A Response to the Responses of John Unsworth and John Rink on "The EVIA Digital Archive Project: Challenges and Solutions"*

Alan Burdette

External Image

Please see:
http://rup.rice.edu/image/cnx_images/shapes-buybutton.jpg

My article on the challenges faced by the EVIA Digital Archive Project, the solutions we have chosen, and key questions with which we are wrestling generated a lively dialog with my respondents. Their responses raise fair questions about our project, but they also warrant a formal response for the sake of clarity. I will address in sequence the issues and questions they raise in an attempt to clear up any confusion and to move the dialog further towards useful solutions.

John Unsworth begins by wondering if in fact three different persons wrote my article. I can confirm that I was its sole author, but his question about different voices in the paper perhaps points to the multi-dimensional nature of the project. It has been our vision to bring more closely together the worlds of field research, archiving, and scholarly publishing. The EVIA Project has never been the pet research of one particular scholar and it has been by necessity highly collaborative in its development. This may account for what John sees as an internal conflict among priorities. While preservation was the primary impetus of the project, what emerged in our year-long planning process was an endeavor to integrate preservation, documentation strategies, research tools, cataloging, scholarly publishing and online access into a systematic whole. This multidimensional approach was not the result of mission creep but was established at the outset and was part of the original development-phase funding.[1] "Mission-creep," as John identifies at the core of his response, is inaccurate both from the standpoint of the ambitious goals set at the beginning of the project, but is also not fair from the perspective of the digital landscape in which this project operates. When the EVIA Project began as a series of discussions in 2001, YouTube was still four years in the future, video streaming was difficult and rare, no best practices existed for *audio* preservation let alone video, and a terabyte sounded like an awful lot of storage space. The EVIA Project has had to push a lot of software, hardware, and infrastructure boundaries during its development and it has needed room to explore solutions, step back and try different routes. The EVIA Project has been less about a bounded one-off digital edition

*Version 1.2: May 14, 2010 4:51 am -0500

†http://creativecommons.org/licenses/by/3.0/

[1] We have had our share of feature expansion in our software development, but the mission of the project has not changed significantly.

and more about systems, workflows, and scholarly practices in an area with few precedents and a need for exploration.

What emerges in Unsworth's response is an argument for putting preservation first. Given the crisis we face in the world of media preservation, I strongly agree. The emerging consensus of media archivists is that we have less than twenty years to digitally transfer and preserve all of the audio and video media that we want to survive into the future. Despite the efforts of such notable institutions as the Library of Congress, Stanford, Indiana University, the National Library of Australia, and others, the infrastructure for saving our media heritage is not nearly large enough to meet the challenges we face. That said, in the world of archiving, preservation and access are inextricable, both from the perspective of the archival mission and practices and from the standpoint of funding. Audiovisual preservation requires different technologies and different skill sets from access development, but they are of a piece. At the Archives of Traditional Music, we do not even accept a collection for deposit if we cannot provide at least some minimal access to it. When we make the difficult decisions about what recordings to preserve and what to hold off for yet another round of funding, questions of its accessibility are figured into our analysis. It also does little good to preserve an item if no one can find it, so the cataloging and collection of metadata are critical dimensions of the digital preservation process. Most funders recognize the interrelated nature of the preservation and access matrix. The sources for audiovisual preservation are woefully small, but the primary one for ethnographic media archives such as ours is NEH, and an applicant can't simply request to preserve a collection without providing access of some sort.

I would like to stress that the EVIA Project is concerned with preservation, but the project was never envisioned as a preservation-only project. Mellon did not set out to fund a preservation-only project and we never saw the project that way. I have foregrounded preservation in my paper because all too often it gets lost or presumed as part of digital arts and humanities projects. For all of the talk about cyberinfrastructure and digital tools for research, unless we work to preserve our audiovisual resources during our generation, they will likely be lost to those of the future.

Unsworth takes the EVIA Project to task for what he perceives as the slow pace of preservation as it relates to the amount of video and the amount of funding. His criticism is well placed but is misleading. While at the time of this writing, we have only made seventy hours of video recordings available online, we have indeed made digital preservation transfers of five hundred hours of video signal under the Mellon funding for the EVIA Project. Of those that remain unpublished, one hundred eighty hours are extremely close to publication and the rest are on their way to online publication and availability. An additional eight hundred hours are being ingested as part of a two-year NEH project and we are on target to complete the transfers and annotation of that material in May 2011. It is true that the scholarly annotation process slows down our ability to make these materials available. The nature of the materials we are dealing with makes them different from many other projects of digital preservation and online humanities scholarship. In distinction from many of the resources discussed at this conference, they are not part of a canonical body with accompanying lines of scholarly discussion and critique. Indeed, within the ethnographic mode, critical analysis of the material is largely absent from the materials in the EVIA Digital Archive Project. Instead, the focus is on qualitative description and cultural analysis based around salient cultural performances. Unsworth argues for "lowering the bar" so that preservation may be unhindered, but the descriptions and cultural analysis that form one of the bottlenecks to accessibility is extremely important. The nature of the ethnographic material argues, in my view, for maximizing the information that the documenting scholar brings to the process. First, a high number of ethnographic scholars work in cultural domains that are poorly documented otherwise, and they may be only one of a handful of scholars who have worked in that place and time. The value of the recordings increases dramatically with the depth and quality of the accompanying documentation. The EVIA Project thus far has worked with living scholars and primarily with collections that had not been institutionally accessioned. If these recordings are digitally preserved only, with little expert description, they are in danger of being curiosities of the future—interesting, but largely inscrutable moving images. An argument we make with the EVIA Project is that we should maximize the value of these cultural documents by giving scholars the tools to share their knowledge and to better collaborate with their research subjects.

That said, we share Unsworth's concern about the scholarly bottlenecks and we have taken several steps to address them. First, we have taken important lessons from each of the summer institutes we have held and have improved our ability to help scholars complete their work. Looking back at our first summer institute, held in 2004, less than a year after programmers had begun work, the annotation software barely functioned, we didn't have any annotation models for scholars to follow, and we devoted a lot of time to discussing what we wanted the end result to look like. Many of the collections accepted for that first summer institute do not yet have complete annotations and remain unpublished. By the time of our last summer institute in 2009—nine months ago—we have dramatically increased the amount of work that gets accomplished and half of the participants have already submitted their work for peer review. Second, we have been reevaluating the path we have chosen and are considering a model in which annotation is more basic and more focused on description, allowing access to move more quickly. Deeper description and cultural analysis would form another layer, manifested in work in various kinds of online journals. In this scenario, the work of more narrative writing falls into disciplinary domains and peer review apparatus outside of the EVIA Project.

Unsworth's description of an alternative preservation and access process is useful for the sake of argument, but is simplistic in its outline and ignores key aspects of the kind of content we address. Ironically, the initial process he describes is not so different from what we do. We send out a call for depositors and select those that are appropriate. Those applications gather very general collection-level information. Unsworth does not indicate a selection process, but I think he would agree that the costs of digital preservation warrant some selectivity. Once we select collections for accession, we collect information necessary for preservation production—basic information about each recording that generates numbers, labels, and information we use for tracking the item. As for digital preservation transfers, Unsworth describes a situation of the future that does not presently exist. While depositors in our case are using largely consumer-level formats such as VHS, Hi-8, and MiniDV, we cannot rely on standard consumer transfer services for digital preservation. It remains a challenge to find vendors who understand digital preservation rather than simply copying tapes to DVD for the family media album. Our present digital preservation format for analog sources is uncompressed video which creates a 125 GB file for every hour of video. While no internationally accepted standards and best practices for video preservation exist, we must follow the models from audio preservation and the best intentions of video preservation work happening at other institutions if we want to make a sincere effort in this direction. Thus, we must have technical metadata collected about the transfer process and run quality control processes on the completed file. Our project relied on the University of Michigan for most of our transfers and commercial vendors in more difficult cases. The costs are significant in either case and so selection is necessary. In addition, as these formats age, the transfer process becomes more challenging. Even MiniDV has proven to be a vexing format from which to get good, error-free transfers.

Unsworth's alternative process and our own diverge more dramatically once we have a digital preservation master. From that master we create a lower-quality copy that is much easier to handle than the gigantic preservation master, transcoding the file from 250 Mbps down to 1 or 2 Mbps. A web-based annotation tool is appealing and indeed, we are currently working on such a tool, but at present there are significant technological barriers to such a technology working very well. In our case, we are not dealing with isolated clips of video like those found on YouTube, but rather entire tape transfers that maintain the archival integrity of the original object. Such an object in most cases demands some sort of content segmentation because it is rarely singular in subject matter. Good segmentation requires frame-accurate tools, and this is fairly difficult to manage with video streamed online, which is typically compressed to the point where many of the frames are missing and must be interpolated. This technology will get sorted out, but it is because of these current limitations that we opted for a desktop-based annotation tool when software development began in 2003.

On the issue of crowd-sourcing, we believe that there is a place for unsolicited response, commentary, and analysis from many different quarters. However, I remind Unsworth and others that ethnographic scholars ethically accept a stewardship role of the recordings they make. Unsworth's critique of academics slipping too easily into elitism is a fair one, but if we look at the models around us—YouTube, or even the Flickr Commons photo collections put up by the Library of Congress and other institutions, there is plenty of evidence for behavior that is reprehensible, wrong, and idiotic. This behavior is annoying within a collection

of eighty-year–old photographs, but in a collection of relatively recent ethnographic video, it has the potential to do great damage to personal and professional relationships. For the depositors who function as scholarly mediators, they have a responsibility to moderate and shield individuals who are the subjects of the video from online abuse. Now, it is true that this does sound a bit elitist and perhaps paternalistic, but I feel fairly certain that all of the scholars I have worked with on this project would be upset to see their research collaborators abused or ridiculed in some online venue. On the other hand, the Library of Congress has seen immense benefit from crowd-sourcing metadata for their photos within the Flickr Commons, and the photographs have gotten public use and access that has been educational and creative and well beyond the use they had seen in the past. We believe that the same potential exists for the ethnographic video we are publishing. We believe that the insights and corrections of other scholars and local experts will add greatly to the research and educational value of the material. We welcome this and are working on tools to do so. So we are not opposed to "crowd-sourced" information, but the esoteric nature of much of the content and the sometimes sensitive ethical relationships embedded in the production of the recordings mean that this crowd-sourcing ought to moderated. We also believe that the recordings are the result of a particular subjectivity—that of the scholar and his or her research goals, theoretical orientation, and personal history. It is important that we gather the scholar's reflections on the recordings they have made so that future viewers will better understand the manner in which they were produced.

Both Unsworth and Rink express concern about the level of detail in the annotations. Because we share that concern, we are actively developing alternative models that will not rely on the detailed and multi-layered annotations that have characterized the project thus far. For example, the AHEYM Project will utilize a more basic form of annotation when it incorporates its materials into the EVIA Project. We are also exploring more basic modes for scholars who want to make their collections available quickly or for collections where the collecting scholar is not available to make the annotations. The metadata structure created by the project is certainly flexible enough to accommodate both less and more detailed approaches to annotation than we have seen thus far.

Unsworth notes that "preservation will always turn on sustainability," but I think this perspective is only applicable in a relatively small number of high-profile objects or collections. I believe it is more accurate to say that preservation has usually turned on the passions of a small group of people. Historically, preservation has rarely been sustainable. Media archiving is almost always an afterthought, archiving of analog media and documents is patchy, at best, and it has been driven by collectors, enthusiasts and researchers with particular avocations or ideologies. Preservation has always turned on institutional good will and patronage in the past and I don't see that changing. Indeed, the digital revolution is exposing the true costs of archiving as we must move from passive to active strategies of archiving. Most archiving is dependent on a leap of faith that materials of little interest today will one day be of great value to a small group of people. Government, institutional, foundation, and individual patronage are the ways in which we have preserved things in the past and they are likely to be the primary engines behind preservation in the near future at least.

Despite my rebuttals of some of Unsworth's key points, I agree that an all-out focus on preservation is warranted. We have perhaps less than twenty years to digitally preserve our legacy video recordings before they disintegrate or become functionally obsolete. Most audiovisual archives remain ill-equipped to handle the looming onslaught of born-digital recordings scholars are creating in the present. The scale of this challenge across the United States requires massive investment from private and public sources. I too am concerned that we are not moving fast enough. One of the challenges we are working to address is how to successfully save our media while at the same time maximizing its value as an academic product.

As I said in my original paper, peer review creates a problem of fixity, which John Unsworth rightly criticized for having a limited view of peer review. John Rink asks if we need peer review at all. Peer review is one convention for adding authority and value to an academic product. In my paper I have perhaps overstated our challenges in the area of peer review. Clearly, scholars write articles, get them peer reviewed, and move forward in their research. They simply write new books and articles as their thinking develops. This ought to be true for video annotations. I described the problems of peer review as one of fixing the content as a static object. The problem of fixity is not just a peer review or publishing problem—it is also an archival problem. My archivist and my cataloger at the Archives of Traditional Music do not want to

see a text in process. They don't want to be continually updating finding aids and catalog records. They only want to do this once, and as director, I only want to pay to do it once. The nature of cataloging and indexing will surely change, but at present, accessioning the materials in our archive and library system is one of the ways we sustain the content and so we must abide by the wisdom of library conventions for locating materials. The risk that young scholars take to invest in digital products is real. This is why peer review is important. Regardless of the possibilities of "crowd-sourcing" and its potential value, I think it is accurate to say that it is not accepted by the professional mechanics of the academy.

As for whether academic societies should shoulder the burden of peer review, I believe that ultimately they should. Peer review is one of the primary activities of academic societies and it is their service to their discipline and their colleagues. I also believe that different kinds of communities can provide various kinds of peer review, but ultimately they are dependent upon structures of authority and dialog. As long as those are in place, then peer review might come from a variety of places. I do not mean to imply that I think peer review is necessary for everything. The EVIA Project will eventually be putting up material that has not been peer reviewed. This material will be designated differently and the user will simply have to rely on his or her own judgment.

John Rink in his response raises related concerns but focuses more directly on funding and sustainability. To address John's questions about funding very directly, we believe that it is possible to sustain the project without the level of funding we had during the extended development of the project. Sustainability includes many areas, not just funding, and I will address each of them in sequence.

Preservation of media objects is a basic activity of the project and we are achieving preservation by adhering to best practices in video preservation as well as we can, given the lack of international standards. The objects themselves are stored in a Fedora repository with preservation masters in Indiana University's Mass Data Storage System. The files are maintained as part of the Archives of Traditional Music and the Digital Library Program. All EVIA Project files are incorporated into the Indiana University library infrastructure for long-term support of digital objects.

We are well aware that our current interface for accessing the video content will likely change in the future and there is a risk that in a worst-case scenario we will not have the funds or the partnerships to keep up with changing web technologies. However, we are doing everything we can to insure that the video segments created by the project remain as accessible as possible for as long as possible. We achieve this by our relationship with the library and a commitment from the Digital Library Project to maintain access. We create Persistent URLs for each video segment so that the access video copies can continue to be located despite changes in servers or even in hosting institutions.

Software development has been the most expensive aspect of the project and like many projects, the basics of keeping the technology running, let alone new feature development, requires regular investments. We are addressing the software development needs through additional grant funding that typically combines content expansion with software development to address basic maintenance and the addition of new features. In the long term, we hope to see video and video tools as part of the Indiana University technology infrastructure, and we are working to realize a larger university effort in this regard.

The EVIA Project has been actively engaged in pushing forward strategies that will address the needs of audiovisual collections on the entire IU Bloomington campus and thus leverage the possibilities of a shared resource pool for preservation and access services. We are making important first steps towards these goals.

Funding for the project is currently reliant on two kinds of sources. The first is Indiana University's Institute for Digital Arts and Humanities, which provides a platform from which EVIA Project personnel can pursue new funding and collaborate with existing projects. The second is grant funding from external sources or internal funding for new initiatives and infrastructure development.

Both Rink and Unsworth raise questions about the high cost of the project relative to the amount of video preserved. The EVIA Project began from ground zero. Preservation has been one part of a larger package that has included software development and scholarly summer institutes. Now that we have built a basic framework and system, we can and have taken on additional projects to supplement both the video delivered and in some cases extend the software capabilities. For example, the AHEYM Project will preserve and annotate eight hundred hours of video—nearly twice the amount preserved over the entire Mellon-funded

cycles—for less than what it once cost to run the EVIA Project for one year. This is possible because we have now built the framework.

It is fair to ask if the EVIA Project is financially stable, but one could ask the same question of the state of California, which is facing a twenty-billion–dollar deficit next year. After the recent economic downturn, it is easy to question if anything is financially stable. My point is that financial stability is relative and not always reliable. To address Rink's question more directly: no, the project is not financially stable but it *is* financially engaged. We have several active grants at present that are funding EVIA-related work, and we have several more in the pipeline. We are situated within Indiana's Institute for Digital Arts and Humanities, which at least gives us a platform and a network from which to pursue further funding. We are actively and successfully working to move key preservation activities into the core functions of special collections support at Indiana University.

The EVIA Project should be considered in terms of how it has built a certain level of cyberinfrastructure. Unsworth stated in an interview with Kevin Guthrie, "Beyond the ACLS Report":

> *I like to think of cyberinfrastructure as the middle layer of a cake. The base layer is all of the hardware and basic operating systems-level technology on the network. Fiber optic cables, storage devices, things like that. The icing is made up of specific applications to serve a particular purpose. Software applications and tools that can be shared by different people for different purposes represent the middle layer of the cake and are what we mean by cyberinfrastructure. It is important to point out that cyberinfrastructure is not just equipment or software, it also includes the human interactions, protocols, standards, work processes, and so on, needed to make the system work and to structure collaborative or related activities... All of these elements are part of the "infrastructure. (Guthrie 2007)*

In many respects the EVIA project was always about cyberinfrastructure. Mellon did not intend the project to be solely about preservation. Mellon nudged us toward the summer institutes and peer review. Through the EVIA project we established coalitions on the Indiana University campus that cut across library, archive, technology, and faculty units. We developed a set of tools that fit the purposes of the project but always kept larger uses in mind. We created a network of scholars who grappled with a new way of non-linear writing and developed some protocols for doing so. We laid the foundation for a set of problems and challenges in video preservation. We tested and retested software with scholars, improving its fit with scholarly needs and desires with each iteration and development cycle. We pushed at our own university's technology infrastructure so that mass storage, video transcoding, and database services would be more responsive to the needs of the arts and humanities. We used the networks and the experience we gained on the EVIA Project to create Indiana's Institute for Digital Arts and Humanities (IDAH). IDAH, in collaboration with other units, has been the home from which we have pursued further-reaching coalitions and external funding that simultaneously expands the content of the EVIA Digital Archive and builds upon the framework created by the EVIA project to improve and expand that very framework. For example, the Ethnomusicology Multimedia Project brings together three different university presses to create a marketing platform that utilizes online audio and video samples for "first books" in ethnomusicology. This project will not only create a way to deliver audio and video samples in conjunction with scholarly books but will also create a web-based annotation tool. We were also a key part of the effort last year to build a coalition of special collections across campus and conduct a detailed survey of all audio, video, and film holdings. We are now set to embark on a preservation planning initiative that we believe will result in a centralized facility for media preservation and digital access at Indiana University. Such an endeavor will create more efficient processes for preservation and access and will dramatically improve our ability to digitally preserve our holdings and those yet to be deposited.

Feedback from scholars participating in the project has been generally positive. Most did not anticipate how much work was involved in video annotation, but at the same time, each successive "class" of fellows has pushed the bar higher in terms of the amount of annotation they provide. Most have found the summer institute to be immensely helpful and enjoyable. They deeply appreciate having the opportunity to share their work in an intense environment with an interdisciplinary group of scholars from a diverse representation

of career stages. As for how the content is getting used online, it is too early to report. Users are exploring the site and some instructors are beginning to use it in their courses. We will have to wait longer to have more useful data.

In closing, I value the discussion instigated by John Unsworth's and John Rink's responses to my paper and appreciate the assessment they have made. My response has been written for the sake of clarity on issues that I do not think were well understood, and to further the larger discussion on challenges that remain to be solved. I hope that my responses will not be taken as defensive but rather in the spirit of moving us towards a greater understanding of the issues faced by media preservation and by collections of ethnographic material. The responses by Unsworth and Rink have provided a welcome outsider perspective and we will certainly take them under consideration as we continue to develop the EVIA Project and its sustainability.

Bibliography

Guthrie, Kevin. *Academic Commons*. "Beyond the ACLS Report: An interview with John Unsworth." December 16, 2007. March 24, 2010 http://www.academiccommons.org/commons/interview/acls-report-interview-john-unsworth.