background

A brief history is necessary to provide some context. A fairly detailed account by a former classical music curator, Timothy Day, does exist and helps to illuminate the early years of the archive. This summary is indebted to his article.[1]

One of the most integral figures in the history of the archive is Patrick Saul. Saul recognised from an early age the significance of sound recordings and what they could potentially offer as resource for future generations. It was his enthusiasm and determination that eventually led to the establishment of the British Institute of Recorded Sound (BIRS) in 1955 – the first true incarnation of a national sound archive in the UK. While early acquisitions tended to focus on private donations of course-groove shellac classical music discs, the policy of the archive provided a clear purpose; to be as comprehensive as the BL was in regards to bibliographic material. The preservation of sounds should not be limited to just human sounds – in fact Saul insisted that 'animal sounds of all kinds should be preserved' and maintained that his favourite recording was 'the mating call of the haddock'.[2]

In 1983, BIRS officially merged with the BL, at which point the archive became known as the National Sound Archive (NSA). It was not until 1997 though that the NSA physically joined the BL in St. Pancras. The NSA would eventually become recognised as the British Library Sound Archive and is at the heart of the British Library Sound and Vision department today. A range of material is found and curated within the archive collections; wildlife and mechanical sounds, pop/jazz, classical, literary and creative recordings, world and traditional music, news and moving image, spoken English, radio, and oral history are all clearly represented. In regards to holdings, the archive is home to approximately 5 million unique recordings found on 1.7 million physical carriers. Around 40 different formats have been identified ranging from early Edison cylinders to contemporary high quality digital audio files.

---

1. Day, T. "The National Sound Archive: The First Fifty Years" in Linehan, A. (ed.) *Aural History: Essays on Recorded Sound*. London, The British Library, 2001, pp. 41-63.
2. Day, T. "The National Sound Archive: The First Fifty Years", p. 45.

Documenting and cataloguing the range of material contained within the archive has presented challenges from the very beginning. From the 1950s though to the 1980s, acquisition registers were the main means of recording information on holdings and were crucial for discovery from a user perspective. These registers would focus on record company names and matrix disc numbers but they were not open to the general public. If a user wanted to find a particular recording, they would usually have to peruse the discographies and record release sheets made available in the archive to identify which disc contained the recording they wanted to hear. After noting the record company name and matrix number for the disc, the user could then make a request, and an archivist would check the registers to see if the disc in question was held in the archive. Full cataloguing of audio material did occur but it was limited by some of the challenges we still encounter today (particularly funding and resources). What material was selected for this treatment was generally left at the discretion of the cataloguer.

Cataloguing guidelines and standards for audio-visual material were initially informed by the meticulous standards created for print-based bibliographic records. It was already recognised though that AV material would have unique requirements. A BL committee in the 1980s had decided not to impose common bibliographic standards and the sound archive was given the opportunity to 'find a solution based on the perceived needs of its users'.[3]

SAMI

In the early to mid 1990s the sound archive introduced the first iteration of the current cataloguing platform, SAMI, with an OPAC going live in 2001.[4] It continues to be one of the key strategic data repositories for the BL. SAMI operates on a system that has both adopted and adapted. It has its own unique flavour of MARC (SAMI-MARC) and has incorporated aspects of AACR2 and RDA into local cataloguing guidance. This approach seeks to accurately describe the physical items being catalogued and the individual recordings contained on them in a consistent manner. It also allows for a degree of flexibility to cater to the needs of various user communities and curatorial areas.

What we would now call an early entity relationship model provides the foundation of this digital system. Three 'libraries' or databases exist within SAMI: Product entries describe physical carriers; Recording entries describe the audio recordings (defined as 'recorded events') found on physical carriers; and Works-File entries can be used to link distinct artistic or intellectual creations that appear on various formats or recordings, be they symphonies, operas, pop songs, theatre performances, or poems.

Working right to left with the example in Figure 1, we can see that the song *We'll Gather Lilacs*, written by Ivor Novello, has a Works-File entry that includes a standardised title and composer credit. A recording of this song appears on one side of a course-groove shellac disc. Two Recording entries exist for this disc - one for each of the recordings present on either side of the disc (assuming of course there are recordings on both sides and there is only one recording on each side). A different recorded performance of the same song appears as track 20 on a CD. This 22 track CD will have 22 Recording entries in total, one for each track.

---

3. Day, T. "The National Sound Archive: The First Fifty Years", p. 58.
4. When the OPAC went live in 2001, SAMI was initially christened 'Cadensa' and this name is still associated with the catalogue.

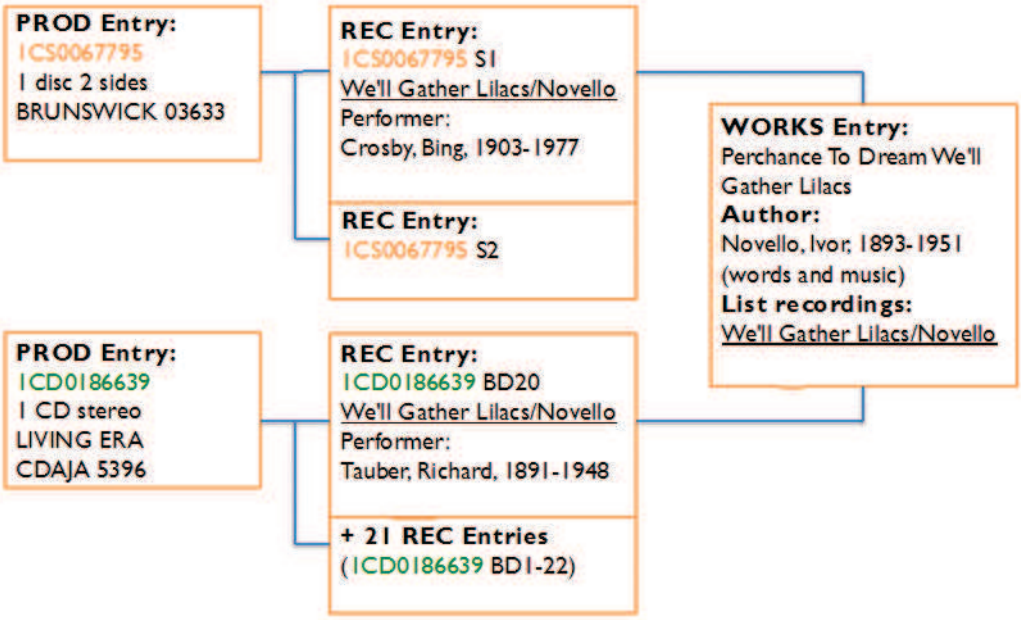

Figure 1. *We'll Gather Lilacs* example showing Product entries, Recording entries and Works-File entries

Product entries provide information about the physical carriers (physical characteristics, record company name, matrix numbers, shelfmark, etc.) while the Recording entries will describe the individual recordings (title, performers, locations, duration, etc.) and are linked to the Product entry via a shelfmark (the '1CD' and '1CS' elements in the example). Nuanced parent/child structures at Recording entry level can also be created that identify distinct parts within larger 'recorded events', such as a radio broadcast of Choral Evensong. These Recording entries could then be linked to Works-File entries. The use of Authority records also helps to provide data consistency across the catalogue, particularly in relation to performers (incorporating standardised identifiers such as ISNI where possible).

Challenges

The challenges involved with cataloguing audio material can be manifold. Some fall firmly into the realm of archive preservation and conservation concerns while others clearly affect the quality of any catalogue records being created. The two predominant archive challenges faced are degradation and obsolescence; is the carrier in a state that it can be handled, played, and read and do we have working equipment that will allow us to listen and capture the audio information found on it?

From a cataloguing perspective, all available resources must be called upon to inform our decisions. An unlabelled spool of open reel magnetic audiotape without a box tells us very little indeed. In such a situation, we rely heavily on what audio information is present on the tape to inform our catalogue records. Additional resources can prove beneficial in providing further metadata, as can drawing on the knowledge and experience of the curators who can sometimes provide supporting collection documentation to aid the process of metadata creation.

It is also important to consider the needs of each curatorial area and this impacts on the standards that are applied. Interview summaries that are usually associated with oral history material have no place within wildlife Recording entries. Noting the country, region, and locality with geographical co-ordinates would be a touch overzealous for a radio broadcast Recording entry but would be of immense interest to a Zoologist researching wildlife sounds. Dates can also prove to be a contentious subject; will dates associated with audio material be catalogued as product publishing or creation date, a recording date or a transmission date? When should a radio programme have a transmission date versus a recording date or both? Considerations for curatorial area, information found in the audio itself, and the carrier they are found on can all affect the approach to capturing and the quality of metadata found in SAMI. Without a doubt the most important attribute though is the ability to actively listen for long periods of time. This helps to develop an understanding of the material being catalogued and how to achieve consistency in approach, particularly when dealing with numerous items within a single collection.

Challenges exist in regards to legacy cataloguing decisions and data. For example how should record labels and their subsidiaries be represented? A historic parent recording company such as Universal can appear in many guises today; Universal, UMG Recordings, UMC, UME, etc. Do we amend legacy records to fit with current trends? Should we use the 24-hour clock to document transmission times for radio broadcasts or the 27 -hour clock that was regularly used by broadcasters? Such challenges, also identified by Timothy Day in his article in regards to cohesive standards across the catalogue, are still present today and discussions continue to refine best practice.

There are also technical challenges ahead. New developments in the preservation and delivery of digital audio material have meant integrating new tools into workflows and future access requirements will need further consideration. Any future work with institutions across the UK will also require additional knowledge and resource sharing. While the digital storage space needed for millions of catalogue records is not insignificant, the additional storage capacity needed for any associated high quality digital audio files and the ingest of this material will continue to grow and remain challenging.

Conclusion

Sound is something that is ever present in our environment and can easily be taken for granted. There is still no legal deposit requirement for audio material in the UK but large-scale national projects such as Save Our Sounds
(https://www.bl.uk/projects/save-our-sounds) do acknowledge the importance and significance of audio material. This multi-faceted project has a number of aims:

- To safeguard sound collections found across the UK and preserve our sound heritage.
- To develop a radio archive that is representative of the UK.
- To develop the digital acquisition capabilities of the BL and ensure that a representative amount of audio published in the UK is acquired digitally.
- To develop IT infrastructures and workflows both internally and with partners across the UK.
- Offer further opportunities for access and engagement with audio collections and continue to promote the resource as widely as possible.

The project has now received generous funding from the Heritage Lottery Fund, which helps to reaffirm the need to preserve audio material. These collections help to inform our understanding of ourselves, our shared history and culture, and their digital preservation ensures that such information resources can continue to be utilised. While it is imperative that at-risk audio is digitally captured before it is potentially lost forever, the accurate cataloguing of this material is just as crucial to ensure it remains discoverable and accessible for future generations.

References

Day, T. “The National Sound Archive: The First Fifty Years” in Linehan, A. (ed.*) Aural History: Essays on Recorded Sound*. London, The British Library, 2001, pp. 41-63.