A Data Management Plan
for the
Little Bighorn Battlefield National Monument

Submitted: January 25, 2011
By: Gregory Colati, Consultant

Little Bighorn Battlefield Visitor Center (NPS Photo)

LIBI Digital Asset Management Plan
January 25, 2012
Submitted by Gregory Colati

Executive Summary

The Data Management Plan for the Little Bighorn National Battlefield National Monument (LIBI) was prepared by Gregory Colati for Sharon A. Small, Curator, Little Bighorn Battlefield National Monument. The purpose of this report is to develop a plan for managing digital assets of the Museum Collection including both Natural and Cultural Resources at the Park that will insure not only long-term preservation but also enhanced access and use by Park staff and visitors in a way that also maintains National Park Service's control over use of the material as much as possible.

The process of developing the plan consisted of writing the plan of work, conducting preliminary investigation both with LIBI staff and staff at the Western Archaeological and Conservation Center (WACC) in Tucson to determine the goals, needs, and the existing conditions, including technology, policy and human resources. Next, a site visit was conducted in early June, 2011 that included in-depth conversations with LIBI staff and first-hand observation of conditions and infrastructure. Finally, the report was prepared with feedback from LIBI and WACC staff.

LIBI's resources are essential primary material for interpreting the site and its history. Better and more flexible access to these resources will make it possible to better present and interpret the site. The strength of LIBI staff is in understanding and interpreting the materials related to the site and the history of the era. Not surprisingly, LIBI does not, and will not, have the staff, infrastructure, or expertise to support sustainable management of its digital historical assets. Building on strengths, it makes sense to invest LIBI staff with higher-level presentation and curatorial responsibilities and centrally manage the lower-level technical and technological infrastructure that supports those activities. The recommendation in this report is for LIBI to partner with WACC to develop a collaborative and distributed digital asset management system where presentation activities take place at LIBI and digital preservation activities take place at WACC. This model can be generalized for use by other National Park Service (NPS) units in the area.

The framework of the plan is to centrally manage digital content files and offer customized user environments based on roles: internal (staff only), on-site public access, and internet public access. These three experiences will draw content and metadata from a central repository for delivery. The first environment to be developed will be the internal staff access environment.

Background of Park

Little Bighorn Battlefield National Monument memorializes the site of the Battle of the Little Bighorn which took place on June 25-26, 1876 between the United States Seventh Cavalry Regiment led by Lt. Col. George Armstrong Custer, and the Sioux and Cheyenne

under the political and spiritual leadership of Sitting Bull who were fighting to preserve their way of life.

Mission and Purpose of Park and Unit

"The purpose of the Little Bighorn Battlefield National Monument is to preserve, protect and interpret the historic, cultural, and natural resources pertaining to the Battle of the Little Bighorn, leaving them unimpaired, and to provide visitors with an understanding of the historical events leading up to the battle, the sequence of activities by both military and Native American contingents on June 24-27, 1876, and the historical consequences of the results of those fateful days."

The Museum Collection is meant to support the general mission of the park. This includes the interpretive plan, as well as external researchers and the more casually interested. Additionally the Museum Collection manager is responsible for the long-term preservation of both physical and digital resources and is the steward of these collections, providing access to information about the collections as well.

How Digital Asset Management Can Support the Mission of the Museum Collection and the Park

Digital Asset Management can support the mission of the park specifically, and the National Park Service more generally by enhancing access to collections and information that supports the work of the Park staff (supports the work of all park staff) and external researchers. Enhanced access to collections and information supports interpretation and exhibit development at the park itself and the interests of casual visitors and researchers.

The following scenarios are a few examples that illustrate how enhanced access to collections and information supports park activities:

Interpretive staff is working to update training manuals with primary resources to illustrate themes in the interpretive plan. They use the staff mode to discover relevant documents and use them in the printed training manual. They search the staff database for potential objects/documents to include. From the resulting list they mark as "favorites" a set of objects for later review and inclusion in the training material. Access derivatives are downloaded and used as needed, high-resolution master files are requested.

Interpretive staff may want to update the Park website to include documents, photos and video that provide information about the Park and that will encourage visitors to come to the Park. They search the staff database for potential objects/documents to include. From the resulting list they mark as "favorites" a set of objects for later review and inclusion in the website. Access derivatives are downloaded and used as needed.

Exhibit staff may want to select digitized content for a multi-media exhibit in the Visitor Center based on the interpretive themes developed in the interpretive plan. They use terms identified in the plan to search the staff database for potential objects to include. They filter their search by specific criteria like format, type, and "digitized." From the resulting list they mark as "favorites" a set of objects for later review and selection. Later, they download access derivatives and object metadata for inclusion in the exhibit.

Some objects require the high-resolution master images. For these images they use a staff online request function.

After first viewing the exhibits, a park visitor may listen to a Ranger talk or take a Ranger-led tour. This may lead to the visitor wanting to learn more about certain people and events discussed. (Perhaps (s)he has discovered someone with the same last name and is interested in more information about that person.) (S)he returns to the Visitor Center and using a public-access station, browses among the digitized historical resources available, each one with descriptions and other metadata included.

A researcher may want to find letters and documents relating to a relative who was involved in the battle and would like to search for that name among the records. They use a full-text search engine to search through digitized documents as well as finding aids and cataloged object-level metadata. After refining his search result, the researcher accesses the online objects directly, and requests copies of the non-digitized material through an online request function.

A historian interested in military equipment may want to survey the various pieces of equipment found at the site. The historian uses a public search function to discover and browse digital representations of collection objects and to discover non-digitized documents from the finding aids. The physical location of all objects is noted in the metadata record and the historian makes an appointment at the appropriate location to view the original objects for research.

A park visitor or researcher may view a collection object and has some information to add to what we know about the object. The visitor's information is verified by park staff and the local metadata record is updated. Later this metadata update is passed along to the main metadata repository so that all records stay in synchronization.

What is Digital Asset Management?

Digital asset management is a system that is a combination of hardware, software applications, policies, and collection objects that supports "good" stewardship of digital objects. Digital objects exist in a continuum of use and management, and comprise not only the "primary content object" but the metadata and associated derivatives as well.

Characteristics:

A digital asset management system will preserve and manage digital assets and associated derivatives, as well as produce information about collection objects that can be used in discovery applications and finding aids.

Briefly a digital asset management system will be:

- standards-based at the repository level, including a standard set of required metadata elements/schema and file formats
- governed by policies
- sustainable over time with appropriate staffing and hardware

- secure—only those authorized to access the system and its different parts are able to do so

Digital asset management systems are built on the foundation of a digital repository. Digital repositories support all of the functions listed above and are themselves based on a set of principles designed to insure preservation and access of these valuable assets.

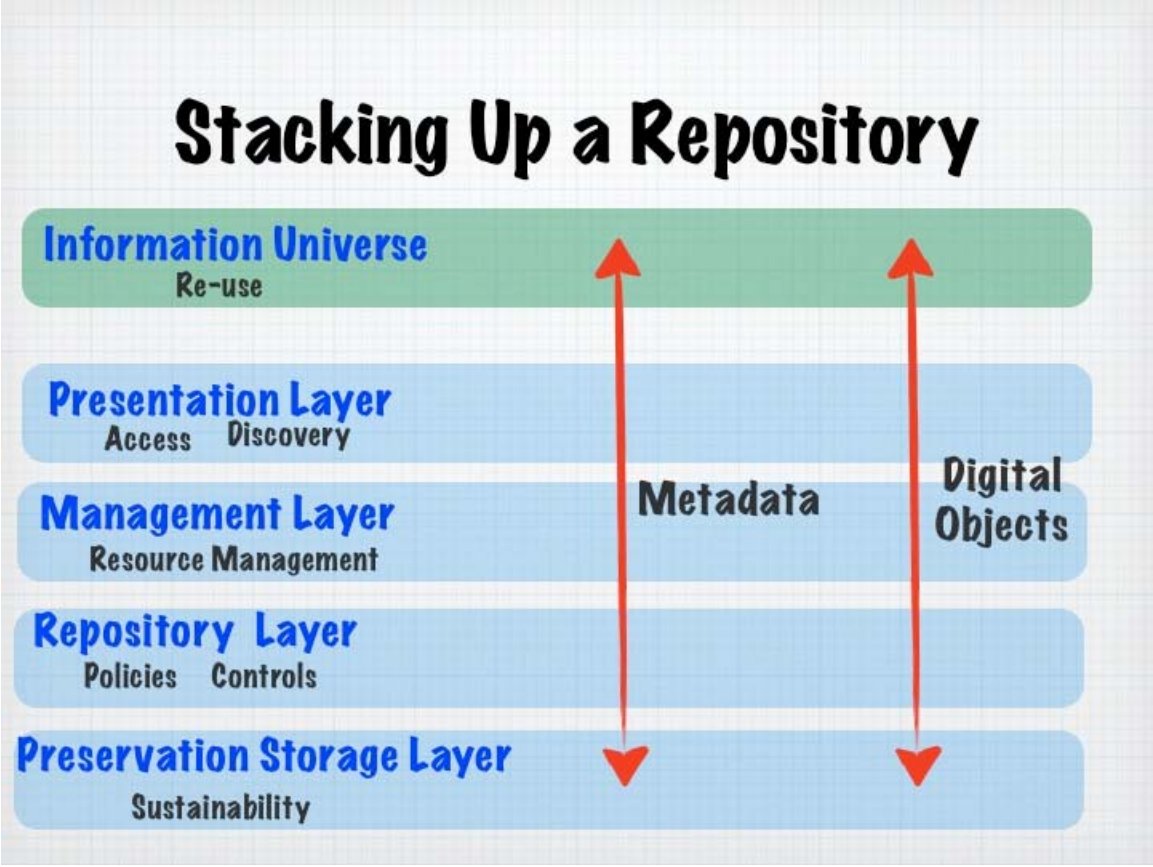
The Center for Research Libraries (CRL) is a leader among digital preservation standards organizations. The CRL maintains the Trustworthy Repository Audit and Checklist (TRAC), the benchmark for determining quality in digital preservation. Their "*Ten Principles" (*http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re) of a digital repository is a distillation of the much longer TRAC document and gives an overview of the basics of digital preservation systems.

- The repository commits to continuing maintenance of digital objects for identified community/communities.
- Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.
- Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
- Has an effective and efficient policy framework.
- Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
- Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
- Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
- Fulfills requisite dissemination requirements.
- Has a strategic program for preservation planning and action.
- Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

The key premise underlying the core requirements is that **preservation activities must be scaled to the needs and means of the defined community** or communities for repositories of all types and sizes.

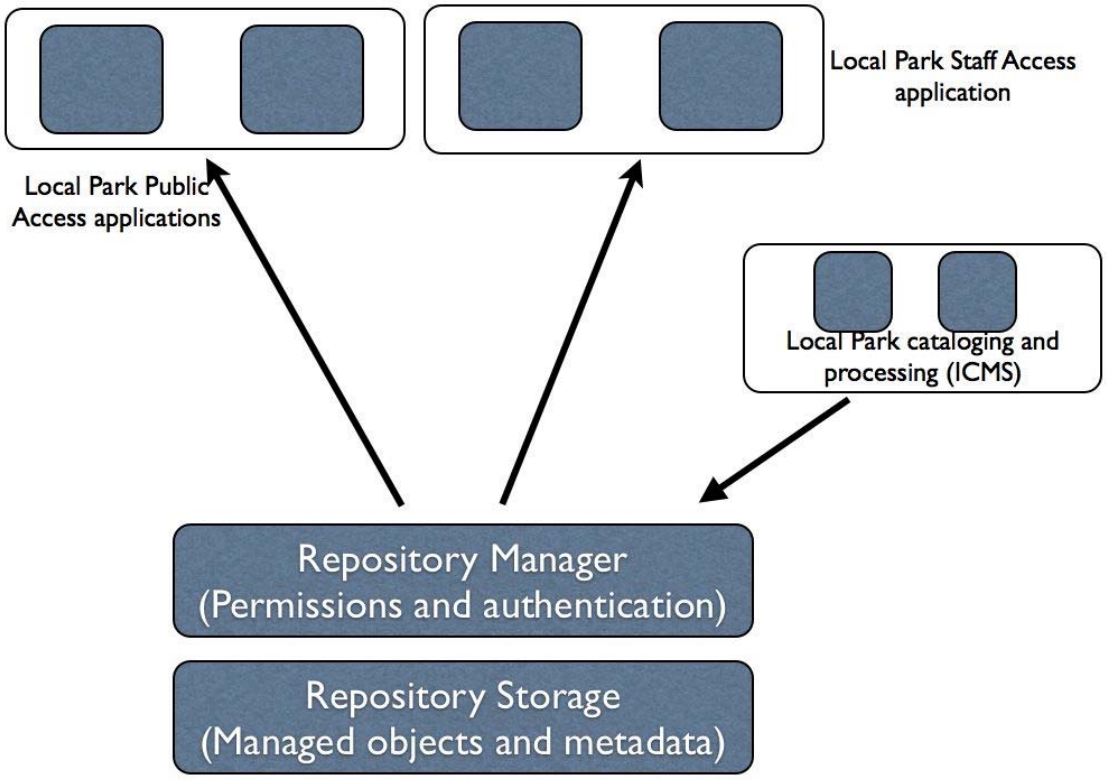
Commonly, digital repositories are made up of a combination of applications and hardware systems that are governed by policies and work together to achieve the goals of the organization. One way to think about organizing a digital repository is to think of it as a series of "layers" that are "stacked" in a way that will move and manage digital objects

and metadata. The figure below outlines the layers included in most digital information systems. These layers may or may not be part of a single hardware/software installation. In fact, it is increasingly the case that these layers are distributed across different units in a larger organization. Top layers tend to be dispersed and locally managed, while lower layers tend to be centrally located and managed to take advantage of economies of scale.



Each layer may consist of different parts as well, for example in the case of the Park Service, there could be a central repository and preservation layer with multiple management and presentation applications connected to it. Since the greatest cost is associated with repository and preservation storage, it makes some sense to centralize these functions if the necessary network connectivity is available to exchange content among the different layers.

For example, in the figure below, the “stack” includes several different presentation applications in the presentation layer, each one dependent on the central content repository for metadata and content objects. Also, the Interior Collections Management System (ICMS), installed at local institutions can feed into the central metadata store, where more costly and technically difficult preservation activities can be carried out and maintained.



Digital Asset Management at Little Bighorn Battlefield National Monument

The outlines of a digital asset management system and the user scenarios can be combined to develop a plan for a digital asset management system at LIBI. The preceding user scenarios assume a set of functionalities within the digital object management system that can be generalized here:

- Separate staff and public access interfaces
- Local and central metadata management with synchronization to and from the local instances
- Request functions for both staff and researchers
- Request fulfillment procedure for both staff and researchers
- Central storage of master primary content objects and metadata.
- Local management of public access tools and their contents.
- A system to authenticate and regulate access to master objects (and metadata?) relating to the collections.
- Preservation-oriented storage based on the CRL ten principles

Existing Infrastructure for Managing Research Assets

Using the layers identified previously it is possible to assess the existing infrastructure at LIBI and the potential for best developing that infrastructure to support activities identified in the previous section.

1. Presentation
    a. Presentation applications and hardware are available and used by staff to disseminate information about the Park.
    b. This same infrastructure can also be used to connect visitors and researchers with research collections.
    c. The Content Management System (CMS) web site application currently available and maintained by central NPS can support upload of content and material for use.
    d. The CMS may not be able to draw directly from the repository for access copies of digital objects. This is an area for more research.

2. Management
    a. Collections
        i. The NPS-wide ICMS software can be leveraged as the primary data source for digital object metadata in the central digital repository. Metadata exported from local ICMS systems will be used as the source for the repository metadata records
        ii. The distributed nature of this system makes management a challenge, but can be addressed with protocols and processes that keep the data in synchronization.
        iii. WACC must create a process based on a standard workflow to export metadata from local ICMS systems and transfer it to the repository for upload into the discovery system. (Ideally, updates to metadata records in the ICMS at local parks would automatically propagate through the system to the master database.)
    b. Administrative records
        i. Administrative records, including contemporary photographs and other documentation that will soon become historically valuable, are not now part of a management system.
        ii. These records are used often by park staff to create informational publications for a variety of uses and are difficult to maintain both administratively and technologically at the local level.
        iii. Some system to manage administrative records, especially those used for marketing, training, and interpretation that can be linked to historical records systems should be explored and developed. **Recommendations on the nature of this system are outside the scope of this project**.

3. Repository
    a. IT staff at LIBI are primarily responsible for managing and maintaining the administrative systems that support day-to-day activities of the Park and park staff.
    b. LIBI IT is not staffed or resourced to implement and manage a digital repository system for research material.

c. Siting repository services at a central location may resolve this problem most efficiently. This option should be explored immediately.

4. Preservation Storage
   a. Similarly, IT at LIBI does not have the resources to implement a local preservation storage system for digital research material.
   b. It also seems unlikely that building such a system at LIBI would make economic or programmatic sense.
   c. Locating preservation activities at a central location that is staffed and resourced for this purpose should be explored as soon as possible.

Collateral Implications and Consequences.

Any program or system that results in more information about the collections and more collection objects available on line tends to drive increased demand for access by both researchers and staff. The Museum Collections manager should be prepared to accommodate more requests and interest than before. Some base funding to accommodate research support (for staff and visitors) should be added to the unit's resources.

Similarly, continued work to make collection information available to support increased requests for both digital and non-digitized material should be supported with base funding.

Recommendations

Technical Infrastructure

Responsibly managing digital assets for research requires a digital asset management program supported by a preservation-oriented repository infrastructure. Preservation-oriented repositories are expensive to create and maintain and benefit most from cost savings related to scale. The larger the scale of the repository the lower the unit costs for maintaining any individual metadata or content object. Another benefit of centralized metadata and object management is increased security and control of digital objects within the repository and their availability to a wider variety of authorized users.

LIBI could leverage the greater resources of the WACC as a central repository of digital objects (metadata, primary content objects, and access derivatives), and manage and maintain only presentation applications locally.

Ideally, network connectivity and application interoperability would make it possible for presentation applications to access the repository in real time for all object and metadata needs. This level of network connectivity and interoperability does not now exist and in the near term, manual methods of file transfer and local storage of access derivatives can meet local needs only if they are accompanied by systematically enforced policies.

Centralized maintenance of an OAIS[1]-modeled and TRAC-compliant preservation system with local control (but not necessarily local implementation and technical support) of presentation layers is the ideal scenario for research data management and preservation in this situation.

In a first step implementation for example an internal access website would be hosted on a dedicated server at WACC and require regular attention from an IT Specialist. Due to the need for dedicated equipment and dedicated support, it would make economic sense for the server to be hosted at a location that currently has the necessary infrastructure, equipment or the means with which to acquire the equipment and has the dedicated IT staff to manage the equipment.

Currently, WACC has a dedicated IT staff member whose primary focus is the equipment located and operated at the WACC facility. This would allow for all the necessary software support as well as any onsite hardware support to be provided by this staff member. The IT infrastructure at WACC provides the best environment for hosting the server as well. The server room operated at WACC is access controlled with two levels of locked doors, surveillance, and alarm service. This room is also temperature and humidity controlled. The server room is equipped with a full size rack infrastructure allowing the equipment to be mounted well above the ground to prevent any damage from a possible flood. The facility is wired to a backup generator with the server room receiving priority from the generator. All the servers at this facility are regularly monitored by the IT specialist and all equipment is configured to provide 24 hour alerts in the event of a failure of any type or magnitude. Multiple staff members are currently setup to receive these alerts. A sophisticated backup plan and hardware is currently in place to support the existing and future servers at the facility.

The main technology cost of implementing a solution at the WACC facility to meet the needs of LIBI for digital asset management is the purchase of a dedicated server. Sufficient rack space and power is available to house this server. WACC staff currently estimate the price of purchasing a server equipped to meet these needs between $8,000 and $10,000.

The WACC facility also offers the advantage of being able to host more parks' data on these servers, allowing for a centrally located and centrally managed system. WACC already has the necessary staff to provide the support for this system. WACC also already has a backup system that performs daily, weekly, and monthly backups which are then stored offsite at a secure facility.

[1] OAIS (Open Archival Information Systems Reference Model — ISO 14721:2003) provides a generic conceptual framework for building a complete archival repository, and identifies the responsibilities and interactions of Producers, Consumers and Managers of both paper and digital records. http://www.dcc.ac.uk/resources/curation-reference-manual/chapters-production/using-oais-reference-model-curation

The cost of implementing a solution at LIBI to meet the needs of this plan is much higher. Beyond the previously mentioned $8,000 to $10,000 hardware cost are numerous other physical and human resources costs not currently budgeted. The facility would need to be able to support a secure, temperature and humidity controlled server room. A one FTE systems administrator, in addition to current IT staff is required to fully support this system. Given the significantly increased data storage requirements, the park would also need to upgrade and expand its backup system to one capable of meeting these needs. Estimates supplied by WACC for backup hardware range from $5,000 to $8,000. The software for the backup system would cost between $1,500 and $2,000. Both of these components require regular attention as well as the media needing to be transported offsite weekly, if not daily, to a secure offsite facility. The cost for the equipment and necessary software alone would cost LIBI between $14,500 and $20,000, recurring every 3 to 5 years. This hardware cost would be in addition to the salary of an IT specialist. The benefit and economical savings in hosting the digital asset management system at WACC is that most of these expenses are already regular expenses at the WACC facility. The necessary backup software is already in place as well as the necessary backup hardware. The IT specialist is already in place at the facility. The only additional cost would need to be the server and an incremental (for WACC) increase in disk storage. The additional benefit to hosting the system at WACC is the ability to expand the system to multiple parks without replicating expensive infrastructure.

Policy Development

A digital asset management plan consists of both a technical infrastructure and a set of policies. These policies insure that digital assets are treated as the valuable assets they are.

It should be park policy that:

- all digital master files and metadata be contributed to the central repository
- no digital master files should be kept on local systems except those awaiting transfer to the central repository
- once digital master files are verified in the repository, copies must be deleted from local systems
- access derivatives should remain on local systems only while active
- inactive access derivatives should be deleted immediately or according to a predetermined schedule

Research Access and Hi-resolution Images:

Research access is expensive to create and maintain, and the cost recovery is a means to support the creation and maintenance of these digital assets. Without some amount of cost recovery it would not be possible to make these resources available. At this time, WACC is not authorized, nor does it have the infrastructure or policy framework to allow open access to high resolution images. Therefore, until such time as it becomes policy to allow open access, and budgetary resources are supplied to create, maintain, and fulfill requests for these resources, cost recovery will continue to be used to support the digital collections. The request and fulfillment process for this activity is detailed below.

Potential Infrastructure:
Central management of a preservation-oriented digital repository (at least among a group of regional parks) could be headquartered at WACC in Tucson. Expanding the use of ICMS can support metadata management in the short run. Some research is required to determine if it is possible to automate the process of metadata update and exchange and while this is an ultimate goal it is not envisioned as an immediate goal. The WACC will require hardware and systems enhancements mentioned previously to manage the quantity of data and provide TRAC-compliant, preservation-oriented storage.

As noted above the staff access system would require a new server hosted at WACC. This server would hold the program data for managing the digital assets. It would be connected to the NPS ESN to provide access to LIBI data from their park via an internal website. This server would be configured with a daily, weekly, and monthly backup plan, utilizing existing backup hardware and software at WACC. This server could be expanded to support other parks as well in addition to LIBI.

Additionally, an on-site public access server that exists separately from the NPS ESN could be connected to a public kiosk terminal. The server could be located anywhere, but should remain separate from the internal NPS ESN system. Remote software could be used for the management of this isolated network. Files would need to be transferred from the internal only database to the public onsite database to grant the public access to new assets. This could be done in a variety of ways, as long as the two systems ultimately remain separate. This public kiosk terminal would be placed in visitor locations to give access to a limited database of digital assets. More research is required to determine the exact configuration and cost of such a system.

The public internet access would require collaboration from the webmasters of the NPS. Current NPS web tools could be leveraged to support this function without the need to purchase dedicated equipment.

Implementation
This plan assumes that the core recommendations above are supported by the park and WACC. The plan can be implemented in phases, each phase supporting a wider scope of management and access.

Phase I: Staff only access, “testbed” content
In this phase three previously digitized collections (Elizabeth Bacon Custer, Seventh Cavalry Documents, and the historic photograph collection) with associated metadata would be ingested into the WACC DSpace application. This application will be made available to staff only for research, discovery, and access and will test the value of the search and retrieval interface, as well as the quality of metadata for retrieval. Staff can view all resources, and request hi-res copies through a separate research request mechanism.

Successful implementation requires (responsible unit in parentheses):

- Identification and export of clean metadata that meets current content standards (WACC)
- Refine and revise request mechanism and procedures workflow to support request and delivery of electronic content (LIBI and WACC)
- Identification of a "authorized person" at LIBI who has authority to request hi-resolution files (LIBI)
- Identification of LIBI staff who will act as testers and reporters relating to the usefulness of the system (LIBI)
- Graphic design and template configuration of DSpace implementation to mirror other official NPS web services (WACC—either through in-house staff or outsourced contract, cost estimate for this piece is beyond the scope of this report.)

Phase II: As the system is refined, it can be made available to public researchers at the White Swan Library. Researchers would be supported by library staff and would request reproductions or digital copies through the library staff.

- Researchers would access the testbed content through the same web interface the staff uses. The workstation would be a public access station not connected to NPS file system services. Supervision by LIBI staff during this activity will enable LIBI to gather feedback from researchers about the features and functions of the system. (LIBI)

- LIBI staff will prepare a request form for researchers to complete to order reproductions. This could simply be a paper form filled out by the researcher that mirrors current practice. (LIBI)

- LIBI staff will communicate the request to WACC through the same channels as in Phase I. A mechanism for delivery to the researcher would be developed that builds on the process developed in Phase I (WACC and LIBI)

Phase III: Public access to low resolution digital copies and to finding aids. Low resolution copies of digital objects supplied to NPS Focus can form the basis for general public access to LIBI digital content. NPS Focus can serve as the primary remote public access tool for preliminary research. Also, catalog records in MARC format for PDF finding aids can be contributed to NPS Voyager with a link to the HTML page for the list of finding aids at LIBI. Providing a link in each record to the HTML page for the list of finding aids rather than to a specific finding aid lowers administrative costs for creating and maintaining links and makes it possible to globally update finding aid records in NPS Voyager when necessary.

- Export metadata from ICMS for upload to NPS Focus (LIBI or WACC)
    - Develop an export template that meets NPS Focus metadata standards
    - Test export and upload with NPS Focus staff (LIBI or WACC and NPS Focus staff)
    - Export could be done centrally from WACC for all participating parks, or individually by each park.

- Export metadata for finding aids records for upload to NPS Voyager (LIBI or WACC)
    - Use collection level metadata in ICMS to create MARC record for NPS Voyager finding aid record (LIBI or WACC and NPS Voyager staff)
    - Test export and upload with NPS Voyager staff (LIBI or WACC)
    - Create finding aid page on LIBI site to link to MARC records (LIBI)

Assessment

During each phase, specific information relating to functionality and service should be collected. Development of this assessment process is beyond the scope of this report however some general suggestions can be made. For example, during Phase I, data on the following activities could be collected:

- time between request and fulfillment of hi resolution digital files
- quality of metadata for discovery—is metadata detailed enough? Consistent in content and use?
- quality of the search mechanism. Does the discovery system provide sufficient and desired functions such as search limiting, faceted results, truncation, etc?

Phase II information gathering could include:

- ability of researchers to perform mediated and unmediated discovery
- time required to support researcher training in the system
- increase or decrease in number of reproduction requests
- nature of reproduction requests—paper or digital?

Phase III information gathering might include:

- increase or decrease in requests for in-person access to collection materials
- increase or decrease in digital reproduction requests

Retrieval requesting:

Increased access to the collections will mean increased use of the collections and increased demands on staff at the White Swan Library. Researchers will arrive with more knowledge of resources available and with more detailed questions. Library staff should be able to identify needed material to the lowest level of description: item number, or collection, box, or folder number. The library staff member is the mediator between the patron and WACC. **WACC will support specific retrieval requests from park library staff only, and will not do research in the collections or identify collection material of potential use to researchers.** This work is the province of the staff of the White Swan Library.

Conclusions

It would be overly ambitious and cost-prohibitive to attempt to develop an OAIS-based, preservation-oriented repository system for NPS research data at LIBI. Collaborative development of a central storage system based at WACC can produce a system to meet

the digital asset management needs of the LIBI in a more cost effective manner. This project can serve as a proof-of-concept for central management and local consumption of digital assets for interpretation and research. Based on the experiences related to this project, especially those related to policy development and enforcement and the ability to exchange data between WACC and LIBI it should be possible to move to the next level preservation system. Meanwhile, valuable digital assets will be consolidated and responsibly managed, and a culture of cooperation will be built that can ensure a continued and expanded trust relationship between WACC and the regional parks.