Interviewee:	**Smith, Plato L.**
Interviewer:	**Valerie Modoux**
Date:	**March 14, 2008**
Category:	**Florida; Florida Archives digitization**
Status:	**Open**
Tape location:	**CD**

Modoux: Can you first tell me a little about your background -what education you have, what are your specializations and such?

Smith: My name is Plato L. Smith II. I have been the head of the Digital Library Center here at Florida State University since June 2005. Prior to coming here, I worked two years at North Carolina State University Library; I worked one year in their systems Department in DH Hill Library, and also worked a year in a college of veterinary medicine medical library. Prior to that, I worked a year at Duke University, at Perkins Library, also in their systems department as a graduate student assistant. And, the specializations that I have; I have worked with relational databases, creating web pages in HTML, some SQL structured query language, working within a UNIX environment, as far as certain open source softwares.

Modoux: Now as far as your job here, what are your main duties?

Smith: My main duties consist of managing all day-to-day operations, project-based activities, and research for FSU Libraries' Digital Library Center (DLC) located in the ground floor of Strozier Library. The FSU Libraries DLC is primarily responsible for scanning, large-format poster printing, and building online digital collections for faculty students, and staff. The DLC receive many digitization services requests from students representing Sports Management, Interior Design, Arts, Foreign Arts, and many other FSU colleges and departments. On average, the DLC receives digitization projects requests anywhere from twelve to over thirty different campus departments to scan and/or print on a monthly basis. Many students request large-format poster prints for class projects, various conferences, and presentations specific to their research disciplines, such as psychology, interior design, art, or the College of Information. I manage two full-time USPS employees and one rotating graduate student assistant. The graduate assistance works 15-20 hours per week. The projects we are currently working on are building digital collections in Digitool such as FSU "Flying High" Circus, FSU Special Collections EAD Finding Aids, and FSU Historical Photographs collections; Digitool is an Ex Libris product and is a digital content management system for managing digital assets such as PDF for text, jpg and tiff for images, mpeg for video, and mp3 for audio. FSU Libraries DigiTool currently contains fourteen digital collections which include the FSU Historical Photographs collection(2500 B&W images), FSU Electronic Theses and Dissertations (ETD) collection (over 11K ETD PDFs), FSU Undergraduates Honors in Major Theses collection (1960 PDFs), FSU Special Collections EAD (Encoded Archival Description) Finding Aids Inventories which contains over 70 EAD finding aids with multiple manifestations in HTML, PDF, and XML, and the FSU Heritage Protocol Collection which focuses on digitizing and preserving historical artifacts of FSU history

including digitized artifacts from when FSU was a women's college.

Modoux: Now, about the digitizing process, can you give me some information about it? For instance, what items you digitize, what tools/programs you use?

Smith: The DLC specializes in digitizing books, documents, and photographs. The DLC uses a large overheard scanner by Tarsia Technical Institute (TTI) to digitize scan and digitize physical/analog objects. The large-format scanner has high resolution anywhere up to 2400dpi (dots per inch). Typically, if patrons will present digitized content via the web, DLC recommends 300, but will digitize from 300 to 600 dpi. However, if patrons will publish in a print publication, then the DLC recommends a resolution higher than 300 dpi. The DLC digitizes theses and dissertations, historical photographs, historical juvenile books, and historical artifacts such as signs, rings, and yearbooks. The DLC also scans 3D object poster boards, designs, and drawings for students' presentations and conferences. The majority of FSU Libraries' DLC online digital collections is digitized content from FSU Special Collections. Currently, the DLC is in the process of identifying a sample of the Napoleon Collection from FSU Special Collections with which to digitize and make available online. The DLC has already digitized one book from the Napoleonic Era which is available online. The DLC, working in collaboration with the head of the Special Collection and Dr. Blaufarb from the Department of History, plans to identify a subset of the Napoleonic Era materials in FSU Special Collections to digitize and make available online. However, grant funding is needed to digitize and make collections online. The DLC is currently working with Dr. Prasad from Biological Science to make a portion of his diatoms collection available online. However, Dr. Prasad maybe interested in working with FSU Libraries DLC to make available online diatoms from his Gulf Coast diatoms collection. DLC is also working with Dr. Cheryl Ward from FSU Department of Anthropology to build additional online digital collections similar to her Sadana Island Shipwreck, Egypt, 1995-1 998 collection (http://diaitool3.lib.fsu.edu/R/?func=collections-result&collection id=1083). The collection contains 87 images of porcelain artifacts from the Sadana Island Shipwreck. The images are displayed online via METS -METS is a meta-data standard for describing digital objects. METS stands for Metadata Encoding Transmission Standards for digital objects and is a metadata standard which is maintained by the Library of Congress. Dublin Core which is a simple meta-data element is used to describe the content and create basic metadata records in DigiTool. Most of the digital content in DigiTool are in .pdf format. The FSU "Flying High" Circus digital collection contains one short b&w historical video DigiTool supports various file formats such as .pdf, .jpeg, .tiffs, .mets, .mp3, .mpeg. DigiTool's strength is that it can accept wide formats, file formats. However, DigiTool's weakness is middleware and metadata standards compatibility issues.

Modoux: Technically, how does it function? Do you have different areas where you digitize pictures, and then texts or audio?

Smith: Actually, for audio, we would work with Chuck McCann, who is the head of the media center. He is in charge of all video and all audio editing. The DLC digitizes text for the

ETD collections, inter-library loan requests, Special Collections projects, and the campus community. DLC digitizes images from Special Collections which make up several of the DigiTool digital collections. Over-sized, large-format, most Special Collections projects/scans and high-resolution scans use TTI reflex 45ei
(http://www.skgrin~es.com/thisweek~l-05-05-532/index.htm)
whereas books scans (b&w) use Minolta PS 7000 open-face, overhead book scanner, and single sheet fed (theses/dissertations) scans use Cannon DR-9080R.

Modoux: You said something about the format, do you do everything in .pdf? Or did I understand it wrong?

Smith: Texts, documents, and some books are saved in PDF format. Images which are web accessible are in jpg (low resolution -300 dpi) format and the archival or master images which are not web accessible and exist on networked attached storage are in tiff format (high resolution -300 + dpi).

Modoux: Now, in relation to the master copy. Can you tell me about the standards that you use to digitize it and then for the backup?

Smith: Tagged Image File Format (tiff) is the file format used for master copies during digitization. The master copies are stored on networked attached storage and are backed up biweekly via tape backup. The DLC will explore Florida Digital Archive (FDA) and Dark Archives in the Sunshine State (DAITSS) at Florida Center for Library Automation along with MetaArchive Lots of Copies Keep Stuff Safe (LOCKSS) are future preservation strategies of digital master/archival copies.

Modoux: What about resolution?

Smith: The DLC's TTI large-format scanner can scan images of high resolution up to 4800 dpi but typically use 600 -1 200 dpi for master copies.

Modoux: Who sets those standards and what are they?

Smith: The DLC sets the resolution standards: master copy .tiff high resolution is +600dpi. The archival copy is at a resolution of +600dpi, and the web copy is at a resolution of 300dpi.

Modoux: Now, can you now tell me about the backup process? Where are the master copies saved, and the web copies?

Smith: The master copies which are on networked attached storage are backed up biweekly via tape backup. The web copies are served from the server and are backed up during regularly scheduled server backup.

Modoux: Is it like a server?

Smith: Yes. The network-attached storage is a non-web accessible server.

Modoux: And do you save what goes on the web somewhere else?

Smith: The DLC is in the process now of using MetaArchives as a distributed digital preservation network to save web copies. MetaArchives consists of six project partners' institutions -Auburn, Virginia Tech, Georgia Tech, Louisville, Emory, and FSU. MetaArchive uses LOCKSS to replicate and emulate web accessible digital content. LOCKSS was developed at Stanford University. MetaArchive institutions must develop a manifest pages and XML plugins which identifies the collection, and gives the LOCKSS software permission to replicate all of the content that is on that page. The DLC is in the process now of allowing the other institutions to replicate what is on the web and/or network-attached storage -whatever we want the LOCKSS software to. All copies, both web and networked attached storage, are backed up to tape twice a week. So yes, the web version is backed up to tape; that is our standard. However, for a more comprehensive and complete digital preservation program, DLC will be using LOCKSS to preserve and replicate our master copies at other institutions.

Modoux: So, to make sure that I understood correctly the technical aspects of this: LOCKSS is the fact that different institutions can mirror what you have?

Smith: Yes, they replicate and emulate. LOCKSS, the acronym LOCKSS stands for "Lots of Copies Keep Stuff Safe" and was developed at Stanford.

Modoux: Thank you! Now, in your opinion, what is -or what are -the safest way to backup digital archival documents?

Smith: Well, I would have to say there is no single way to backup digital archival documents but there are multiple ways with the best way contingent on current IT infrastructure and resident technical expertise ... The safest way to backup digital archival documents is with traditional tape backup with your server and a distribute digital preservation similar to LOCKSS or another preservation strategy such as FDA or DAITSS.

Modoux: Yes. Now, some institutions preserve them only on external hard drives, or DVDs and CDs. Do you have any thoughts on that? Is there a reason why you do not use those here?

Smith: I do not believe that this is optimal and anything can happen to a hard drive -it can get magnetized -and there is limited life on a disk, anywhere from five to ten years, depending on what type of disk, the metal of construction. So, just having them on a standard floppy disk, or a hard drive.. . Yes, that is good for quotability and backup, but that should not be the main preservation archival backup. Yes, that is good, but anything can happen, it can get lost.

It can be part of preservation, but it should not be the only preservation strategy. Tape, removable hard drive, and then a distributed digital preservation network should be used to preserve digital content. Of course, with a distributed digital preservation network you are going to need electrical source, servers to maintain that sort of hard drives.

Modoux: By "tape", you mean internal harddrive? That c

Smith: Tape is not an internal hard drive but removable cartridge with magnetic tape. You can use a system such as Veritas, which you have some magnetic tape that you can remove to distributed locations.

Modoux: To wrap this up, could you tell me what your opinion is on the digital format? Do you consider it as a tool to make some items available to users, or as a preservation tool? Or would you consider digital as an archival format?

Smith: I would have to say all of the three. Having the correct description for the digital content will make it more accessible and discoverable. For instance, with Dr. Cheryl Ward's collection. Her collection existed on CDs stored in her office and offline. I contacted her and we met, and then I was able to create a METS format for her digital content, used Dublin Core to describe the collection, and then also work with a cataloger to further describe it according to Library of Congress subject headings. FSU Libraries further enhanced the meta-data to describe the collection, to build MARC record in the OPAC, (online public access catalog), and upload to OCLC WorldCat for worldwide discoverability. Currently, anyone around the world can now discover Dr. Cheryl Ward's Shipwreck collection which previously existed off line on CDs in her office but now exists in the OCLC WorldCat, FSU Libraries OPAC, and DLC institution repository (IR). Yes, you need standards for descriptive meta-data elements -one to describe the information and also to make digital content accessible and discoverable on the web. DLC uses Dublin Core to describe digital content in the institutional repository which creates a metadata record with a persistent URL (PURL) to created a MARC record in the OPAC -Online Public Access Catalogue and used a the PURL from institutional repository in an 856 link-field, within the MARC record to deeplink from OPAC to IR for resource discovery. This metadata mapping allows anyone searching FSU Libraries catalog to access digital content in the institutional repository from the 856 link-field, which is the PURL of IR digital content to create access from the MARC record. This project is very good because it is a collaborate effort between the researcher, a digital initiative librarian, and a cataloguer to increase discoverability and accessibility to faculty research. The researcher, digital librarian, cataloger all had to work together to describe and create access to online digital research content. Another deliverable of this project in addition to creating an online digital collection, is the future publication of article describing this project in OCLC OSS: IDLP due in early fall 2008. The researcher is excited because she says this digital collection project is a personal and career accomplishment for her that allow other researchers, faculty students, staff, to access her digital collection online. Yes. Archiving digital content is necessary. Some of the master copy .tiffs do not have extensive meta-data and thus need to descriptive meta-data elements to describe the archival copy.

Descriptive metadata, established metadata standards, and open file formats are imperative for digital curation, which is the promotion of future access to legacy digital content.

[End of Transcription]