Lessons Learned in Archiving: A DIY Perspective

Caleb Sayan; Blue Rider Design, Visual Archiving Solutions; Portland, OR (USA)

Abstract

In May of 2009 I embarked on an ambitious project to digitize and catalog the Andrea Aranow Textile Documents collection. Comprised of 26,000 pieces of antique textiles and original artwork drawn from sources worldwide, ranging from the 19th to late 20th century. It was used commercially by weavers and surface designers as source of inspiration for both fashion and interior products.

The project's goal was to find a permanent home for the collection at an institutional through acquisition in order that it continue to be an important design resource far into the future. By having a collection that was digitized, it was theorized that the collection would be more appealing as a potential acquisition.

Prior to starting the project I had a good familiarity with the source material having worked with it for over five years, maintaining it and working with the designers who utilized it. In addition, as the son of the collector (Andrea Aranow), the project was deeply personal and important to me. I wanted to avoid its becoming flat, cold or lifeless in its digital form.

This paper attempts to describe my approach and the lessons I learned, both in process and technique, throughout the three year project. Though my approach and background may differ from many in the archiving field I hope my experience proves useful as I strongly believe physical archives can thrive and be utilized in new exciting ways in an increasingly digital world.

Introduction

The archiving project began with my attending the IS&T Archiving 2009 conference where I gathered information, discussed strategies, and met experts from throughout the world. Prior to that, I had no background in archiving or information science and all of my imaging knowledge came from being an avid photo enthusiast. Throughout the three year project I was aided greatly by dozens of part time workers and volunteers from various specialities, who helped provide the knowledge and expertise necessary. The project was completely self-financed, through revenue garnered from it its use as a commercial resource and supplemented by funds of the author and the collector, on a tight budget.

The vision for the archiving project from the outset was that the digitized collection would offer greater utility and accessibility, thus expanding its potential user base and the way it could be used. In adding this utility to the collection we hoped that an institution, museum or academic, would see value in acquiring the archive. It had previously had been used only by a very small targeted audience of professional designers, and I felt that if made more readily available the collection could serve a much broader audience including students, educators, textile enthusiasts and designers. To meet the needs of this wider potential audience it would not only have to be beautifully digitized and meticulously cataloged but it would also need to be contextualized, conveying the vibrancy of not only the physical archive but also of the collector herself. Successfully conveying this attribute would not be a technical feat but a creative one. Surveying existing digital archives revealed that the this facet of documentation was often overlooked or not effectively presented.

Through research and planning the project was divided into three phases. The first phase would involve choosing a file format that would endure and digitally capturing the collection, aiming for the highest quality and linking the digital assets to their physical counterparts. The second phase would be cataloging each piece comprehensively with both its physical and visual attributes. This necessitated deciding which, if any, meta standard to use to create a hierarchical taxonomy sufficient for describing the entire contents of the collection and documenting the taxonomy and concepts so they could be understood by others. The last phase would involve creating a database and interface to access the rich information and imagery, one that would be powerful enough for an expert but inviting to a novice. While not a formal phase in the project, it was understood that even if the project was successful technically it might not reach its potential if it only documented the objects in the collection. Without including contextual information about the collector, her influences and how the collection was assembled the project would be incomplete. Thus contextual information from the collector was gathered throughout the project and plans call for continuing this aspect into the future.
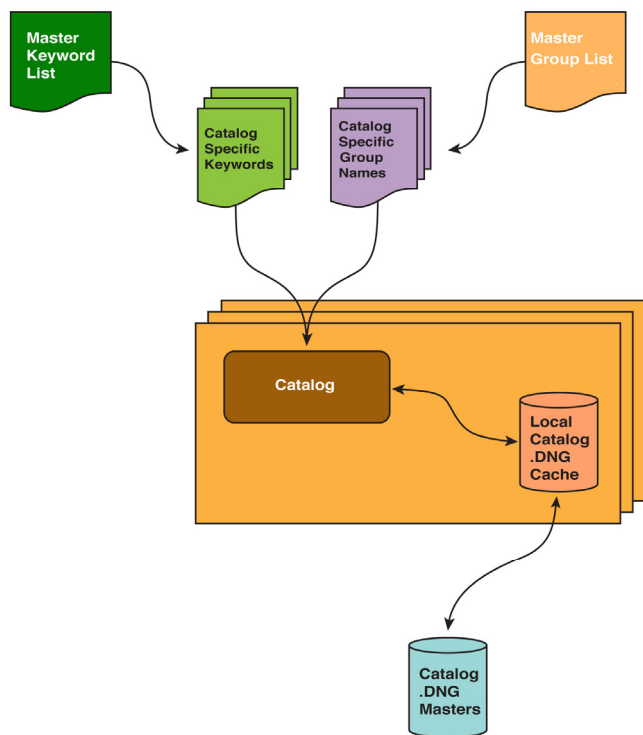
The Capture

A critical decision made during the first phase of the project was whether to digitize the collection using a digital scanner or digital camera. Digital cameras were chosen based on several criteria, the most prominent being cost and time. Items in the collection ranged in size from a few inches to several feet. Large scale scanners above A4 size were extremely expensive, and furthermore, the time each scan took to complete would be significantly longer than the time taken to photograph an item. After testing several camera bodies a identical pair of consumer cameras were chosen that featured a full frame sensors and high resolution. The digital cameras were coupled with a macro lenses that offered little distortion.

Choosing a file format to preserve our photographs was another important early decision. After careful analysis and consideration the Adobe DNG file format was chosen. The primary reason this format was chosen was because of its ability to have metadata embedded in the actual file rather than as a sidecar file. Additionally the DNG file format offered the ability to store the original raw file within, effectively housing a backup file in each DNG. Though we had very few instances of corrupt files, in the instances of corruption it was possible to extract the original RAW file and save us from having to re-photograph items.

Choosing an application to store the DNG files was another critical decision made after careful analysis. A consumer image management application was chosen, primarily because of its ability to work with large image collections and to use the

hierarchical tagging critical to our cataloging phase of our project. We tested several applications and each would have been sufficient for image processing needs. However, the application we chose had the ability to handle both image processing and the projects cataloging needs. Using one application for both purposes thus minimized the complexity of our work flow. The application we chose is primarily used for housing and editing images, it is not a professional cataloging tool or networked. Thus special care was needed devise a workflow to use the application for the project's purposes.

Figure 1. Work flow for handling image files and cataloging using a non networked application

Two stations were used simultaneously to digitally capture the collection. Each was set up to capture differently sized objects, which were lit using compact fluorescents lighting kits. Color checker charts were used to measure color temperature at regular intervals and a color checker chart was included in every image, so that color values could be restored in the event they were ever lost or changed by accident. A given station was operated using a camera tethered to a computer and each item photographed was affixed with a unique ID barcode. This unique ID was used as the filename of the image so that the filename and barcode were identical.

About 10,000 items in the collection were of a set size and mounted on paper. The remainder of collection consisted of variously sized items and preparing those items to be photographed proved to be the biggest challenge of the capture phase of the project. We began by photographing the items that were mounted on paper, which was a mistake. Had the varied sized items been photographed first, the hardest technical and logistical problems

would have been solved earlier, avoiding a disruption in workflow flow later on. Once varied sized objects began to be prepared for photography, it quickly became apparent that there would need to be tight coordination between preparation and photography. Daily items were carefully sorted daily by size and prepared, arranged carefully so that they would remain pristine until photographed.

The number of people necessary to execute this was greater than anticipated and we were aided by the contributions of volunteers and interns. Maintaining and tracking all of the items in the collection after they were photographed and given a unique ID posed additional challenges that were not anticipated. Prior to being digitized, the collection had always been well maintained and organized by various criteria; items could freely migrate with little consequence. However, once an item was given a unique ID and assigned a specific location great care had to be taken ensure its valid location. After the photography was completed a full audit was conducted resulting only .00019 of the collection being unaccounted for.phase of the project. We began by photographing the items that were mounted on paper, which was a mistake. Had the varied sized items been photographed first, the hardest technical and logistical problems would have been solved earlier, avoiding a disruption in workflow flow later on. Once varied sized objects began to be prepared for photography, it quickly became apparent that there would need to be tight coordination between preparation and photography. Daily items were carefully sorted daily by size and prepared, arranged carefully so that they would remain pristine until photographed.

Figure 2. Last textile photographed June 3rd, 2010.

The number of people necessary to execute this was greater than anticipated and we were aided by the contributions of volunteers and interns. Maintaining and tracking all of the items in the collection after they were photographed and given a unique ID posed additional challenges that were not anticipated. Prior to being digitized, the collection had always been well maintained and organized by various criteria, items could freely migrate with little consequence. However, once an item was given a unique ID and assigned a specific location great care had to be taken ensure its valid location. After the photography was completed a full audit was conducted resulting only .00019 of the collection being unaccounted for.
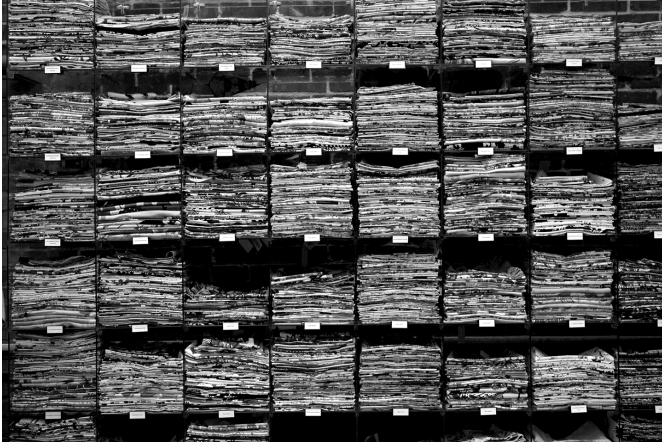
Figure 3. Textiles from the archive organized after photography and cataloging at Blue Rider Design in Portland.

Cataloging

Cataloging the collection was by far the biggest challenge of the project. Especially difficult was structuring the cataloging and taxonomy. The majority of our collection was textiles, but also included were large amounts of artwork (original designs for textiles) as well as printed material. Through our research we found well defined taxonomies for artwork, but not for artwork in the context of textiles. Additionally the project faced two competing sets of interests: those of the current user base and those of new potential users. The first set of interests were those of design professionals who had utilized the collection and supported it economically. These users cared most about the visual characteristics associated with items of the collection, characteristics that were largely subjective. The second set of user interests we tried to account for were those of new potential users, who might care about visual characteristics but also want to know more objective information about pieces in the collection.

Ultimately a custom taxonomy was created to best describe the collection we were documenting rather than using a pre-existing taxonomy. We created a thesaurus of our terms and a controlled vocabulary, using the Art & Architecture Thesaurus (AAT) produced by Getty [1] and The Textile Museum Thesaurus [2] as our primary resources. A data dictionary was created documenting our concepts and defining our fields. The thesaurus and data dictionary for the project were housed in a wiki along with many of our working documents and forms. The project's wiki was an invaluable tool throughout our entire project. It greatly helped to centralize the project's materials, offering revision control thus the ability to monitor changes over time.

The taxonomy created was primarily comprised of two sections: the first filled with objective information, the second with subjective visual information. An example of an objective category was *country of origin* indicating the country where an item from the collection was produced. A related but subjective category was *cultural aesthetic* which described where an item appeared to be from rather than where it was actually from. Both categories are important, but perhaps to different type of users. A designer might want to search for items that had an *African* aesthetic as opposed to only viewing items that were produced in Africa. Conversely a researcher might only want to search for items which were produced in *Africa*.

Each piece in the collection was effectively cataloged twice by two people with different skill sets. We drew heavily from The Textile Museum Thesaurus for the structure of the objective textile taxonomy, though we had make changes where our collection differed from theirs. The subjective, or visual characteristic side of taxonomy, was based on experience working with design professionals and their interaction with the collection since its creation in 1987. Consistently applying subjective metadata was a considerable challenge. In order to achieve consistency, it became clear that accurate definitions our concepts also needed be augmented by clear visual examples. The completed hierarchical taxonomy contains over eighteen key fields, each with multiple sub categories and thousands of terms. The depth of taxonomy makes it unique by amplifying the ways it can be queried and used.

Creating and maintaining such a large taxonomy was quite a challenge. Despite our best efforts to anticipate all the potential terms beforehand, in the course of cataloging we inevitably found items which required new words to be added. In designing a system, it was imperative for it to be flexible enough to incorporate additions while ensuring its integrity. Having several cataloging stations deployed simultaneously and using software that was not networked posed serious challenges to maintaining the taxonomy. In order to make the system work, additions to the taxonomy were documented and monitored by a dedicated administrator.

Another major challenge faced during the phase was the interaction with the catalogers themselves, which had not been anticipated. Cataloging candidates with the domain expertise were chosen, but due to budget constraints, none had cataloging experience and all had to be trained. Many came from backgrounds in textile design and were used to processes that were creative. Cataloging is not a creative endeavor; it is very regimented, and it was incredibly difficult over a long period to maintain consistently high level standard. An important lesson learned was that people who have the domain expertise required may not make the best catalogers. Utilizing domain experts without cataloging experience, required significant management and was one of the most difficult aspects of the project.

The Database, Application and Interface

As the project completed the cataloging phase, attention turned to finding or creating a database, application and interface to interact with the digitized collection. It was understood that if there wasn't a compelling way to display both the photographs and the metadata that the project would not fulfill its potential. Surveying institutional digitized collections and their interfaces there were several which had impressive collections and source material but the user experience of navigating and searching was not a compelling one. In looking at commercial web interfaces, there were many dynamic examples that displayed stunning visuals, nearly instantaneously, but lacked depth in helping users gain real understanding of subject matter.

After a search was conducted to find an existing solution to act as a database and interface for the collection, it became clear that to create the experience envisioned a custom solution would be needed. A requirements document was created outlining the parameters of the database, application and interface and recommending that it be created using open source tools where

possible and be browser based. After distributing the requirements document several meetings with various interactive design firms, of varying sizes, were set up. Although the firms expressed interest in the scope of our project, the proposals they returned ranged widely, were not detailed in technical approach, and all were priced well above our modest budget.

Thus, unable to afford a firm to build the application, the only other option was to hire contractors to build it in close collaboration and to the specifications we envisioned. This approach brought the responsibility of managing the entire process of creating our application and interface and assembling the team to execute it. Ultimately three individuals were chosen who had not only the technical skills necessary but were also engaged by the project's goal and subject matter. The application team consisted of a backend developer, front end developer and a user experience designer. Each team member has a full time job and is taking on the application as a side project. At this time (March 2012) the application is in the midst of development, hence its outcome cannot be described. However, the first iteration of the application is planned to be completed and demonstrated in June at the Archiving 2012 Conference in Copenhagen, Denmark.

Archiving As a Creative Process

Every archive and archiving project is different. From what is contained to what information you choose to capture there are myriad of choices that come to define any archiving project. From the outset vision for the project was not only to comprehensively catalog and index the collection but also to add contextual and supplemental for depth and vitality.

In this particular project, the use of video and audio interviews [3] as well archival photographs and articles were used along with the primary materials of the collection. Styled photographs showing select groups of textiles were also taken to supplement the individual views of each piece in the collection.

Figure 4. Example of styled photograph from the project.

The collection represents the collector's passions and aesthetics and may ultimately become her legacy. There aren't any fields in a taxonomy that can accurately describe the essence of the collection, for that information has to come from the collector.

Another layer of our project that is distinct from its cataloging is a curated view of the collection is. Having the collection digitized and indexed makes it a great tool for users to search if they know what they are looking for, but curation is another important layer of discovery that will add to user's overall experience. Overtime curation is likely to play an increasing role with the digitized collection.

Conclusion

While technical innovation and cataloging rigor are important to any archiving project they should not be the measure of success. Digital archives should strive to be as dynamic and engaging to users as possible. They should not be viewed as flat one-way portals but more as rich immersive experiences where users not only interact with collections but actively participate, engage and ultimately shape them.

Throughout the past three years our project faced numerous obstacles that were ultimately overcome, providing an example of a serious independent archiving project on a modest budget.

My hope in discussing our approach and experience digitizing the Andrea Aranow Textile Document collection is that it stimulates others working on other archiving projects, whether they be personal, academic or institutional. My other great hope also to find a permanent home to my mother's collection so that it lives on inspires and is utilized by a large diverse user group for a long time to come.

References

[1] The Getty Research institute:
http://www.getty.edu/research/tools/vocabularies/aat/index.html
[2] Cecilia Gunzburger, The Textile Museum thesaurus, 2005
[3]Aaron Rayburn, 2011: http://vimeo.com/31315023

About the Collector

Andrea Arnow is a Designer, Ethnographer and Entrepreneur; she received a degree in Cultural History from Brown University(1967). In 1967, she went to New York to open a custom leather and snakeskin apparel shop in the East Village. A few years later, Aranow began traveling worldwide to build a textile library of living cultures. She spent fifteen years abroad, creating ethnic textile collections for museums some of which are housed at The British Museum, The Royal Scottish Museum, and The Costume Institute at the Metropolitan Museum of Art. In 1987 Aranow returned to New York to start Andrea Aranow Textile Documents, a business selling textiles and artwork from her international collection to fashion companies as inspiration for new designs.

Author Biography

Caleb Sayan received a BA in International and Comparative Politics from Trinity College(2000). In 2009, he founded Blue Rider Design LLC and Visual Archiving Solutions to house, maintain and digitize the Andrea Aranow Textile Documents collection.