COLLECTIVE COOKBOOK:
BEST PRACTICES IN VISUAL ARCHIVING BETWEEN
NEWS ORGANIZATIONS AND INDIVIDUAL COLLECTIONS

A Professional Project
presented to
the Faculty of the Graduate School
at the University of Missouri – Columbia

In Partial Fulfillment
of the Requirements for the Degree

Master of Journalism

By

MATTHEW HELLMAN

Dr. Keith Greenwood, Committee Chair & Project Supervisor

Brian Kratzer, Committee Member
David Rees, Committee Member
Edward McCain, Committee Member

DECEMBER 2016

TABLE OF CONTENTS

      Background & Personal Investments

      Activities

      Establishing a Structure

      Building Personal Knowledge

      Reference Documents

      Interview Transcripts

      Project Proposal (Includes Proposal T.O.C. & Sections)

LIST OF TABLES & FIGURES

Transcripts (In APPENDICES section)

PREFACE: ACCESS & PERMANENCE

"There has always been a trade-off between access and permanence. When we first started the graphic record, we could pound a rock against a rock and make a picture, and that picture would last a long time, but it was not very accessible. You would either have to carry the rock around and show people or they had to come and see the rock. Then we (humanity) moved to clay and tablets. These objects were more portable but also more easily broken.

Eventually we moved to paper, which we learned could last for hundreds or thousands of years if you keep it in the right environment. But it burns, can be eaten by insects, can be torn up... Today, technology has brought us into the digital realm. This digital stuff is so new that we are just beginning to learn how to grapple with preserving it. One fragmented file or disc on a hard drive and we lose quite a lot of content. On the other hand, access is fantastic because we can go all over the world with it...

...But we are on the opposite end of the rock."

*— **Edward McCain, MU Digital Curator of Journalism***

CHAPTER 1: INTRODUCTION

It has long been a mindset in the media industry that the content we produce is the first draft of history. As soon as ink hits the paper or an image is exposed and saved in the camera, history is being transcribed. Before beginning my master's program at the University of Missouri, I had never really put much thought or weight behind the value of archiving. Maybe it was because I was too young to understand the importance of it or maybe it was because I was used to a semi-decent archive structure at my undergraduate paper and internships I worked at. Working with the Columbia Missourian photo archives, my eyes were opened to how crucial a good archiving system is and why it should not be taken for granted. With modern technology producing so much content, quality archiving infrastructures have never been more crucial. As the written word and the visual element become less permanent, preserving them becomes more prominent.

In 2011, an organization known as the Newspaper Archive Summit hosted a convention that gathered archivists, librarians and key newspaper members to discuss the "preservation of newspaper content for future generations" (Carner et al, 2014, p. 11). Their goal: to create a policy for maintaining and protecting historical information that cannot be replaced. In 2014, a study funded by the Donald W. Reynolds Journalism Institute (RJI) sought to identify how digital-only newspapers dealt with the issue of archiving in today's emerging media market (Carner et al., 2014). Researchers Dorothy Carner, Edward McCain and Frederick Zarndt specifically looked at two kinds of newspaper models – the legacy (print)

1

papers that also published online, which we called "hybrid" and the newer model "online only" publications.

What did they find?

Modern archiving is a huge cost and huge time commitment, and most news organizations have either limited or drastically reduced the people that they have working in their news library (Hedstrom, 1997). By archiving, I am referring to the process of preserving "unique content" in a manner that is organized and accessible (Schmidt, 2009). Adding to this limited librarian role, the growing issue for most news organizations is that more of their news content is being published digitally.

Over the past four years, I have had the opportunity to explore how different newsrooms manage and maintain their photography archiving systems. Beginning at the Baylor Lariat newspaper in Waco, Texas, my understanding of archiving and digital content management originated when we had to undergo upgrades due to lack of server space for our visual content. As modern technology progresses, so does the amount of storage space necessary to manage the increasing quantity and quality of content. This was my first stint with local, in-house servers and the limited use they served. The beauty of the Lariat was that they really prepared me for what came next: the Columbia Missourian and their disarray of storage units they called an archive.

Before I began my master's program back in 2013, Edward McCain, digital curator of journalism at the Donald W. Reynolds Journalism Institute and MU Libraries, began to direct his focus towards developing a program for the Missourian to preserve their digital content. He brought this conversation to Tom

Warhover, the Columbia Missourian's managing editor, and Missourian Director of Photography Brian Kratzer. Together the three began to discuss plans for digital preservation and visual content management. Beginning in January of 2014, Greg Kendall-Ball, a master's candidate working at the Columbia Missourian, played an important role in helping design and implement a new workflow for the photo department. His work largely informs many parts of the present workflow, including the archives.

In August 2014, in what happened to be twist of fate, I arrived and found the Missourian archives in disarray. Soon enough I made it my personal mission to fix them. The Missourian newspaper, when I arrived, had an unorganized archiving system consisting of 12 separate almost-decade-old hard drives, hundreds of CDs and three local servers that weren't organized in any manageable fashion. At the Lariat, everything was consolidated and organized on a single server with an established workflow. The Missourian presented me with the opportunity to assist in building an infrastructure they could more easily sift through and find the content they needed, and so it began.

I began in Fall 2014 by trying to get a grasp on where all the Missourian content was and where it needed to be. After consolidating the various hard drives and merging the CD files into a single location, I began a reorganization process that classified the content first by year, then by semester and finally by person and event. This process lasted through the fall semester and into Spring 2015, where I began to assess the metadata within these archived files. With direction from Edward McCain and Brian Kratzer, we began to develop a workflow and metadata template that

could be incorporated into the staff photojournalism class in order to better organize visual archive content. Using the same workflow and metadata style, I also began to correct older content files to fit the same mold. The back half of Spring 2015 came with a whole new obstacle. The Missourian switched from an in-house server to a new campus-housed server that was larger and contained greater redundancy and a constant monitoring staff. It became my responsibility to make sure all of our old content was migrated to this new server.

Fall 2015 had me focusing on getting the servers transferred over. By the end of the semester, all digital photography content was consolidated on the new server known as Pre- Digital Asset Management System, or Pre-DAMS for short. The Pre-DAMS server functions as a local digital archive, but also serves a bigger role as the preparation drive for visual content waiting to be integrated into the Missourian's new digital asset management system still in development.

Fast forward to present day: the Missourian currently uses none of the old servers, hard drives or CD's but now operates off of a single consolidated archive that is organized by dates, semesters, months, weeks and days, with every new file added containing proper metadata and caption information. Older files from previous storage systems are being updated with proper metadata to fit the present style structure but it is a slow process still under way. The Missourian is also now evolving to a new server that involves everything being stored on a local RAID device and also backed up within Amazon Glacier's long-term storage services that will be searchable through a ResourceSpace digital asset management software nicknamed MOchive. The ultimate goal for this massive endeavor is to get the

Missourian from an almost non-existent archiving format to a digitally consolidated, managed and searchable database that increases production efficiency, availability and economical value for the news organization, and the process is gradually making great strides. Nevertheless, there is still much work that needs to be done.

My most recent work has focused on evaluating the best image resizing software before we begin uploading the vast amount of digital content to MOchive and creating a resizing and upload workflow that can be implemented to get content prepared and into cloud storage. Resizing images is necessary in order to conserve space and save money for the Missourian. Most of the images in our archive aren't of high importance so there isn't much need to maintain an extremely large version of them. With a series of image resizing trials at various sizes and quality levels, it was determined that the best software to use for this process was Photo Mechanic, a batch photo asset management software that allows for mass file processing at once.

Through these experiences, I have begun to learn how to identify weak points in archival systems and improve the existing assets to adapt better methods of finding the content news organizations need at a faster and more reliable rate. This knowledge has directed me to pursue a project that seeks to gather the best archiving practices from individual archival projects to newspaper organizations. This project also provides an example of the skills I possess and could potentially provide to future employers in regards to their archiving and digital asset management scenarios. For this study, digital asset/content management refers to the role of maintaining and correcting workflow errors such as metadata and file information as well as any consolidation processes that may be necessary.

Although this project provides me with a better understanding of digital asset management and preservation, it is important to understand that those skills and recognition gained are not the reason I chose to do it. From the beginning, I have always had an affinity for being organized. My personal archives are redundant and also backed up on cloud storage themselves. I understood the stress of losing a lot of digital content early in my experience when I lost my laptop during a car accident. Some important things can't be replaced and a simple small accident could be devastating to digital files that aren't properly preserved. Edward McCain's system design has great potential for the media industry. To see it grow and succeed, it has to begin by being tested and implemented successfully. That task alone requires diligence, cooperation and most importantly collective backing and effort from everyone in the department in order to prosper. That is what drives me to do this project. I want to see a future where at least Missourian photography preservation isn't a concern or an obligation, but rather a regularly incorporated and significant piece of the daily workflow.

Henceforth, the purpose of this project is to identify the aspects of news publication archiving that rely solely on the visual photography content produced at media organizations and individual collections. This project will then analyze what each subject found successful or not and compare why things did or didn't work for one that did or didn't for another. Finally, the ultimate purpose of this study is to present the best practices for any visual archive whether they be personal or organization-based.

CHAPTER 2: FIELD NOTES

Activities

Over the expanse of this project, most of the physical component has involved restructuring the Columbia Missourian digital photography archives in preparation for the files to be uploaded to MOchive, the Missourian's ResourceSpace-based digital asset management software. The process of restructuring involves correcting file naming conventions to the most current format, correcting basic metadata information such as captions, keywords, location information and photographer information, and consolidating any missing or misplaced files into the local digital archive. Over the present semester (the last 10 weeks) I have taken weekly, or on occasion bi-weekly, field notes describing my tasks, accomplishments and challenges faced each week as well as a to-do list for the following week. My field notes can be seen below:

WEEK 1: 8/29 - 9/3

Good afternoon everyone. As this is my project semester, I will be sending you weekly reports that will include any changes, progressive actions taken, setbacks and any other relevant information related to my project. Feel free to provide commentary and or suggestions along the way. I look forward to seeing what the development looks like by semester's end. Here we go.

Week one tasks completed:

- Transferred spring and summer semesters from PhotoIngest to Pre-DAMS.

- Summer was about 600GB and the remainder of spring was about 200GB; 4.1TB available on PhotoIngest
- Files were sketchy-er back then since no real archiving system was in place but I organized it best I could. Seems like we only kept the final edited versions back then, not the dumps.
- NOTE: We have 5.37TB left on Pre-DAMS; definitely time to start getting things into MOChive so we can free up some of that space

- Met with Edward and Sophia on Thursday and reestablished where we were as far as uploading content to MOChive
    - Long way to go getting content uploaded. It is by far the slowest part of the process.
    - Looking into concerns regarding upload problems. (Ed I don't believe it is a Photo Mechanic problem since the For_Upload folder is showing up with the same number as the original folder. Something is wrong with the uploader on MOChive. Will definitely test it out on Tuesday.)
    - Reviewed captioning and keywording for summer semester and made any alterations necessary (Brian, aside from one photographer, the new keywording/captioning metadata process seems to be working well and everyone seems to be following it.)
    - Set a schedule for reviewing captioning/keywording of current photo staffers. My first check will be Tuesday followed by every 3 weeks onwards. This should be less of a concern now as it has been engrained into the ADOP's and editor's heads through previous semesters.

Week two to-do list:

- Test MOChive uploader and attempt to figure out what is causing the

   backup/issues

- Continue fixing keywording and captions in 2014 files

Alright well that's it for this week's report. Keith if you want me to format these

differently let me know. Google Docs perhaps? Also, this report will be sent out

every Monday from now on. I meant to do it then but got caught up with pay stuff

for the semester.

WEEK 2: 9/4 - 9/10

Overall week two was fairly short week as I was out of town for the weekend

starting Thursday. Nevertheless, quite a bit happened and here is what was

accomplished.

Week two tasks completed:

- Repaired IPTC data to 1999 files and begin adjustments to 2000.

  ▪ As Ed recommended, we have **3.89/4.23 TB available on PhotoIngest** (The

     server we use for daily workflow in the Missourian photo department.) We

     have **5.36/18.18 TB on Pre-DAMS**. (To put things in perspective, the entire

     2015 year took up 3.06 TB. The sooner we get files permanently on MOchive

     and in glacier, the sooner we can start freeing up space.) SIDE NOTE: Brian, I

     cleared off the Multimedia hard drive early last week and opened up 2.25

     TB/3 TB. That should hold us over through at least 2017.

- Files were sketchy-er back then since no real archiving system was in place but I organized it best I could. Seems like we only kept the final edited versions back then, not the dumps.
- Met with Sophia on Wednesday and cleared up why the folder counts weren't matching up with images in the •for_upload folder.
    - Since editors went through last semester and color classed images from the entire take, some of the ones they chose weren't tagged by the photographer (sometimes we find stuff in the entire take that photographers didn't think were their best)
    - I just had Sophia go back and tag all the color coded ones before proceeding to convert them to for_upload files.
- Performed the 1st round of ingest evaluations on photographer folders.
    - Thanks to me beginning this in the spring, all the proper ingesting techniques have rolled over to fall and things are really going well as far as metadata is concerned. Very minor fixes and nothing that couldn't be corrected easily.

Week three to-do list:
- Continue repairing metadata for 2000-year files and 2014-year files
- Check to make sure photographer made corrections to individual dump folders
- Meet with Ed and Brian on Thursday to discuss new uploading method for MOChive (possibility of setting up a queue for continuous uploading. If we can swing this queue system then it will really streamline the uploading process. Especially since we can prepare •For_Upload folders quickly.)
- Follow up with Sophia afterwards.

Sorry it took so long to get this week's notes out. Been a busy weekend and was dealing with payroll/financial aid/summer term issues. Should be back on schedule now.

WEEK 3: 9/11 - 9/18

Week three presented new challenges in the form of a new queue system and defining the archive structure and workflows. This queue system really needs to happen. It will play a key role in the future of MOChive. Here is what happened:

Week three tasks completed:

- Met with Brian and Ed about creating a new queue system (called StaticSync)
    - We are still working on details as to what level of folder should be dropped in to this queuing system. Personally, I am leaning towards being able to queue up individual event folders rather than entire semester folders all at once because subfolders vary in their hierarchy (specifically with MU football game folders) and could confuse the process. Hopefully we will have some sort of system hammered out soon and we can then bring Sophia into the process. For now, I will assume that we will just have her keep doing what she is doing.
    - Football game shoots contain 1st quarter, 2nd quarter, 3rd quarter, 4th quarter, halftime and pregame subfolders rather than having all of the photographers coverage lumped into one folder. This is mainly because photographers can shoot upwards of 3000+ images per game.

- Finished updating 2000 Spring, 2000 Summer and 2000 winter metadata, naming and folder structures to reflect our current style.

- Began work on 2000 Fall (this is proving more time-consuming than I projected, but I will get it figured out and moving faster.)

- Continued updating 2014-year files (More current years are already pretty well in sync with what we want metadata wise, but the Missourian editors back then clearly weren't as strict and vigilant on it as we have been this past year.)

Week four to-do list:

- Continue updating metadata for 2000-year files and 2014-year files

- Create Google workflows for Missourian photo department and MOChive in one place

  ▪ Also include full IPTC fields chart showing what metadata fields are currently in use by the Missourian

- Create Hierarchy document that shows the current structure of Pre-DAMS and begin updating prior year folders to fit within the specific structure laid out

- Follow up with Sophia on meeting details and progress

WEEK 4: 9/19 - 9/25

Overall, the workflow is looking better as we progress. Things are slowing down as far as editing 2000-fall and 2014-year metadata since more files are being dealt with. Still one of the slowest parts of the process but a necessity nonetheless. I will keep working at it. Here was the week's lineup:

<u>Week four tasks completed:</u>

- With Edward's guidance, we have an all-in-one workflow laid out that shows the start of the ingest process all the way through the MOChive uploading process. Future things to include will be any MOchive specific processes that we will need to have for different roles.

    ▪ I am still working on the folder structuring for "zz" labeled folders. This will be something I will want to sit down with Brian about and see what he wants to do, mainly because we don't really have a structure set up inside handout or AP folders. When we receive handouts or AP pictures (at least in past semesters), editors would just dump them in there to be processed for print or web. They didn't really rename or structure the files in any systematic format.

- Progressed on Fall 2000. The main hold-up here is the file naming structures as I have to go in and rename each image. Unlike 2010+ years, the 2000 folders don't have the entire dumps but just selects from various shoots. So the problem there is having to go to individual files and re-caption and re-keyword them one by one. That is why I am still working on them.

- Continued correcting metadata information on 2014 files.

<u>Week five to-do list:</u>

- SPECIAL NOTE: I am working MPW this week as the vortex crew member so I am working remotely from Cuba, MO. On that note, I will be focusing primarily on the continued updating of metadata for 2000-year files and 2014-year files.

- I will follow up with Sophia today and see how uploading is going, but for the most part nothing has changed in her process as of yet.

- I will continue to look into folder structuring options for "zz" folders.

WEEK 5: 9/26 - 10/2

Week 5 was spent at Missouri Photo Workshop so it wasn't nearly as interesting but things did get done.

Week five tasks completed:

- Finished updating the folder structuring document. I have decided to take a different directional approach and am going to go through all of the years and fix naming conventions first before working more on keywording and metadata information.

  - My reasoning for this switch is that naming conventions differ widely across the 15-year digital span from 1999 to 2014 and that needs to be corrected so when we do get them into the system nobody will be confused by different naming conventions.

    - On this note, I have formatted all the naming conventions for 2014-2016 and have begun working on the 2013 files. I will continue working my way backwards on this.

  - Keywording and captioning are still extremely important, but naming conventions are imperative to the structuring of MOchive and henceforth must come first.

- Sophia is still uploading files, but after 5 weeks she still only has one person's folder for one semester in the system.

  - The automated system really needs to happen as soon as reasonably possible otherwise my project will not be able to progress much past the current phase.

  - **Side note:** After 5 weeks into the semester, we have used up approximately ½ TB of space on the PhotoIngest server. We had **3.89/4.23 TB** available on PhotoIngest and we now have **3.41/4.23 TB** remaining.

- As far as keywording goes, I will finish my current work on 2014 then divert all my focus to naming conventions and folder structuring first so everything has continuity.

Week six to-do list:

- Finish updating metadata for 2014 files. Almost done, just need to finish a few folders in the fall semester.

- Update 2013 year to the current folder structure illustrated in the Google docs.

- Perform the 2nd round of ingest evaluations on fall 2016 photographer folders

- I will be out of town Thursday through Sunday this weekend as I will be headed to Texas to spend a bit of time with my family (I had a birthday yesterday). If I have extra time down there, I will continue the folder structuring process remotely. Finally got my PC computer working with the remote access.

*Edward, whenever you have new information on StaticSync just let me know. If you need anything specific give me a shout.

WEEK 6-7: 10/3 - 10/15

I decided to lump these two weeks together as they consisted of much the same

work. Here you go.

Week six and seven tasks completed:

- The biggest jump here was I finished resetting the folder structures all the way

  back to 2009.

    ▪ Thanks to the help of a new software I found called Batch Rename Utility,

      that made it simple to rename entire semester's worth of folders fairly

      quickly. (Previously I had to do each photographer's folders one by one.)

    ▪ The only thing that remains to be restructured is the images within each

      folder. As far as I can tell, there isn't a way to batch rename them in every

      folder all at once, so I am taking every shoot for the entire semester into

      photo mechanic and batch renaming them that way, folder by folder.

    ▪ Keep in mind, this part is just to have the proper structuring in place for

      MOChive. All these folders still need to be checked for proper metadata,

      especially on the ones selected for publication.

- Finished ingestion evaluation follow-ups (I've started to record notes on routine

  problems and concerns here since these evaluations are becoming fairly important

  to illustrating the integration progress of the new ingest and folder structuring

  format.)

- Prepared a quick guide to ingest structuring and metadata for Jackie Bell at her

  request so she could include my evaluations as part of the photographers'

  professionalism grades.

- One of many good effects of these evaluations. Having this as part of their professionalism grade forces upon them the importance of good ingesting technique and provides an actual deterrent for sloppiness.

Week eight to-do list:

- Finish restructuring the naming conventions for images in the 2010-2014 folders. Not sure how long this process will take but I am working as fast as I can. The semester's end is drawing near.

- COMMITTEE MEETING: We are past the mid-point of this semester and the research phase needs to be looked at considerably if I am to actually have a project done before semester's end. That being said, it looks as if Wednesday or Friday at 2pm are the best times to meet.

  - Other things to look at in this meeting are the progress on StaticSync and what roles need to be established with ADOPs or editors so all of our progress doesn't just stop when I leave.

- Look into 2001-2009 archives and see what needs to be fixed. I haven't looked at these much yet because the heftier folders content-wise were after 2010. (Tackle the big fish first.)

WEEK 8: 10/16 - 10/22

Following the committee meeting. It looks as if things have finally narrowed down to a target that is doable.

Week eight tasks completed:

- COMMITTEE MEETING NOTES: Changed the direction of the final project to be a "Best Practices" analysis and composition. We determined that minimum 4 interviews (2 organizations, 2 individuals) are needed for a reasonable comparison and research quantity.

  ▪ Ed has emailed five people (Jack Dykinga, Mike Meiners, Chris Steppig, Peter Krogh and Rick Wiley) as potential contacts.

  ▪ I have heard back from Mike and Jack. I have sent both of them time options and am waiting to hear back.

  ▪ As for interview questions, see the following list:

  https://docs.google.com/document/d/1HysDQRgG4zte1StLJMAxLTeRQV0zwz9KtKyrLanlRg0/edit

  ▪ Please add to this list if you think I am missing any important points. I do pose additional questions based on interview answers.

- Skills Component Update

  ▪ Have begun looking at "prior to 2009" files to see what needs to be fixed. Begun renaming conventions for all folders to make everything follow the same style as the current semester.

- Note: problems with file naming before 2005. Images only have dates and sequence numbers. Not sure what caused this, but need to look at DROBO to see if this is a constant problem or if it is something new.
- Finished fixing naming conventions for all folders after 2009. Some images still need renaming but that requires a folder by folder check. For the most part, those files are good as far as structure goes.

Week nine to-do list:

- Hopefully make contact and get interviews arranged. From this point out, I may be updating reports on a bi-weekly basis instead of weekly since things are going to get really repetitive.

- Need to continue renaming conventions for 1999-2009 files. Things are shaping up, but I really need go back and make sure EVERYTHING is transferred over from DROBO.
  - Haven't logged into DROBO in a while so hopefully that still works. Will perform a check next week.

- Follow up on project report formatting with David and Keith so I am on track to finish before or by Monday of thanksgiving break. More updates to come. As long as the interviews get in, I don't foresee this being a problem.

WEEK 9-10: 10/23 - 11/6

Well plenty to report here. First off. Things are progressing well with the report and with the interviews, although not as much as I'd like. Nevertheless, everything is still doable and I hope to have more to report by the end of the week.

Week nine and ten tasks completed:

- Got folder renaming and structuring fixed for everything back to 1999. Working on fixing image naming conventions all the way back now.

- Stumbled across some interesting problems outside of the wrong naming convention issue from last week. It seems not everything from DROBO was copied over.

    ▪ The CD archive has a bunch of raw images (The 2005, 2006 and 2007 fall dumps) that weren't moved over. I will get that done this week for sure.

- Lots of project updates. I now have 2 out of 4 interviews needed, and with good sources too.

    ▪ Jack Dykinga gave me lots of good information and pushed me to think outside of just the newspaper mindset with his archiving practices. It is nice to see that he is preserving RAWs and has a good process and that will be a good comparison to the JPG-only mindset that most newsrooms maintain.

    ▪ Michelle Jay, photo archivist of the Boston Globe, agreed to be my second source. The Globe is a great option since they are only a couple years into their archiving practices, just like the Missourian. They will make for a great comparison to our own practices and the things they are doing differently including monetization, their treatment of RAWs and JPGs, their filing

system, etc. I interviewed her on Monday and I am stoked to transcribe that interview for my report.

- I have begun formatting my project report. Already finished the Title Page, Abstract, Table of Contents, Introduction and begun working on the professional skills component. I expect to have it completed (up to the research section) by the end of the week.

Week 11 to-do list:

- Continue folder structuring. Still need to get image names fixed. I expect this is going to go on past the project deadline and even past the defense, but since I am here until Dec. 17 I will keep working to get it as far as I can before leaving.

- Get the CD archives copied from DROBO to Pre-DAMS.

    - It looks to be about 500 GB worth of content. Not a lot in retrospect, but seeing as the server is **down to 5.2 TB/18.18 TB** and we still need to get Fall 2016 copied over from PhotoIngest eventually, I can't stress enough about getting StaticSync up and running ASAP. Pre-DAMS looks to only last through 2017 unless we start getting things into MOchive.

WEEK 11: 11/7 - 11/12

Week 11 has been primarily focused on managing the research portion of my professional analysis. I have 3 of the 4 interviews I need and will be getting the final one soon. As for the project report, I have created a folder in Google Docs that I am saving the sections (as laid out by the MU Project Report Checklist) for committee review. If you have any questions, don't hesitate. *Note: as the last official project report to be included in my paper, I'm including things to do in the sub-bullets.

Week eleven tasks completed:

- Organized project report - Finished abstract, keywords, Chapters 1-4 (with space allotted for this field note and for the final interview) and attached the proposal.

   - The only things I have left to do are the actual project analysis (working on it today and tomorrow) and the project alteration updates that follow the proposal

   - Updated all committee members on the folder. Awaiting feedback from Keith before sending out final project report. (I intend to send it out Friday if he gives me the go-ahead)

   - Will begin scheduling defense with committee members as soon as Keith says I can.

   - Edward, I will update Dorothy after I get the fourth interview today. I just need to hammer out this analysis right now since it is on deadline.

- Performed 3rd round of ingest evaluations

   - Overall, the ingest evaluations are the heart of the progress for future semesters. These evaluations not only keep the photographers from making

more mistakes, but they also guarantee a correct archive each semester. Not to mention refreshing the concept of archival practices amongst the photographers, editors and ADOPs.

- Expecting corrections to be made by Friday by the photographers.

Final report note: I wanted to include this as the last thing because it is most important after I am gone to follow through with the following things:

1. Keep doing ingest evaluations every 4-6 weeks of each semester. I can't even express how vital this has been to making the current archives easy to maintain. Brian, appoint this task to an ADOP or make it an editor beat. It will ensure good archive practices.

2. StaticSync needs to happen (the server is filling up quickly, but I have been informed the situation is manageable.) We aren't even done with the semester and Fall 2016 is **1.75 TB** in size. As a final update:

   a. **PhotoIngest has 2.48/4.23 TB**

   b. **Pre-DAMS has 5.18/18.18 TB**

   c. **DROBO has 5.98/17.57 TB**

3. **According to Edward, the server is expandable and can be added to or re-allocated for more space if needed. Regardless, that shouldn't be a reason to slow down on this project as it is still extremely valuable.**

4. I will finish getting content off DROBO after Thanksgiving and my defense. We can be done with it by the end of the semester.

CHAPTER 3: EVALUATION OF SKILLS COMPONENT

Establishing a Structure

Over the past 10 weeks, much of this project has evolved from a fast-paced scenario to a slower, more in-depth process. What started out as a plan to have MOchive up and running by the end of Fall 2016 semester ended up with reverting back to basic folder restructuring and building primary workflows for how the archives are organized. The first question that must be asked is why did the plan change?

MOChive Concerns & Solutions

Various unexpected problems and realizations led up to the delayed integration of MOchive. Primarily, the archival team, comprised of Sophia Zheng, Edward McCain and Matt Hellman (myself), vastly underestimated the sluggishness of the upload process into MOchive. After four weeks, we were still working on the inputting all of the files from the first person of the Spring 2016 semester. This backup is primarily caused due to the MOchive uploading system requiring a manual uploading of files through a drag-and-drop format. The system itself can process the files just fine, but the manual process of actually moving the files to the system and then waiting for them to upload really limits the productivity of our archivists, especially when only working on it for a maximum of two hours a day (the amount of time Ms. Zheng is allotted for her work study). The secondary cause is the severe lack of manpower as Ms. Zheng is the only one uploading files to

MOchive. The reason for this is so that I can devote my time to restructuring the other 15 years of digital photo archives.

While the MOchive delays have caused problems, it was fairly easy to identify a solution. What is needed is an automated system in which the user can simply drop folders into a queue that automatically processes them constantly 24 hours a day. In essence, a queue box with automated coding built into it. This scenario has already been looked into and a software has been identified that does exactly that. StaticSync will simplify the process by allowing for constant uploading to MOchive without the user needing to be present throughout the entire upload. Ideally, the Missourian would have a system in which StaticSync would automatically pull the archives directly from the server when they are fully ready to be uploaded, but due to the status of the archives and the lack of metadata on many files, this system isn't ready for implementation yet. Regardless, StaticSync is a step in the right direction for future semesters. As mentioned before, it could eventually be directly integrated into the archives, but even without direct integration, StaticSync can still be utilized by Assistant Directors of Photography in the Missourian photo department to continue the uploading process on a weekly or monthly basis. It is a slower department integration process than directly linking StaticSync to the archive folder, but this process allows for archival development and accuracy that direct integration wouldn't necessarily permit. Full automation and integration directly to the archive server would be faster, but user error could make such a system become the cause of problems rather than the solution. Such a scenario is a possibility due to the next current process this project has been revolving around: archival structure.

Archive Structure Development

Building an archive is a long, tedious process. I know this because it has been one for me over the past two and a half years. Organizing and consolidating all of the Missourian's visual content into one location and then having to move that entire collection to another server has really engraved on me the importance of quality structuring. Over the past ten weeks, the development of these archives has been slow with quite a few setbacks, and there are still challenges to face. It is important to recognize that many of the files being handled originally came from hundreds and hundreds of CDs that were created back in a time when individual file names were the metadata and descriptive means of understanding the content out of a batch of images. To that extent, this project has proven a great success. With a collective archive ranging from 1999 to the present day, the infrastructure of the system has been established so that every folder reflects the present structure used by the current Missourian photography staff. The beauty of having such continuity amongst file names is that 1) accessibility with the archive itself is simplified for the user (any photographer, editor or ADOP searching for content) to a recognizable form, and 2) when uploading to MOChive in the future, everything will fit the same process without the need for adjustments to the coding structure or confusion or future users searching for content on MOChive.

With every success, though, comes hardships and challenges that had to be overcome. In this case, there were three significant concerns. The first major issue was that renaming folder structures was extremely time consuming. For example, to re-label one student's folders, the process involved going folder by folder and

changing the date at the front, the initials at the back, removing unnecessary underscores and other symbols and then batch renaming the files within that folder to fit the same folder name, each person's taking upwards of 5-10 minutes for the entirety of the process. To put this into perspective, each semester consists of roughly 15 staff photographers that need this correction made. Furthermore, each semester contains roughly 40 volunteer photographers that needed to undergo the same process. After about six weeks of repeating this process over and over, progress was limited to maybe two years of the archive. The solution to this problem was simple when I finally had to start doing things remotely from my PC. Windows systems have software called Bulk Rename Utility that can be downloaded, which allows for batch folder renaming (SEE Reference 1 in Chapter 4). Mac OS has similar software, but due to administrative access on work computers I was not able to test any of them. However, not only does this tool batch rename folders and files, it can change parts of the name like date order and initials without changing the unique sections. Instead of spending weeks on one year, I was able to batch rename six years of archives within a week. The next problem came to light almost instantly after the batch renaming came into play.

From 2008 to 2016, archive structures generally followed the similar fashion of YEAR→TERM→PHOTOGRAPHER→ASSIGNMENT DUMP→image.jpg. However, prior to 2008 folder structures are more sporadic with some photographer folders being split into multiple subfolder structures (SEE References 2-3 in Chapter 4). To understand why this happened, it involves understanding staff management over the past decade. Brian Kratzer, the current Director of Photography for the

Missourian, didn't take the position until around the 2009-2010 years. At that time, the archives really show an increased emphasis in some form of standardized structuring that continues today (with present-day modifications). Before that, the primary form of archiving was through CD preservation, multiple discs pertaining to each individual photographer (SEE Reference 4 in Chapter 4). Not only did these CDs need to be copied and consolidated into the archive, other than being sorted by photographer and month, the CDs contained no other form of structuring for the images besides file names. Through prior-stated processes, this concern was addressed, although it was a very timely process.

The final crucial setback, one that is currently in the process of being addressed, is the event labels within the file names of semesters prior to 2004 (SEE Reference 5 in Chapter 4). Without event labels, each file must be inspected in Photo Mechanic or some other batch editing software to see what the caption offers as far as image descriptions. Then an event label must be added to the file name along with the date, initials and sequence number. This will prove to be the timeliest of all processes as each semester can contain upwards of 500,000 pictures in those particular years. This process may have to be placed on hold due to the lack of manpower able to do such extensive research. The concern here is that by not getting this research done, it requires placing a hold on uploading these years to MOChive. Future research in this area would be ideal to discover if there is a possible way of either automating this process or finding a way to process these files in a batch edit format.

Building Personal Knowledge

Even before the birth of this study, increasing awareness of the need for digital preservation was always the primary goal. Digital preservation can be constituted as the preservation of all forms of media (text, images, graphics, designs, videos, etc.) in an organized digitally managed system. The introduction cited the concept of born digital news and the convention known as Dodging the Memory Hole. This project was meant to branch off of their fundamentals of "saving born digital news" by focusing on a single aspect of media content: photography. On a personal level, this project holds specific meaning to me since my educational and experience background has always been in photojournalism. Not only has the physical component demonstrated the fragility of digital content, but also it has educated me on the full depth of effort that is required to create and maintain a substantial archive.

Building an archive takes plenty of attentiveness and persistence throughout the process, but it also takes patience. While the software limited me physically in the time it took to process transfers, metadata and folder structuring, that time was exponentially increased with the amount of faculty deliberating and process approval involved. This must not be misinterpreted as a negative aspect. Progress throughout the process is important, but this project has emphasized to me the importance of accuracy over haste. Throughout the restructuring process alone, at least ten meetings were necessary to approve various aspects like what to preserve, what resolutions of that file should be preserved, what information is necessary to have embedded, at what process level we should consider keeping a picture, etc.

These meetings and all the consulting was necessary not only to verify what information was needed to be preserved for accuracy sake, but also more importantly to keep everyone in the process engaged in this preservation project. In reflection, the heart of the matter isn't the archive itself but rather maintaining an active and engaged presence to continue the process. The archive cannot survive if there isn't anyone to promote its continued growth and maintenance. If I were to offer one piece of advice to follow this project, it would be to maintain an archive manager role at the Missourian so this project doesn't fade into the background of regular production and out of the realm of important priorities.

As for me, the other primary influence of this project is the need for me to better maintain my own archives. Having learned so much about proper forms of archiving, it only makes sense that I put this knowledge to use and reorganize my own archives so I don't suffer the same fate the Missourian did many years ago when they lost a substantial portion of their archives to hard drive disk failure. I have gained a significant value for the concept of permanence and how easily corruptible a digital file is. (This report was actually set back ten pages just the other day due to a glitch in Microsoft Word's saving.) It is about time the media industry started increasing their recognition of the value of history and how we want to maintain our place in it. Do we as a profession just want to create it or do we want to make an organized effort to protect it?

CHAPTER 4: PHYSICAL EVIDENCE

For this project, the physical evidence is the Columbia Missourian photography archive itself. In chapter 1, I defined my role and involvement during the construction and modification of the archive. The following reference materials are screen captures of software to modify the archives' structure, physical and digital forms of the archive as it existed throughout the duration of my participation in its restructuring and one of the key issues faced during the process. The final reference is the workflow I created and reflects the current layout of the Missourian archival format. The following reference material is included to aid in the understanding of this professional project and any studies produced.
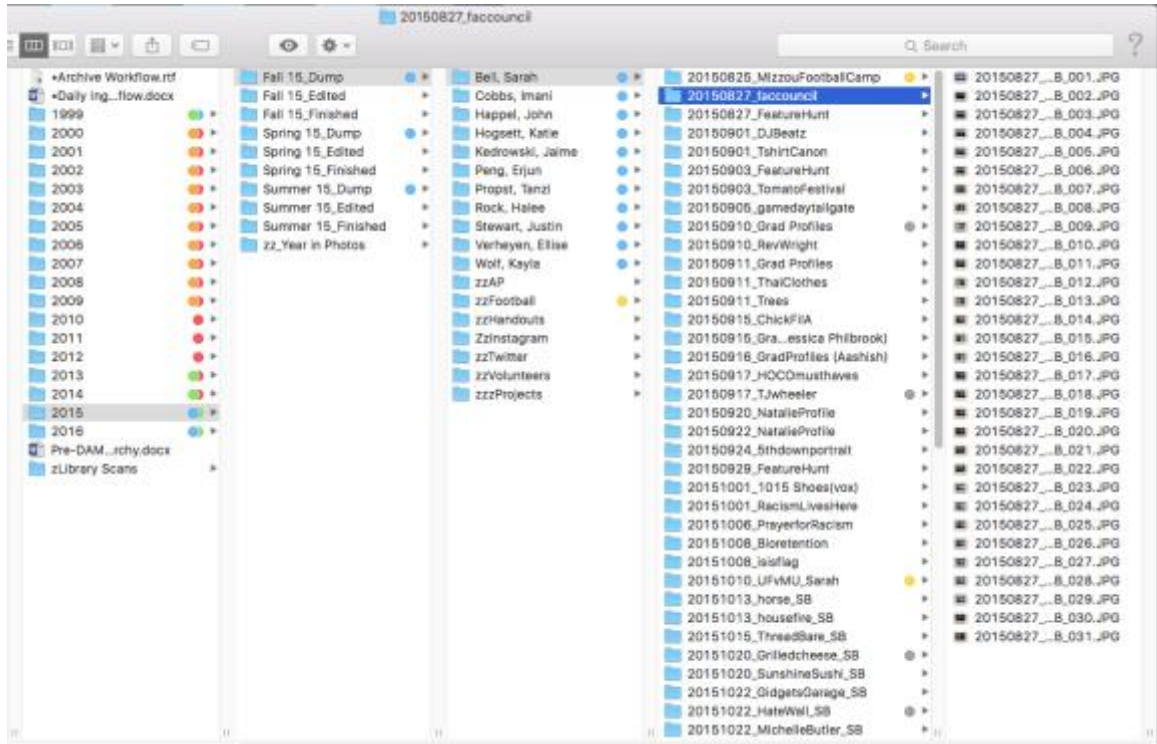
Reference Documents

i. Reference 1: Bulk Rename Utility Software
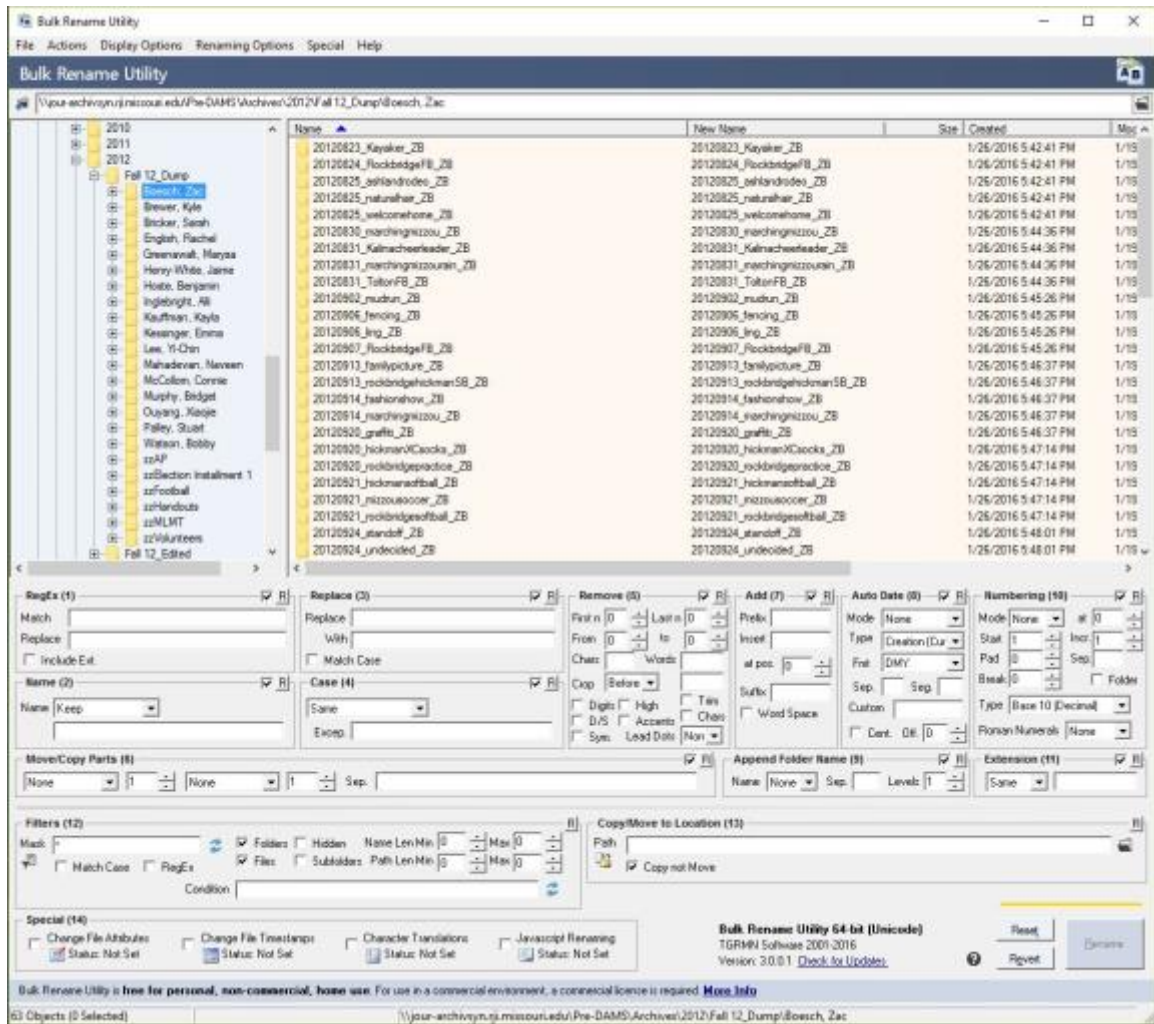
ii.   **Reference 2: Missourian Folder Structure Prior to 2008**
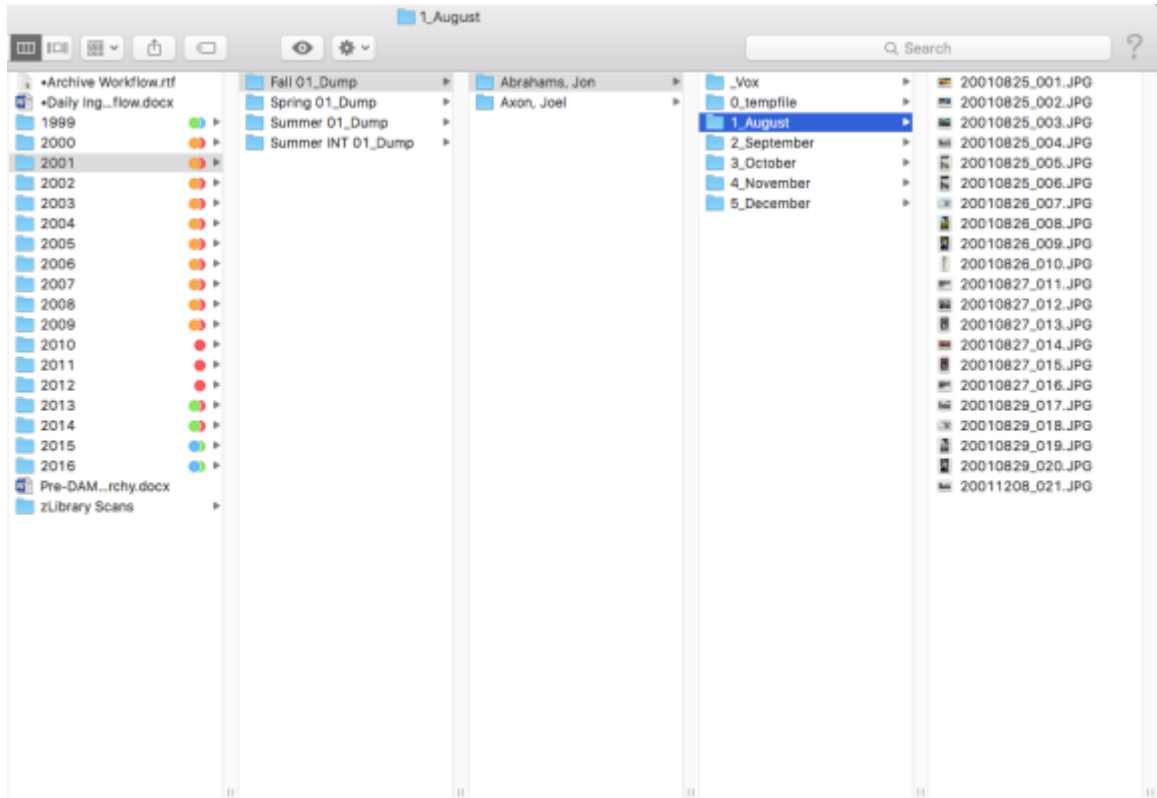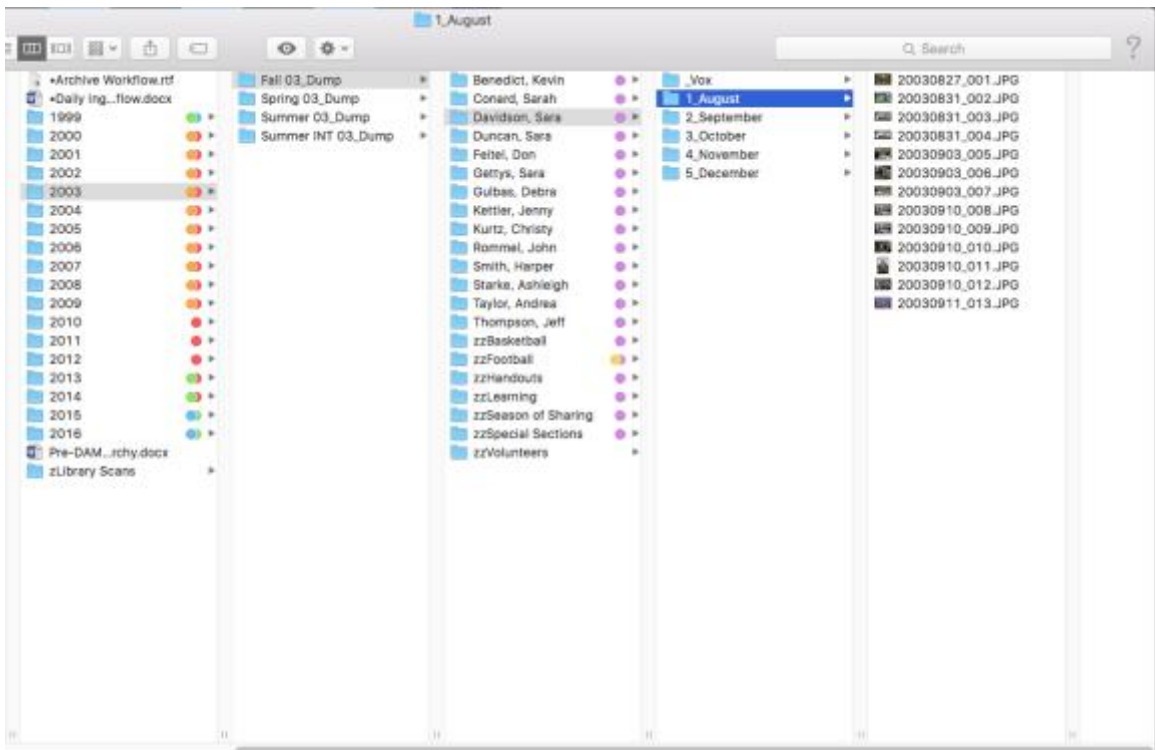

iii.   **Reference 3: Missourian Folder Structure Subsequent to 2008**
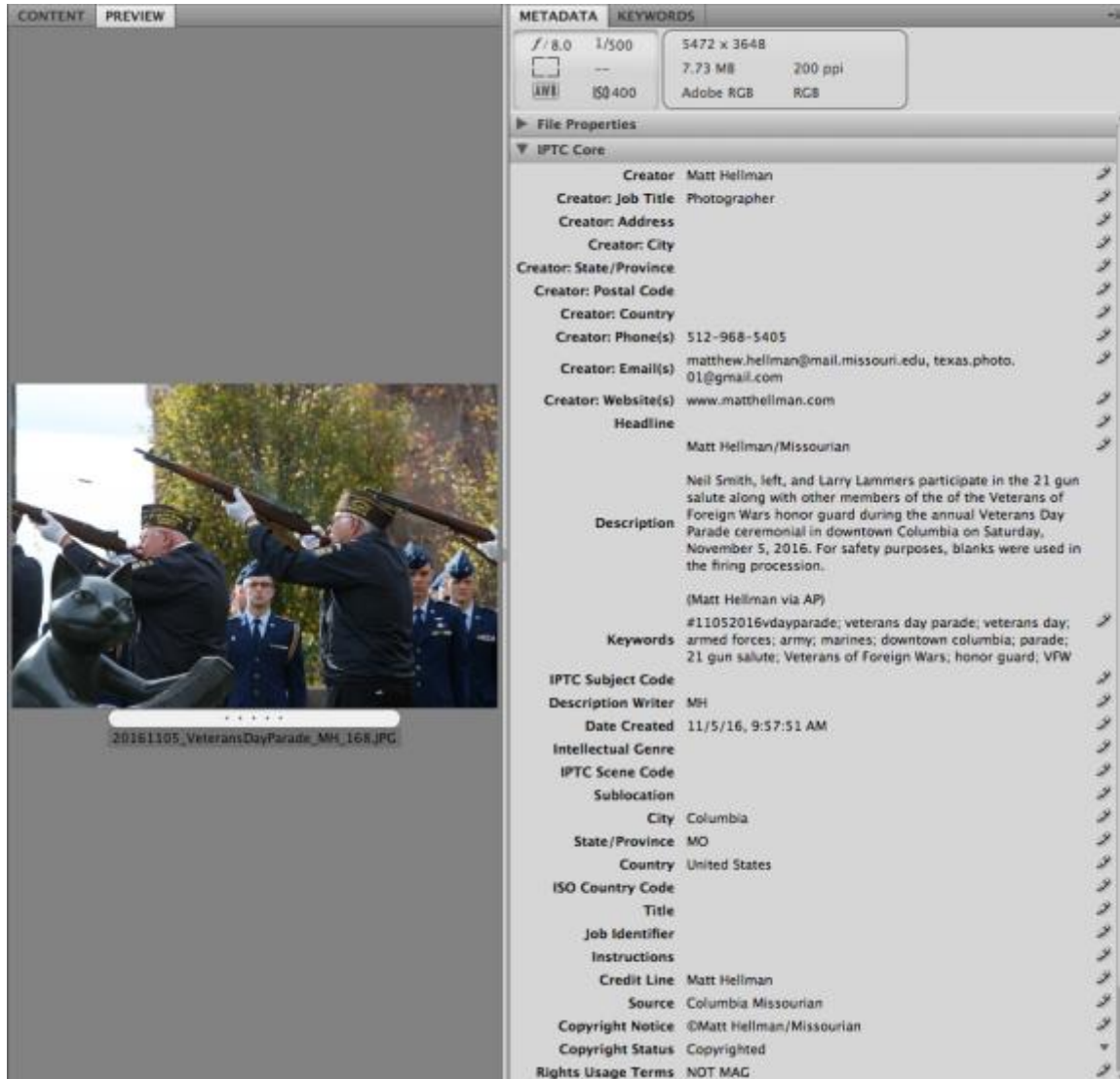
iv. **Reference 4: Missourian CD Archive**

v. **Reference 5: Naming Convention Problem**

vi.   **Reference 6: Current Missourian Folder Structure Workflow**

A	B	C	D	E	F	G	H	I	J	K
					Folder structure model - for StaticSync					
PhotoIngest										
	Semester(Fall, Spring, Summer)_YY_Dump									
		PhotogLast, PhotogFirst								
			YYYYMMDD_slug_PhotogInitials(2 or 3 char)							
				YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG						
				1st Quarter						
				2nd Quarter						
				3rd Quarter						
				4th Quarter						
				Halftime						
				Pregame						
		zzVolunteer								
			PhotogLast, PhotogFirst							
				YYYYMMDD_slug_PhotogInitials(2 or 3 char)						
					YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG					
		zzHandouts								
			YYYYMMDD_slug							
				YYYYMMDD_slug_*PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG						***If Applicable**
				Varies by Semester						
		zzFootball								
			#_MUvs(Opponent)							
				YYYYMMDD_slug_PhotogInitials(2 or 3 char)						
					YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG					
				Selects						
					YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG					
		zzAP								
			File Naming Varies							
		zzInstagram								
			YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG							
		zzTwitter								
			YYYYMMDD_slug_PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG							
		zzArchives								
			YYYYMMDD_slug_*PhotogInitials(2 or 3 chars)_SeqNo(3char).JPG							***If Applicable**

vii. **Reference 7: IPTC File Metadata Example**

viii. **Reference 8: Staff Workflow/Manual**

https://docs.google.com/spreadsheets/d/1mlnuOTilRsI9Z6rtij3ZN07sUev1m

hrJDeqHYYnwNG8/edit - gid=0

CHAPTER 5: PROFESSIONAL ANALYSIS

There is no doubt that photo archiving and preservation is essential. Newspapers and media outlets have long been acknowledged as the first rough draft of history. While there are many different approaches to archiving, both individual photographers and news organizations maintain some collective to protect their visual history.

Such is reflected in the practices of four different professionals and their methods, but for all the differences there are also similarities. Here is some solid advice from these four professionals who have had plenty of experience working with their specific archive projects: preserve the highest quality file available, have good backups, get organized as soon as possible if not before you begin archiving, use your archives for monetization and engaging your audience and viewers, archive everything and maintain good metadata practices. The advice that follows comes from two photographers who primarily deal with individual archives and two others who mainly work with news organization archives. Each one of the four professionals interviewed holds a different background in photojournalism and deals with a different format of archives.

With photographers coming into the profession every day and new technology increasing the amount of images produced, any bit of photography archival advice helps.

Pulitzer prize-winning photographer Jack Dykinga, who is also known for his landscape and nature photography, said his best advice is to just preserve the highest quality of an image possible.

"You can always dumb it down. I can spit out jpegs at will, but why would you want to have something so eliminating, in terms of color range and nuance of colors. I want the original file to be as broad as possible," Dykinga said.

For example, he mentioned that some magazines typically use 300 megabyte TIFF files which are much larger than what a newspaper would generally print. He also mentioned the scenario of publishing an image on a billboard that is 50 feet long, saying that a JPG image wouldn't be able to handle that size and keep its quality at the same time.

"Biggest box of crayons is a way to look at it," Dykinga notes. "To me, you should want infinite versatility."

"JPGs are worthless. What am I going to do with a 1.5 megabyte JPG? I can't modify it. I want a layered file I can always go back to. [Newspaper photographers] are printing on toilet paper. I'm not in that world. You can always dumb it down," he says.

Dykinga won his Pulitzer for feature photography in 1970 while working at the Chicago Sun-Times and has since shifted his focus to landscape and nature photography, producing various photography books and habitually having his works featured in National Geographic and Arizona Highways.

Dykinga also advised on having reliable back-ups for archives. United Press International (UPI) contributing photographer William Greenblatt agrees on this stance, acknowledging his wariness of digital storage.

Greenblatt advises that having an archive he can physically access is a better process than just keeping the digital copy.

"This way… stuff is done. It's there. Nothing is going to disappear. Digital is nothing but air, so I guess going from something physical like film to digital, by putting it on a CD I have it. I have something tangible," Greenblatt said.

Greenblatt notes that hard drives have a tendency to disappear or for things to fail such as files getting corrupted or memory disks crashing. For his personal archive, he maintains quarterly books that include a CD or DVD for each shoot organized by date.

The downside of Greenblatt's archive is the ease of access. While he is able to search through the selects on UPI's database, his physical archives tend to take a lot of time to navigate through.

For example, Greenblatt mentioned that Jack Buck's wife asked for pictures of every famous person Buck was on the field with, and Greenblatt replied that it would be a six-month research project just to find all of those images since he didn't have much of a digital catalog.

"The ideal thing would be to, if I shoot it, to cross reference it for the subject so I could just go to the computer and find it, but I have little need for that. If somebody walks up it just takes a little time to find it," Greenblatt admitted.

Greenblatt's needs are based off of convenience, timeliness and the concept of wanting to hold on to physical, tangible assets. As a regional photographer who doesn't work in a group setting, he is the only one to rely on his archives and therefore is the only one who really needs to understand how they are organized.

"I started shooting digital in '98-'99 so I've got 20 years of negatives, and to start going through negatives and to start putting them by subject I would have to have three people doing it eight hours a day… and I might never need it," he notes.

In contrast, Dykinga sells his work on a regular basis and has value in maintaining a decent, sortable back-up. He is able to sort his images by digital catalogs in Lightroom where he has created all of his databases that link to his two back-up drives.

"I have two file copies. 'Photography' contains all of my digital work and 'Scans' which is both the Django scans and Hasselblad scans. They are in separate Lightroom files. One of each is backed up in the G-RAID via TimeMachine, and then manually backed up into a J-BOT through a terabyte drive that is kept off the computer," Dykinga notes.

Unlike Greenblatt, Dykinga has all his images cross-referenced so he can easily find them through metadata information and dates digitally.

This trend of referenced and cataloged archiving becomes more prominent as the collective gets bigger or more photographers contribute.

Only after a few years did Michelle Jay, the Boston Globe's photo archivist, learn the importance of good preparation and organization.

Jay has been at the Globe for around 2½ years. The project she and her colleague are working on has only been active since she just before she arrived.

"The riskiest thing digitally is we were just scanning stuff into our internal system and not pushing it out to the [company-wide] system for reasons that were not explained to me in my earlier role as the lowest member on the totem pole," Jay says.

Jay recommends getting organized first before beginning an archival project or direction shift while also having a clear mind in terms of subjects so things can be prioritized.

"We initially tried to mitigate some corrections that would need to happen by not taking the risk of putting [the pictures] in, but the risk of not putting images in now is that nothing is in and we have probably 7,000 images that are just sitting around and waiting for me to get to them."

Her advice mainly stems from the Globe's recent decision to relocate from their current building to a few floors in an office building downtown, leaving many questions revolving around what happens to all the content of the physical archive.

"We hit a point where we were like, 'what happens to our archives,' and the project became less about scanning and more about just getting stuff done. Before we sacrificed speed for the sake of file size and things being very complete, and within the last year it has been like scan it and get it in," she says.

The shift in location has caused Jay to adopt a "Don't worry about perfection, just get it done" mentality. Due to a limited new office building, she wonders what happens when they need something from the archive they haven't scanned.

"We were scanning in RAW, and then every iteration was being saved as a JPEG just for space purposes. So we had one RAW and one JPEG. That's what we were doing for close to two years, but since the beginning of 2015 we've been scanning just as JPEG's in the kind of second iteration of this project. I'm technically the only archivist we have (we went from 2 1/2 to 1) so we not only lost people but we lost some of our funding," she notes.

She hopes the digital archiving will help offset the costs of daily production since the Globe has monetized their archives.

"Anywhere where a newspaper can make money is great without having to spend a lot of money," she says. "Monetization is always a good thing in this industry."

Jay believes people are interested in much of the archival content yet to be published. She advocates the importance of engaging readers and audience members in any way that could benefit organization.

"A great thing for me would be to public outreach a little more in an attempt to sell more prints to get that notice out there more and educate Boston. We have a lot of interesting stuff. Everything from like a raccoon jumping out a window to escape a fire to pictures of the great molasses flood of the 1900s," she says.

According to Arizona Daily Star photo editor Rick Wiley, a rich history of photographs is its own reward. His organization was lucky enough to have merged with the Arizona Citizen within the last decade and in doing so received their print and digital archives as well.

"If you have a rich history of photographs at your organization or newspaper, think about making those available to the public in some way," Wiley says.

He creates then-and-now galleries that the Daily Star publishes on their website where photographers go out and shoot locations based off of "really interesting scenes" from 50 or 60 years ago.

"That is source material that would be gathering dust otherwise but is now producing page views and ad revenue for the newspaper," he mentions. "And it's only because somebody, the librarians, had the foresight 50 or 60 years ago to actually archive the material, put captions on it and make it searchable in the old physical archive."

The difference at the Daily Star is that Rick Wiley only preserves the selects that his photographers send in but requires them to have full metadata information before being uploaded to their Merlin platform. Wiley believes in the importance of making, "sure that [the photograph] has all information about the event in each photograph."

While this form of archiving differs from the "save everything" position the other three professionals take, Wiley feels his style fits the needs of his paper.

"You need to be judicious about it and always teach photographers what the mission of newspaper is and help them understand what is important to edit for saving, for future use," Wiley states.

He still agrees with his peers on a point of maintaining an organized and structured archive.

"If I can't find your photographs then you are not doing yourself any favors. If you're not getting published, you're not going to have a job, so make sure you're photographs are able to be found," said Wiley.

All four photographers agreed on the importance of good archiving technique.

"The true value is in maintaining a database and cross-referencing via captions or keywords so you can resurrect anything. If you can file it by date, you can always retrieve," said Jack Dykinga when reflecting on his own organization practices.

"I just keep everything," Greenblatt says. He recalls the days of film when other photographers had boxes of reels that they couldn't recall when or where they were shot and he is thankful he didn't have that problem.

"I've got it in order and I was good about cutting everything up and labeling it," he says.

While their archival practices may stem from different backgrounds, the combination of their advice is meant to be a starting point for any individual photographer or media organization to get a foothold in archival practices.

Learning from our fellow professionals, quality visual archiving and preservation should be of high importance for any visual creator or media organization.

As with any guide, once you have the basic idea, that is the point in which modification and adaptation can flourish based on individual or organizational needs. The fact that so many organizations choose to function without backup

servers, in-depth metadata or a clear and definitive infrastructure in the digital age is only detrimental to the preservation of the first rough draft of history.

"I have millions of images," Greenblatt states. "I keep telling my kids when I take them down into the [archive] room, 'This is your inheritance right here.'"

Interview Transcripts

i.   **Transcript 1: Interview with Jack Dykinga**

How does your archive process function?

> I import [files] directly into Lightroom. I have multiple external G-RAID units
> that currently add up to about 40 terabytes. One is a G-Raid and one is a J-Bot
> and they are redundant.

What is the ultimate goal you have for your archive? Preservation?

Monetization?

> Well monetizing is kind of an interesting word. These days I send images for
> publication to agents, I sell directly to publications and I sell to fine arts and
> European arts. Basically my whole system is predicated on largest highest
> quality files possible and then dumbing it down depending on where it goes.

What image quality do you tend to preserve? RAW? JPG? TIFF?

> I save .PSDs and .TIFFs along with Nikon tiffs.

Do you market your work through a website or contact the organization

directly?

> It's a combination but I happen to be represented in England by the BBC film
> library. German and German fine arts print seller in Germany and a book
> publisher in Germany so I seem to be gravitating towards the European
> market. The domestic one is pretty bad.

Regarding your personal archives, how do you feel about the security and preservation status?

I'm not worried at all actually. I have two files. "Photography" contains all of my digital work and "Scans" which is both the Django scans and Hasselblad scans. They are in separate Lightroom files. One of each is backed up in the G-RAID via TimeMachine, and then manually backed up into a J-BOT through a terabyte drive that is kept off the computer. The interface is usb3 and theta.

Any difficulties during setup of this system?

No. It is pretty cut and dry. To me it is a really good idea to have a mix of a back-up via TimeMachine and then also through your drag and drop on a regular basis to the external. So there are two different techniques for two different locations.

Why backup TIFFs and PSDs instead of jpegs?

JPGs are worthless. What am I going to do with a 1.5megabyte JPG? I can't modify it. I want a layered file I can always go back to. [The newspaper photographers] are printing on toilet paper. I'm not in that world. You can always dumb it down. I can spit out JPGs at will, but why would you want to have something so eliminating, in terms of color range and nuance of colors. I want the original file to be as broad as possible. Biggest box of crayons is a way to look at it.

A typical 4x5 scan comes in at 900 megabytes and your final file comes up to 2 gigabytes. If you were trying to make a 60 x 80-inch print from a JPG, you would be in a little trouble.

Which kind of clients do you have?

When asking about fine arts prints, we are talking about $18,000 a print for large edition, you know that is 120 cm. Those are the big ones and very limited edition. The opposite of that would be web use, which is down to 72 dpi and an 8 x 10 inch size limit maybe smaller. You can designate the output of Lightroom for any way you want to go. Whether you want a color space of RGB or sRGB or 16-bit or 8-bit, it's all convertible. That's why Lightroom is so good. Lightroom does process, but its true value is its ability to become a card file for wherever you go.

What are your opinions on how news organizations operate with their archives?

Generally speaking, the newspaper world is probably the lowest grade quality. Distributing to magazines publishers use 300 megabyte files and they are RGB TIFFs. They are still pretty high quality. It is probably about 100 times better than what you would see in a newspaper.

Do you see yourself adapting to cloud preservation in the future?

That is another fallacy that computer companies are passing around. Cloud is very limited in terms of moving big files. The cloud is good for iPhone photographers.

If you are loading 29 megabyte JPGs then fine but if I get back from my shoot and I've shot files at 120 gigabytes, that's just raw TIFFs. It is a different world really. My philosophy is still valid because I have friends that go the other way. What if I want to make a billboard in Times Square that is 50 feet long? You think a JPG is going to handle that space? No. If it was Lightroom you can designate a file for a specific market. You can specify size, specify color space and spit it out. To me, you should want infinite versatility. News media and print media in general are kind of behind the curve. You got Ken Geiger at National Geographic and he has been instrumental in bringing them up to speed in digital but I'm not sure that newspapers are where they should be but then again it's a fading market so they are probably not going to spend money. They are not where they should be in terms of valuing images. That's an opinion. It's just the nature of the beast.

It seems so silly because Lightroom is perfect. It is a good program for processing. If you are processing you can see 14 stops of latitude with a good Nikon camera, how can you beat that?

What is the basis of your filing system? What metadata fields do you use?

The true value is in maintaining a database and cross referencing via captions or keywords so you can resurrect anything. You can file it by date. You can always retrieve. I file by date and location.

I do pretty in-depth captions. Since I am dealing with natural history, wildlife and landscapes, it is pretty specific. In general, the idea is to start from the wide-general description and then narrow the search by hierarchical

keywords. For instance, you could be in Arizona>Tucson>Sonora

Desert>Santa Catalina Mountains thereby narrowing the search. You could

also include synonyms.

[END TRANSCRIPT 1 INTERVIEW SESSION]

Transcript 2: Interview with William Greenblatt

What do you do as a photojournalist and where?

> Well I am a staff photographer with United Press International (UPI) in St. Louis. I cover all of Missouri and parts of Illinois, Arkansas, Kansas… wherever I need to go really. I've been with UPI since 1980. I worked for the Missourian when I was in the J-school.

Could you provide me with a brief description of the structure of your archive?

> Back in 1968 or 69 when I started doing this, I had to come up with a way of cataloging what I did and I just decided to do everything by date rather than subject because 9 times out of 10 if you know the date of the event then you can find it. So I just decided to do everything by date […] with negatives and now with DVDs I burn a DVD for every event that I do. I'm sure there are better ways, cheaper ways of doing it, but I didn't want to send [pictures] to any sort of cloud or anything like that. That has come up recently as another way of storing and I just wanted to have everything in hand.

> I burn a DVD for every event, label it, put it in a book and I have four books for a year by quarter. The only bad thing is somebody will say, 'I need a picture of so-and-so.' Okay. Where was so-in-so? I need a date where I might have photographed them. I guess I could go through negatives and start cross referencing everything but the amount of time that it would take… I don't think it would be beneficial.

I started shooting digital in 98-99 so I've got 20 years of negatives and to start going through negatives and to start putting them by subject, I would have to have three people doing it 8 hours a day and I might never need it. If somebody wants something from the film days then give me a month and I'll find it. Sometimes it takes that long to find something. But having to look at film on a loop through a light table, even looking for a digital image you have to put the DVD in and start combing through all that stuff and that takes time as well.

The ideal thing would be to, if I shoot it, to cross reference it for the subject so I could just go to the computer and find it, but I have little need for that. If somebody walks up it just takes a little time to find it. For example, after Jack Buck died his wife called me and she goes, 'I'd like a picture of everybody who,' Jack Buck was always on the field with either a movie star or a recording artist or somebody famous, she wanted to put a book together with all those types of people, and I told her, 'you are talking about a six month project to find that type of stuff.' I gave her a few things and that was good enough, but I just do it by date and usually 9 times out of 10 we can find everything that needs to be found.

What kind of metadata do you include on your images?

My photos that go out to UPI, about 3000-3500 per year, all have a caption on them. There's keywords on them, 5-10 for every photo that I send, but personally if I am trying to look something up I will just go to the UPI website and put the subject name in there because usually I transfer something from

everything that I do. Then I can find it that way. Sending to UPI is just like

sending to any kind of wire service: you have to put in keywords. I don't

reference them myself because it's like newspapers or websites when they

are using photos the keywords will be in there for [them].

Alright so you basically use UPI as the CMS navigation for your archives?

Yes because I don't remember. Somebody will as me for a picture of

something and I will say 'I don't know… let me check' so I will go and check

the UPI website and put in the name and notice I've covered them several

times. Then I can go back to that CD and find even more.

What do you consider the benefits of CD preservation versus hard drives?

I make a lot of pictures, so I don't think a hard drive could handle everything,

and this way… stuff is done. It's there. Nothing is going to disappear. Digital is

nothing but air so I guess going from something physical like film to digital,

by putting it on a CD I have it. I have something tangible. I don't know what is

going to happen to a hard drive. Things fail, things disappear. This way I

actually have it, I can put my fingers on it and I don't have to depend on

anything else like that.

On the back side, I have four huge books per year so you have to have a place

to store those things. I have a room that is temperature controlled and it's in

the basement so water won't leak on it and nothing will happen to it. I can

just page through a book and see what I've done. It's right there at my

fingertips. I have millions of images. I keep telling my kids when I take them

down into the room, 'This is your inheritance right here.'

Do you have any goals to ever create a digital collective out of your archives?

No I mean the University of Missouri – St. Louis has expressed interest in my collection. Whether they get it or not I haven't decided. I plan on doing some books but no I think it is going to stay how it is.

Do you have any preservation advice for other photographers?

I'm sure there is stuff out there that I don't know about like cloud and I just... this is a system I came up with myself and I was never really interested in trying to look up other systems. I'm so far into it now that there is no way I'm going to change it.

I just don't know about all the different ways to store stuff. I know there is a lot of other people that just use [the images they need] and then just trash the rest of the images. I keep every image. With film I used to use what I needed and transfer to UPI two frames on either side of the one I wanted then just throw all the rest away. Somewhere in the early 80's I decided to save that stuff. I started keeping everything.

I wish I knew more about updated storage systems but all I know is what I do.

Why did you decide to start keeping the outtakes?

I just thought someday I might need this stuff. I had a boss that said 'You don't need to keep that stuff. The best stuff that is being used is what we are keeping and the rest is nothing,' but I just felt differently. So I just started keeping the rest of the film, putting a date and a subject on it and putting it in

a notebook. It's so much easier just to keep it than have to edit out all of the [bad stuff]. To have to edit it, that is just valuable time that I don't have. And it is digital so what difference is it going to make. Someday someone is going to look at that stuff [...] but it's not going to take any more space on a CD than it would if it was gone so I just keep everything.

As general advice, would you recommend photographers keep everything?

I would keep everything. There is a lot of photographers who did keep everything and never cut it up, labeled it and threw it in a box. A lot of people will tell me they have a huge box of film that they don't know when it was shot or what it is. I don't have that problem. I've got it in order and I was good about cutting everything up and labeling it.

I just wish I knew more about storage systems. Being 62 I could care less. It's just more I have to do. It is just a time thing that it is easier to just burn three CDs and put them away and move on to the next thing. I shoot 3-4 things a day so it's not like I have a lot of free time to do that.

[END TRANSCRIPT 2 INTERVIEW SESSION]

How long have you been with the globe?

I have been here a little over 2 1/2 years.

Was the archive established at that point?

Not really. There had been previous attempts to digitize the Globe's archives before that, years before I came along. My manager... had been working on the archive approximately three days before I started.

Could you provide me with a brief description of the structure of the Globe's archive?

We have a giant library that has [shelves] full of folders of photos. There is no probably no way to put a number on how many are down there. Those are organized by various things. There is a whole section for sports, then we have people folders and subject folders and even a specific, just Massachusetts area. They are alphabetical by whoever was filing that folder just kind of decided what folder that picture was going into. It's an extremely imperfect system down there, but we have been here forever and know where everything is somehow.

Digitally it is slightly complicated. At some point we switched over to Methode before I got here. Methode is our content management system here. For photo purposes we have like a picture dock which does hold a certain amount of time, I think it is like 3 months, and then we have the archive section of it that holds an undetermined amount of space. I'm not sure where

any of it is stored, but I'm pretty sure [it is an off-site server.] That came into

existence way before I got here, but that's like the company wide content

management system and archive. Our Washington DC office could search that

archive and so could we and so could someone that was overseas as well. It is

a company-wide server.

In addition to that, the project that I work specifically on, we have our own

server and our own content management system that is internal just to my

co-worker and me. We created a content management system to basically

record all the information from the print and then output it to be able to

place it in Methode. We go down, grab a folder, and scan it. It saves in

Filemaker Pro. It is an independent program that basically lets you build

content management systems from the ground up. We built our own

database in Filemaker that is independent of the company wide one and so

we input all the information in there like the caption on the back, the date,

photographer, all the relevant stuff. Then we can export the metadata from

there to the digital file that we then put in our company wide system.

What quality of file do you preserve?

Initially we were scanning RAW files. Our document size was 20-25 because

we sell high quality archival prints. We were scanning in RAW, and then

every iteration was being saved as a JPEG just for space purposes. So we had

one RAW and one JPEG. That's what we were doing for close to two years,

but since the beginning of 2015 we've been scanning just as JPEG's in the

kind of second iteration of this project. I'm technically the only archivist we

have (we went from 2 1/2 to 1) so we not only lost people but we lost some of our funding. The editor who proposed our project to our publisher took a buyout and things, I don't want to say crumbled from there but, fell apart and then the Globe is moving buildings in the undetermined future. We are going from our own building that was built specifically for the Globe that houses our archives to two floors in a building in the heart of Boston. We hit a point where we were like, 'what happens to our archives,' and the project became less about scanning and more about just getting stuff done. Before we sacrificed speed for the sake of file size and things being very complete, and within the last year it has been like scan it and get it in. Make sure the info is as correct as you can get it, but don't spend a week researching a folder and photo. If you don't have the information, just move on.

What processing software does the Globe use?

We export the files from Filemaker Pro and then we use code replacement in Photo Mechanic to rewrite all the metadata. I think because the rest of the department uses Photo Mechanic, we try to stay as close to them as we possibly could.

Since your time as the Globe's photo archivist, what sort of difficulties have you encountered during development?

A lot. The information on the back being incomplete or non-existent. For example, the other day it was the 50th anniversary of Bobby Orr's 50th NHL game and our sports department came to us and said, "Hey we have a photo that ran in our paper. It's a great photo. Can you find it?" and my answer was

no because I'd gotten a request for Bobby Orr stuff early so I already knew what was in our print library and a lot of pictures down there just had a stamp with the photographer's name. So it was like no date, and you could gather some information, but with no other information we have this stack of prints now that are just meaningless basically. That's the biggest thing we faced.

Also storage space, digitally and physically. The room I'm sitting in currently is a mess because we have probably 16 boxes of negatives back here and one table is just full of prints that I am not entirely sure what they are. Digitally there is only a finite amount of space we have and our server that we had allotted was very slow so we switched off of the server and went to external hard drives. It was an uphill battle there.

And then a lot of conflicting or verifying information from 20 years ago can be tough.

What were some major successes in the archiving process?

Probably the ease of transferring data from the Filemaker Pro server onto the files and then being able to distribute them out. The globe has a partnership with Getty Images where we are considered a partner paper so everything of the Globe staff we can put on Getty Images to license and our archive does go up there. Looking at our Getty sales numbers, we do sell a decent amount of archive photos of random things, even things you wouldn't expect we would sell. The Getty Partnership would probably be one of the

biggest successes, and also how we can transfer all the information to disseminate it quickly.

What is the fundamental goal of the Globe's archives? Preservation? Monetization?

Both really. Anywhere where a newspaper can make money is great without having to spend a lot of money. Monetization is always a good thing in this industry. Also just preservation. At least once a week an editor will be like, 'hey do we have this' and when we do it is awesome Typically once a week I am in the library looking for something and I'll just scan the rest of the folder because there is probably a good chance that someone else is going to come looking for it in a month or two.

What are some of the more risky endeavors the Globe has taken with their archives?

Well an old photo technique was literally painting on photos to silhouette things or dodge and burn in very strange ways. Not knowing how to get any of that stuff off because sometimes we need the print behind it too so we have taken water bottles or cotton balls that are wet to photos. One time we used silly putty to get something off. We had this spray stuff that at one point fogged up the whole office so that we had to leave for the day. Probably the riskiest things we do involve actually handling the print. Typically though we take a scan of the print before we actually touch it at our usual specs so in case we do ruin the photo we at least have a record of it for later.

The riskiest thing digitally is we were just scanning stuff into our internal

system and not pushing it out to the globe system for reasons that were not

explained to me in my earlier role as the lowest member on the totem pole.

We initially tried to mitigate some of the corrections that would need to

happen by not taking the risk of putting it in but the risk of not putting it in

now is that nothing is in and we have probably 7000 images that are just

sitting around and waiting for me to get to them.

What do you see as potential goals for the archive?

Our grand plan would be to get as much of the stuff we could possibly get in

or get at least digitized before we move as possible. We do have an Instagram

account just for our archival project. We modeled it off of the Chicago

Tribune's Instagram. It was wildly popular when I had time to manage it, but

a great thing for me would be to public outreach a little more in an attempt to

sell more prints to get that notice out there more and educate Boston. We

have a lot of interesting stuff. Everything from like a raccoon jumping out a

window to escape a fire to pictures of the great molasses flood of the 1900s.

What sort of questions or concerns are you looking to address?

The big thing is that we are moving buildings and our space is shrinking. The

big question is what will happen if we need something from our photo

archive when we move and we haven't scanned it in? That is pretty much

constantly on my mind and I would say that's probably the biggest challenge

right now. I think there was talk with the university but that was the last

thing I heard.

Is there anything else you would like to add?

We don't deal with negatives, we only deal with prints. We consider our negatives our outtakes folder and so we don't scan any of them unless we absolutely need something that we cannot find in our print library and we consider our print library our selects folder. I hate when they make me go look for negatives. I probably do it once a month and try to get out of it as much as I can. The number of boxes we have is astronomically high and extremely scary to look at. That is the one thing we haven't addressed and probably never will address is, 'what do we do with our negatives and how do we deal with them in terms of our outtakes?'

Also we are missing a lot of stuff that has gotten stolen, thrown out, misplaced or never returned. We will lock up a lot of stuff that people would walk away with. It is prints and topics that are extremely important. The breadth of what we are missing is fairly high.

Can you provide three pieces of archiving advice for other archival projects?

1) Get organized first and then start your project.

2) Don't worry about perfection, just get it done.

3) Have a clear mind in terms of subjects so that you prioritize.

[END TRANSCRIPT 3 INTERVIEW SESSION]

Could you provide me with a brief description of the structure of the Daily Star's archive?

It is actually in two pieces. The digital archive is a Merlin archive that is hosted by Merlin in Cambridge, Mass. We FTP everything to them and then we access it over the Internet. The newspaper went 100% digital in 1997 or 1998, and the newspaper they were still shooting film but they were scanning photos and saving those into Merlin. And they still have the negatives too. So we kind of have a digital archive back to the mid-90s for argument sake. Also when the Tucson Citizen closed, we inherited their digital archive as well as their physical archives. So those photographs are also transferred into our Merlin archive and those date back to maybe the late 90's.

The second part of the archive is the old physical archive which houses is the Daily Star negatives and prints that date back to the early 70's. They never had a professional hard or paid photo staff until the early 70's. Before that they would hire people under contract or freelancers. So we have physical prints and negatives from that time and then the Tucson Citizen's physical archive and their stuff goes back a little bit farther. We have photos from the Citizen that date back to the 1950's and 60's both print and negatives and we frequently, depending if we have a story about somebody who has died and we're doing an obit we might go back and look for their photographs if they

are a well-known person from Tucson history. Frequently we will find that in the Daily Star or Citizen archives. At that point we scan that and move it into Merlin so it is in the digital archives.

About 4-5 years ago we inherited and using the Citizen archive we actually started actively re-editing negatives and prints and scanning them into Merlin. If it was topical or somebody was doing story about something that has some historic connection or maybe I will do a Throwback Thursday gallery where I will just stumble across something that is interesting and I will re-edit the negatives and scan maybe a half dozen or ten and make a gallery for our website. And all those photos go into our website as well and into our content management system for print and web so they are in two places.

When tapping into the old archives, it's pretty much just for if it is something we think is topical. For example, last year was the 100th anniversary of the University of Arizona homecoming. I went back and pulled photographs from every homecoming back as far as I could which was around the mid-1950's and then before that I work with the University of Arizona special collections library and accessed their old photographs through They have all their yearbooks scanned as PDFs and online so a lot of that I got permission to extract photos off of the yearbook. So between special collections and the Star and Citizen's archives I was able to put together an image for each of the 100 years of U of A homecoming.

What is the daily process the Star uses? What quality of file do you preserve?

The photographers for the most part edit their own photographs and then they send them to the newspaper and they frequently depending on the assignment when they are going to need a lot of photographs for future use and when they don't need very many. When there is a project, often times I become involved in the editing process, but pretty much they edit or anything that we tend to publish will go into the Merlin archive at which depending on the situation (say they have 10 photos of an assignment) we may publish and print one or two. The other photographs may end up as a photo gallery online or we could end up using some of the other outtakes with other stories down the road a year later or so it a notable person or politician.

So the Star doesn't save the full outtakes, just the selects?

Right. That's my philosophy. That goes back to me when I worked at the East Star Tribune in Mesa, we started using digital cameras and storage was real hard to come by in the mid-90s we made a decision then that there's no reason to save every single photograph. There is no value in it because a lot of photos are out of focus or they are over exposed... it's just a waste of disk space so we didn't see any value in it at all. The photographers are good at spotting things that might have publication value in the future so they will put them in Merlin and so we only have a good quarter of cropped, toned,

edited photographs from which to choose and I don't have to get involved in that.

What software does the Star use to process photographs?

We use Photo Mechanic for ingesting and editing and captioning quite often. We also use Photo Mechanic to FTP images into Merlin from wherever they are. The photographers will pull them into Photoshop and after editing put them into a Transmit folder and use Photo Mechanic to FTP them back to our archive. That is pretty much it for our workflow.

What metadata is included in the processing phase?

Our website has three fields that it needs: a headline field, a caption field and a byline field so we make sure those fields are populated at all times. We don't usually use keywords unless it's an operational function that does something in our web system. For example, all photographs that go into our web system default to a 'Buy this' sort of options so you can buy the photo as a photo reprint. You can click on it and order an 8x10 for personal use. Obviously if we don't own the photograph, we can't sell it so we have a keyword called #nosale that we put on handout photos. Or a photographer right now we are doing a photo story on a WWII veteran, we may copy some of the veteran's personal photographs and on those photographs that they edit they'll put a keyword #nosale that turns off that [purchasing] option online. So basically the only reason we use keywords is for stuff like that. We don't use them for search function. In the Merlin archive everything is indexed so we can find anything we need without keywords.

What do you see as potential goals for the archive like expansion or modification?

Not right now. Our contract with Merlin, it's maybe a couple thousand dollars a month, they upgrade the archive and upgrade the storage every year. They also make necessary upgrades as needed. We benefit from all the changes that they are doing because it is part of our contract. As far as any other expansion plans, we really don't have any because we don't have as many photographers as we did 5 or 6 years ago (we used to have 11 and now we are down to 5) so we are not saving as many images as we used to.

There is no physical need for more storage space per se. The photographers are producing more images but not as many as the staff in the past, but we are not saving video to Merlin. We tried that for a while, we had to pay an extra fee for a month, and they decided it wasn't working out and we stopped doing it. Right now we are storing video basically through Youtube and through our content management system online.

Eventually we will move towards a digital only/digital-first platform, well I'm almost halfway there. We looking to move photos directly from the camera to the archive. [...] We are looking to move stuff faster through the system. That doesn't change the archiving functionality basically because they are still going in the archive and we can still access them through the archive. But this is more of a speed issue of getting the images from the camera to the reader.

Is there anything else you would like to add?

I'm really happy with Merlin. They are incredibly responsive and their IT has been fantastic. We've had very little down time with them and no slowdowns. It is very robust and secure.

For anybody starting an archive system, all I can say is do it because most newspapers are a historical record for the community and I think we've lost sight of that as we've, well we laid off our entire library staff about five or six years ago, so Merlin archive is the defacto visual archive now. There is nobody maintaining it but me and Merlin so Merlin is a librarian unto itself. It is extremely valuable because we can draw on it. We have the advantage over say TV stations of having this rich visual archive that they don't have and so when somebody dies or there's an event that happens that requires some historical significance, then we can draw from what we have and really show people 20 or 30 years ago and give them some perspective on things. That is immeasurable. That is something that nobody else in our market can offer. It is worth the money and the time that we spend to maintain this archive, to offer it to the readers, because it back in page views and ad revenue.

Can you provide three pieces of archiving advice for other archival projects?

1) From my personal experience being an editor for 20+ years, photographers don't need to archive everything you shoot. You are wasting disk space. I think you need to be judicious about it and always teach photographers what the mission of newspaper is and help them understand

what is important to edit for saving, for future use. Once that happens, once

they get into these best practices, then it's almost an automated system. The

servers know exactly what to do, they move the photographs directly into the

archive system and they become available for everybody to see. Those

photographs are vetted. In other words, they've been toned, we're proud to

publish them, they're in focus, they are good looking photographs and we

don't have somebody with no knowledge of photography picking a

photograph that we wouldn't want to see in the paper. So I think that is

important that you edit before they go into the archive.

2) Make sure that [the photograph] has all information about the event in

each photograph. Four of us covered two high school playoff games last

night. In every caption they write, "XX vs XXXX high school on,' and the date,

'at XXXX high school in Tucson, AZ. XX won 68-43,' and that is in every

photograph. They are not going to get lost in the netherworld. They can be

grouped together. [...] Everything is just easy to find. That is an important

procedural thing that all photographers need to know. If I can't find your

photographs then you are not doing yourself any favors. If you're not getting

published, you're not going to have a job, so make sure you're photographs

are able to be found.

3) If you have a rich history of photographs at your organization or

newspaper, think about making those available to the public in some way.

Like we do through me creating galleries for them to look at or using those

old photographs as source material for, we're doing like a then-and-now

69

gallery. We will find really interesting scenes from 50 or 60 years ago and then we go back and photograph what's there now. We run those galleries online and they are immensely popular. Readers just love them. So that is source material that would be gathering dust otherwise but is now producing page views and ad revenue for the newspaper so it turns out they are very valuable and it's only because somebody, the librarians, had the foresight 50 or 60 years ago to actually archive the material, to put captions on it, to make it searchable in the old physical archive where I can find it so we are benefiting from it today.

[END TRANSCRIPT 4 INTERVIEW SESSION]

Project Proposal

FILLING IN THE GAPS:
AN EXAMINATION OF THE IMPLEMENTATION OF ARCHIVING
AND METADATA PROCESSES IN A DAILY NEWSROOM

A Master's Project Proposal
Presented to
The Faculty of the Graduate School
University of Missouri – Columbia

By
MATTHEW HELLMAN

Dr. Keith Greenwood, Committee Chair & Project Supervisor
Brian Kratzer, Committee Member
David Rees, Committee Member
Edward McCain

TABLE OF CONTENTS

PREFACE: ACCESS & PERMANENCE

"There has always been a trade-off between access and permanence. When we first started the graphic record, we could pound a rock against a rock and make a picture, and that picture would last a long time, but it was not very accessible. You would either have to carry the rock around and show people or they had to come and see the rock. Then we (humanity) moved to clay and tablets. These objects were more portable but also more easily broken.

Eventually we moved to paper, which we learned could last for hundreds or thousands of years if you keep it in the right environment. But it burns, can be eaten by insects, can be torn up... Today, technology has brought us into the digital realm. This digital stuff is so new that we are just beginning to learn how to grapple with preserving it. One fragmented file or disc on a hard drive and we lose quite a lot of content. On the other hand, access is fantastic because we can go all over the world with it...

...But we are on the opposite end of the rock."

— Edward McCain, MU Digital Curator of Journalism

INTRODUCTION

Background

In 2011, an organization known as the Newspaper Archive Summit hosted a convention that gathered archivists, librarians and key newspaper members to discuss the "preservation of newspaper content for future generations" (Carner et al, 2014, p. 11). Their goal: to create a policy for maintaining and protecting historical information that cannot be replaced. A study funded by the Donald W. Reynolds Journalism Institute (RJI) in 2014 sought to identify how digital-only newspapers dealt with the issue of archiving in today's emerging media market. This survey specifically looked at two kinds of newspaper models – the legacy (print) papers that also published online, which we called "hybrid" and the newer model "online only" publications.

What did they find?

Modern archiving is a huge cost and huge time commitment, and most news organizations have either limited or drastically reduced the people that they have working in their news library (Hedstrom, 1997). By archiving, I am referring to the process of preserving "unique content" in a manner that is organized and accessible (Schmidt, 2009). Adding to this limited librarian role, the growing issue for most news organizations is that more of their news content is being published electronically, and only electronically.

Over the past four years, I have had the opportunity to explore how different newsrooms manage and maintain their photography archiving systems. Beginning

at the Baylor Lariat newspaper in Waco, Texas, my understanding of archiving and digital content management originated when we had to undergo upgrades due to lack of server space for our visual content. As modern technology progresses, so does the amount of storage space necessary to manage the increasing quantity and quality of content. This was my first stint with local, in-house servers and the limited use they served. The beauty of the Lariat was that they really prepared me for what came next: the Columbia Missourian and their disarray of storage units they called an archive.

Before I began my master's program back in 2013, Edward McCain, digital curator of journalism at the Donald W. Reynolds Journalism Institute and MU Libraries, began to direct his focus towards developing a program for the Missourian to preserve their digital content. He brought this conversation to Tom Warhover, the Columbia Missourian's managing editor, and Missourian Director of Photography Brian Kratzer. Together the three began to discuss plans for digital preservation and visual content management. In August 2014, in what happened to be twist of fate, I arrived and found the Missourian archives in disarray. Soon enough I made it my personal mission to fix them. The Missourian newspaper, when I arrived, had an unorganized archiving system consisting of 12 separate almost-decade-old hard drives, hundreds of CDs and three local servers that weren't organized in any manageable fashion. At the Lariat, everything was consolidated and organized on a single server with an established workflow. The Missourian presented me with the opportunity to assist in building an infrastructure they could more easily sift through and find the content they needed, and so it began.

I began in Fall 2014 by trying to get a grasp on where all the Missourian content was and where it needed to be. After consolidating the various hard drives and merging the CD files into a single location, I began a reorganization process that classified the content first by year, then by semester and finally by person and event. This process lasted through the fall semester and into Spring 2015, where I began to assess the metadata within these archived files. With direction from Edward McCain Brian Kratzer, we began to develop a workflow and metadata template that could be incorporated into the staff photojournalism class in order to prevent future content from being disorganized. Using the same workflow and metadata style, I also began to correct older content files to fit the same mold. The back half of Spring 2015 came with a whole new obstacle. The Missourian switched from an in-house server to a new campus-housed server that was larger and contained greater redundancy and a constant monitoring staff. It became my responsibility to make sure all of our old content was migrated to this new server.

Fall 2015 had me focusing on getting the servers transferred over. By the end of the semester, all digital photography content was consolidated on the new server known as Pre-DAMS.

Fast forward to present day: the Missourian currently uses none of the old servers, hard drives or CD's but now operates off of a single consolidated archive that is organized by dates, semesters, months, weeks and days, with each file containing proper metadata and caption information. The Missourian is also now evolving to a new server that involves everything being stored on a local RAID device and also backed up within Amazon Glacier's long-term storage services that

will be searchable through a ResourceSpace digital asset management software nicknamed MOchive. In essence, it has been a long process to get the Missourian from an almost non-existent archiving format to a digitally consolidated, managed and searchable database that increases production efficiency, availability and economical value for the news organization.

My most recent work has focused on evaluating the best resizing software before we begin uploading the vast amount of digital content to MOchive and creating an resizing and upload workflow that can be implemented to get content prepared and into cloud storage. With a series of image resizing trials at various sizes and quality levels, it has been determined that the best software to use for this process was Photo Mechanic, a batch photo asset management software that allows for mass file processing at once.

Through these experiences, I have begun to learn how to identify weak points in archival systems and improve the existing assets to adapt better methods of finding the content news organizations need at a faster and more reliable rate. This knowledge has directed me to pursue a project that seeks to demonstrate the development and implementation of metadata and content management techniques in a newsroom, a workflow that is becoming more sorely needed than ever with the rapid growth of technology over the last few decades. This project will provide me with a valuable foundation of professional analysis that I can base my experience emphasis off of. It will also provide an example of the skills I possess and could potentially provide to future employers in regards to their archiving and digital asset management scenarios. For this study, digital asset/content management

refs to the role of maintaining and correcting workflow errors such as metadata and file information as well as any consolidation processes that may be necessary.

Although this project provides me with a better understanding of digital asset management and preservation, it is important to understand that those skills and recognition gained are not the reason I chose to do it. From the beginning, I have always had an affinity for being organized. My personal archives are redundant and also backed up on cloud storage themselves. I understood the stress of losing a lot of digital content early in my experience when I lost my laptop during a car accident. Some important things can't be replaced and a simple small accident could be devastating to digital files that aren't properly preserved. Edward McCain's software has great potential for the media industry. To see it grow and succeed, it has to begin by being tested and implemented successfully. That task alone requires diligence, cooperation and most importantly collective backing and effort from everyone in the department in order to prosper. That is what drives me to do this project. I want to see a future where at least Missourian photography preservation isn't a concern or an obligation, but rather a regularly incorporated and significant piece of the daily workflow.

Henceforth, the purpose of this project is to identify the aspects of news publication archiving that rely solely on the visual photography content produced at small and large media organizations and determine the feasibility and effectiveness of digital cloud archiving versus local server storage archiving when scaled to a larger archive systems, such as that of large regional and national newspapers. This project will analyze the implementation process of the MOchive system through a

case study with the Columbia Missourian photography department and archives in

order to examine the success and viability of cloud-based storage systems and

management software in the news organization's daily production workflow.

PROFESSIONAL SKILLS COMPONENT

Physical Component & Qualifications

As this project focuses on studying the implementation process of MOchive within the Columbia Missourian newspaper, the physical contributions I plan to incorporate must reflect development within the archive itself.

My qualifications for this project include the prior two years of assisting with the development of the Missourian archive infrastructure to the present state of consolidation and appropriation. Under the direction of Missourian Director of Photography Brian Kratzer and Edward McCain, digital curator of journalism at the Donald W. Reynolds Journalism Institute and MU Libraries, I have been able to assist in the development and implementation of a workflow for the continued maintenance of the Missourian photography archives and for future archival incorporation of content preceding the previous 15 years. Furthermore, I have previously developed improved workflow techniques for organizations including the Baylor Lariat newspaper, multiple individual business archives and the web-based archiving format for www.chisholmcrossing.com, a Waco-based digital downtown information source for the local and greater McLennan community.

My work aspects will include editing image metadata spanning over the entire decade of collective Missourian photographer works that have already been digitized and consolidated into the archive. Weekly, this project will consist of full-time (40 hours per week) digital asset management practices including a top-down system of migrating the most recent content to MOchive and working backwards

until the entire span through 1999 is in the DAMS system. The process will also include updating and incorporating metadata where necessary to maintain consistency of archives. As this project will be my primary focus, I will spend beginning few weeks (2-3 depending staff retention) with ingest/metadata training for the new staff members. This process will consist of weekly routine check-ups to ensure accuracy in their metadata files. Further checkups will be established throughout the semester (on a quarterly basis, approximately every four weeks) in order to maintain consistency and continuity. I will also continue editing metadata in the older archives (beginning with 2014 and working backwards) by filling in missing keywords, captions and photographer information where it is lacking, and I will continue implementing and facilitating the migration of the Missourian archives to the MOchive system and Amazon Glacier long-term storage service. Finally, it will be my commission to continue promoting and championing the incorporation of MOchive into the daily workflow of the photography department as the primary archive-searching and reference tool. The implementation phase of MOchive will depend on how quickly content can be uploaded. Early expectations can place initial implementation procedures at early October. By initial implementation procedures, I am referring to introducing the staff editors and assistant directors of photography to MOchive and how the digital asset management system works. The timeframe for the extent project can be set at approximately five months ranging from August 2016 to December 2016, pending MOchive and maintenance/coding development. A visual layout of the Project timeline is illustrated in Appendix A.

Director of Photography Brian Kratzer and Edward McCain will maintain the supervisory role of this project. McCain, while not an official graduate faculty committee member, will fill in informal role as co-supervisor as he is the developer of the MOchive software this project is based off of. Field notes will be shared with project committee members by both email and in person for assessment. These notes will include weekly progress in f metadata editing of the old archives as well as observations of staff retention, and problems and resolutions, any MOchive testing and development results and the overall progress of the project timeline or any setbacks that may cause delay or alteration to said timeline. Corrective measures can be put into place on a case-by-case basis. I will maintain constant communication with Edward McCain and Brian Kratzer as to the status of the Pre-DAMS and PhotoIngest server space and any storage concerns that may present themselves if necessary.

It can be expected that the database built on MOchive as well as the local server can serve as the "abundant physical evidence" required of this professional project. Finally, the dissemination of this project would be best formatted as a report published through the Reynolds Journalism Institute research website and incorporated in next collaborative of Dodging the Memory Hole, an outreach collaborative dedicated to preserving historical news content in modern digital formats.

SETTING AN ANALYTICAL FOUNDATION

Defining the Topic

As established in the preface, preservation is becoming a greater concern as the journalism and news industry adapt to the plethora of modern technological advancements available. The topic of this paper focuses on how news organizations can utilize cloud-based storage opportunities to improve photography preservation techniques while also minimizing the costs required in maintaining such visual archives. Optimistically, this project will serve as an example or model for other organizations to build off of the structure being established by the Columbia Missourian in cooperation with the Reynolds Journalism Institute and University of Missouri.

Research Questions

Since this project seeks to identify a method for integrating a digital-based visual archive preservation system in a daily functioning newsroom, the following research question will be addressed:

RQ1: What is the most productive and efficient implementation method (characteristics supported by the diffusion of innovations theory) for establishing a cloud-based digital asset management system workflow to ensure survival of a newsroom (organization unit) visual archive system without limiting preservation or access?

This research question creates a basis to identify and address factors that could influence the implementation of digital archiving technologies within news organizations. RQ1 is significant because, as more and more born digital photography is created, more storage space is required for that content. Backing up locally is susceptible to data loss or degradation. Sure, redundancy can be put into place, but preserving it in a cloud-based system would allow for the entirety of that content to secure across a network that is naturally redundant. News organizations face the same risks since and many don't have a consolidated local server with built-in redundancy software that they can rely on or a process for integrating web based preservation options. One outcome of this project would seek to evaluate the workflow method for integrating a visual archive preservation system in a daily functioning newsroom.

Identifying and understanding the implementation process provides a foundation for other organizations to base their future archiving techniques off of. Through the process of implementation, considerations including preparation time, monetary expenses and staff development can be analyzed to help create a workflow that could benefit other organizations in their endeavors to build off of this system's success.

Research Structure

This study deals with the justification of implementing digital archiving techniques into the workflow of visual newsrooms. An overview of the research subject is presented in the theoretical framework section. It is important to establish

the research areas reflected in this study. Primarily in the realm of digital

preservation exploration, this study also reflections on innovation implementation

in an organization environment.

Based off of the research areas, the remainder of this chapter focuses on

defining the theoretical framework used to qualify this study and an analysis of

prior literature and its association in the context of digital preservation

implementation. The final section proceeds by addressing the research methodology

including the process of professional analysis and validity concerns.

Establishing a Theoretical Framework

In this section, a theoretical framework is established through the analysis of

former literature pertinent to the concepts of digital archiving and curation. Many

studies have been recognized involving the implementation of digital technologies

within newsrooms along the model provided by the diffusion of innovations theory.

With the advance of the digital era, innovative technologies including web-based

platforms could present an answer to the long held conflicting concepts of access

versus permanence.

Diffusion of Innovations theory can be traced to its roots in communication

studies. Everett Rogers first published on the theory in 1962 and the book is now on

its 5th edition. Rogers defines diffusion as "the process in which an innovation is

communicated through certain channels over time among the members of a social

system" (2003, 5). Take, for example, the integration of convergence practices

between major news outlets such as the Dallas Morning News and TXCN (Singer,

2004). In Singer's (2004) study, she analyzed how the Dallas Morning News merged convergence practices including video and broadcast production with other news organizations like TXCN in order for both groups to benefit from the news value of the content. Before proceeding though, we must first understand how Rogers interpreted innovation as well. According to Rogers, "an innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (2003, 12). In the previous example, the innovation would be convergence practices.

Generally, the bulk of Rogers' theory research was directed towards *the individual* and his/her adoption of a new innovation. Due to the modernization and incorporation of technology, there has been a shift in research trends involving the diffusion theory. Lately, the amount of research involving a *unit* rather than an *individual* has increased. The reference of unit generally applies to Rogers' approach to an organization as a whole in his theory. Examples of this can be seen in numerous studies by Bruce Garrison involving the diffusion of online research in newsrooms (2000, 2001). While his studies proceeded to last over long-term periods spanning years, the result still allowed for his findings to reveal that, although newsrooms involve a certain level of complexity (especially amongst larger newsrooms), "growth occurred earliest and fastest at larger daily newspapers" with the financial resources, individuals and staff that supported the innovation (Garrison, 2000). Garrison's research would seem to prove the diffusion a success in specific contexts, particularly those where the organization has the infrastructure and resources that would support innovative processes. Still, existing

literature has yet to demonstrate any semblance of focus as to where news

members and organizations stand in the implementation portion (rather than

adoption) of the diffusion theory.

Figure 1: Five Stages in the Innovation Process in Organizations
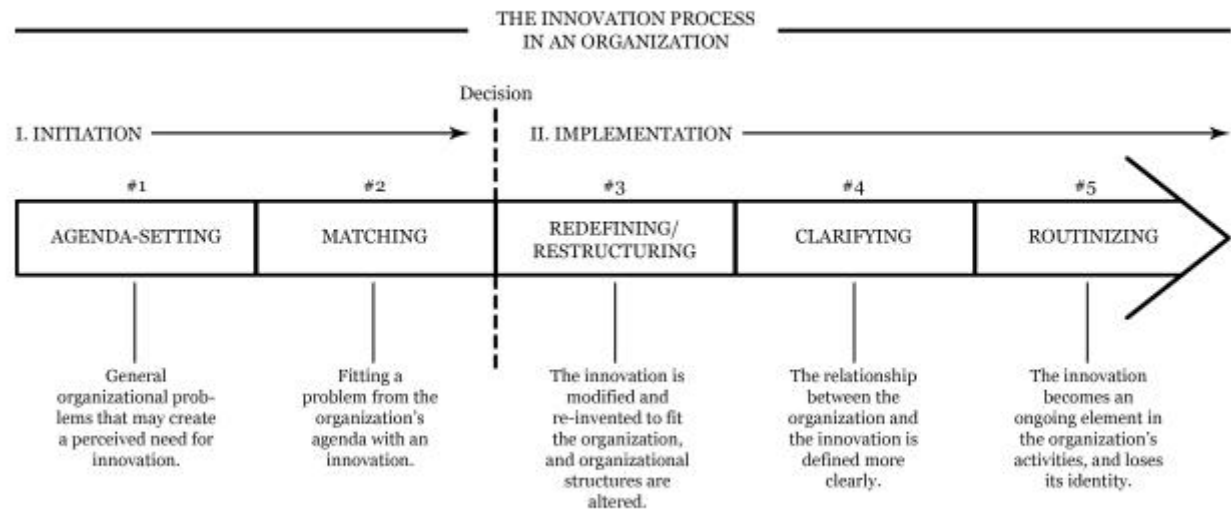
Figure 1: Stages of Innovation Process in Organizations (Rogers, 2003).

As illustrated in Figure 1, Rogers breaks down organization innovation into

two separate sections: initiation (generally referred to as adoption) and

implementation (2003). The initiation section, consisting of two stages (agenda-

setting and matching), entails all of the preplanning and information gathering that

builds up to an organization adopting an innovation. The second section,

implementation, is divided into three stages which can also be considered the

primary variables to monitor for this project's case study:

Redefining/Restructuring – Here the innovation or new technology is

adjusted to fit the parameters of the organization. However, the organization

must change as well to be suitable for the innovation.

87

Clarifying – After the innovation is adjusted into the organization, in must be put to use frequently and in a greater capacity to improve the comfort level among members. Rogers notes that this stage cannot be done too quickly as it could end up getting reject for being too great of a change to handle (2003).

Routinizing – The final stage consists of maintaining the use of the innovation as a consistent aspect of workflow to the point where it is no longer an innovation but rather a custom.

Even though a new technology completes all five stages of the innovation process, there is still a concern for sustainability; different factors can affect the strength of an innovation taking hold (Rogers, 2003). In essence, there is a level of acceptance among organization members that must take place before an innovation can be fully realized and included. Here Rogers is offering a brief reflection on the negative aspects of this theory expressed by other researchers.

For example, there is the perspective that diffusion studies discount the complexities that are a part of the process of innovation acceptance (Micó et al., 2013). Micó et al., comes to this position based on a case study of a convergence initiative within a Catalonia public broadcasting group. The results of that study displayed inconsistencies in the diffusion theory due to the problematic build of tension and lack of an "innovation champion" to help "foster the adoption of convergence" between the Catalonia organizations (2013, 134). The *innovation champion* is a reference to Rogers' *champion* or an "individual who throws his or her weight behind an innovation, thus overcoming indifference or resistance that the new idea may provoke in an organization" (2003, 414). Lacking this individual, the

Catalonia initiative was unable to "foster the adoption of convergence" amidst other factors both internal and external. (Micó et al., 2013, 134) Upon reflection, Mico's study doesn't necessarily disprove Rogers' research on diffusion, but moreover it identifies how lacking specific important factors can severely influence the success of an innovation. Essentially, Rogers' diffusion theory, due to the practicality of its application, allows it to endure the test of time, now going on four decades in the field of communication.

While the research of this study focuses directly on a case study of the Columbia Missourian photography archives, other studies can provide insight into similar instances of innovation. Certain types of innovations like metadata and digital filing systems build into the aspects of content management systems allowing for a functional database with multi-variable search-ability. Based off of the depth of this theoretical framework, the following literature review offer insight into how prior researchers and organizations have addressed many of the innovations and their concerns related to new technologies.

Understanding Diffusion Innovation in Prior Literature

Since just before the turn of the millennium, digital technology has begun to play a more important role in the newsroom and the media industry at large, with an increase in the amount of assets at journalists' disposal as well as the amount of outlets and methods for connecting to audience members. While the growth of these digital platforms and technological advancements have generally been considered widely accepted innovations (Ekdale et al., 2015), there have been many instances

where the industry has suffered or fought back to preserve different industry

standards or better methods.

In this chapter, a literature review applicable with the theoretical framework

and in relation to innovation implementation and adoption will be presented. The

initial section will present an understanding of existing information on the

acceptance of innovations in newsrooms while also presenting scenarios where

innovations failed and the conditions that resulted in such circumstances. The

following segment reflects on addressing the concerns of RQ1 and how it

corresponds to the diffusion of innovations theory in terms of visual archiving

innovations and the concerns analyzed in prior literature. Here, it is important to

note that, while prior literature is abundant with material on innovation adoption in

terms of media communication and digital technology, there is a significant lack of

research that pertains to visual innovations and photography archival and digital

asset management technologies.

Innovation Trends in Newsrooms: With the overwhelming acceptance of digital

technology by the general public, newsrooms have slowly accepted innovation

adoption to maintain audience approval.

New technology has been an increasingly forward motion for the news

industry, establishing a two-way road between journalists and readers.

Nevertheless, many news groups have been slow to adapt the process, even though

it has become an apparent necessity (Pavlik, 2013). While innovation does increase

amount of production and effort required of news organizations (something many

would consider a potential negative aspect), innovation also improves many qualities of the industry in terms of accessibility and efficiency.

"Innovation is the key to the viability of news media in the digital age" (Pavlik, 2013, 190). Pavlik's research focused on the need for innovation in the newspaper industry if it were to survive in the future. His study, which analyzed large papers including the New York Times, Wall Street Journal, The Guardian and the New York Magazine, presented three key results where innovation should be targeted: engaging public with quality news, engaging public on digital and social media platforms, and adopting methods that are best for the digital age (190). Pavlik notes that although revenue is the main battle the industry faces, many of the organizations in his study have developed in the area of original web content and advertising to help offset the lack of traditional print audience and engage the growing network population where they are: online.

On the other hand, contrasting Pavlik's (2013) study of how organizations adopted innovations to adapt to the growing digital news front, Wilson Lowrey's (2011) study takes a 'devil's advocate' approach to news industry innovation by approaching the industry through the failures that have been exhibited in terms of modern connectivity to readers. Lowrey's research indicates that the innovation process in many of the newsrooms he studied failed due to the fact that leaders and managers in the newsroom, when confronted with innovation, tend to "retreat" to "institutionalist tendencies," or trends that mimic the rest of the industry. In essence, Lowrey's findings reveal that organizational failure to adopt an innovation can be summed up as a failure on the innovation champion's part to progress with

the innovation, here the "champion" being the newsroom manager. His findings also revealed that innovations did succeed when "correlated with ownership needs and with weak-tie engagement with readers" (Lowrey, 2011, 75).

In a more recent study, Ekdale, et al., (2015) found that "journalists can use new technologies to interact directly with members of the community, and community members, in turn, use social media as a new way to suggest, correct, praise, complain, and otherwise try to influence coverage decisions" (2015, 24). Again we are brought back to this two-way connection between readers and journalists. Their study is based on data from a mid-sized newspaper with a 50,000-readership circulation that has undergone multiple innovation processes involving new technologies within the last few years of the research timeframe. The rebuttal they present here is that innovations that tend to fail are the innovations that "hurt the quality of the news product" or were not communicated well from company management (Ekdale et al., 2015, 26). Both Lowrey and Ekdale demonstrate the same concern of innovation failure in their research studies. This is especially interesting considering their projects were four years apart.

Overall, both research studies reflect a strong correlation with that of Rogers' diffusion of innovation theory in the areas of full organizational participation and the need for a champion to be backing the innovation to improve the parameters of its success. In a related study, Mieneke Weenig (1999) examines the effects of informal versus formal communication amongst employees and the organization throughout the innovation process. Weenig discovered that "formal communication sources contributed more to the process of information diffusion, whereas informal

communication sources (especially the employees' strong ties), were more influential on attitudes and adoption intentions" (p.192). His research sheds light on the implementation process with regards to how thoroughly the implementation was handled. In essence, the flow of communication can be is better received in a formal capacity from the organization in regards to the actual implementation process, but the informal conversation and communication aspects employees share improves the overall push to adopt the innovation in the first place.

Bruce Garrison's (2000) research also took to studying diffusion as it pertains to innovation in daily newspapers over a five-year period. In his results, he found that "it is evident that the leading newspapers in the USA believe the web has to be part of their effort to reach readers" (Garrison, 2000, 100). Even though Garrison's research is directed at the study of innovating online research tools in news organizations, the trend of new technology adaption can be applied in regards to forward motion in terms of digital archiving and editing techniques as well, since web-based technologies are becoming the best way "to interact with viewers and readers" (Garrison, 2000, 88). Weenig's (1999) results further provide a vital step to acknowledge when beginning the observation phase of this research. The importance here is that in this projects' observation, the structure of the photography department is based off of a hierarchy of ADOPs above editors and editors above photographers, and communication will play an important role in how well received the MOchive software is throughout the implementation and retention processes.

Refining Visual Innovation: Archival and digital curation innovation requires an answer to the question of preservation in media industries in order to be successful.

Before diving into the complexities surrounding digital asset management and archiving in newsroom environments, a foundation for how such technologies became a concern must be established. At the turn of the millennium, digital imaging could be considered an innovation in many respects with the increase of computer editing software, the incorporation of digital cameras in newsrooms and the need for archive systems to preserve these no longer physical files. John Russial (2000) took a more refined approach in his study analyzing the effects digital imaging had on the production workflow of journalists. Russial's (2000) research revealed that even though "65.7 percent of photo editors said workload was 'much' or 'somewhat heavier'" (p. 76), the majority of papers didn't increase staffing due to the increased production (metadata and archiving) work associated with digital imaging technologies. Although this concept seems relatively insignificant, Russial makes a point to emphasize that this fluctuation of increased workload could be due to the fact that "digital imaging is so widely used in newspapers, photojournalists feel it is important to develop those skills to ensure that they will be able to advance in their careers" (p. 80). Here Russial is acknowledging that the adoption of digital imaging as an innovation fell seemingly on the shoulders of photojournalism as a necessity to progress not only in their careers, but on an individual level as well, due to the fact that digital imaging was a nationally accepted emerging technology.

In their 2003 research, Shahira Fahmy and Zoe Smith acknowledge many of the pros and cons of digital imaging as it has been innovated in the modern media

industry, back when digital photography technology was considered a newer fashion. While their results can be seen as a bit dated, it still provides relevance since "digital imaging has freed the industry from the time consuming chemical processing and has replaced film in many, if not most, newsrooms" (p. 82) and has become the pinnacle of modern journalism practices. Based off their research the innovation of digital imaging gives photojournalists more time in the shooting process, increased awareness of their own work and increases the access factor of images by being able to transmit from the field, decreasing the delay (p. 93). On the other hand, Fahmy and Smith identify the most impactful disadvantage of digital imaging: limited storage. One important factor is that they present a unique perspective of one of the fallouts of limited storage: the concern that due to the lack of space, photographers "may be forced to delete images on location" (p. 93). Fahmy and Smith qualify this concern with the fact that accessibility has technically always been a concern due to the deterioration factor of older forms of preservation. Still, new technology presents a rapidly growing concern in terms of digital preservation, an area of emphasis that many researchers and news industries seek to address.

Review & Summary

In light of the modern age of technology, the media industry doesn't have many other options than to embrace the innovation of digital technology or be lost in the pages of their own history. As the audience migrates to the digital realm, so must the content in order to remain competitive and relative. Prior research has shown us that this adoption of innovation, while tedious and burdensome, can also

lead to many modern benefits. Especially in the journalism industry, the Internet has provided news content almost instantaneously for distribution and redistribution to the point where media industries can now communicate, collaborate and coexist by thriving together like never before. The question remains: at what cost though? While this review predominantly focused around the adoption of innovations as the key measures of study, many obstacles and future concerning scenarios were presented. While the news industry is on the forefront of instant accessibility, there is a looming concern on how best we can safely preserve our content.

RESEARCH DESIGN & METHODOLOGY

Introduction

The objective of this research was to examine the factors that affect

implementation of web-based archiving systems by professional news organizations

using the framework provided by the diffusion of innovations theory. This section

introduces the analytical and practical component of this project. Here, the

information collection process, the method of analysis and concerns of accuracy and

validity are addressed. As a professional analysis project, this study does not require

nor fall under the parameters set by the University of Missouri Institutional Review

Board as a scholarly research study. In the following sections, this study will offer a

background for the specific content under examination and then present the two

methods of data collection that this research will primarily emphasize. This section

will predominantly identify the parameters of this project and the methods in which

this case study will be researched.

Research Background & Design

The physical component of the Missourian Photo Archives case study has

been in progress for some time now. Edward McCain, RJI's Digital Curator of

Journalism, and the Columbia Missourian archiving staff have reorganized the

paper's archives into a single local server that has access off campus through a File

Transfer Protocol (FTP) login. Before McCain began this project, the archives were

stored on either CDs, print copies or spread across multiple independent hard

drives, each in various levels of degradation. Such a scenario poses significant risk factors that could be a detriment to not only the newspaper, but the community as well since such archives provide a deep visual history for the city of Columbia.

The present platform is a local area network (LAN), or closed server, in the sense that the user must be connected directly to the server's network needs login credentials to gain access. Specifically, the archiving process has included over a year digitizing documents, CD's, migrating external hard drives and a smaller local server all into one central location (a new NAS server housed in Reynolds Journalism Institute) that is automatically accessible and searchable by any members of the visual staff at the Missourian.

The next phase is to implement and test a new digital asset management system (MOchive—used in all future references) that will allow the staff to search metadata information and keywords to promote an improved search-ability system for archival content they need. To provide a bit of understanding, MOchive, a ResourceSpace based software, allows the user (be it the archivist, photo editors, photographers, director, etc.) to sort through massive amounts of content uploaded to it by the organization. This sorting feature is run through ResourceSpace's "Simple Search" menu (see Method Content 2) that searches the metadata sections embedded in any uploaded files.

The intention here is that content will be reachable "through a browser based system that you could log into basically anywhere you want" (McCain) outside of LAN restrictions. It is important to note that administrative access to this system is strictly maintained and limited to key members of the media organization

and system managers. This study focuses on the previously discussed phase. Particularly, the research methods involve examining the implementation process of this phase into the Columbia Missourian Newsroom. The following section takes a look at the research methods involved.

Method

This case study features a combination of direct/participant observation and both individual and group semi-structured interviewing as the methods of research. Multiple data collecting techniques assist in maintaining the utmost validity and accuracy throughout this project's analysis component.

Direct Observation

The primary method of this study will be direct observation, particularly in the case study scenario of the Columbia Missourian newsroom as it is the constant variable within this research project. The Missourian, as the host organization for MOchive, includes a daily functioning staff of approximately 8-10 photographers, 5-7 photo editors and two assistant directors of photography (ADOPS). (For this purposes of this study, Brian Kratzer, the Director of Photography, will be included in the interview phase primarily rather than direct observation. This is in an attempt to focus the observation on staff members with direct involvement in daily workflow without the fundamental administrative or archivist roles.) The observation research and analysis will be divided into 3 categories:

- Factors affecting implementation integration

- Factors affecting implementation retention

- Factors affecting overall success/failure of DAMS

Each category reflects a critical research point in the process of implementing an innovation. Integration, not to be confused with adoption from Rogers' (2003) "initiation" phase as the innovation has already been adopted, reflects rather the restructuring sub-phase of Rogers' implementation process. Here, the subjects will be studied on the speed (simplicity of search), efficiency (navigating the system and its search capabilities) and consistency to which they utilize MOchive system on a daily basis. These units of measurement will be factored in reference to the subject's understanding of the software. As the process of integration is a new feature, it is important to assess any drawbacks and technical errors that may be addressed in the future.

Retention focuses on Rogers' (2003) clarifying sub-phase. At this point, MOchive must be slowly pushed more and more by editors and ADOPs as the main archival and workflow database from which to operate from (but not to slow to the point of needing to re-learn every week.) The key focus here will be analyzing the muscle-memory of photographers and photo editors in utilizing the search function and workflow of the MOchive system. Questions to be considered in this observation category include:

- What functions continuously cause the most difficulties?

- How active are the editors and ADOPs in promoting MOchive as a resource system?

The final category will be an observation of the factors influencing success or problem growth in the previous two phases. The main focus here will be to observe how problems are dealt with amongst editors and photographers when concerns and confusion do arise. This category will also focus on the archivist's engagement as a "champion" (Rogers, 2003, 134) and how such a role affects the outcome of this case study.

Semi-structured Interviewing

For this section research study, the focus is placed on two methods of interviewing that will act as subject review analysis to understand the effectiveness of the MOchive system. The importance of individual interviews is that it primarily allows for the subject to understand his or her specific understanding of the topic and formulate their own opinions and perspectives without external influence outside the depths of their own experiences and knowledge (Knox & Burkard, 2009).

Stage 1 will involve individual in-depth semi-structured interviews of photographers, photo editors and ADOPs. This stage of analysis involves questioning participants for their interpretation of implementation situations so that they can help assess where the problems occurred individually and when communicating with their peers. Questions would be grouped into sections of personal experience in interpretation. Personal experience interview questions allow the participant to analyze and interpret their own backgrounds in the scope of a specific topic (Knox & Burkard, 2009), directly relying on their individual

knowledge to assess the potency of the situation or experience under examination. Personal experience questions also allow the participant to get grounded in the interview prior to expanding into more intensive scenarios. It allows the researcher to understand the basic human settings of the subject, putting them "in the role of experts who teach the researchers" about the scenario in question (Schulze, 2007). In essence, this form of study provides insight into how interviewees interpret flaws and problems through the framework of trial and error. Subjects define the problems for themselves through their own capacity of rationalization. Such form of research allows for bias to be minimized.

Stage 2 will involve having administrative members of the news organization as well as MOchive developers participate in a direct individual semi-structured interview. This stage is mostly a reflective process as the sources and line of questioning will be directed towards understanding and evaluating the implementation process as a whole rather than the actual production and individual aspects addressed in stage 1. Sources for this stage include, but are not limited to:

1) Brian Kratzer – Director of Photography, Columbia Missourian – Mr. Kratzer, as director, is essential because he operates at the highest level of authority in the implementation process, overseeing all of the photographers, photo editors and ADOPs throughout the course of this innovation.

2) Edward McCain – Digital Curator of Journalism at the Donald W. Reynolds Journalism Institute and MU Libraries – Mr. McCain has been an essential part of this study from day one as it is his developmental software (MOchive) that acts as the innovation this study is analyzing. McCain has held an overarching "innovation champion" role throughout the process by pressing the Missourian and Reynolds Journalism Institute to adopt

this new technology and make the necessary preparations for its eventual implementation in the news organization workflow.

3) Rob Weir – Director of Digital Development, Columbia Missourian – Mr. Weir has been a key individual in the maintenance and management of current Missourian content servers since before Edward McCain began focusing on restructuring the Missourian archives. With his perspective on the position of the archive within the overall Missourian structure, Mr. Weir's knowledge could prove valuable in the analysis process.

The series of questioning in stage two will be formatted based on each subject's strengths pertaining to the study. General overarching questions will address the effectiveness of the MOchive software as a whole both before and after the subjects have had a chance to interact with it.

Improving Validity & Accuracy

Analysis of Mixed Methods Approach

Following the proposed methods, analysis will commence involving comparative studies of the interview data with findings received in the observation phase. The use of multiple methods of data collection enables this project to reflect a greater amount of valid and accurate results that future studies can rely on (Church, 2001). For this study, validity holds higher importance as it is based mostly on the interpretations of the researcher as much as the participants, if not more. Negative case sampling is already incorporated in as each interview opinion has two sides of position and can be interpreted in multiple ways, notwithstanding that reflective responses are based on each subject's individual experiences and as a group can be generalized rather than generalizing off of an individual account. In essence, both

positive and negative reactions will be recorded and assessed in the analysis phase

to better reflect an accurate understanding of the effectiveness of the MOchive

implementation process. As for the research, due to the heavy reliance on

participants, this study would strongly benefit from participant feedback, a form of

validity that incorporates "feedback and discussion" between the researcher and

participants to help clarify possible interpretations and conclusions through

"verification and insight" (Burke, 1997). As the participants can claim the source of

knowledge the researcher used, having them review and reinterpret their own

insights could hold great value. It is important to note that the researcher must be

careful that they do not attempt to replace scholarly interpretations with their own

bias.

The researcher must also maintain a sense of self-reflexivity throughout the

study. That is, being in such close proximity to the interest area and the field of the

topic being studied, self-awareness, or "critical self-reflection" (Burke, 1997), plays a

vital role on the researcher's part so that he does not incidentally insert his own

biases. Researchers are not judgmentally based, but should understand their

position on the subject and maintain an intellectual mind free of sway. To do this,

the researcher should submit to a peer review of his findings in order to address the

concerns of bias presence before submitting the analysis for approval.

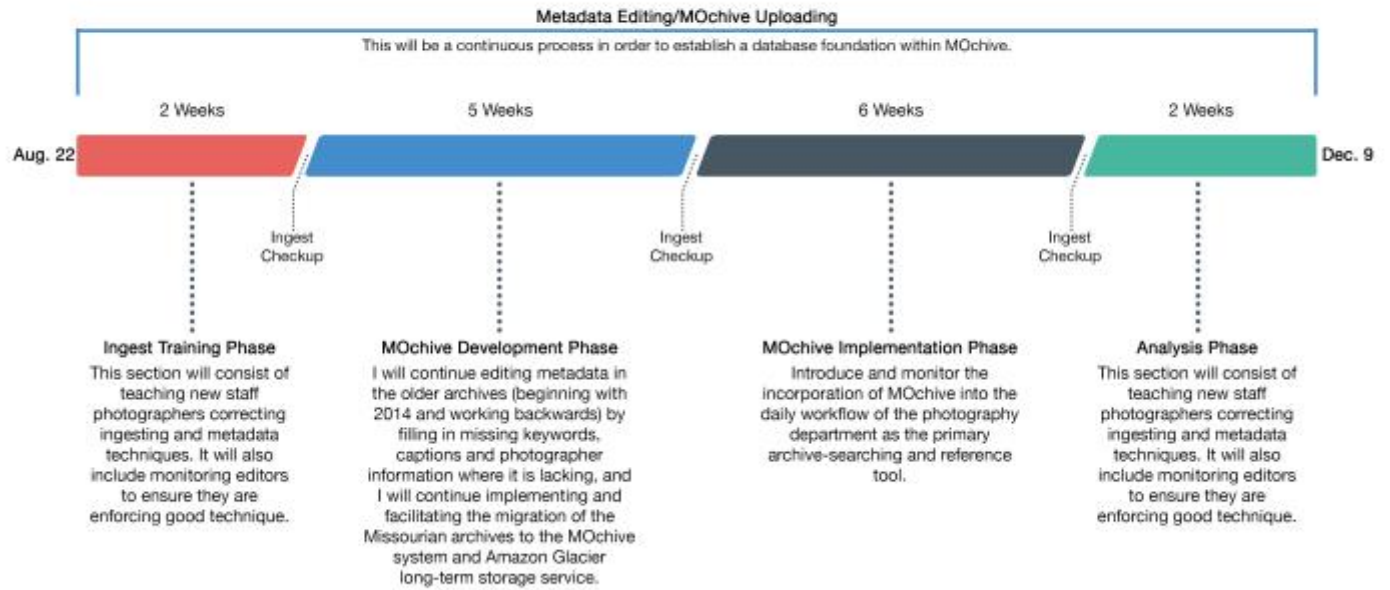*Research and Analysis Limitations*

While Rogers diffusion theory has been in the academic realm for quite some

time, modern technology has only recently allowed researchers to approach the

issue of digital metadata preservation (Alemneh, 2009). Studies by Edward McCain and Kathleen Hansen have begun to shed light on the topic, but without case studies, digital preservation still presents many dark areas that provide no context, understanding or solutions to the growing problem. With technology ever changing and the archival state of media organizations varying to such an extraordinary degree, the findings of this study has the opportunity to at least provide a physical foundation and case study that future research can be based off of.

This case study, while establishing an analytical foundation for future research, is only as strong as the limitations it is constrained to. Primarily, the greatest limitation is the organization the research is tied to as the Columbia Missourian, even though it is a professional medium-sized news publication, is unique in its methods of content production and staff development. Specifically, the Missourian operates off of a rotating staff of photographers that changes each semester, therefore requiring the necessity to repeat the training process constantly. In time, the process could be developed within fundamental level classes in order to prepare students for the staff position, but such an option isn't feasible at the present time. The other significant limitation is the concern that content preparation for MOchive may take significantly longer than expected due to various factors including upload speed, server speeds, metadata processing and staff development. This is a process that must be addressed as it arises since there is a factor of unpredictability that must be acknowledged and adapted to, as previously stated in the professional skills component section.

Appendix A

Project Timeline – Fall 2016

Metadata Editing/MOchive Uploading

This will be a continuous process in order to establish a database foundation within MOchive.

| 2 Weeks | 5 Weeks | 6 Weeks | 2 Weeks |

Aug. 22 Dec. 9

Ingest Checkup   Ingest Checkup   Ingest Checkup

Ingest Training Phase

This section will consist of teaching new staff photographers correcting ingesting and metadata techniques. It will also include monitoring editors to ensure they are enforcing good technique.

MOchive Development Phase

I will continue editing metadata in the older archives (beginning with 2014 and working backwards) by filling in missing keywords, captions and photographer information where it is lacking, and I will continue implementing and facilitating the migration of the Missourian archives to the MOchive system and Amazon Glacier long-term storage service.

MOchive Implementation Phase

Introduce and monitor the incorporation of MOchive into the daily workflow of the photography department as the primary archive-searching and reference tool.

Analysis Phase

This section will consist of teaching new staff photographers correcting ingesting and metadata techniques. It will also include monitoring editors to ensure they are enforcing good technique.

References

Alemneh, D. G. (2009). *An examination of the adoption of preservation metadata in cultural heritage institutions: an exploratory study using diffusion of innovations theory* (Doctoral dissertation, UNIVERSITY OF NORTH TEXAS).

Burke, J. R. (1997). Examining the validity structure of qualitative research. *Education*, *118*(2), 282-92.

Carner, D., McCain, E., & Zarndt, F. (2014). Missing links: The digital news preservation discontinuity.

Church, A. H. (2001). Is there a method to our madness? The impact of data collection methodology on organizational survey results. *Personnel Psychology, 54*, 937-969.

Ekdale, B., Singer, J. B., Tully, M., & Harmsen, S. (2015). Making Change Diffusion of Technological, Relational, and Cultural Innovation in the Newsroom. *Journalism & Mass Communication Quarterly*, 1077699015596337.

Fahmy, S., & Smith, C. Z. (2003). Photographers note digital's advantages, disadvantages. *Newspaper Research Journal*, *24*(2), 82.

Garrison, B. (2000). Diffusion of a New Technology On-line Research in Newspaper Newsrooms. *Convergence: The International Journal of Research into New Media Technologies*, *6*(1), 84-105.

Greenwood, K. (2006). *Pictures and pixels: digital photographic archives at newspapers, photographic agencies and libraries* (Doctoral dissertation, University of Missouri--Columbia).

Hedstrom, M. (1997). Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, 31(3), 189-202.

Kitzinger, J. (1995). Qualitative research: introducing focus groups. Bmj, 311(7000), 299-302.

Knox, S., & Burkard, A. W. (2009). Qualitative research interviews. *Psychotherapy Research*, *19*(4-5), 566-575.

Lowrey, W. (2011). Institutionalism, news organizations and innovation. *Journalism Studies, 12*(1), 64-79.

May, T. (2011). *Social research: Issues, methods and process*. Buckingham: Open University Press.

Micó, J., Masip, P., & Domingo, D. (2013). To wish impossible things*: Convergence as a process of diffusion of innovations in an Actor-Network. *International* Communication Gazette, 75(1), 118-137.

Pavlik, J. V. (2013). Innovation and the Future of Journalism. *Digital Journalism, 1*(2*)*, 181-193.

Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe.

Rogers, E. M. (2003). *Diffusion of Innovations.* Simon and Schuster.

Russial, J. (2000). How digital imaging changes work of photojournalists. *Newspaper Research Journal, 21*(2), 67.

Schmidt, S. (2009). Split Personalities: A Librarian in the Archives. Society of American Archivists. http://www.archivists.org/saagroups/lonearr/ publications/Schmidt_SplitPersonalities.doc.

Schulze, S. (2007). The usefulness of reflexive photography for qualitative research: A case study in higher education. *South African journal of higher education*, *21*(5), 536-553

Singer, J. B. (2004). Strange bedfellows? The diffusion of convergence in four news organizations. *Journalism Studies, 5(1),* 3-18.

Weenig, M. W. (1999). Communication Networks in the Diffusion of an Innovation in an Organization1. *Journal of Applied Social Psychology, 29*(5), 1072-1092.