2011-01

HathiTrust and the Challenge of Digital Audio

Beers, Shane; Parker, Bria

https://hdl.handle.net/2027.42/83276

HATHI TRUST AND THE CHALLENGE OF DIGITAL AUDIO

Shane Beers and Bria Parker, University of Michigan, USA

Introduction

As a digital repository for the nation's great research libraries, HathiTrust (http://www.hathitrust.org/) brings together the immense digital collections of partner institutions. As HathiTrust makes every effort for the repository to conform to the characteristics of a Trusted Digital Repository,[13] a tremendous amount of work has gone into developing ingest functionalities that analyze Submission Information Packages (SIPs) to determine that they meet a number of standards. The standards include the technical aspects of the digital image files in a SIP (such as resolution, well-formedness, compression type, color and bit depth), descriptive elements of the SIP (including PREMIS preservation metadata and image header metadata), and structural metadata that explain what the digital image files represent and allow software tools to display the images correctly.

Initially, the majority of SIPs being deposited into HathiTrust were books that had been digitized by Google, Inc. The specifications Google uses in its digitization package were worked out collaboratively with Google library partners, resulting in a tightly controlled technical and descriptive SIP. The validation environment employed in HathiTrust was developed around the ingest of book packages digitized by Google and other sources, checking for agreed upon qualities. For some time, this ingest process has worked well in verifying SIPs against set standards, allowing content into the repository when compliant, and reporting when something failed.

The close relationship between Google's standards and the HathiTrust ingest verification process made perfect sense when the majority of SIPs were in a single format coming from a single source. However, the scope of digitization at HathiTrust partner institutions is much broader, as the University of Michigan Library seeks to incorporate time-based media, initially audio, into the repository. Thus, expanding the capabilities of HathiTrust to preserve and provide access to these materials was a logical, and highly desirable, direction to pursue.

Project overview

The audio digitization project at the University of Michigan Library originated out of the realization of the fragility of the various recordings in Library collections, the increasing obsolescence of the technology needed to play the formats, and, most significantly, the desire to preserve the unique items that are a part of the Library's collections. The majority of libraries in the University Library system have sound recordings in their collections. Of these collections, the majority are published commercial recordings. Therefore, digitization and digital preservation would be desired only for the unique items of high research value, and would not be performed systematically. The libraries with particularly valuable sound recordings are the Special Collections Library and the Music Library. Beginning in 2007, their collections were examined and the working group conducted preliminary research on audio preservation. The audio collections were prioritized for preservation based on the research value of the information held on the recordings, as well as the format and physical condition of the recordings.

Collection managers and the Library's Department of Preservation and Conservation contacted the Library Dean's group in early 2008 to solicit support for an audio digitization project. In the FY 2009-2010 budget, funding was provided for a pilot digitization project and a year-long term position was created in July 2009 to lead the organization and execution of the project. As the Library planned to digitize a small number of recordings, the working group decided against purchasing digitization equipment, as the expertise necessary to operate them was not available or too costly to obtain. The group instead drafted a Request for Proposal (RFP) and

13 TRAC. (2007). Trustworthy Repositories Audit & Certification: Criteria and Checklist. Center for Research Libraries and OCLC. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

put the audio digitization project out to bid in October 2009.

Developing the RFP for audio digitization involved an in-depth examination of the current standards and best practice s in audio digitization and digital audio preservation. Several sources were consulted, including IASA TC03[14] and TC04,[15] Sound Directions,[16] and CDP Best Practices documentation.[17] The group also examined project documentation from Columbia University's audio digitization project, and from other universities and institutions.[18] A set of technical specifications for the pilot audio digitization project were decided upon:

- Broadcast Wave Format (BWF) file
- 96 kHz sampling rate
- 24-bit sampling size
- Preservation Master File (an unprocessed flat transfer).
- Production Master File (some post-processing/clean-up of sound)
- Access copy (on audio CD)
- Metadata (forthcoming Audio Engineering Society (AES) technical metadata standards in a METS wrapper, including Dublin Core and PREMIS events.

While developing the technical specifications, the highest priority was creating an Archival Information Package (AIP) that would preserve essential audio information and its associated metadata in the HathiTrust repository. However, throughout the planning and development of this project a number of challenges were encountered.

Problems faced

The HathiTrust shared digital repository was designed to validate, ingest, store, and provide access to digital text and image files. This project is the first instance of time-based media, such as audio, being stored in the repository. While the Library gave financial support to the project, obstacles still existed. First, standards and best practices for digital audio preservation are still being developed. Second, many of the technical, structural, and descriptive characteristics of the audio content did not match the previously developed standards for materials in HathiTrust. Additionally, the philosophies behind the current specifications for text and images and the desired specifications for audio were different and, at times, difficult to reconcile.

One problem faced was the lack of well-developed and widely adopted specifications for digitizing and preserving digital audio. While there has been much development in this area, we had to rely more on best practices and locating a critical mass of institutions doing the same type of work, instead of relying on published standards. Acquiring institutional buy-in was therefore more difficult, as a large set of standards was not available to justify our specifications.

In addition to lacking established digitization standards, metadata standards for digital audio are also nascent. There have been previous attempts to develop metadata schemas (Library of Congress Audio Technical Metadata Schema, New York Public Library Rodgers and Hammerstein Archives data dictionary), but it was felt that what was available was not adequate for capturing all the information that was desirable to collect and preserve. Both the

14 International Association of Sound and Audiovisual Archives and Dietrich Schüller. The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy ([Canberra] IASA Technical Committee, 2005).

15 International Association of Sound and Audiovisual Archives and Kevin Bradley. IASA-TC04 Guidelines on the Production and Preservation of Digital Audio Objects: Standards, Recommended Practices and Strategies ([Canberra]: IASA Technical Committee, 2009).

16 Mike Casey and Bruce Gordon, Sound Directions: Best Practices for Audio Preservation ([Bloomington]: Indiana University ; [Cambridge]: Harvard University, 2007).

17 CDP Digital Audio Working Group. Digital Audio Best Practices (2006).

18 Columbia University Libraries Oral History Audio Preservation Project, accessed October 24, 2010, https://www1.columbia.edu/sec/cu/libraries/bts/mellon_audio/index.html.

technical and process history metadata for digital audio is more detailed than for what is currently implemented in HathiTrust for text and images. The team examined metadata used in other projects, including the Library of Congress schema and NYPL data dictionary, and opted to use the forthcoming Audio Engineering Society (AES) metadata schemas. While the AES schemas are not final, they are certainly the most comprehensive. The next challenge was to incorporate information described in these schemas and other essential metadata elements into a METS document, including descriptive and preservation metadata about the recordings. As this document would include substantially different information than the METS for other HathiTrust packages, changes would be required to the ingest process that verifies information sent to the repository.

It was particularly difficult to balance a strong desire for consistency in repository packages with what were considered proper digital audio preservation practices. As current digital preservation practices for audio and textual materials differ, there was hesitation in adopting what were considered appropriate digital preservation specifications for the digital audio materials in HathiTrust. For example, there was concern about the number and size of digital files preserved for each AIP. Storing a production master in addition to preservation master was not the practice for text and image packages in the repository. For digital audio materials, it is considered good practice to store a preservation master, which is an uncorrected capture of the original recording. The preservation master can be re-processed in the future, instead of a single processing session permanently determining the audio data in the archival file. It is also considered good practice to store a production master that includes post-processing, such as normalization and some sound clean-up, with the goal of clarifying the information on the recording. Without knowing the audio quality of the recordings, it was difficult to say whether or not post-processing would be a vital part of the project, or if the production masters would be significantly different from the preservation masters. These concerns made storing both preservation and production masters for the pilot project especially important to the project team.

The HathiTrust repository software uses server-side processing to create access derivatives that are displayed to users as they interact with book packages via a Web browser. Therefore, these access files are not stored along with the other files in the individual book packages. For audio, the original intention was to store an access master in the repository in mp3 format in addition to the preservation and production masters. However, access methods for streaming audio and displaying metadata to end-users via the HathiTrust Web interface did not exist, and were not a part of the pilot project. Without streaming access a decision about a method for giving users access to the content had to be made. As copyright is a potential issue with some collections, having streaming access was not a priority for the pilot until more work could be done to ascertain copyright status for these recordings. Although streaming access was not a task outlined in the pilot project, the lack of these methods has delayed ingest of digital audio SIPs because of the inability to actually provide the information to users.

In addition to not matching existing repository conventions, storing two sets of BWF files per AIP created issues related to repository storage space. While the average size of a text AIP is around 50 MB, the average size of an audio AIP for one hour of content captured at a 96 kHz sampling rate, and 24-bit depth is expected to be at least 4 GB (2 GB per hour-long BWF file each for preservation and production master, plus the much smaller mp3 files). Audio is therefore more costly to store than the digital book packages currently stored in HathiTrust. However, a finite number of unique recordings are to be digitized, and a systematic digitization of the entire collections of the Library is not planned. It is hoped that this will lead to controlled and manageable storage costs.

As there was little institutional experience or expertise in the digitization and digital preservation of audio, there was no existing workflow in place to guide the project. Additionally, the differences between book and audio formats meant that the workflow used for vended book digitization could not be easily adapted to audio materials. Determining the workflow for what would occur when digital audio files were received from vendors was difficult to establish, and

is still not finalized. While digital audio files will undergo the same or similar events as image files, it took time to come to a decision on the path the files would take and what programs they would be run through en route to being ingested into the repository.

Creating ingest and validation methods for this workflow also presented difficulties, as the current routines were built completely around images and Google-based content and could not be easily adapted for audio. While models existed, entire new processes had to be created to process and validate audio files. This involved selecting what characteristics of the files were to be validated, such as file format, as well as what metadata values were required in the AES schemas, such as sampling rate and size. The characteristics to be validated were selected based on the technical specifications for audio digitization. The technical characteristics, overall validity and well-formedness of the files could be validated by the current methods employed in HathiTrust, but the additional (automated) validation processes for the metadata needed additional programming.

Solutions developed

To overcome these challenges, the team developed a number of solutions and compromises for ingesting audio packages into the repository. Stakeholders of both the repository and the project participated in several meetings in order to come to a compromise on how audio would be stored in the repository.

The AIP structure for audio was based on best practices for audio found in Sound Directions. The AIP includes the preservation and production masters and a METS file that includes descriptive, technical, provenance, and structural metadata. Originally, three sets of audio files (preservation, production, access) were to be included in each preservation package, but instead a CD copy was used for access. Access masters will not be stored in the repository, as they could be easily derived from the production master if a new CD was needed, or when streaming delivery capabilities became available. Thus, only preservation and production masters will be stored in the repository. A .txt file of any notes made about the transfer by the vendor is included, which can be helpful in giving specific information about problems encountered during the playback and capture of the audio, such as areas where the sound drops out or is garbled. It is considered good digital preservation practice to use lossless compression for compressing and storing digital audio. However there is still some concern about this due to the large size of the files, and employing a lossy compression method is currently in discussion.

A METS document was developed that would fit the needs of audio content while simultaneously meeting repository specifications. While technical and process history (or provenance) metadata could be documented using the AES schemas, there were other types of metadata to include in the METS document in each AIP. At the time the project began, descriptive metadata was not included in the HathiTrust METS. Digitized books commonly had catalog records associated with them, and instead of including a descriptive XML schema such as MARCXML, the METS document included a link to the catalog record. However, this descriptive metadata does not exist for audio collections, as the majority are part of archival collections and extensive descriptive metadata has not been collected. Finding aids in Encoded Archival Description (EAD) do exist for many of the collections, but there is no simple way to link the METS document to the specific sound recording within the finding aid. Additionally, any bibliographic records in the Library's catalog are for the entire collection and linking to them would not be especially useful. Including descriptive metadata in METS was therefore a necessity. A local database was already in the process of being created for project and item tracking, so it was employed to record basic descriptive information. This is then exported and transformed to create Dublin Core metadata in XML for the METS document.

PREMIS preservation metadata also needed to be included in the HathiTrust METS document for audio, as was the practice with book materials. It was decided that audio would use the same preservation events in use for textual materials. These include capture, fixity check, message digest calculation, validation and ingest. Additionally, the team considered including preservation events for more specific events, such as post-processing and qualitative quality control. However, post-processing and qualitative quality control are not performed on every audio file in a SIP, but only a percentage. This posed a problem, as the current PREMIS implementation included an event for the entire package, not for each individual file, which made it difficult to indicate what event applied to which file. The team decided to not break from this convention. Instead, a FileList extension schema, which had been developed for use in the HathiTrust book packages, was employed in the PREMIS metadata to indicate which files in the AIP had undergone a particular event. If no FileList is present in an event, the event applies to each file in the AIP. Using this FileList extension allows us to assign an event, such as qualitative quality control, to only the file or files to which the event applies, rather than to all the files in the package. An example PREMIS event using the FileList extension is below:

```
<PREMIS:event>
  <PREMIS:eventIdentifier>
    <PREMIS:eventIdentifierType>UM</PREMIS:eventIdentifierType>
    <PREMIS:eventIdentifierValue>manual quality control1
    </PREMIS:eventIdentifierValue>
  </PREMIS:eventIdentifier>
  <PREMIS:eventType>manual quality control</PREMIS:eventType>
  <PREMIS:eventDateTime>2010-03-13T18:11:00</PREMIS:eventDateTime>
  <PREMIS:eventOutcomeInformation>
    <PREMIS:eventOutcomeDetail>
      <PREMIS:eventOutcomeDetailNote>files manually reviewed
      during quality control process
      </PREMIS:eventOutcomeDetailNote>
      <PREMIS:eventOutcomeDetailExtension>
        <HT:fileList status="reviewed">
          <HT:file>am000001.wav</HT:file>
          <HT:file>am000002.wav</HT:file>
          <HT:file>pm000001.wav</HT:file>
          <HT:file>pm000002.wav</HT:file>
        </HT:fileList>
      </PREMIS:eventOutcomeDetailExtension>\
    </PREMIS:eventOutcomeDetail>
  </PREMIS:eventOutcomeInformation>
  <PREMIS:linkingAgentIdentifier>
    <PREMIS:linkingAgentIdentifierType>AgentID</PREMIS:linkingAgent
    IdentifierType>
    <PREMIS:linkingAgentIdentifierValue>UM</PREMIS:linkingAgentIdent
    ifierValue>
    <PREMIS:linkingAgentRole>executor</PREMIS:linkingAgentRole>
  </PREMIS:linkingAgentIdentifier>
  <PREMIS:linkingAgentIdentifier>
    <PREMIS:linkingAgentIdentifierType>tool</PREMIS:linkingAgentIden
    tifierType>
    <PREMIS:linkingAgentIdentifierValue>WaveLab 6</PREMIS:linking
    AgentIdentifierValue>
    <PREMIS:linkingAgentRole>software</PREMIS:linkingAgentRole>
  </PREMIS:linkingAgentIdentifier>
</PREMIS:event>
```


After the final decisions regarding the various metadata were made, a METS creation practice document was created to outline how to include and locate the metadata in the METS. For example, the descriptive and technical metadata are stored in the same directory as the audio files. The preservation metadata comes from multiple sources. Metadata about the capture event is pulled from particular fields in the vendor-supplied technical metadata and/or the file header. Metadata documenting the fixity check and message digest calculation events originate in the programs that run those events. Qualitative quality control is documented in a tab-delimited spreadsheet that is stored in the directory, while validation and ingest event metadata is generated during the ingest process.

Based on this METS creation practice document, a process was developed to pull together the metadata supplied by the vendor, the descriptive information from the Library database, and the preservation metadata to create a HathiTrust METS document for audio. As the creation of the HathiTrust METS for book packages is automated, the audio METS needed to be as well in order to fit into the workflow. All metadata is stored in the same directory as the completed vended digital audio files and technical metadata. The metadata is used to populate a boilerplate HathiTrust METS document during the ingest process. Likewise, the creation of the PREMIS metadata is automated and uses the metadata found in the audio file header or the vendor-supplied METS document.

The validation of the audio packages was modelled after what was currently in place for text, in order to take advantage of the expertise and existing technical infrastructure. HathiTrust was already using the JSTOR Harvard Object Validation Environment (JHOVE) output to validate file characteristics, including information from the file header. During the audio validation process, JHOVE data is output in XML. This data is used to validate file characteristics and ensure that the file is well-formed and valid. As a part of this process, JHOVE outputs metadata from the bext header of the BWF file, as well as AES metadata about the digital audio file in XML, which validates the metadata in the METS file against the metadata stored in each audio file. This also allows for the opportunity to ensure that the package matches our technical specifications. While a new validation process did have to be developed, applying the same philosophies and approaches from text to audio materials provided a measure of continuity between the two formats. The validation specifications can be found in Appendix A.

Lessons learned

Building strict repository validation and ingest routines is important, but can make ingesting new content types difficult. Trying to fit new content and data formats into the same models that already exist can present several challenges. It is important to find a balance between the needs of the repository and the needs of the particular collection, content, and format. Current practice for one format may not necessarily be sound preservation practice for another. Trying to fit all types of content into a single, strict method or routine may prove detrimental for other formats or content types.

If community standards for a particular media type are not fully developed, making decisions on how to correctly preserve the content can be difficult. As HathiTrust is designed to be a long-term preservation repository, its policies and procedures related to digital images have been informed by standards that have been refined over many years, along with the experiences of a large number of earlier projects. As standards for digital audio are not as accepted and there are fewer projects to use as examples, it is more difficult to make a strong argument for or against particular preservation strategies and policies.

It is important to make access a vital part of the planning process when developing a digitization project. As the HathiTrust repository was primarily developed to support text and images, there was no expertise in delivering streaming audio content via the repository web interface. As CDs satisfied the access portion of this pilot project, developing an access system was put on hold. As there was not a clear idea of what streaming audio access in the repository would look like, it was often difficult to come to a decision on what needed to be collected or kept, because the contents of the Dissemination Information Package (DIP) were unknown. Initially, access to the audio in the repository may be a simple descriptive metadata record displayed to the user (similar to what exists for text and images) accompanied by a message that the item can be listened to at the originating library, along with a link to the finding aid so the user can view more of the collection to which the item belongs.

Large-scale digital audio preservation requires different resources than large-scale digital image preservation. These resources include researching and developing new digitization specifications and metadata standards; developing a RFP to send to outside vendors; purchasing new hardware for internal digitization or quality control; hiring new staff or training existing staff; allocating larger amounts of hard drive space; and developing new methods for ingest, validation, storage, and access. As the capture of audio is done in real-time, as is any manual, qualitative quality control that is performed on the digital audio, significant amounts of staff working hours must also be incorporated into the budget.

We don't even know if any of this works yet! Implementing new policies and technical procedures is not a straightforward task, and therefore takes some time. Even with a large amount of research, discussion, and development talent, issues that compromise the preservation viability of the digital audio packages may arise in the near or long-term. Pilot projects can be risky, even with the best intentions and the strongest talent behind them.

Conclusion

The pilot audio digitization project posed several challenges that could not always be solved by developing specifications based on conventions for textual materials. However, it was helpful to look to repository convention for guidance and modeling when making decisions about preservation package structure, metadata, validation, and ingest. Ideally, the work done for this pilot will provide a model for the ingest of digital audio content from other HathiTrust partners into the repository, and provide a framework for introducing other time-based media formats into the repository.

Appendix A: Validation specifications

Metadata Type	Element/ Attribute	Required or Recommended Value	Validation Recommendations
<bext>	description	Item number from collection	validate that exists and contains data
	originator	University of Michigan Library	validate that value is "University of Michigan Library"
	originatorReference	filename (barcode/track.wav)	validate that this is in correct format ; validate that a file with this name exists in directory, and that a primaryIdentifier with this value exists in METS document supplied by the vendor.)
	originationDate	yyyy-mm-dd	validate that exists and contains data
	codingHistory	none required	validate that exists and contains data
AESX098B	analogDigitalFlag	ANALOG, PHY_DIGITAL, FILE_DIGITAL,	validate that the analogDigitalFlag value in a file's JHOVE output matches the value in that file's <techMD> section. In <sourceMD>, analogDigitalFlag must be ANALOG UNLESS <aes:format>=DAT or CD, then analogDigitalFlag="PHYS_DIGITAL"
	format		The <aes:format> in each <techMD> section must match the <aes:format> from the JHOVE output for each file. There is no required value in <sourceMD> sections, but must exist and contain data.
	audioDataEncoding	"PCM audio in integer format" or "PCM"	validate that this field exists in <techMD> sections where analogDigitalFlag="FILE_DIGITAL" and matches JHOVE output for that file.
	useType	<useType> restrictions: ORIGINAL_MASTER, PRESERVATION_MASTER, PRODUCTION_MASTER	Within any <techMD> section, if <aes:primaryIdentifier> has "am" prefix, then <useType> must be PRESERVATION_MASTER. If <primaryIdentifier> has"pm" prefix, then <useType> must be PRODUCTION_MASTER. Within any <sourceMD> section, <useType> must be ORIGINAL_MASTER
	primaryIdentifier	filename in barcode/track.wav format: 39015087083500/am000001.wav	Validate that all <aes:primaryIdentifier> values in <techMD> sections exist as filenames in the directory, and that all .wav filenames in the directory exists as <aes:primaryIdentifier> values.
	checksumValue	must exist and contain data	validate against submitted file
	checksumKind	md5	validate that value is md5
	numChannels	no required value	validate that <aes:numChannels> value from the JHOVE output for a given file matches the <aes:numChannels> value in that file's <techMD>.

Metadata Type	Element/ Attribute	Required or Recommended Value	Validation Recommendations
	sampleRate	96000, unless original format is DAT or CD, then sampleRate should match that of original object	If <sourceMD>/<aes:format> is "CD" or "DAT", then <techMD>/<aes:sampleRate> must have the same value as <sourceMD>/ <aes:sampleRate>. Otherwise, <techMD>/ <aes:sampleRate> must be 96000. Also, match to each file's JHOVE output to ensure that file does in fact have the properties indicated in the METS
	bitDepth	24, unless original format is DAT or CD, then bidDepth should match that of original object	If <sourceMD>/<aes:format> is "CD" or "DAT", then <techMD>/<aes:bitdepth> must have the same value as <sourceMD>/<aes:bitdepth>. Otherwise, <techMD>/<aes:bitDepth> must be 24. Also, match to each file's JHOVE output to ensure that file does in fact have the properties indicated in the METS

THE UK SOUNDMAP: AN AUDIO CROWDSOURCING EXPERIMENT

Richard Ranft, Head of Sound & Vision, The British Library
Paper presented at the IASA-AMIA 2010 Conference, Philadelphia, USA

Why a sound map?

Certain kinds of sound collections are amenable to presentation online using a map interface. Plotting the geographical position of recordings on a map gives a visually attractive and intuitive way to compare and group neighbouring recordings and may be a better alternative to text-based search and browsing. For example, many of the 3.5 million recordings in the British Library's sound collections contain geospatial data: locations where recordings were actually created, music or dialect origins, or place-names referenced to in the metadata. In 2001 the British Library launched *Listen to Nature*,[19] a website showcasing 400 nature sounds from around the world from its existing wildlife sound collections. The site uses static web maps built by BL staff and is occasionally added to. In 2007 the Library created the *Sounds Familiar* British dialects map,[20] similarly designed and built in-house. Listeners may submit new dialect recordings of their own by mailing them on CD to the British Library. The recordings are then digitized and eventually added to the *Sounds Familiar* web map. Our *Archival Sound Recordings* website, launched in 2007, makes extensive use of interactive maps[21] to show the locations of music, environment, oral history and dialect recordings that can be played directly from the map interface. Additional recordings are added from time to time to the maps. However, none of the aforementioned websites allow direct, real-time upload and interaction by contributing recordists. That has become a more recent phenomenon of a participatory culture, shown by the flourishing number of web maps that are updated by the general public, many of them using the Google Maps API.[22] Examples include *Radio Aporee*,[23] launched in 2006, which allows anyone to upload geotagged environmental sounds worldwide under a Creative Commons License; and there are now numerous other websites of sound maps of Montreal, Berlin, Seoul, Barcelona and other cities.

An acoustic survey of Britain using geotagged sound samples

In 2009 British Library sound archive staff began tests for a new kind of field recording project to aggregate user-generated digital audio content using mobile phones. Named the UK SoundMap, the project represents a radical departure from the more traditional, curator-led professional archival practices we were used to. Hitherto, these involved a drawn-out sequence, beginning with acquisition, then professional accessioning, cataloguing, preservation and eventually, if funding and copyright terms allow, online access. The UK SoundMap instead uses an informal community of guided but untrained mobile phone users and field recordists to capture and describe their environmental sounds, then enable near-instant public sharing on a dedicated website: in effect, contributors as curator-publishers.

The initiative arose when one of our curators, Isobel Clouter, secured seed funding via the Noise Futures Network,[24] an interdisciplinary network of UK academics, urban planners, sound artists and noise pollution experts interested in Britain's acoustic landscape who need large audio datasets for research**.** The project aim is to create an acoustic snapshot of the UK in 2010-11, using low-cost technologies to archive a significant body of audio data, around 3,000 recordings, of sufficient quality so that it can be analysed by the Noise Futures Network and other researchers. Sounds that are likely to disappear from our environments, including everyday noises that are often overlooked in archival acquisition programmes, are being

19 www.bl.uk/listentonature
20 www.bl.uk/soundsfamiliar
21 http://sounds.bl.uk/Maps.aspx
22 See for example http://googlemapsmania.blogspot.com/
23 http://aporee.org/maps/
24 http://www.noisefutures.org/