Calculating All that Jazz: Accurately Predicting Digital Storage Needs Utilizing Digitization Parameters for Analog Audio and Still Image Files

Calculating All that Jazz

Accurately Predicting Digital Storage Needs Utilizing Digitization Parameters for Analog Audio and Still Image Files

Krista White

Library professionals and library assistants who lack computer science or audiovisual training are often tasked with writing digital project proposals, grant applications or rationale to fund digitization projects for their institutions. Much has been written about digitization projects over the last two decades; digital storage has been highlighted as a central feature of any digitization project, especially the need to purchase additional storage mechanisms to house digitized collections. What is missing from the library science literature is a method for reliably calculating digital storage needs on the basis of parameters for digitizing analog materials such as documents, photographs, and sound recordings in older formats.

Much has been written about digitization projects over the last two decades, and digital storage has been highlighted as a central feature of any digitization project. What is missing from the library science literature is a method to reliably calculate digital storage needs on the basis of parameters for digitizing analog materials such as documents, photographs, and sound recordings in older formats.[1] Library professionals and library assistants who lack computer science or audiovisual training are often tasked with writing digital project proposals, grant applications or providing rationale to fund digitization projects for their institutions. Digitization projects involve purchasing additional storage mechanisms to house files for preservation and access. Digital project managers need tools to accurately predict the amount of storage for housing digital objects and estimate startup and ongoing costs for such storage.[2] To make those predictions, they must decide which standard their organization will use to create archival masters for long-term access and/or preservation because the standards they apply will affect digital file sizes. This paper provides two formulae for calculating digital storage space for uncompressed, archival master image and document files and sound files. The two formulae presented provide parameters for digitization that will also aid digitization project managers to make informed decisions regarding digitization standards and equipment purchases for their projects. Formulae for 3-D scanning and moving image (video) objects would be a valuable addition to the field, but are beyond the scope of the current study.

The first part of this paper lays out the method for the formulae for predicting the digital storage needs of analog objects, which depends on their media types and characteristics. The second section, the literature review, demonstrates

Krista White (kwhite2@rulmail.rutgers.edu) is a Digital Humanities Librarian at Rutgers University Newark.

Manuscript submitted July 20, 2015; returned to author for minor revision October 5, 2015; revised manuscript submitted October 28, 2015; accepted for publication November 6, 2015.

The author wishes to thank Isaiah Beard, Fernanda Perrone and Robert Nahory for their assistance and advice in the research and writing of this paper.

aspects of digital project management, contextualizing the environment in which librarians and digital project managers must predict digital storage needs, including costs, professional debates about digitization as a preservation tool, and varying best practices and standards documents that complicate project implementation. The third section of the paper introduces the formulae, the experiment design, and the results of testing the formulae for accuracy and reliability. In the final section, the results of the experiments and the elements of the formula for still image and document storage calculations are contextualized using experiences reformatting the transcripts for the Jazz Oral History Project (JOHP) at the Institute of Jazz Studies (IJS) at Rutgers University. The appendix at the end of the essay defines terms to help those new to digitization navigate specialized terminology used here.

The JOHP is

> a collection of audio tapes for 120 oral histories of seminal pre-Swing Era and Swing Era jazz musicians recorded between 1972 and 1983. The JOHP was initiated in 1972 by the Jazz Advisory Panel of the Music Program of the National Endowment for the Arts. Musicians sixty years and older (as well as several younger artists in poor health) were interviewed in depth about their lives and careers. The taped interviews range in length from 5 to 35 hours each and are accompanied by typewritten transcripts. They have been consulted by hundreds of scholars and writers producing articles, books and dissertations, in addition to frequent use by producers of radio and television.[3]

The process of digitizing the nearly 26,000 pages of transcripts for ingestion into RUcore, the Rutgers digital repository, is underway to make the transcripts and audio files of the JOHP publicly available online. Research on calculating digital storage needs occurred simultaneously with the JOHP digitization project because other oral history projects were being submitted to the libraries for digitization and digital storage of the JOHP needs for these projects also needed consideration. In conjunction with advice from the Rutgers University Libraries' (RUL) Digital Data Curator, the research presented here helped in the evaluation of the digitization standards and processes used for digitizing the JOHP and helped to highlit how much storage space the team saved on the institutional repository servers.

Method

The first part of the method involved a literature review, combing computer science literature on digital storage, library science and archival studies literature regarding digital libraries, and professional literature on standards and best practices for the digitization of materials. During the course of the research, the author discovered formulae for calculating digital storage on the basis of the characteristics of analog materials. These formulae surfaced in older instructional and do-it-yourself literature on multimedia object creation and in online multimedia and computer science literature designed for high school and undergraduate students. The second part of the method focused on testing the found formulae to determine whether they were reliable and accurate. The first experiment tested the accuracy of formulae for predicting file sizes when digitizing still images and documents. A second experiment tested the accuracy of the formula for predicting file sizes of digitized audio recordings. The findings on the formulae and accuracy were then applied to the work digitizing the JOHP transcripts. The JOHP example demonstrates how project managers can use the formulae discussed to make decisions about purchasing equipment and evaluate digitization standards to meet the needs of their projects and institutional goals.

Literature Review

The project began with a literature search for ways to calculate digital storage needs of digitized, analog objects. That search met with no success. This may be because librarians depend on their information technology specialists to supply such information. It is absolutely appropriate for librarians to rely on experts from areas like information technology (IT), which typically fall outside the domain of the profession, to help them calculate storage needs for digitization projects. However, IT professionals, even within library systems and IT groups, are not always familiar with digital preservation best practices. If they are familiar with digital preservation practices, IT professionals often present storage figures in absolute terms, assuming fixed values for digitization variables that may or may not be appropriate for long term preservation. In his 1997 book, *Practical Digital Libraries*, Lesk estimated thirty megabytes of storage for every hour of compressed audio, one megabyte for a page of uncompressed, plain text (bitmap format), and three gigabytes for two hours of moving image media.[4] Lesk gives no estimate for raster images such as TIFF, JPEG, or GIF images, or for the storage of uncompressed, archival master digital files; rather, he is concerned with providing figures that represent the most economical memory and storage options for delivery of objects in a digital library system. Jordan estimates storage needs for raster image files of 90 megabytes for uncompressed raster image files, 600 megabytes for one hour of uncompressed audio recording, and "nearly a gigabyte of disk space," for one minute of uncompressed digital video.[5]

In *The State of Recorded Sound Preservation in the United States*, the National Recording Preservation Board quotes a figure of 100 gigabytes (GB) of storage for 100 hours of audio tape.[6] Calculations given by these information technology professionals and standards organizations are accurate (if approximate in some cases), but they assume fixed rates for many variables in the digitization process that may not suit a particular institution's needs or the chosen digitization standard for a project. Those variables can be adjusted to alter both file size and quality, which affects the choice of digitization standard, the combination of variables used in a chosen standard and the quantity of digital storage required.

Among librarians and archivists, the issue of digital storage is taken quite seriously for digitization and digital library projects. In the newly released *Association for Recorded Sound Collections (ARSC) Guide to Audio Preservation*, Lacinak's chapter provides an overview of the issues related to digital storage, providing an in-depth example with guidance on decision making in that domain of digital initiatives.[7] Other literature in the field discusses storage as the platform for stable, long-term storage of digital assets. The section on storage in Hodge's paper on a lifecycle framework emphasizes its importance as a mechanism for long-term preservation.[8] Hooper's audio e-reserves project at Tulane features stable, digital storage for the new, master files as a cornerstone of the project, as do Pastine, Bayard, and Lang in a similar project at Temple University to create an e-reserves system for digital images for courses and general discovery.[9]

Other library science literature frames digital storage as a basic resource of digital projects whose capacity will need to be enlarged to accommodate digital library initiatives. Jones' paper on the creation of a history portal for digital objects related to the history of Michigan lays the issue out plainly, "Planning for storage needs is an ongoing task. . . . The increase in storage capacity made necessary by (Making of Modern Michigan) is only a fraction of the increase needed to handle the expansion of our own digitized collections."[10] This was the case with Pastine, Bayard, and Lang's project at Temple University, and was a feature of Maurya's paper on the challenges and hopes for digital library services in India.[11]

The suitability of migrating existing materials into digital format for preservation purposes has been contentious for more than a decade.[12] However, Arthur and her colleagues have argued that digitization be accepted as a preservation format.[13] In concert with Hodge et al., Arthur and her colleagues have highlighted the importance of digital storage as a final destination for files migrated from older, analog formats.[14] Some individuals and organizations may not consider reformatting analog materials to digital files as a long-term, viable preservation strategy, but for many projects, including the JOHP audio tape recordings, this option is the best for analog materials that have reached the end of their useful life. In these cases, library science literature demonstrates the importance of storage in digitization initiatives, but there is no guidance in any of these sources for a method to estimate the amount of storage a given project requires.

Knowing the cost of digital storage—whether starting a new project or expanding on existing storage infrastructure—is crucial to digital project planning. Costs of digital storage are generally framed in terms of dollars or cents per storage unit. Planning for storage costs, therefore, requires knowing how much storage a project will need. In an older example, Lesk's 1990 report to the Foundation of the American Institute of Conservation for Conservation at the American Institute for Conservation of Historic and Artistic Works provides a detailed chart of the cost of various digital storage formats for library materials.[15] More recently, Lazorchak compiled an excellent bibliography on "Digital Asset Sustainability and Preservation Costs."[16] Echoing Kenney and Personius, two papers featured in Lazorchak's bibliography, one by a group at the San Diego Supercomputer Center and the other by Sanett, provide an overview of the costs of different storage media plus maintenance, labor, infrastructure software licenses utilities and floor space associated with digital storage hardware.[17] Lesk's paper is the only one that provides actual numbers for magnetic hard disk costs (what is often called "server storage"), quoting a price of $4,000 per gigabyte in 1990.[18] Smith's informal but comprehensive bibliography on the cost of hard drive storage space claims the cost in 1990 was $9.00 per megabyte ($9,216.00 per gigabyte) and shows current costs per gigabyte of storage to be $6.33 as of July 2013.[19] Digital storage comes in many forms, from gold CDs to magnetic hard drive arrays connected to networks (also known as "server storage" or "cloud storage"). Costs vary by storage format and must be sustainable. Despite the downward trend in the cost per gigabyte, storage media must be periodically replaced as hardware gets old and experiences failures or the format becomes obsolete. That translates into ongoing, permanent costs for storage mechanisms in every digital project. Even small costs can be burdensome to cultural heritage institutions working with limited budgets. The ability to plan for costs related to the growth of digital assets hinges on an organization's ability to accurately estimate the amount of digital storage for current and future objects in a collection.

Further complicating the work of digital project managers and directors in estimating digital storage needs is the existence of multiple standards and best practice documents for proper stewardship of archival digital materials because standards used to digitize analog materials directly affect file sizes. The Library of Congress' Federal Agencies Digitization Guidelines Initiative includes recommendations and resources for digitizing still images and advice for preparing the digitization environment, file format comparisons, digitization workflows and overall stewardship

recommendations.[20] It is a deeply technical document that institutions can use to evaluate and create their own digitization standards. The Bibliographical Research Center built upon the Colorado Digitization Project's work to create the Best Practices for Digital Imaging. Their document includes a nuanced, understandable explanation of the digitization process, recommendations for decision making for staffing, and training and software and hardware considerations for digitization projects, plus concrete parameters for digitizing analog materials.[21] The Smithsonian Institution Archives and the World Digital Library both have simple webpages detailing digitization standards for their collections.[22] The Colorado Digitization Project wrote a guide to best practices for digitizing analog audio sources and the Association for Library Collections and Technical Services (ALCTS), a division of the American Library Association whose Preservation and Reformatting Section's (PARS) mission includes, among other things, the preservation of library materials, created a guide for digitizing all types of analog objects according to format.[23] The Digital Preservation Coalition has a standard for digitizing moving image media, which differs significantly in file format, bit rate, and color recommendation from that of the standard set out by the Consortium of Academic and Research Libraries in Illinois (CARLI), while the Federal Agencies Digitization Guidelines Initiative's Audiovisual Working Group is still formulating its standards for video/moving image materials.[24] Individual institutions, especially those with digital repositories, may create their own guidelines. Rutgers has locally developed standards for digitizing analog documents, images, audio, and moving image materials that are based on independent review and testing of standards set by other bodies.[25] Utilizing standards is crucial to creating stable, long-term digital surrogates of older archival objects, but the existence of multiple standards, even when closely matched, may be confusing to the uninitiated digital project manager. Standards bodies do not provide insight into how standards affect digital storage needs or provide guidelines that would help library professionals choose appropriate settings for parameters within those standards. These variables profoundly affect the amount of digital storage necessary for a project.

Reviewing older instructional and do-it-yourself literature on the creation of multimedia objects resulted in the discovery of mathematical formulae for calculating digital storage needs for analog still images, documents, audio and moving image recordings. Many of these sources were rightly concerned with explicating the process of digitization and monitoring variables to insure quality.[26] Only a few were concerned with the practicality of determining the size of digital files in the final output of the digitization process. Three resources yielded formulae for calculating the file sizes of digitized still images and documents: Tally's *Avoiding the Scanning Blues*, Note's *Managing Image Collections: A Practical Guide* and Cunningham's formula for digital, bitmap still images.[27] Tally's formula omits the essential element of the physical dimensions of the scanned image, which is crucial in calculating file sizes for photos and documents; Both Note's and Cunningham's formulae include image or document size, scanning resolution and bit depth. The formulae in these two sources are essentially the same.[28]

Cunningham's webpage and Johnson, Gault, and Florence's *How to Digitize Video* were the two sources that elucidated formulae for calculating files sizes for digitized, analog audio recordings.[29] The formulae in both sources contained the same elements necessary for calculating storage sizes: length of the original audio recording, sampling rate, bit depth, and number of audio channels.[30]

Three resources in the literature review contained formulae for predicting uncompressed digital file sizes for moving image (video) objects. Rice and McKernan's formula seemed incomplete. They added an extra, unnecessary number for RGB color, which should be accounted for in the bit depth value and their formula lacked any variables to account for sound in the moving image recordings.[31] Cunningham's and Johnson, Gault, and Florence's formulae contained mathematical elements that included frame rate, frame size, bit depth, and length of recording.[32] The combination of Cunningham's two separate formulae for calculating uncompressed digitized moving image files and for uncompressed audio files is identical in content to Johnson, Gault, and Florence's for calculating file sizes for uncompressed audiovisual materials.[33] Johnson, Gault, and Florence combine two formulae for calculating audiovisual materials; they present one formula for the moving image portion of an audiovisual file and another formula for the sound portion of the video file.[34] Cunningham presents his moving image formula separately from the audio formula and does not make clear if they should be combined to calculate the size of digitized audiovisual materials.[35]

None of the formulae proposed by authors listed in the literature review provided supporting evidence of their effectiveness. This required experimentation to test the accuracy and reliability of the formulae. Though the literature review produced formulae for still images and documents as raster files, audio recordings and moving image recordings, the scope of the current paper is limited to testing and explicating the formulae for the reformatting of still images and documents and analog audio files into digital formats. The complexity of the processes behind digitizing analog moving image or video, plus that for scanning as-yet-unmentioned 3D objects, requires its own experimentation and analysis beyond the scope of the current work.

Also absent from the literature reviewed are formulae and experiments for predicting file sizes for born-digital media in all formats. Many tutorials with formulae are available online, created by instructors for high school and

undergraduate-level computer science courses.[36] As with the formulae for calculating storage space of analog-to-digital reformatting procedures, the formulae presented for born-digital objects do not contain data on their reliability or accuracy. In the context of digital exhibits and the archival preservation of born-digital objects, calculating storage space for uncompressed, born-digital objects would be invaluable, but is beyond the scope of this study.

$$\text{File Size in Bytes} = \frac{\text{image width} \times \text{image height} \times \text{resolution}^2 \times \text{bit depth}}{8}$$

Figure 1. Formula for Calculating File Sizes of Uncompressed, Still Images

Image/document size: 2x3 inches, 8.5x11 inches, 11x17 inches.
Resolution: 150 ppi, 300ppi and 600ppi
Grayscale image bit depths: 8 bits, 16 bits
Color image bit depths: 24 bits, 48 bits

Figure 2. Experiment Variables Used to Test the Accuracy of the Still Image Formula

Experiment Design: Accuracy of the Still Image Formula

Figure 1 displays the formula for calculating the file size of uncompressed, unedited still images in bytes, suitable for use as archival master files.[37] Image scanning experiments were performed using an IBM PC with Windows 7 operating system to test the reliability of the formula. An Epson Expression 10000XL scanner and the native EsponScan software, version 3.49A were used to digitize still images. Images were captured as uncompressed TIFF files in accordance with digital, archival practices set out by standards bodies mentioned earlier in this essay. Images were scanned using a combination of variables in each scan, as shown in figure 2.

Variables were chosen on the basis of the digitization standards for still images and documents in the "BCR's CDP Digital Imaging Best Practices," "Minimum Digitization Capture Recommendations" and in "Digitizing Analog Documents and Images."[38] Any variables that do not match those standards were chosen to create atypical file combinations that would test the limits of the still image digital storage calculation formula.

Bit depth was separated into grayscale and color categories because of the fundamental difference between digital capture of grayscale versus color imagery. The two most common archival standards of 24 bits and 48 bits were used to capture color images.[39] The combination of variables resulted in twelve scans per image, with a total of thirty-six images scanned at various document sizes, bit depths and resolutions. Each file was assigned a unique ImageID that indicated its size, scanning resolution, bit depth and whether it was scanned in color or grayscale. For instance, one file was labeled "Si85x1160024C." Si indicated that it was a still image, "85x11" indicated that the original document was 8.5 by 11 inches in size, "600" indicated that it was scanned at a resolution of 600 pixels per inch (ppi), "24" indicated that it was scanned at a bit depth of 24, and "C" indicated that it was scanned for color.

Once the scans were complete, two methods were used to obtain the *measured file sizes* (labeled Ai_1, Ai_2, Ai_3, Ai_4). The first instrument used to obtain *measured file sizes* was the Windows Explorer details Pane of Windows 7 (Ai_1, Ai_3). When a user clicks on a file to highlight it in the Windows 7 operating system, the details pane displays metadata about that file, including the file size. The second instrument used to obtain *measured file sizes* was Media Info (Ai_2, Ai_4), an open source software program that displays technical and source metadata about multimedia files.[40]

All data about the scanned images and documents from the experiment were entered into a Microsoft Excel spreadsheet. Excel's calculate function was used to anticipate the *calculated file sizes* in kilobytes (Ci_1) and in megabytes (Ci_2) using the uncompressed still image file formula.[41]

The differences (Di_1, Di_2, Di_3, Di_4) between the *calculated file sizes* (Ci_x) from the formula and the measured file sizes recorded from Windows Explorer and Media Info in both kilobytes (Ai_1, Ai_2) and megabytes (Ai_3, Ai_4) were calculated in Excel.

$$Di_x = Ai_x - Ci_x$$

Looking at numerical differences between file sizes is useful, but does not provide the lay user with a sense of the value of the differences (Di_x) between the calculated values (Ci_x) and the measured file sizes (Ai_x). A pure mathematical difference would not be an informative measurement of the accuracy of the formula, since different sized files would not produce comparable, uniform variations. To that end, the Percent Difference (Pi_x) between the calculated file size and the measured file size was calculated to show the percentage of the measured file size represented by the difference (Di_x) between the measured file size (Ai_x) and the calculated file size (Ci_x).

$$Pi_x = |Di_x \div Ai_x| \times 100$$

There were discrepancies between values for some of the measured file sizes reported by Windows Explorer and Media Info. Adobe Photoshop CS6 was used as a control to

Table 1. Abbreviations for Still Image File Variables in Calculations by Instrument and Unit of Measurement

Variable	No Instrument (KB)	No Instrument (MB)	Windows Explorer (KB)	Media Info (KB)	Windows Explorer (MB)	Media Info (MB)
Calculated File Size	Ci_1	Ci_2				
Measured File Size			Ai_1	Ai_2	Ai_3	Ai_4
Differences between File Sizes			Di_1	Di_2	Di_3	Di_4
Percent Difference between File Sizes			Pi_1	Pi_2	Pi_3	Pi_4

Table 2. Calculated File Size Compared to Actual File Size of Digital Image Files. Brackets | | indicate absolute values.

Still ImageID	Size (in)	PPI	Color/ Gray	Bit Depth	Calculated File Size (MB) Ci_2	Actual File Size (MB) Ai_4	Difference (MB) Di_4	Percent Difference (MB) Pi_4
Si2x315024C	2x3	150	C	24	0.386	0.395	0.010	2.102
Si2x315048C	2x3	150	C	48	0.772	0.781	0.010	1.123
Si2x330048C	2x3	300	C	48	3.090	3.100	0.010	0.326
Si2x330024C	2x3	300	C	24	1.545	1.550	0.010	0.326
Si2x360048C	2x3	600	C	48	12.360	12.400	0.040	0.326
Si85x111508G	8.547x11	150	G	8	2.017	2.020	0.000	0.129
Si85x1115024C	8.547x11	150	C	24	6.052	6.060	0.010	0.129
Si2x33008G	2x3	300	G	8	0.515	0.516	0.000	0.124
Si2x315016G	2x3	150	G	16	0.257	0.258	0.000	0.124
Si2x31508G	2x3	150	G	8	0.129	0.129	0.000	0.124
Si11x1730024C	11x17	300	C	24	48.151	48.200	0.050	0.102
Si85x1160016G	8.548x11	600	G	16	64.564	64.600	0.040	0.056
Si11x1760048C	11x17	600	C	48	385.208	385.000	|-0.208|	|-0.054|
Si85x1130024C	8.547x11	300	C	24	24.209	24.200	|-0.009|	|-0.036|
Si85x113008G	8.547x11	300	G	8	8.070	8.070	0.000	0.006
Si11x176008G	11x17	600	G	8	64.201	64.200	0.000	|-0.002|

compare measured file sizes. The comparison between measured file sizes reported by Windows Explorer (Ai_3), Media Info (Ai_4) and Photoshop CS 6 (PS_2) in megabytes revealed Media Info to be the preferred reporter of measured file sizes because file sizes measured in Photoshop matched Media Info's measured file sizes more often than they matched measured file sizes in Windows. For the sake of consistency, only files measured in megabytes are reported in this study.[42]

Because the aim of the experiment was to test the reliability and accuracy of the still image digital storage formula, the *absolute* valuesof Di_x and Pi_x were used.[43] The rationale for this choice is that the most desirable value for determining the accuracy of the still image digital storage formula is zero. Therefore, all values produced in the experiment are evaluated as more, or less, accurate by their distance from zero. See the appendix for the definition of *absolute value*.

Still Image Experiment Results and Discussion

Table 2 compares a sample of the data from sixteen of the thirty-six total files created in the experiment. The table compares the *calculated file size* in megabytes of scanned images using the still image (Ci_2), the *measured file sizes* of the files as reported by the Media Info software in megabytes (Ai_4), the *difference* between the calculated files size and the measured file size (Di_4), and the *percentage of the measured file size* that the difference between the calculated and measured file sizes represents (Pi_4). The results in the table represent a spread of files that demonstrate the least amount of accuracy (highest Pi_4 values), results that demonstrate some error between calculated and measured file size (median Pi_4 values) and files demonstrating the least amount of error (small or no Pi_4 values). Negative numbers indicate that Ci_2 was larger than Ai_4.

The 2.0 percent and 1.0 percent errors displayed by

$$\text{File Size in Bytes} = \frac{\text{length of recording in seconds} \times \text{sampling rate in Hz} \times \text{bit depth} \times \text{number of channels}}{8}$$

Figure 3. Formula for Calculating File Sizes of Uncompressed Audio Recordings

Length of Recordings were 30 seconds, 60 seconds, 90 seconds
Sampling Rates at 44.1kHz, 48kHz, 96kHz, and 192kHz
Bit Depths of 16 bits and 32 bits, with the exception of sound recorded at 192kHz
(for which the Adobe Audition software would only record at 16 bits and not 32 bits)
Channels Single and double channels (mono and stereo)

Figure 4. Experiment Variables Used to Test the Accuracy of the Audio Formula

Si2x315024C and Si2x315048C are small and represent statistical outliers in the current dataset. Comparing the rest of the data collected demonstrates that the calculated file sizes differ less than 0.5 percent from the measured file sizes in 94.0 percent of the sample size. The large differences displayed for Si2x315024C and Si2x315048C may be the result of the relatively small file sizes of each; any Pi_4 value will represent a larger absolute percentage of Ai_4 because of the small file sizes. All indications are that the formula for calculating uncompressed, still image digital file sizes is based on original object dimensions, scanning resolution and bit depth is accurate and reliable enough for common use.

Experiment Design: Accuracy of the Audio Formula

Figure 3 displays the formula for calculating the size of uncompressed digital audio files based on recording length, digitization sampling rate, bit depth and number of channels. The experiments were performed on an IBM PC running the Windows 7 operating system. Equipment included a Sony TC-WE475 Stereo Dual Cassette Deck and the audio was transcoded from a standard, commercial-grade music cassette using a Creative Labs Sound Blaster Converter as the Analog to Digital (ATD) device. The audio feed for the single channel audio files of the experiment were captured using Audacity software downloaded for free from the Internet.[44] The dual-channel audio files were captured using Adobe Audition 3.0 software.

To test the formula for calculating uncompressed analog to digital audio conversion, the audio was recorded as uncompressed .WAV files from the same ninety seconds of a commercial music tape using the variables shown in figure 4.

This set of variables was chosen because they contain values recommended by both the Rutgers *Sound Object Archival Standards* and by many the standards bodies mentioned in this paper.[45] Any variables that do not match those standards were chosen to create atypical file combinations that would test the limits of the audio digital storage calculation formula. Sample audio files were recorded for ninety seconds and then copied and cut down those files into sixty- and thirty-second lengths for each sampling rate, bit depth and channel combination, creating a total of forty-two sound files. Lengths of recordings were chosen to provide a variety of sample sizes and because they were intervals that were easy to produce in the editing software suites used in the experiment.

Sampling rates were chosen based on the entire range of possibilities for digitizing sound recordings available in each of the two software suites used in the experiment. A sampling rate of 44.1kHz is the CD quality standard for digital audio recordings, and thus was set as the lowest sampling rate in the experiment. A higher sampling rate of 96kHz is recommended by various standards bodies. Bit depths of 16 and 32 were chosen to represent extremes. Half of the files were recorded with one channel and half of the files with two channels. This decision provided more than one value for the channels variable, but kept the number of created samples at a manageable level for the experiment.

Each file was assigned a unique AudioID which indicated the combination of variables used. For instance, one file was labeled "Au30-44-16-001." Au indicated it was an Audio file, "30" represented the number of seconds in length, "44" indicated the 44.1kHz sampling rate, "16" indicated the bit depth, and "001" indicated the number of channels.

All data about the different audio recordings were entered into a Microsoft Excel spreadsheet. The calculation function in Excel's calculation function was used to calculate the anticipated file size utilizing the uncompressed audio file formula in megabytes (Ca_1).

For the audio files produced in the experiment, all of the *calculated* (Ca_x) and *measured file sizes* (Aa_x) fell into the megabyte size range. Discussion and examples of the results of the audio file formula tests will, therefore, be limited to those measured in megabytes. Already having determined that Media Info was the preferred *instrument* for the measurement of measured file sizes in the portion of the experiment dedicated to still images, only results comparing measured file sizes as reported by Media Info are presented.

The differences (Da_1, Da_2, Da_3, Da_4) between the calculated file sizes from the formula (Ca_x) and the measured

Table 3. Abbreviations for Audio File Variables in Calculations by Instrument and Unit of Measurement

Variable	No Instrument (KB)	No Instrument (MB)	Windows Explorer (KB)	Media Info (KB)	Windows Explorer (MB)	Media Info (MB)
Calculated File Size	Ca_1	Ca_2				
Actual File Size			Aa_1	Aa_2	Aa_3	Aa_4
Differences between File Sizes			Da_1	Da_2	Da_3	Da_4
Percent Difference between File Sizes			Pa_1	Pa_2	Pa_3	Pa_4

file sizes recorded from Windows Explorer and Media Info in both kilobytes (Aa_1, Aa_3) and megabytes (Aa_2, Aa_4) were computed using Excel's calculating function.

$$Da_x = Aa_x\text{-}Ca_x$$

To provide consistency in method and presentation of the data, the Percent Difference (Pa_x) between the calculated file size and the measured file size of audio files in the experiment was calculated to demonstrate the relative accuracy of the formula.

$$Pa_x = |Da_x \div Aa_x| \times 100$$

As in the experiment with the still image digital storage formula, the results of the experiment for the audio digital storage formula will be discussed in terms of the *absolute values* of Da_x and Pa_x.

Audio Experiment Results and Discussion

Table 4 contains results comparing the *calculated file size* (Ca_2) from the formula for calculating digital storage needs from an inventory of analog audio materials, the *measured file sizes* as reported in Media Info (Aa_4), the *difference* between the calculated file size and the measured file size (Da_4) and the *percentage of the measured file* size represented by the *difference* between the calculated file size and the *measured file size* (Pa_4). A sample of sixteen files were chosen that represent results with the least amount of accuracy (the highest value of Pa_4), results that demonstrate some error (median Pa_4 values) and files that represent the least amount of error or no error (small or no Pa_4 values). Negative numbers indicate that Ca_2 was larger than Aa_4.

Of the forty-two files tested, the calculated file sizes are always less than 1 percent different from the measured file sizes. The largest absolute Pa_4 is for the 90-second clip recorded at 192kHz with a bit depth of 16 and 2 channels; the difference represents an absolute Pa_4 value of 0.88 percent of the measured file size. The file with the smallest Pa_4 value was 90 seconds long, recorded with a sampling rate of 96kHz at 32 bits with two channels; its Pa_4 value was 0.03 percent of the measured file size. The absolute median Pa_4 value of measured file sizes in the data set is 0.12 percent. The absolute mean Pa_4 value of measured file sizes as reported by in the dataset is 0.23 percent. Unlike the still image digital storage formula, there are no instances of absolute Pa_4 values that are extremely high or extremely low when compared with the absolute Pa_4 values of other files in the experiment. There were no files for which Ca_2 exactly matched the measured value.

The data show that the formula for calculating file sizes for uncompressed .WAV files from analog audio sources is extremely reliable. Examining the files with the ten highest absolute Pa_4 values and the ten lowest absolute Pa_4 values indicates that the formula is most accurate when using shorter recordings with a sampling rate in the 48kHz or 96kHz range at a lower bit depth. For the purposes of planning digital storage, the errors in the data are so small—less than 1 percent in all cases—that they are of no real concern. The trend toward slightly less accuracy with larger file sizes will only be proven or disproven with a much larger sample set.

Applications of the Still Image Formula in the Jazz Oral History Project

The high accuracy of the formulae in these experiments indicates that they can be used reliably when attempting to calculate storage needs for an audio digitization project. As the project manager for the JOHP, and as a regular consultant on digitization projects for teaching faculty at her institution, the author has found the still image and audio formulae invaluable in calculating storage needs and evaluating both digitization standards and equipment for projects.[46]

When working with a collection of analog documents, images and audio recordings, the physical dimensions or duration of objects in a collection are predetermined. The other elements of the audio and still image formulae, bit depth, resolution/sampling, color type (for still images and documents), rate and number of channels (for audio files), are variable depending on the standards used. Project managers can use the still image and audio formulae to determine how much storage space they will need to house the

Table 4. Calculated File Size Compared to Actual File Size of Digital Audio Files. Brackets | | indicate absolute values used in analysis.

AudioID	Length (sec)	Sampling Rate (KHz)	Channels	Bit Depth	Calculated File Size (MB) (Ca)	Measured File Size (MB) As_4	Difference (MB) Da_4	Percent Difference (MB) Pa_4
Au90-192-16-002	90.00	192.0	2	16	65.91	66.50	0.58	0.88
Au60-44-16-001	60.00	44.1	1	16	5.04	5.09	0.04	0.85
Au90-192-16-001	91.00	192.0	1	16	33.32	33.60	0.27	0.82
Au90-96-16-001	89.00	96.0	1	16	16.29	16.40	0.10	0.63
Au60-192-16-001	60.00	192.0	1	16	21.97	22.10	0.13	0.58
Au90-96-16-002	91.00	96.0	2	16	33.32	33.50	0.17	0.52
Au90-44-16-001	90.00	44.1	1	16	7.57	7.61	0.04	0.52
Au60-44-32-001	59.94	44.1	1	32	10.08	10.10	0.02	0.16
Au60-96-16-001	60.00	96.0	1	16	10.98	11.00	0.01	0.12
Au60-96-16-002	60.00	96.0	2	16	21.97	22.00	0.03	0.12
Au90-48-32-001	90.00	48.0	1	32	16.47	16.50	0.02	0.12
Au30-48-16-002	30.28	48.0	2	16	5.54	5.54	0.00	|-0.07|
Au30-44-32-001	30.00	44.1	1	32	5.04	5.05	0.00	0.06
Au30-44-32-002	30.00	44.1	2	32	10.09	10.10	0.01	0.06
Au60-44-16-002	60.00	44.1	2	16	10.09	10.10	0.01	0.06
Au90-96-32-002	90.00	96.0	2	32	65.91	65.90	0.01	0.03

archival quality digital surrogates in a collection. Once the amount of digital storage has been calculated, if the budget of the project disallows purchasing enough storage to house the entire collection, digital project managers can make strategic decisions about which objects would benefit the most from digitization on the basis of the original document and recording conditions, user interest, and institutional mission. They may also choose to utilize a different standard if adjusting bit depth, scanning resolution, or sampling rates would enable digitizing the entire collection. Digitizing the JOHP transcripts served as a case study which confirmed the usefulness of the still image formula.

Because the JOHP files will be housed in RUcore, the RUcore standard, "Digitizing Analog Documents and Images," was used as the guideline for the project.[47] This standard falls well within the spectrum of other digitization standards developed by bodies both national and regional. Many of the transcripts from the JOHP are on older typing paper and, in some cases, have a yellowed appearance. To capture the look and feel of the original documents to provide the user with an experience as close to handling the physical pages as possible, the RUcore standard for capturing color images was chosen. That standard requires a resolution of 600 ppi, the use of the RGB colorspace with 24-bit color, and outputting files in TIFF file format. The project began with digitization of one of the longest single transcripts, the interview with jazz great Maxine Sullivan, totaling 775 pages. This provided project staff with a robust sample for testing workflows and the digitization standard.

Plugging a height of 11.5 and a width of 9 for the page sizes (to accommodate the edges of the paper) into the still image formula results in a file size of 106.6 MB per page. It would require 80.16 GB of space to store 770 such pages, which is quite a lot of storage for a single document that is part of a larger collection. The JOHP collection contains 25,995 pages of transcripts. At the chosen bit depth and resolution for the project, the still image formula indicates a total size of 2.64 Terabytes (TB) for all JOHP transcripts. After fine-tuning the scanner settings, the scanned area for each page was adjusted to 8.82 by 11.10 inches. This results in a per-page file size of 100.84 MB and a total storage size of 2.50 TB for all 25,995 pages of transcripts.

The total storage capacity of RUcore is currently 55 TB, expandable to 15.5 Petabytes; 2.50 TB is approximately 4.5 percent of the total current storage on the RUcore servers. Each interview transcript is accompanied by approximately three to five hours of audio files, which add additional storage requirements. While the RUcore servers can handle such volume, it is always wise to try to conserve as much storage space as possible to save on maintenance, upgrade and labor costs for stewardship over the life of the data being stored.

After considering the size of the digitized Sullivan transcript, the JOHP staff determined that transcripts should be scanned at the 600 ppi required for color documents, color type and bit depth were changed to 8-bit grayscale

instead of 24-bit color in the interest of lowering the size of files for each page of transcripts.[48] The desire to scan them in 24-bit color was an aesthetic choice that needed to be altered to accommodate stewardship of the entire collection of both transcript files and audio files in the collection. Providing the "look and feel" of the original transcripts in digital format would have been pleasant for users, but was set aside in favor of storage economy. Recalculating the file sizes for the entire collection with adjusted scanning parameters revealed that the changes would result in a 66 percent reduction in the necessary storage capacity for the transcript files. This reduction took the total storage needed for all 25,995 pages down from 2.50 TB to 853.21 GB or 0.83 TB, a much more manageable storage requirement for the entire collection, which demonstrated the efficacy of the formula.

Conclusion

In July 2015, AVPreserve, a consulting firm that helps institutions manage and implement digital library projects, released, "Quantifying the Need: A Survey of Existing Sound Recordings in Collections in the United States."[49] In the report, Lyons, Chandler, and Lacinak estimate that there are 254,159,631 preservation-worthy audio holdings in US collections, and that the market cost of the digitization process for these items would be more than twenty billion dollars, "which does not include the costs that will be associated with . . . ongoing storage of digital files for preservation and access."[50] Assuming a very conservative estimate of 5 minutes per audio recording, with CD quality bit depth of 16 and sampling rate of 44.1 kHz in stereo (two channels), using the audio calculation formula, we know that purchasing storage at the current, consumer rate of approximately one dollar per gigabyte will require an extra $12,526,408.00 for the purchase of storage media alone. AVPreserve's survey does not indicate the average length of preservation-worthy audio recordings in its survey; the cost of storage media could be much, much higher. In an era of shrinking academic and cultural heritage budgets, purchasing digital storage to house and preserve these audio objects will be no mean feat.

The author has found that the still image and audio formulae are valuable tools for anticipating digital storage needs and for helping faculty outside the library evaluate their equipment for digitization projects. As the experiments demonstrate, the formulae for still image and audio recordings are extremely accurate. They will prove invaluable to digital archivists, digital librarians and the average user in helping to plan digitization projects, as well as in evaluating hardware and software for these projects. An understanding of the parameters of digitization contained in each formula—bit depth, color type, scanning resolution, sampling rate and audio channels—provides insight into both the quality of a digital image or sound file and provides guidelines for project managers to evaluate best practice standards and digitization equipment. Digital project managers armed with the still image and audio formulae will be able to calculate file sizes using different standards to determine which standard will suit the project needs. Knowing the parameters of the still image and audio formulae will allow managers to evaluate equipment on the basis of the flexibility of the software and hardware before purchase. Using the still image and audio calculation formulae in workflows will help digital project managers create more efficient project plans and tighter grant proposals.

Future work in the area of calculating digital storage needs to be done. Discovering or developing formulae for uncompressed, archival quality files produced by 3D image scanners and for digitizing analog moving images (video) would add significant value to the library, archival, and cultural heritage professions. Further research into predicting file sizes for born-digital objects, as well as calculating the file size savings when converting digital multimedia files from uncompressed to compressed formats would benefit the literature. Such formulae will enable even more accurate, additional projections to be made for a greater variety of projects.

References and Notes

1. Sam Brylawski et al., eds., *ARSC Guide to Audio Preservation* (Eugene, OR: Association for Recorded Sound Collections; Washington, DC: Council on Library and Information Resources, 2015), accessed May 29, 2015, www.clir.org/pubs/reports/pub164, 223.
2. Brylawski et al., 224.
3. "Jazz Oral History Project," Institute of Jazz Studies, Rutgers University Libraries, accessed April 25, 2015, http://newarkwww.rutgers.edu/IJS/OralHistory.html.
4. TechTerms.com, s.v. "Megabyte," accessed April 1, 2015, http://techterms.com/definition/megabyte; TechTerms.com, s.v. "Gigabyte," accessed April 1, 2015, http://techterms.com/definition/gigabyte; Michael Lesk, *Practical Digital Libraries: Books, Bytes, and Bucks* (San Francisco: Morgan Kaufmann, 1997), 81.
5. Mark Jordan, *Putting Content Online: A Practical Guide for Libraries* (Oxford: Chandos, 2006), 100, 107, 93.
6. Ron Bamberger and Sam Brylawski, *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (Washington, DC: Council on Library and Information Resources, 2010), 81–82, www.clir.org/pubs/reports/pub148.
7. Chris Lacinak, "What to Do After Digitization," in *ARSC Guide to Audio Preservation*, edited by Sam Brylawski et al. (Washington, DC: Council on Library and Information Resources, 2015), 127–52, www.clir.org/pubs/reports/pub164.

8. Gail M. Hodge, "Best Practices for Digital Archiving: An Information Lifecycle Approach," *D-Lib Magazine* 6, no. 1 (2000), www.dlib.org/dlib/january00/01hodge.html.
9. Lisa Hooper, "Building an E-Audio Reserves Program at Tulane University: Methods and Procedures," *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 22, no. 3-4 (2012): 181–96, http://dx.doi.org/10.1080/1072303X.2012.723671; Maureen Pastine, Ivy Bayard, and Carol Lang, "Digital Diamond: Temple University Libraries' IMLS Grant," *The Bottom Line* 14, no. 2 (2001): 76–84.
10. Ruth Ann Jones, "Empowerment for Digitization: Lessons Learned from The Making of Modern Michigan," *Library Hi Tech*, 23, no. 2 (2005): 205–19.
11. Pastine, Bayard, and Lang, "Digital Diamond"; Ram Nath Maurya, "Digital Library and Digitization," *International Journal of Information Dissemination & Technology* 1, no. 4, (October 2012): 228–31; Jones, "Empowerment for Digitization."
12. Peter J. Astle and Adrienne Muir, "Digitization and Preservation in Public Libraries and Archives," *Journal of Librarianship & Information Science* 34, no. 2 (2002): 67–79; Yan Quan Liu, "Best Practices, Standards and Techniques for Digitizing Library Materials: A Snapshot of Library Digitization Practices in the USA," *Online Information Review* 28, no. 5 (2011): 338–45.
13. Kathleen Arthur et al., "Recognizing Digitization as a Preservation Reformatting Method," *Microform & Imaging Review* 33, no. 4 (2004): 171–80.
14. Arthur et al., "Recognizing Digitization as a Preservation Reformatting Method"; Hodge, "Best Practices for Digital Archiving"; Hooper, "Building an E-Audio Reserves Program"; Lacinak, "What to do After Digitization."
15. Michael Lesk, *A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access* (Washington, DC: Commission on Preservation and Access, 1990), http://cool.conservation-us.org/byauth/lesk/lesk#digital.
16. Butch Lazorchak, "A Digital Asset Sustainability and Preservation Cost Bibliography," *The Signal: Digital Preservation* (blog), June 26, 2012, http://blogs.loc.gov/digitalpreservation/2012/06/a-digital-asset-sustainability-and-preservation-cost-bibliography.
17. Richard L. Moore et al., "Disk and Tape Storage Cost Models," San Diego Supercomputer Center, (University of California San Diego: La Jolla, 2007), http://users.sdsc.edu/~mcdonald/content/papers/dt_cost.pdf; Shelby Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs Across Cost Models and Cost Frameworks," *RLG Diginews* 7, no. 4 (2003), http://library.oclc.org/cdm/singleitem/collection/p267701coll33/id/366.
18. Lesk, *A Report of the Technology Assessment Advisory Committee*.
19. Ivan Smith, "Cost of Hard Drive Storage Space," Nova Scotia's Electric Gleaner, April 3, 2014, http://ns1758.ca/winch/winchest.html. Note: This source is quoted in other, online articles about digital storage, and, despite its unprofessional appearance, diligently collates and tracks its sources from a wide variety of literature including primary sources such as computer hardware catalogs and computing magazine articles.
20. Don Williams and Michael Stelmach, eds., *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* (Washington, DC: Federal Agencies Digitization Guidelines Initiative, revised August 2010), www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf.
21. Colorado Digitization Project Digital Imaging Best Practices Working Group, "BCR's CDP Digital Imaging Best Practices Version 2.0," (Aurora, CO: Bibliographical Center for Research, 2008), http://mwdl.org/docs/digital-imaging-bp_2.0.pdf.
22. "Digitization Standards for Still Images," Smithsonian Institution Archives website, accessed June 5, 2015, http://siarchives.si.edu/services/digitization; World Digital Library, "WDL Digital Image Standards" (Washington, DC: World Digital Library), accessed June 5, 2015, http://project.wdl.org/standards/imagestandards.html.
23. Colorado Digitization Project Digital Audio Working Group, "Digital Audio Best Practices Version 2.1," March 25, 2015, http://sustainableheritagenetwork.org/content/digital-audio-best-practices-version-21; Ian Bogus et al., "Minimum Digitization Capture Recommendations" (working paper, Preservation and Reformatting Section, Association for Library Collections and Technical Services, American Library Association, 2013), www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations.
24. Richard Wright, *Preserving Moving Pictures and Sound* (Helsington, York, UK: Digital Preservation Coalition, 2012): 17; CARLI Digital Collections Users' Group, Standards Subcommittee, "Guidelines for the Creation of Digital Collections: Digitization Best Practices for Moving Images" *Digital Collections: Best Practices for Digital Collections* (Champaign: Consortium of Academic and Research Libraries in Illinois, 2014), 4, www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_video.pdf.
25. Isaiah Beard et al., *Sound Objects: Recommended Minimum Requirements for Preservation Sampling of Audio* (New Brunswick, NJ: Rutgers University Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-Audio.pdf; Isaiah Beard, *Digitizing Analog Documents and Images: Recommended Minimum Standards for Archival and Presentation Datastreams* (New Brunswick, NJ: Rutgers University Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-ScannedImagesDigitalSurrogates.pdf; Isaiah Beard et al., *Video and Moving Image Objects: Recommended Minimum Standards for Archival and Presentation Datastreams* (New Brunswick, NJ: Rutgers University

Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-Audio.pdf.

26. Alan P. Kefauver and David Patschke, *Fundamentals of Digital Audio* (Middleton, WI: A-R Editions, 2007); Arch C. Luther, *Using Digital Video* (Orlando, FL: Academic Press, 1995); Francis Rumsey, "Appendix I: From Analog to Digital and Back," in *Digital Audio Operations* (Boston: Focal Press, 1991), 215–29.
27. Taz Tally, *Avoiding the Scanning Blues: A Desktop Scanning Primer* (Upper Saddle River, NJ: Prentice Hall, 2001), 67; Margot Note, *Managing Image Collections: A Practical Guide* (Oxford, UK: Chandos, 2011), 54; Forrester High School Computing, "Bitmapped Graphic Calculations," accessed April 2012, http://forrestercomputing.wikispaces.com/Bitmapped+Graphic+Calculations.
28. Techopedia, s.v. "Resolution," accessed April 1, 2015, www.techopedia.com/definition/2743/resolution.
29. Forrester High School Computing, "Calculations (sound)," accessed April 30, 2012, http://forrestercomputing.wikispaces.com/Calculations+%28Sound%29; Nels Johnson, Fred Gault, and Mark Florence, *How to Digitize Video* (New York: Wiley, 1994), 28, 31. Note: Despite being a resource for digitizing video, this book contained a formula for calculating digital file sizes for audio recordings as part of the calculations for video recordings.
30. Cambridge Dictionaries Online, s.v. "Track," accessed April 1, 2015, http://dictionary.cambridge.org/dictionary/british/track; Brylawski et al., *ARSC Guide to Audio Preservation*, 229.
31. John Rice and Brian McKernan, eds., *Creating Digital Content: Video Production for Web, Broadcast, and Cinema* (New York: McGraw-Hill, 2002), 18.
32. Johnson, Gault, and Florence, *How to Digitize Video*, 28; Forrester High School Computing, "Calculations (Video)," accessed April 30, 2012, http://forrestercomputing.wikispaces.com/Calculations+%28Video%29.
33. Forrester High School Computing, "Calculations (Video)"; Johnson, Gault, and Florence, *How to Digitize Video*, 28.
34. Johnson, Gault and Florence, *How to Digitize Video*, 28–29.
35. Forrester High School Computing, "Calculations (Sound)"; Forrester High School Computing, "Calculations (Video)."
36. Eddie Woo, *Calculating File Size: Audio*," MPEG4 YouTube video, 10:32, posted June 4, 2013, www.youtube.com/watch?v=0ctX3z7yzuU; Eddie Woo, *Calculating File Size: Images*, MPEG4 YouTube video, 7:06, posted June 4, 2013, https://www.youtube.com/watch?v=6jIhaAkzOvo; Eddie Woo, *Calculating File Size: Video*, MPEG4 YouTube video, 4:21, posted June 4, 2013, https://www.youtube.com/watch?v=lju4RzSIkqA ; Kenneth R. Koehler, "Storage Requirements," in *Elementary Computer Mathematics*, 2002, http://kias.dyndns.org/comath/44.html.
37. TechTerms.com, s.v. "Byte," accessed April 1, 2015, http://techterms.com/definition/byte; The still image formula has been presented previously at AUTHOR, "Digital Imaging Specification and the Management of Digital Storage Needs" (presentation, Image Resources Interest Group, ALA Annual Conference, Las Vegas, NV, June 28, 2014), http://dx.doi.org/doi:10.7282/T3MG7MXQ; and at Krista White, "Calculating All That Jazz: Linking Technical Specifications to the Management of Digitization Projects" (poster presentation, Research Data Symposium, Columbia University, New York, New York, February 27, 2013), http://hdl.handle.net/10022/AC:P:19180. The formula, as presented was adapted from Tally, *Avoiding the Scanning Blues*; Note, *Managing Image Collections*; and Cunningham, "Bitmapped Graphic Calculations."
38. Colorado Digitization Project Digital Imaging Best Practices Working Group, "BCR's CDP Digital Imaging Best Practices"; Ian Bogus et al., "Minimum Digitization Capture Recommendations"; Isaiah Beard, *Digitizing Analog Documents and Images*.
39. TechTerms.com, s.v. "Bit," accessed April 20, 2015, http://techterms.com/definition/bit; Brylawski et al., 224.
40. "Media Info," MediaArea, 2015, accessed August 3, 2012, https://mediaarea.net/en/MediaInfo.
41. TechTerms.com, s.v. "Kilobyte," accessed April 1, 2015, http://techterms.com/definition/kilobyte.
42. All data and the full description of the experimental design and results are available at URL TBA.
43. Oxford English Dictionary, s.v. "absolute value," accessed October 14, 2015, www.oed.com/view/Entry/679?redirectedFrom=absolute+value#eid4714829.
44. "Download," Audacity, accessed June 27, 2012, http://audacityteam.org/download.
45. Beard et al., *Sound Objects*; Colorado Digitization Project Digital Audio Working Group, "Digital Audio Best Practices"; Bogus et al.,"Minimum Digitization Capture Recommendations."
46. Only the still image formula is used as an example here; the digitization of the audio files was completed in 2005, some seven years before the author began digitizing the JOHP transcripts.
47. Beard, *Digitizing Analog Documents and Images*.
48. This compromise for the standard was made at the suggestion of Isaiah Beard, the Digital Data Curator, who has been an integral part of the JOHP work.
49. Bertram Lyons, Rebecca Chandler and Chris Lacinak, "Quantifying the Need: A Survey of Existing Sound Recordings in Collections in the United States" (New York: AVPreserve, 2015), https://www.avpreserve.com/papers-and-presentations/quantifying-the-need-a-survey-of-existing-sound-recordings-in-collections-in-the-united-states/.
50. Ibid., 19–20.

Appendix. Definition of Terms

Absolute value: "the value of a real number disregarding its sign" where a sign indicates negative or positive value.[1]

Analog/Born analog: a device or system that represents media as continuously variable physical quantities. Analog media cannot be displayed on a computer or uploaded as files without transferring them into digital format.[2]

Audio: sound recordings of any variety.

Bits: "A bit (short for "binary digit") is the smallest unit of measurement used to quantify computer data. It contains a single binary value of 0 or 1."[3]

Bit depth: a unit that measures the amount of information recorded for each pixel in a still image or each sample in an audio file. Bit depth indicates the amount of information about the color of a pixel in an image or the sound level of the wave in a sound file.

Born digital: any recording (or file) that was digitally encoded at the point of creation.[4]

Bytes: "A byte is a unit of measurement used to measure data. One byte contains eight binary bits, or a sequence of eight zeros and ones."[5]

Channels/Tracks: "a part of a magnetic strip onto which sound can be recorded, with several tracks on one magnetic strip."[6] In digital sound, tracks are referred to as "channels." In analog and digital recordings, multiple tracks or channels are usually "mixed down" to create mono (one channel/track) or stereo (two channels/tracks) in the final version of a recording.

Gigabytes (GB): a unit of measurement for data equal to 1024 megabytes.[7]

Kilobytes (KB): a unit of measurement for data containing 1,024 bytes.[8]

Megabytes (MB): a unit of measurement for data equal to 1024 kilobytes and containing 1024^2 or 1,048,576 bytes.[9]

Raster graphics: "Computer graphics employing pixels as the display elements, storing data regarding the component pixels for a given image."[10]

Resolution: "In the computer and media industry, resolution refers mostly to display resolution and the number of picture elements (pixels or simply dots) that can be displayed both horizontally and vertically by a screen. Resolution in this case will then refer to how many pixels the display can produce horizontally (width) and vertically (height). This measure also applies to digital images."[11]

Sampling rate: "how many times per second a continuous (analog) signal is sampled during the digitization process."[12]

Still image/Document: objects such as photographs, letters or manuscripts.

References

1. *Oxford English Dictionary*, s.v. "absolute value," accessed October 14, 2015, www.oed.com/view/Entry/679?redirectedFrom=absolute+value#eid4714829.
2. Sam Brylawski et al., eds., *ARSC Guide to Audio Preservation* (Eugene, OR: Association for Recorded Sound Collections; Washington, DC: Council on Library and Information Resources, 2015), accessed May 29, 2015, www.clir.org/pubs/reports/pub164, 223.
3. TechTerms.com, s.v. "Bit," accessed April 20, 2015, http://techterms.com/definition/bit.
4. Brylawski et al., *ARSC Guide to Audio Preservation*, 224.
5. TechTerms.com, s.v. "Byte," accessed April 1, 2015, http://techterms.com/definition/byte.
6. Cambridge Dictionaries Online, s.v. "Track," accessed April 1, 2015, http://dictionary.cambridge.org/dictionary/british/track.
7. TechTerms.com, s.v. "Gigabyte," accessed April 1, 2015, http://techterms.com/definition/gigabyte.
8. TechTerms.com, s.v. "Kilobyte," http://techterms.com/definition/kilobyte.
9. TechTerms.com, s.v. "Megabyte," accessed April 1, 2015, http://techterms.com/definition/megabyte.
10. Art and Architecture Thesaurus Online, s.v. "Raster Graphics," accessed June 16, 2015, www.getty.edu/vow/AATFullDisplay?find=raster&logic=AND&note=&english=N&prev_page=1&subjectid=300215280.
11. Techopedia, s.v. "Resolution," accessed April 1, 2015, www.techopedia.com/definition/2743/resolution.
12. Brylawski et. al., *ARSC Guide to Audio Preservation*, 229.