The River Street Anthology: A Digital Curation Work Plan

Daniel Goodman

Caitlinn Grimm

Alexandra Hoyos

Lewis West

San Jose State University

Dorothy Waugh

INFO 284-15

May 8, 2019

The River Street Anthology is an effort started by Matt Jones to record and preserve Michigan music (Milo, 2016a). Jones and his small team travel the state and record short fifteen to thirty minute performances of local artists, setting up shop in his basement, old schoolhouses, churches, and other venues that will have them (Milo, 2016a). “Jones wants to travel to every corner of the state and record as many singer-songwriters, storytellers, poets, musicians, rappers, composers and overall local characters, as he possibly can… even if it takes forever” (Milo, 2016a). He has now traveled over 2000 miles across the state (Milo, 2016a) and recorded over 200 artists (Coombe, 2016). The initial anthology project focused on the music itself, but as the collectors continued working, they realized that audio recordings of the music were not enough, “Context is essential to a good story, and in order to make The River Street Anthology multi-dimensional, measures are being taken to round out the experience,” which includes adding video recordings, photographs, writings by the musicians about their song selection, and sketches of the recording sessions (The River Street Anthology, n.d.).

Archiving and preserving these musicians and their work allows people in the future to hear Michigan’s history from the people themselves in an engaging way. Descendants may be able to hear their ancestors, musicians find inspiration, and researchers find new artists to study, finding new ways to enjoy and use Michigan’s digital heritage. Jones recognizes this need for “permanent archiving” (Milo, 2016a), and has partnered with the Archives of Michigan to preserve the stories of Michigan that Jones is recording and provide access to them online (Milo, 2016b).

It is unclear, however, how Jones will transfer the files to the Archives, and how the Archives will work to curate and provide access to all of the materials in the Anthology, other than through their website (Milo, 2016b). This work plan is a proposal for how the best practices of digital curation could be applied to the curation of existing and future Anthology materials. These materials consist of the over 200 WAV audio files recorded from a Tascam DP-02 digital 8 track (TEAC American, Inc., n.d.). Other materials include digital photographs assumed to be in TIFF or JPEG, digital video recordings assumed to be in .mov, writings assumed to be either analog or in .doc(x) or PDF, and analog artworks. The work plan will outline how the authors, acting as the Archives of Michigan curators, plan to receive the materials, digitize analog objects, curate the materials, and guide Jones in planning his future creations to be preservation-ready.

Developing this work plan conceptualizes the future processes needed to curate the Anthology materials throughout their life-cycles, guiding daily practices, managing risks, and gaining intellectual control early on (Oliver & Harvey, 2016). The work plan can also serve as its own record of what the Archives and curators have done and how they planned to curate the collection, supporting the development of policies and procedures, and serving as evidence for future users and curators (Oliver & Harvey, 2016). By following the plan’s guidelines for receiving, preserving, storing, providing access, and extracting and assigning metadata, the materials themselves will be further protected from inaccessibility and irretrievability, while also making the analog materials more accessible through digital surrogacy.

**Create or Receive**

**Receive Existing**

The 200-plus existing anthology materials were created by Matt Jones and his team of videographers and artists, and as such, Jones was largely responsible for recording any provenance and other information to be included in the description and metadata of the materials. Jones has a working capture procedure to use his Tascam recorder, a single microphone, and that the musician only has one take in the fifteen to thirty minute sessions to focus on the song or songs of their choosing (The River Street Anthology, n.d.). During the session, basic information about the musician is recorded, and photographs and video may be taken.

A donor interview will be conducted prior to file transfer to clarify exact media and file types to be transferred, to discuss copy and intellectual property rights or access/use restrictions, and decide on essential characteristics for preservation. As previously noted, the digital objects are created in a number of file formats and types, including WAV, TIFF, JPEG, .mov, .doc(x), and PDF. The Tascam digital 8 track can save items to its internal hard drive, can transfer items to a computer or usb drive, and can burn directly to CD (TEAC America, Inc., n.d.). Assuming that the WAV files were transferred to a computer or external hard drive, rather than CD, these files, along with the documentation (writings, photographs, and video files) will be transferred to the Archive, to be added to the digital collection via secure, remote access to the Networked Attached Storage (NAS) discussed further in a later section of this paper. Once in custody of the Archive, preservation copies will be made and converted to their recommended preservation formats (Oliver & Harvey, 2016) of WAV, TIFF, .mov, or PDF (Smithsonian Institution Archives, n.d.), if needed. A unique identifier and basic metadata on provenance, date received by the Archive, file format and size (Oliver & Harvey, 2016), will be added to each file.

**Continuing Creation**

**Digitization.** Part of the Anthology’s documentation are writings and artwork. It is possible that some of the writings are done by hand, in which case they should be accepted following procedures for analog items, along with the sketches or paintings that may be accepted. However, to provide better access to these analog materials and collocate them with the rest of the Anthology collection, analog writings and artwork should be digitized. Paper objects, including short writings and sketches will be scanned and saved as PDFs and TIFFs respectively. Paintings or other 3D objects less conducive to scanning will be photographed and saved as TIFFs. These new digital objects will then be added to the collection in the digital asset management (DAM) program and NAS, along with their basic required metadata and integrated into the digital preservation procedures. Unique identifiers for these items should be assigned and also noted in the description of the original analog items. The original analog items will then be passed to the appropriate archivist to be integrated into the preservation workflow for non-digital materials.

**Accruals.** Jones plans to continue his work in preserving Michigan’s musical present for the future, and as such accruals are expected. As suggested in Oliver and Harvey (2016), the Archives has worked with Jones, as the data creator/collector, to outline the best practices for his continued work and creation of Submission Information Packages (SIPs) that are curation-friendly. According to these guides, Jones and his team will continue to record audio in WAV and video in .mov. The designated photographer will save images as TIFFs, while all of the digital writings and essays submitted to Jones will be saved as PDFs. Jones will have secure access to an upload folder in the Archive’s NAS, to which he will add new files as they are created. Jones will also have each musician/participant sign a donation agreement outlining their rights, the rights they are transferring or granting the Anthology and the Archives, and/or any restrictions they would like placed on their recordings. These agreements will be uploaded to the NAS at the same time as the files for those artists, and have their information added to the associated recordings’ metadata.

**Appraise and Select**

The selected assets carry archival value by capturing a separate, yet important, aspect of the experience and history of the music and culture of Michigan. Each asset is a part of a full set of assets that provide context and information on a separate part of Michigan music culture. Carefully curated, these sets of assets offer a snapshot of Michigan history and “is creating a map in real time of the Michigan musical landscape” (The River Street Anthology, 2015).

The digital assets can be divided into two groups: the music itself, and all of the accompanying materials that bring context and detail to the music. The music is central to the collection and needs to be retained. The collection of accompanying materials, such as documents, videos, photos and artwork, will require regular assessment of their value to the collection. Determining which of these assets are essential or significant to bringing context to the music will be a challenging process. While the original donors may find all of the material to be relevant and worth keeping, retaining such a large amount of the assets is impractical for the storage and curation of the collection (Oliver & Harvey, 2016).

The team of archivists will plan on a maximum amount of accompanying assets, such as 20gb or no more than 80 files, for any given musical recording. This maximum limit will serve as a general guideline to assist with selecting material for disposal. Additionally, assets related to a musician’s personal history that do not directly relate to the songs being collected are not relevant to the collection.

All assets should be considered irreplaceable, since the digital copies in the collection may be the only ones reliably stored. These assets are also being stored for the long-term, as they carry historical merits for the community and state of Michigan (The River Street Anthology, 2015). The Archive has mechanical rights to store and curate the collection indefinitely (The River Street Anthology, 2015), but has no obligation to any individual assets. The determining factor on whether or not to keep each item is strictly how relevant and valuable it is to the collection and to contextualizing the music. Before disposing of any assets, the archivists will contact the donor, in case the original musicians want to retain the assets.

**Preservation Actions**

The methods for preservation most appropriate to the River Street Anthology project involve the normalization of the various file formats and types by transferring media to the recommend preservation formats of WAV, TIFF, .mov, or PDF (Smithsonian Institution Archives, n.d.) and migration to newer and more stable storage media (Oliver & Harvey, 2016). Emulation is not deemed an applicable method for these files, as the technology that stores/accesses the collection has no significance to the experience of the user. Additionally, technology preservation is irrelevant to this collection and would potentially take up a lot of physical storage space, capital, and time (Oliver & Harvey, 2016), therefore migration is the cheapest and easiest solution to maintain and preserve assets. To ensure authenticity, reliability, and usability of the files, preservation metadata and representation metadata will be applied (Oliver & Harvey, 2016). Other methods include ensuring that file formats are accessible, regular backups, multiple copies, disaster recovery plans, data cleaning, checksums, validation, and virus protection (Oliver & Harvey, 2016).

Regarding these particular files, the same preservation method would apply to all, as they already exist in or will be normalized to widely accessible formats, making other preservation methods unnecessary. This project is working closely with the content creator/collector to ensure best practices in file formatting moving forward. The OAIS Reference Model’s guidelines for successful preservation methods provides access to an audit and certification checklist based on the OAIS framework “which represents best current practice and thought about the organizational and technical infrastructure required to be considered trustworthy and capable of certification” (RLG-NARA Task Force on Digital Repository Certification, 2007, p. 2). In addition to documenting preservation strategies, mechanisms will be in place for “monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable” (RLG-NARA Task Force on Digital Repository Certification, 2007, p. 31).

Challenges specific to integrity and authenticity can arise as bits are lost when migration of data is implemented and repeated. Furthermore, to safeguard data or digital objects that might be problematic in the future, this plan will be followed, reviewed, and revised regularly to ensure best practices are implemented. Beginning with the best quality file formats possible and standardizing data formats for long-term preservation and future accessibility is a priority, as well as documenting the process when migrating data, therefore, long-term preservation should not be particularly challenging. Moreover, performing proactive migrations to avoid obsolescence of hardware along with long-term planning to stay on top of shifts in file format best practices will be practiced to preempt potential challenges that may arise.

Preservation metadata includes administrative data dealing with the provenance of a resource and its archival management; the metadata required to support long-term preservation of digital assets expands across descriptive, technical, and administrative data (Oliver & Harvey, 2016). Preservation metadata that will be added to this collection’s objects include information regarding the who/when/where of creation, object transfer, and chain of custody (Oliver & Harvey, 2016). Technical information, like file format and software information, required to ensure access and use will also be included (Oliver & Harvey, 2016). Unique and persistent identifiers are necessary to identify the files; a rights statement should be included to delineate access and use.

Information Packages are imperative to the long-term preservation of digital objects by containing the information required to preserve a digital object, access, and understandability (Lavoie, 2014). The OAIS Reference Model has three components to Information Packages that support long-term preservation: Submission Information Package (SIP), Archival Information Unit (AIP), and Dissemination Information Package (DIP) (Lavoie, 2014). The AIP is the fundamental piece to preservation with the metadata that describes the content information and preservation description information of a digital object (Lavoie, 2014).The inclusion of preservation metadata provides the necessary information to establish provenance, history, context, and access to the digital object. This information can ensure long-term accessibility by giving enough detail to identify the digital objects and the information needed to open the files. It ensures the integrity of the digital object by documenting the chain of custody, and confirming fixity by tracking changes. Thus, supporting documentation of actions taken is evidence of both following best practices or errors made. The OAIS Reference Model highlights best practices for Preservation Descriptive Information (PDI) which includes Provenance Information, Reference Information (identifier), Fixity Information (checksum), Context Information, and Access Rights Information (Lavoie, 2014), and these will be followed.

**Storage**

The first step in determining a storage solution for an archive is figuring out the storage volume required. We are receiving 200 sets of assets that range from 30 minute audio files, to video recordings, still assets and more. As stated previously, there is a limit of 20GB per set of musical recording and accompanying assets. While we can guess that not every artist has the full 20GB of files, for ease of calculation and to ensure enough space our initial estimates will be 20GB per artist. 200 artists over a period of four years would net out to 4 TB at a rate of 1TB per year. In the perspective of modern storage devices, this is minimal and allows us to build a strong server for relatively cheap that can ensure security, storage capacity, and durability/reliability.

The Archive will purchase and maintain a NAS system to store all digital assets. The NAS will be set up in a Redundant Array of Inexpensive Disks (RAID) 1 or RAID 10 configuration to ensure reliability of the hard drives through parity. This will also allow us to easily expand storage space at a relatively low cost. Through free software, fixity checks will automatically occur on a monthly basis to ensure all assets on the NAS are accurate and unchanged/uncorrupted.

While RAID-configured NAS systems are fantastic at providing reliable storage, to ensure the preservation of the Archive’s assets, additional steps will also be taken. Amazon Glacier is an extremely low cost solution to cloud storage, “designed to deliver 99.999999999% durability” and a cost of “$0.004 per gigabyte per month” (Amazon S3 Glacier, n.d.). If we estimate the current collection at 4 TB, that would be a total of $192.00 a year for storage on this platform. The Archives will add each artist’s set of assets once they are finalized and ready for preservation. The uploaded assets will also include an exported copy of the files’ metadata in .csv format. With this parallel solution, if the NAS system is destroyed in an event like a fire, the Glacier assets will still be safe and accessible.

Through the NAS, we will also implement a DAM system. Luckily, on-premise solutions tend to be more affordable. Canto (On Premise Digital Asset Management, n.d.) and AssetBank (AssetBank, n.d.) are two highly reputable potential DAM providers that offer cheap on-premise DAM licenses. We can expect to pay roughly $4000-$7000 a year for this license (AssetBank, n.d.), but the price may be negotiated lower. These platforms also have out-of-the-box transcoding solutions to take our hi-res files and output a lower-res web-friendly file (Brightcove, n.d.; Configuring video download options, n.d.). There may also be cheaper/free but less robust solutions worth exploring such as XnView MP (XnView MP, n.d.).

On-premise solutions mean that security and access to the DAM/NAS is easily configurable. The DAM will reference a folder structure in the NAS. Jones will have access to an upload section/folder to place new assets. As previously mentioned, we have worked with Jones to configure an appropriate SIP so that archiving is predictable and seamless. We have also encouraged Jones to keep all assets he creates on his own external hard drives for redundancy. We will also configure access to users from various partner organizations around Michigan as well as web administrators who are responsible for maintaining and updating web assets.

**Access, Use, and Reuse**

There are several significant properties that should be preserved to enable efficient access to and use/reuse of the collection. These properties include the content of the files; contextual information that can identify creators, locations and times; and structural information that can aid with proper use and access. Most of the files will be static, so behaviors like linking and updating will not be relevant. Some artistic documents may also require additional properties for appearance, such as use of white space, fonts, colors, etc. These properties should fit into carefully defined standards for better access and sharing (Oliver & Harvey, 2016).

One of the key ways the Archives will enable access to the collection is by maintaining a website promoting the music with its accompanying materials (The River Street Anthology, 2015). By using the latest technology and linking together music and context, the website will be a vibrant source of access to the collection. Master and high-resolution copies of digital objects will be restricted to view-only access on the website so that the public can not download the high-res, un-watermarked assets directly from the site. Transcodes will be provided for the website that will decrease loading times while still displaying digital content in an accurate and presentable manner. Measures will also be taken within the website’s coding or the embedded media player to restrict saving and downloading assets to a user’s computer.

The Archives will also have small, rotating exhibits in the physical space of the Michigan History Center, which shares a building with the Archives. Visitors to the museum will be able to walk around the space, listening to the music and viewing or reading some of the contextual assets. The digital curators will rotate the exhibit on quarterly intervals, highlighting different musicians around Michigan. The curators will determine whether to group together artists by theme or geography for these exhibits.

The Archives will also share its collection with partner organizations around Michigan. Since the value of the collection extends beyond the confines of the museum and archive, the curators should find partners with which to share the collection (Oliver & Harvey). Organizations that may be suitable partners could be other museums and libraries around Michigan that may be promoting an exhibit featuring similar work. Other likely partners could be venues that may be holding concerts with musicians featured in the collection, and may benefit from utilizing the historical assets. The material may also be shared with printed publications like newspapers or journals, or scholars researching musicians in the collection if in compliance with donor agreements.

The curators will need to be careful with how the objects are shared, so that the sharing does not impact any musician’s ability to monetize their original work. While the Archives has mechanical rights to the assets (The River Street Anthology, 2015), it must still retain restrictions to protect the rights of the musicians. When sharing or allowing access to the collection, the curators must be very clear about the standards and use rights of the collection.

**Metadata**

Descriptive metadata is necessary to find, identify, select, and obtain the record. This type of metadata ensures intellectual and physical control over files/objects for long-term preservation (Oliver & Harvey, 2016). It functions as the keys to identifying, locating, accessing, understanding, and using the digital objects (Oliver & Harvey, 2016). Additionally, it provides context and tools to use and understand the files (Oliver & Harvey, 2016). Moreover, descriptive and representative information is required to understand and render both the digital material and the associated metadata. It helps users to distinguish items and select the ones that work best for them. Administrative metadata, included in the description information, provides information about how files were created and stored, including its use, management, and encoding process over time, as well as information about intellectual property rights (Oliver & Harvey, 2016). Relational or structural information describes how these multiple aspects of digital objects are arranged through the description of the data and its location (Oliver & Harvey, 2016).

Representation information is comprised of structural metadata regarding the technical information needed to use the files, including file format, information about OS, hardware dependencies, encryption, compression, etc (Oliver & Harvey, 2016). Semantic information is another aspect of representation information that “defines relationships among objects or parts of objects,” (Oliver & Harvey, 2016, p. 77) for example an audio file that contains data about what album contains it. Lastly, additional representation information is needed to understand the digital material regarding storage media, encryption, hardware/software or other technical information (Oliver & Harvey, 2016).

Description and representation information to attach to each file/object varies depending on content type of the digital object; the following are brief examples of descriptive and representation metadata to be captured.

* Audio Recording: artist/band name, location, date, song title, duration, genre, description, technical specifications of file, format, rights, subject.
* Film/Video: artist/band name, location, date, duration, description, technical specifications of file, videographer, rights, subject.
* Photograph: artist/band name, location, date, description, technical specifications of file, rights, subject.
* Fine Art: artist, medium, size/dimensions, date, description, technical specifications of file, relation to analog original, location of analog original, rights
* General Metadata per Artist/Band: artist/band name, history/description, origin location, songs/albums, genre of music, years active, label, website, members, past members, aliases.

The digital objects to be acquired consist of audio recordings (WAV), videos (.mov), images/photos (TIFF), short writings/essays (PDF), sketches (TIFF), paintings/3D objects (TIFF), which center on music, and materials providing contextual information. Hardware and software requirements can impede access to material, therefore it is pertinent to collect technical metadata from digital objects that will grant usability. This consists of technical metadata essential to the technology used to create the digital objects, the hardware and software and the formats in which files were created (Oliver & Harvey, 2016). This includes preservation information about the technology used during transfer, when the transfer occurred, as well as the file formats migrated to and from, which all provides the additional technical context necessary to ensure accessibility to users and track changes. Describing how compound digital objects are structured and how each component relates to the others to create the whole (Oliver & Harvey, 2016).Technical metadata is imperative to network management issues and should include information associated with the identification of intellectual property rights, interoperability, and potential obsolescence (Corrado, 2014).

**Conclusion**

Through the process of creating this work plan, the Archives has both encountered and anticipates encountering numerous challenges. The first and most apparent challenge from the onset of this work plan has been how to reconcile the already existing assets with the new standards moving forward. The original scope of the RSA project, upon its inception, was audio recordings. Slowly more types of content were included. Compared to the new standard set of assets being created, the oldest sets seem incomplete. Once new files types were added to the collection, what happens when there are too many assets? The selection process for these sets of assets is a challenge in and of itself. We set a limit of 80 files or 20GB per set, but it will be a tough call to determine what makes the cut or if there are needs to allow for more files/storage.

The new standards also mean that digitization will play a large role in preserving physical objects. Where are those objects? How were they stored? What quality are they currently in and what information was lost as a result of digitizing those objects, potentially years after their creation? Finally, on the subject of the previous assets, what happens when those file specifications are not aligned with the new standards? Determining what quality was lost along with how to potentially preserve unusual file formats will be a difficult hurdle. Is there a way to increase the quality of those assets (such as re-digitizing files) or are we stuck maintaining the sub-standard file formats because they are the highest potential quality?

Varying curator skill sets and knowledge led to the need to explain ideas and thoroughly communicate to avoid or clarify misunderstandings. The different technical skills of the work team meant diverse ideas and suggestions for storage and transfer that needed to be researched and negotiated until consensus was reached. In the future, these decisions may be re-evaluated in light of new technical expertise or the existence of better technology, standards, or best practices.

There is also the very real challenge of financing the project long-term. Maintaining a working storage system, DAM license, and web domain while potentially supporting a team of artists, photographers, filming crew, and other various expenses requires a consistent and non-negligible cash flow. Hopefully the project will gain traction with web views, donations, and other grass-roots funding to support the continued work on the project. Perhaps the quality of work enabled by this work plan along with Matt Jones’ promotional efforts will bring in enough interest and community support to make this project successful and self-supported.

The River Street Anthology project is an important and captivating effort worthy of the highest quality planning and preservation strategy. This work plan provides industry best practices following the DCC Curation Life Cycle Model; using this plan, as outlined, would ensure that the history and context of music in Michigan will have a permanent, complete, and accessible home in our archives.

References

Amazon S3 Glacier. (n.d.). Retrieved from <https://aws.amazon.com/glacier/>

AssetBank. (n.d.). Retrieved from <https://www.assetbank.co.uk/features/>

Brightcove. (n.d.). Retrieved from <https://www.canto.com/integrations/brightcove/>

Configuring video download options. (n.d.). Retrieved from <https://support.assetbank.co.uk/hc/en-gb/articles/115005116947-Configuring-video-download-options>

Coombe, D. (2016, September 30). Photos: The River Street anthology. *Michigan Radio*. Retrieved from <https://www.michiganradio.org/post/photos-river-street-anthology>

Corrado, E. M., & Moulaison, H. M. (2014). Metadata and metadata for digital preservation. *Digital Preservation for Libraries, Archives, & Museums* (pp. 111–42). Lanham, MD: Rowman & Littlefield.

Lavoie, B. (2014). *The Open Archival Information System (OAIS) Reference Model: Introductory guide* (2nd ed.). York, UK: Digital Preservation Coalition. Retrieved from <http://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>

Milo, J. (2016a, February 20). ‘River Street’ captures Mich. music in historic proportions. *Detroit Free Press*. Retrieved from <https://www.freep.com/story/entertainment/music/2016/02/20/river-street-captures-mich-music-historic-proportions/80543204/>

Milo, J. (2016b, July 23). Archives of Michigan partners with music preservationist. *Detroit Free Press*. Retrieved from <https://www.freep.com/story/entertainment/music/2016/02/20/river-street-captures-mich-music-historic-proportions/80543204/>

Oliver, G., & Harvey, R. (2016). *Digital curation (2nd ed.)*. Chicago, IL: Neal-Schuman Publishers.

On Premise Digital Asset Management. (n.d.). Retrieved from <https://www.canto.com/cumulus/cumulus-on-premise/>

The River Street Anthology. (n.d.). Retrieved from <http://seekingmichigan.org/product/riverstreetanthology>

The River Street Anthology. (2015, July 7). Status update [Facebook post]. Message posted to <https://www.facebook.com/riverstreetanthology/posts/1462979677353223>

RLG-NARA Task Force on Digital Repository Certification. (2007). *Trustworthy Repositories Audit and Certification: Criteria and Checklist*. Chicago: Center for Research Libraries. Retrieved from: <http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf>

Smithsonian Institution Archives. (n.d.). Recommended preservation formats for electronic records. Retrieved from <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>

TEAC America, Inc. (n.d.). Tascam DP-02/DP-02CF digital portastudio: *Owner’s manual*. Montebello, CA: Author. Retrieved April 21, 2019 from <https://tascam.com/downloads/tascam/37/E_DP-02&02CF_web.pdf>

XnView MP. (n.d.). Retrieved from <https://www.xnview.com/en/xnviewmp/>