Archival Vistas: Improving and Increasing Access To the Digital Holdings of the South Carolina Department of Archives and History

By

J. Bryan Collars

Certified Public Manager Project

Class of 2018

Table of Contents

Archival Vistas: Improving and Increasing Access
To the Digital Holdings of the
South Carolina Department of Archives and History

Problem

The South Carolina Department of Archives and History (SCDAH) has over two terabytes of digital content in its holdings. While two of anything on the surface sounds inconsequential in this instance it is not. These two terabytes comprise over 770,000 individual digital objects that need to be managed and accessed by the archival staff. Currently the SCDAH manages access and control to these digital objects in an ad hoc manner that funnels through a single individual. The changing methods of identification of the digital objects over time and the bottleneck of access through just one staffer has led to inefficient retrieval times for a digital object. What is the best possible solution to the management and access issues for these digital objects?

Background

The South Carolina Department of Archives and History is charged by the Public Records Act (Title 30, Chapter 1, South Carolina Code of Laws) with the administration of public records of the state of South Carolina and further with the selection and preservation of the permanently valuable records of the state. In addition the archives seeks to stimulate the research, study, and use of historical documents in the areas of South Carolina history and genealogy. An important function of the department in this regard is providing the taxpayers and general public with access to the record in it custody. Access means more than just allowing the public to view the

records. The archival collection of the department includes 29,000 cubic feet of paper records and 24,234 reels of microfilm divided into 11,084 distinct record series or filing systems. Since 2003 the digital holdings of the department can also be added to the archival collections. This material represents the historical records of over 200 past and current state agencies, departments, boards, commissions and other administrative organizations, 60 past governors, 108 county or municipal governments, and a small amount of private manuscript material. A vital component of accessing the records is establishing intellectual control over the records in the holdings. In the archival field, intellectual control is defined as the creation of tools such as catalogs, finding aids, or other guides that enable researchers to locate materials relevant to their interests. In the digital age this also includes the management and access to the ever growing number of digital surrogates and born digital objects in the archives holdings. Establishing and maintaining intellectual control of the analog (paper and microfilm) holdings is one of the functions of the Archival Processing Unit of the Archives Division within the Department of Archives and History. Whereas this same function for the digital holdings resides with the Electronic Records and Imaging Unit of the Archives Division.

The SCDAH was founded in 1905 and for the first forty-five years of its existence it was led by a single person who for the majority of those years was the sole employee of the department. During this period the department's holding was small, although incremental growth occurred. Maintaining intellectual control of the holdings was accomplished utilizing a subject catalog system developed by department's sole employee. By the mid 1960's, the complexity of state government and the department's archival holdings had grown and the department established a system of arrangement and intellectual control based on national and international archival standards that is still used today. This system established record groups

based on state agencies and local governments. A record group included all the records created or collected by a single agency or local government entity. The record group could be divided into sub-groups to differentiate between different offices, division or branches within an agency or local government, such as the Division of State Human Resources as a subgroup of the Department of Administration record group or the Clerk of Court subgroup of a county record group. A single filing system within the agency would be identified as a record series within that record group,

such as the death certificate record series of the Public Health Statistics and Information Services subgroup of the Department of Health and Environmental Control record group.

The archival staff of the department maintained intellectual control of the expanding holdings via a series of notebooks that tracked the when a record was added to the holdings and identified its source of origin within state government. The notebooks or Summary Guide as it is called listed the Record Groups, Subgroups, and individual records series within those groupings. The Summary Guide though only tracks the records at an aggregate level. It does not record or maintain control of individual documents. This limitation has unintended consequences with the introduction of digital objects in the early 2000s.

In 2002 the SCDAH recognized the emergence of digitization as a process that state agencies were adopting to deal with their record retention issues. Agencies and local governments had begun to digitize their records, both permanent and non-permanent, and many of these bodies were seeking advice and guidance from the SCDAH in regards to the digitization process. The department soon realized that it lacked the knowledge and first-hand experience to respond to these queries authoritatively. The SCDAH needed to get into the digitization game. Doing so would give the department the first-hand experience and knowledge to respond to the

questions being directed at the archives. So in the summer of 2002 the SCDAH initiated the Digitization Pilot Project.

The Digitization Pilot Project started as a modest test project that was staffed by a Front Line Supervisor and a single Processing Archivist both assigned to the project on a part time basis. The project also benefitted from a Library Services Technology Act Grant that supplied state of the art computers, scanners, and software allowing the project to start off at the most advanced level possible. The next hurdle to overcome was, "What to digitize?" Should the project digitize office files? Or, should the scanning focus on the archival holdings? The project staff after some deliberation decided to do both. The decision was made to digitize photographs of National Register Properties of four South Carolina Counties (from the office files of the SCDAH's State Historic Preservation Office) and the archival series, S126088 Confederate Pension Application of 1919 found in the records of the Comptroller General. With these decisions made the project got underway and over the course of the next eighteen months the project scanned over 250 photographs of historic properties and 9,827 Confederate Pension Applications. Each application consisted of two images and coupled with the photographs resulted in just under 20,000 files that needed to be intellectually and physically controlled.

Thankfully the project was able to fall back on years' worth of practice at the Archives and utilized the series level naming convention used to maintain intellectual control of the archival holdings at large. As noted above, the series level naming convention used by the Archives is an aggregate control tool and was less than satisfactory when it was applied to the digital images files of the project which needed control at the item level. Eventually it was decided that the Archive's On-Line Records Index (ORI) would be the basis for the intellectual control of the digital image files. The ORI is the department's item level index of over 300,000

documents from ante-bellum South Carolina. The key number applied by the index for a document could also be used to name the digital surrogate of that document making intellectual control a breeze. This process worked well for those documents that were digitized that were also indexed on the ORI, but what about items not indexed on the ORI? The fateful decision was made that these digital images would be named at the discretion of the staffer doing the scanning.

By 2005 the Digitization Pilot Project had blossomed into a full blown Digitization Program area at the Archives. However, the transition from Project to Program was accompanied by the loss of a staffer resulting in a single person scanning and maintaining intellectual and physical control of the digital images. The naming convention established at the inception of the Digitization Pilot Project remained viable so long as the number of files to be maintained was relatively small or at least was included on the ORI. Fast forward twelve years and nearly 750,000 additional digitized items later, and the process has broken down. The staffer who created the file naming system is no longer the sole individual digitizing items. Each of the three staffers who digitize use slightly different approaches to file naming. The end result is a mish mash of methods to storing the digital files and retrieval of the digital images by anyone other than the original scanner is difficult. (See illustration #1) The vast bulk of the digitization was accomplished by the original staffer leading to the situation where, when that staffer is absent or unavailable, retrieving digitized records is next to impossible.

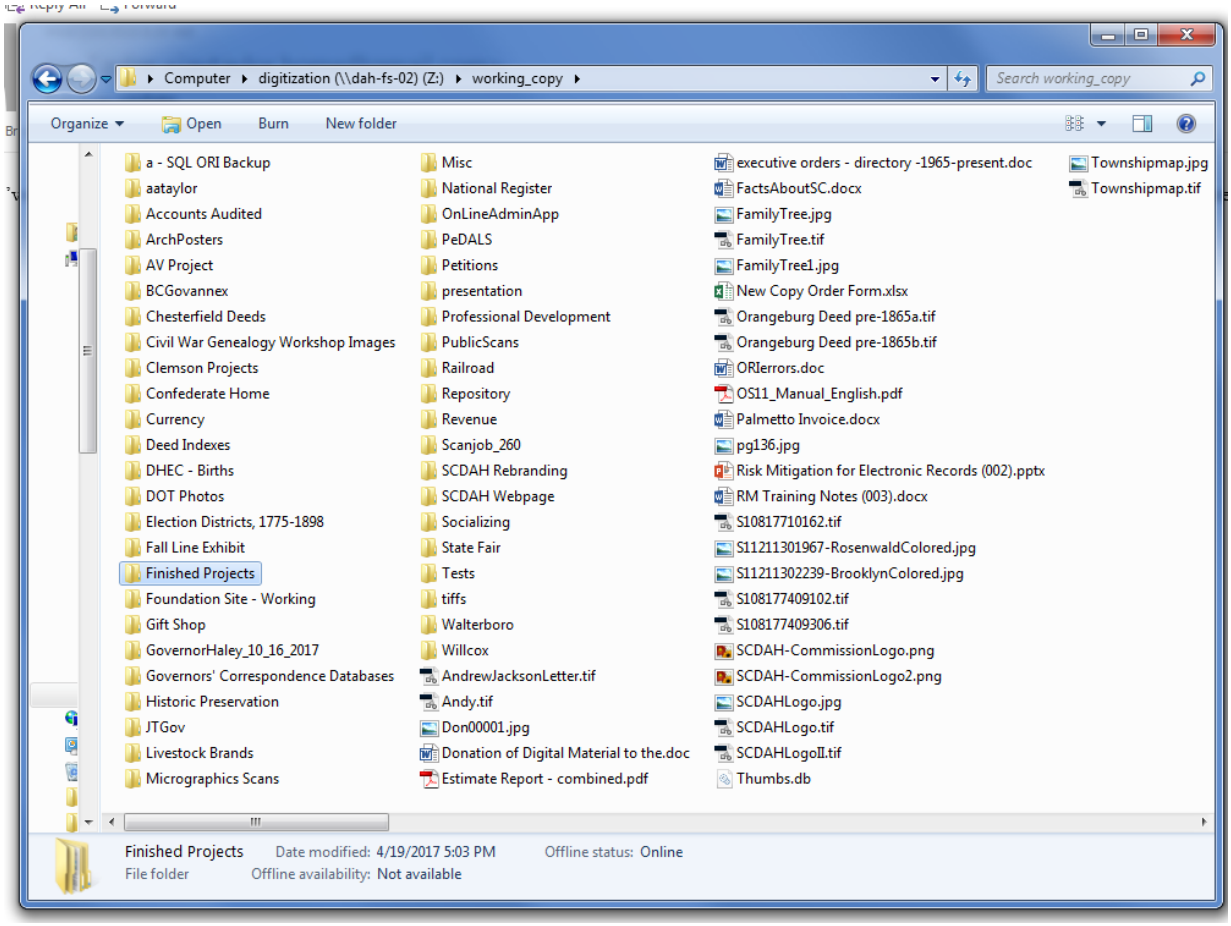


Illustration #1

Current Process

1. Digital imaging is assigned to one of the three staffers who scan and store the resulting images following a pre-determined naming convention. This process minimizes the "losing" of digital files but doesn't eliminate the issue as the staffer digitizing the material still has the liberty to use non-standard methods of naming the files.
2. If the material scanned is a map, then a notation that a digital surrogate exists is added to the map database.
3. Access is still dependent on a single staffer leading to a bottleneck preventing access to some digitized records.

Data Collection

The archival staff at the SCDAH consists of four reference staff, four archival processing staff, and three electronic and imaging staff. All staffers are expected to be able to work reference and perform basic archival processing. The reference staff engages with the public, the archival staff arranges and describes the papers holdings, and the electronic and imaging staff manages the digital holdings. Regardless of the staff's duties, each has a need to at a minimum be familiar with the SCDAH's digital holdings. At the onset of the project a survey was conducted to determine the archival staff's knowledge of the agency's digital holdings. Questions on this initial survey tested if the staff knew where to find the agency's digital holdings and identified at what level the staff believed they had permissions to access the digital holdings. (Appendix 1)

Once the above had been analyzed four staff were selected to be a test group for the remainder of the project. The four staff selected for the test group were chosen based upon their knowledge of the digital holdings. One had very little knowledge, two had some knowledge, while the last was experienced with the digital holdings. The group was then given the references to four different items in the digital holdings and asked to locate them within the digital holdings. They recorded and reported the time it took for them to find the items, or even if they found them at all. These initial results formed the baseline of the second group of data collected, retrieval times for digital holdings. This group repeated this test, seeking different images, once a week for six weeks. (See Appendix 2) The data collected here would show the effects, if any, of training, changes in digital image file naming conventions, as well as the increase in the knowledge base of the Test Group.

Concurrent with the testing being conducted with the Test Group, the Archival staff were surveyed to determine the four preferred options to be used as a file naming convention for the

file titles of the digital holdings. (See Appendix 3) These results were then fed back to the Test Group and utilized by that group to record the effects of the differing file naming conventions on access and retrieval times. The preferred file naming convention that resulted in the most efficient image retrieval is candidate number one to become the standard file naming convention for the South Carolina Department of Archives and History. The Test Group then tested this new file naming convention to insure it was more efficient than previous naming conventions and document any time savings on access and retrieval. (See Appendix 4)

Data Evaluation

The data that was evaluated offered up many anticipated results and most importantly several that were unexpected. The first set of data that was collected sought to measure the archival staff's knowledge of the digital holdings overall. The single biggest surprise of the project was discovered at this point of the eleven staffers surveyed, seven (74%) of the staff could not identify which directory on the network they could find the digital holdings of the SCDAH. (See Appendix 1) It's clear that some basic training was necessary to bring these staffers up to date on where all the holdings of the SCDAH reside. The second part of the initial survey sought to identify the barriers to access. For the 74% who didn't even know where the digital holdings resided, 5 out of 7 (71%) felt that access to Digitization Directory was not in their job duty; 1 out of 7 (14%) indicated they did not have the proper permissions to even enter directory where the digital holdings are stored should they know how to access it. The final staffer who claimed to not know how to access the Digitization Directory noted that they didn't access the digital holdings due to the difficulty in finding the specific file they were looking for. The questions must have confused this staffer. The other staffers who noted they utilized the Digitization Directory and accessed the digital holdings rated their ease of access as found in Appendix 1. After analyzing these results it's clear that training of the entire staff needs to occur.

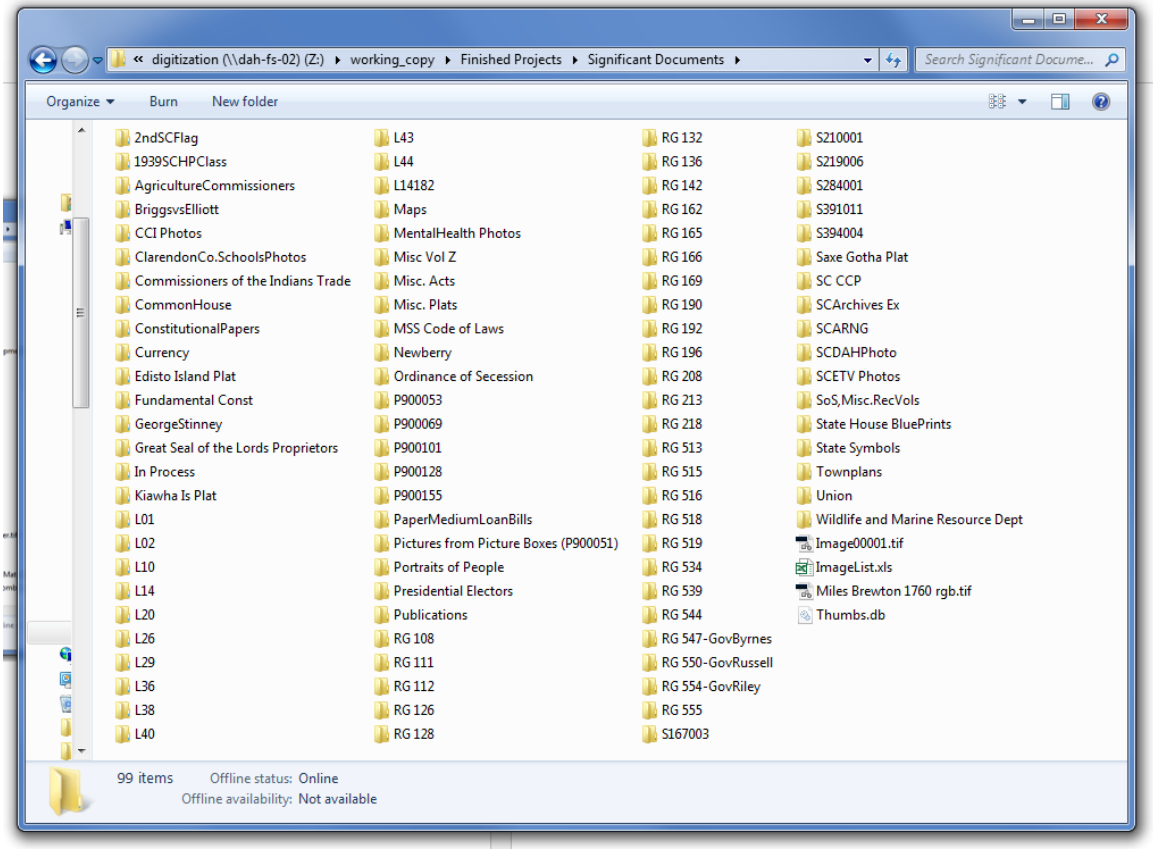
The next set of data was gathered over the course of six weeks and sought to identify which of four different file naming conventions was the easiest to use and it was the most efficient in the retrieval of the digital holdings. (See Appendix 2) Each week the Test Group was given references to four different digital files and asked to find those files on the Digitization Directory. The Test Group members reported the time it took to locate each file which was then given a numeric number based on the following parameters: >15 min = 1; 16-30 min = 2; 31-45

min = 3; 46min+ = 4. After the above numbers were assigned the values were then averaged to create a final score for that week. After six weeks the weekly scores were averaged to find the ending average. The lowest average score indicates the most efficient file naming convention for useful retrieval of the digital holdings.

As the Test Group was working finding digital files on the Digitization Directory, each week the remaining archival staff were exploring different file naming conventions. Over the course of several weeks it became evident that three file name types were clearly the favorites of the staff. (See Appendix 3) The favorite coming out of this process was a file title based upon the ORI index number which is 19 character long (ex. S126088000000212000). Such a string of numbers is certainly unique but when 60 or more are displayed on a screen the numbers blur and are at times difficult to discern. The second example is a text based naming system. This type of naming style certainly allows the viewer to discern what the contents of a file is but when many files are present the files may not present themselves in a systematic matter. The third and final file name finalist was a combination of the first two. This blended the ORI numeric system with the text based system that allows the staff to determine the content of a file while at the same time insures the hierarchical structure of the numeric system. The final step of the project was to test these file names against one another to determine which was the more efficient. (See Appendix 2-wks 4-6)

Conclusions

In the end of the project it is clearly evident that the staff needed adequate training more than anything else to more effectively begin using the digital holdings of the SCDAH. Having three-quarters of the staff not even having the knowledge of where to look for the images is certainly bigger than even a new naming convention. At the end of the Evaluation Phase the decision was made to commit to the Combo file naming convention going forward and start the process of retrospectively renaming older files as needed. The process of retrospectively going back and changing file titles where appropriate is a laborious task but doing so increases the access and availability of the records to the entire staff and through the staff to the public at large. (See Illustration 2) Overall the project was successful and provided useful information on the archival collection at large and specifically on the need for increased staff training.



Appendix 1

Knowledge of the Digitization Directory and How to Access It

	Yes	No
Reference Staff	0	4
Archival Processing	1	3
Electronic Records and Imaging	3	0
Total	4 (26%)	7 (74%)

Of Those that had no Knowledge of the Digitization Directory
What was Their Reason

No Job Duty Requirement	5 (71%)
No Permission	1 (14%)
Too Difficult to Use	1 (14)

Of Those that had Knowledge of the Digitization Directory

Found it Easy to Access	1 (25%)
Can Find what They Are Looking for but it Takes a While	2 (50%)
Knows how to Find Digitized Items	1 (25%)

Appendix 2

	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Avg. Value
19 Character Number	1.25	1.5	1.0	1.0	1.5	.75	1.15
Text Only Name	1.75	1.75	1.25	1.5	1.75	0	1.33
Combo	1.25	1.25	1.25	1.25	1.0	1.5	1.08
Other	1.25		2				1.6

‘>15 min=1 16-30 min=2 31-45 min=3 46+=4

Appendix 3

File Naming Preferences from the Archival Staff

19 Character Number	5 (45%)
Text Only based System	3 (27%)
Combination of the Two	3 (27%)

Appendix 4

Improvement in Access Times to Digital Holdings

	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Avg. Value
19 Character Number	1.25	>	>	>	>	>	1.15
Text Only Name	1.75	>	>	>	>	>	1.33
Combo	1.25	>	>	>	>	>	1.08
Other	1.25	>	>	>	>	>	1.6

‘>15 min=1 16-30 min=2 31-45 min=3 46+=4