



Smithsonian

## Unit Digitization Plan

for

**NMNH**

Unit Name

Approved by:

\_\_\_\_\_  
Unit Director (Signature)

\_\_\_\_\_  
**Kirk Johnson**  
Unit Director (Print Name)

\_\_\_\_\_  
Date

(Copies of this plan are to be submitted to the Digitization Program Office. See UDP User Guidelines for submission process.)

# Unit Digitization Plan

## Table of Contents

|   |           |
|---|-----------|
| <b>1. Executive Summary.....</b>  | <b>3</b>  |
| <b>2. Strategic Context .....</b>   | <b>5</b>  |
| <b>3. Scope.....</b>  | <b>9</b>  |
| <b>4. Goals .....</b>   | <b>11</b> |
| <b>5. Capacity.....</b>   | <b>26</b> |
| <b>6. Priorities .....</b>  | <b>36</b> |
| <b>7. Relationships and Dependencies .....</b>                                  | <b>41</b> |
| <b>8. Policies and Standards .....</b>  | <b>43</b> |
| <b>9. Unit Digitization Program Administration and UDP Plan Management.....</b> | <b>46</b> |
| <b>UDP/DAMP Glossary of Terms .....</b>   | <b>48</b> |

## Revision History

| Version | Date       | Pages affected | Description of Change | Author/Reviewer                      |
|---------|------------|----------------|-----------------------|--------------------------------------|
| 1.0     | 04/11/2013 | All            | First draft           | NMNH Digitization Steering Committee |
|         |            |                |                       |                                      |
|         |            |                |                       |                                      |

## 1. EXECUTIVE SUMMARY (MANDATORY)

The Executive Summary provides a brief, “at-a-glance” snapshot of the most important information in the UDP.

*Consider what information you would like to communicate to senior administrators when writing this section. Also consider that it may be easiest to complete this section at the very end, when you have content from other sections that you may wish to “cut and paste” into an Executive Summary narrative.*

The collections of the National Museum of Natural History (NMNH) document the most diverse and complex knowledge systems known: Earth's natural history. Since opening in 1910, we have amassed collections numbering 127.5 million objects. This represents almost 93% of the Smithsonian's holdings. We have already independently and cooperatively undertaken the digitization mission and as such are represented well to the research community and public at large with approximately 6 million digital records available through the NMNH's [collections site](#), the [SIRIS](#) online catalog, and the Smithsonian's [Collections Search Center](#). Despite great strides in contributing to build a digital presence for the Smithsonian Institution, the size and complexity of the NMNH collections pose uniquely difficult challenges compared to other units. The reader will find that the following unit digitization plan is as complex as the collections it documents.

Despite the obstacle of vast numbers, the overwhelming opinion of the NMNH research and collections community is that digitization efforts should be driven by quality rather than quantity, and compromising the prestige and reputation of the Institution was not an option in meeting the digitization mission. We estimate that approximately 16.5<sup>1</sup> million descriptive and surrogate records are needed to adequately represent the 127.5 million objects in the collection. These records would be the minimum effort

---

<sup>1</sup> This figure was based on an incremental goal; in 2013 an extensive analysis will be done to improve the estimate of anticipated records, an outside number for this total is ~33 million.

to demonstrate the overriding priority of a research and collections management plan of both maximizing local and collaborative expertise and the delivery of the most value-added data. This will be an incredible investment of time and resources. The plan then is subdivided into three incremental goals in hopes to make otherwise untenable aspirations not only feasible, but manageable, and allow for agile development in the face of almost certain financial limitations. These broad goals are: 1) target digital record creation for identified objects and their associated components, 2) provide digital collections information to the general public in novel and informative ways, and 3) create unit-wide infrastructure and methodology to support digitization efforts among all departments. In addition to the overarching goals, we determined that our target audience can be separated into three external groups (researchers and scholars, K-12 students and the general public) and a fourth, internal group representing the unit staff.

Each project priority and audience targeting decision is based on one of the three broad goals. For example, we have highly prioritized the need to generate descriptive records and to digitize at risk specimens to meet the first broad goal and for the research/scholar audience. Additionally, we have highly prioritized the need to make collection materials available for wider educational purposes including collection materials selected to facilitate exhibition activities, which meets the second broad goal and reaches our other priority audiences of K-12 students and the general public. As a rule, projects predetermined to have a lower priority are less targeted, less finite, and address our potential response to a spontaneously arising activity or unforeseen revenue stream. We attempted to address all contingencies and all the complex facets of NMNH.

In order to fulfill our third broad goal, infrastructure and communication, we will strongly leverage the wide-ranging skills of our staff. To centrally manage and coordinate the diverse digitization activities at NMNH, a steering committee was established in October 2011. The NMNH Digitization Steering Committee (DigiComm) has an inclusive membership that has representation from all scientific collecting departments, Education & Outreach, Office of Exhibits, Smithsonian Institution Libraries and onsite Affiliated Agencies. It is co-Chaired by the heads of Collections and the Informatics Branch (NMNH IT) to ensure a broad and comprehensive perspective on digitization. Research scientists will need to collaborate with IT professionals to develop new technologies for rapid digital data capture. Collection managers will have to inform and be informed by exhibitors to enhance communication of difficult concepts to target audiences. They also form the bridge between digitization planning at the institution and unit levels. Despite other considerations, meeting the order of magnitude of our digitization needs will be impossible without proper planning and development. Above all else, the fulfillment of our digitization plan depends on the people and the processes.

## 2. STRATEGIC CONTEXT

The Strategic Context section articulates how digitization supports unit and pan-Institutional goals, and ensures that digitization efforts serve these overarching commitments.

### 2.1 SI Unit Vision & Commitment (MANDATORY)

*Briefly explain how your unit's vision, as articulated in your unit strategic plan, is supported by digitization.*

NMNH aims to make existing electronic collections records available via the web, and eventually to digitize all scientifically valuable items in our collections as well as materials from our archives. Natural history collections, however, often consist of assemblages of essentially identical items with identical data—for example potsherds from the same pot or multiple individuals of the same species or mineral with identical collection data. In such cases, the assemblage is known as a “lot” and would be digitized as a single “item”. In other cases, one digital record is created for each physical object. Viewed in this way, we estimate that NMNH’s 127.5 million physical specimens will result in approximately 16.5<sup>2</sup> million records that therefore represent all the scientifically distinct items in our collections. Clearly, reaching the 16.5 million record target within the three years covered by this plan using existing resources is unrealistic. Therefore, this plan sets achievable goals with a comprehensive, yet synoptic, digital overview of the entire collection (including backlogs), and aims to complete digital inventories of particularly important and worthwhile subsets of the collection. The collections will continue to grow due to the museum’s ongoing research and collections activities. Accordingly, the target will move each year.

As a guiding principle, NMNH has decided that it is more important to create good records (both descriptive and surrogate) in the first instance than to create many records of lackluster quality or completion. In other words, ‘better is better’, rather than ‘more is better’. Higher quality records contain greater content such as sustainable digital surrogates or geo-referencing data – and these ‘enhancements’ directly increase their use and re-use by other initiatives. While the decision to choose quality over quantity affects projections of what can be accomplished by the Museum each year, it ensures that the records that are created or migrated are readily usable by the public with a minimum of re-work. Value-added records benefit all users, including the numerous aggregators of NMNH content as detailed in Section 7.2.

---

<sup>2</sup> This figure was based on an incremental goal; in 2013 an extensive analysis will be done to improve the estimate of anticipated records, an outside number for this total is ~33 million.

## **2.2 Background: Unit Digitization Efforts to Date (RECOMMENDED)**

*When did your unit first begin its digitization efforts? For what purpose(s)? How have these efforts grown, evolved or been built upon? Do you have a formal unit digitization program?*

In 1963, NMNH established a committee dedicated to Automated Data Processing (ADP) which later became a formal program and would eventually morph into the NMNH IT office. ADP provided support, guidance and computer resources for researchers and collections managers in order to meet NMNH's growing needs in a digital environment. The benefits offered even by the early computers and software available at that time were clear and the museum began to investigate and plan for the eventuality of transitioning from paper based collections management to databases and electronic storage.

In 1967, NMNH received a grant from Health Education and Welfare (HEW) to explore the application of computers to the management of biological, cultural and mineralogical collections. This grant funded the first attempts at creating a collections management system. Beginning in 1970, NMNH provided the funding for OCS (now OCIO) to hire three programmers who dedicated their time to the creation of a new digital catalog system using the lessons learned by the HEW project team who provided input into the design. The resulting system was called SELGEM (Self Generating Master) and was made freely available to the global museum community (we would call it OpenSource today). SELGEM ran off a mainframe, managed centrally by the Smithsonian. NMNH personnel, and later the entire SI, would take punch cards to the Castle and run them to update and create new records – all with a 64kilobyte maximum.

NMNH began the lengthy migration from SELGEM to INQUIRE in the early 1990's. Then in 2002, NMNH initiated the migration to the current collections management system, EMu. During the migration to EMu, all original data was preserved as-is in a legacy field, so it is possible to see the original SELGEM and INQUIRE records where present.

As for digitization beyond catalog records, the museum has a long tradition of using new technology as it becomes available. It would be very difficult to determine when the first photographic prints were digitally scanned, or the first born digital images were taken. However, there are several examples of this content still extant within NMNH systems that date to the 1970's and beyond.

To centrally manage and coordinate the diverse digitization activities at NMNH, a steering committee was established in October 2011. The NMNH Digitization Steering Committee (DigiComm) has an inclusive membership that has representation from all scientific collecting departments, Education & Outreach, Office of Exhibits, Smithsonian Institution Libraries and onsite Affiliated Agencies. It is co-Chaired by the heads of Collections and the Informatics Branch (NMNH IT) to ensure a broad and comprehensive perspective on digitization.

### 2.3 Unit Alignment With SI Digitization Strategic Plan 2010-2015 (MANDATORY)

How does the unit's strategic plan align with the Smithsonian's Digitization Strategic Plan 2010-2015 in its goals to:

- Provide access to SI collections, research, and programs by creating, managing and promoting the Institution's digital assets? (See Page 11 of the Strategic Plan);
- Integrate digitization into its core functions? (See Page 12, Objective 3 of the Strategic Plan);
- Secure sufficient resources and build capacity to create and sustain digitization activities? (See Page 13 of the Strategic Plan).

If your unit's plan does not align with the goals of the SI Digitization Strategic Plan, state which goal(s) are not addressed.

[The NMNH Strategic Plan](#) covering 2010-2015 is in strong alignment with the SI Digitization Strategic Plan. Additionally, the SI Strategic Plan goals are supported by the activities of the [NMNH Digitization Steering Committee](#), the [NMNH Collections Program](#) and policies, as well as through the extensive support provided by the [NMNH Office of Information Technology](#) (the largest Unit-based IT office).

The NMNH Strategic Plan supports the SI Digitization Plan goal to "provide access to SI collections, research, and programs by creating, managing and promoting the Institution's digital assets". NMNH Strategic Goal 1: "Explore and Interpret Nature and Culture," Objective: "Build and Improve stewardship and accessibility of collections" directly addresses these aspects of the SI Strategic Plan in the follow statements:

- Improve collections conservation status, based on existing profiles and periodic assessments.
- Augment the digital collections records for primary type specimens to contain appropriate imagery and geo-referenced data.
- Digitize collections of national significance and make them available online.
- Create a virtual natural history collection on the Web, including objects on display in our exhibitions, in support of education programs.

As part of the plan, NMNH identified ten areas of Organizational Change that should be addressed to assist in achieving the stated goals, including:

- Expanded Access: New technologies create new opportunities to capture and share knowledge. We will continue to digitize our collections to reach broader audiences virtually, and we will also improve and expand physical access to our exhibitions and programs for all on-site visitors.

- Streamlined Procedures: We will work to improve the efficiency and effectiveness of our work by proposing and adopting new procedures and technologies to simplify our activities and extend their impact.

NMNH has both a Strategic Plan as well as a [Digitization Strategic Plan](#). The former focuses on the key concepts and programmatic priorities, while the later focuses on the infrastructure needed to support digitization, priority projects, and workflows. Therefore, the NMNH Strategic Plan does not directly address the SI Digitization Strategic Plan proposal to “integrate digitization into [our] core functions”, or the “secure sufficient resources and build capacity to create and sustain digitization activities”; instead those concepts are directly addressed in the NMNH Digitization Strategic Plan (2010-2015). The preexisting Digitization Strategic Plan is now superseded by this UDP. NMNH further supports the integration of digitization into the core functions of the museum through the efforts of the NMNH Digitization Steering Committee. Additionally, the IT Office supports NMNH’s needs regarding resources for digitization workflows. Documentation can be found in the annually updated Informatics Overview Document.



### 3. SCOPE

The Scope section ensures that there is clarity about the specific materials and activities covered in the UDP. All questions in the UDP need to be answered consistent with the digitization scope outlined in this section.

#### 3.1 Scope of Plan and Description of Collections (MANDATORY)

*This plan must address digitization of SD600 collections but it may also address digitization of other collections or materials that are not covered under SD600 (for example, publication materials, collections used in educational programs, conservation materials, performances, events, etc.) It should **not** address business records of the institution: these are covered under SD501 Archives and Records of the Smithsonian Institution.*

*In this section, broadly characterize the materials that are covered by this UDP and briefly identify what you will be digitizing in each of these categories.*

##### 1. SD600 collections

- Collection records
- Descriptive records
- Finding aids
- Specimens/objects
- Taxonomic inventories
- Documentation such as field notebooks and ledger books
- Illustrations
- Archival materials

##### 2. Non-SD600 materials

- Conservation records
- Illustrations
- Maps
- Publications and manuscripts
- Exhibit materials

### 3.2 Out of Scope (RECOMMENDED)

*Apart from business records (see 3.1 above), briefly note which materials explicitly are **not** covered by this UDP.*

Research Data will not be included until SI has an official OpenAccess policy in place.

Collections Transactions records (i.e. registration records) are considered to be business records by NMNH and are therefore excluded.

Tissue records will not be included in this UDP until NMNH completes its plan for the management of tissues and tissue records.

## 4. GOALS

The Goals section clarifies what your unit wants to achieve in the area of digitization, and how you'll get there. It identifies your digitization goals, the activities you'll undertake to meet these goals, the audiences your digitization activities address, and what impact you'd like to make with your digitization efforts.

### 4.1 Goals (MANDATORY)

*Articulate your unit's digitization strategy and goals by presenting your unit's key milestones in a high-level 3-year digitization roadmap.*

As of early 2013, NMNH has over 6.14 million catalog records and just over 1,000,000 digital surrogate records<sup>3</sup> in its Research and Collections Information System (RCIS) and plans to continue the migration of records from remaining legacy systems to RCIS; to capture new records and digital images; and to enhance and standardize data in current records. All scientific collections have available records through the [EMu public web](#). Approximately 16.5<sup>4</sup> million records are needed to adequately represent all of the near 127.5 million objects and specimens. This includes the National Anthropological Archives (NAA) and Human Studies Film Archives (HSFA), which collectively hold 13,394 cubic feet of material, represented by approximately 11,000 descriptive records and 85,000 surrogate records that are made available through the Smithsonian Institution Research Information System ([SIRIS](#)) online catalog and feed into the [Collections Search Center](#).

As a guiding principle, NMNH has decided that it is more important to create good records (both descriptive and surrogate) in the first instance than to create many records of lackluster quality or completion. In other words, 'better is better', rather than 'more is better'. Higher quality records contain greater content such as sustainable digital surrogates or geo-referencing data – and these 'enhancements' directly increase their use and re-use by other initiatives. While the decision to choose quality over quantity affects projections of what can be accomplished by the Museum each year, it ensures that the records that are created or migrated are readily usable by the public with a minimum of re-work. Value-added records benefit all users, including the numerous aggregators of NMNH content as detailed in Section 7.2.

---

<sup>3</sup> NMNH counts all digital surrogates including images of catalog cards, specimen labels, ledger scans, field notebooks and illustrations as they have legitimate uses within the museum community and reflect a large portion of our digitization efforts. If the DPO definition of "surrogates" is used, then the number of NMNH digital surrogates is ~500,000, a more than 50% reduction in the reporting of the actual digitization effort.

<sup>4</sup> This figure was based on an incremental goal; in 2013 an extensive analysis will be done to improve the estimate of anticipated records, an outside number for this total is ~33 million.

Tension naturally exists when planning for either what can be accomplished ‘reasonably’ with existing resources, and what should be accomplished separate from any resource considerations. NMNH has chosen to aspire to the ultimate target regardless of financial limitations while identifying 3-year incremental goals. These incremental goals are drawn from the CDRS, NMNH Strategic Plan, and identified risks, and are contingent on funding.

Three goals establish the roadmap with milestones to digitizing strategic portions of the collections, enabling public uses of digitized records, and ensuring that the necessary infrastructure of hardware, software, systems and expertise is available to support these actions.

**Goal 1: Create a digital record for identified collections and their associated components for all items in target collections: move relevant collections from level 1 to levels 2 and 3.**

NMNH’s departments have digitized their collections to varying degrees. Some have a basic record for virtually every item but need images; others would not benefit from images per se but could enrich some or all records. Others still need to create basic records for the most important specimens. The framework is intended to guide the museum so that standards are eventually met, regardless of the starting point of each department or identified collection. As a general rule, all records should be at an acceptable level of completion (level 3). Specific targets are identified based upon comparisons of Importance vs. Information rankings in NMNH’s Collections Assessment, and by the direction of NMNH leadership. An important element of Goal 1 is to digitize records so that species representation across NMNH’s holdings can be known.

### ***SD-600 Collections***

**a) Create new digital records in RCIS and SIRIS.** Descriptive records include a variety of information depending on the level of granularity of the catalog in question; currently NMNH generates records at the specimen/object level (Type specimens and Specimen/object catalogs) and at the collection level (species inventories for biological science collections and archival finding aids and MARC Machine Readable Catalog records for both the NAA and HSFA archive). Creation of new records might be accomplished by migration of legacy datasets, and by direct data entry into the system.

#### *Anthropology targets:*

- Development of a global online finding aid for North American Language materials as part of Recovering Voices Initiatives.  
Year 1: 20 %. Year 2: 50%. Year 3: 70%

- Migration into EMu of 34,650 conservation treatment records from the Anthropology Conservation Laboratory (ACL). Funded through CIS-IRM Pool 2012. Year 1: 80 %. Year 2: 20%. Year 3: 0%
- Creation of 1067 new descriptive records from the Valdivia collection from Ecuador. Year 1: 58%. Year 2: 28%. Year 3: 14%

*Botany targets:*

- Create 104,000 new descriptive records of important plant families that are staff specialties (Euphorbiaceae, Araliaceae/Apiaceae, Onagraceae, Malvaceae/Bombacaceae and Tiliaceae). Year 1: 20%. Year 2: 40%. Year 3: 40%
- Create 223, 915 new descriptive records for major floras projects (Lichens, Pacific Islands, New World Grasses, and Biodiversity of Mexico). (~47 %). Year 1: 20%. Year 2: 40%. Year 3: 40%

*Education & Outreach Targets:*

- Creation of 10,000 EMu specimen records for Q?RIUS collection. These include records cataloged on paper by Naturalist Center staff and employees. Year 1: 33%. Year 2: 33%. Year 3: 33%
- Creation of 3000 EMu specimen records for E&O collections in the Discovery Room, Exhibit Cart Program, and Forensic Laboratory. Many of these records currently are in an Access database. Year 1: 33%. Year 2: 33%. Year 3: 33%

*Entomology targets:*

- Migration of a backlog of 25,000 records of type specimens identified in legacy systems (parasitic wasps, moths, butterflies, flies, beetles and true bugs). Year 1: 25%. Year 2: 75%.
- Capture of descriptive records to complete the cataloguing of the type specimens of 1,600 ants and 1,900 moths. Descriptive records are available for close to 50% of the ants and 85% of the moths, this project will complete 100% of the cataloguing of these two taxonomic groups. Funding requested CCPF13 and NMNH Student collection internships. Year 1: 20%. Year 2: 80%
- Migration of 5,500 descriptive records of specimens of valuable butterflies. Year 1: 100%
- Migration of 130,000 descriptive records of the Species Inventory identified in legacy systems from across the collection. These records display a varied degree of completion, and most of them are expected to need standardization in the near future. Year 1: 50%. Year 2: 50%
- Complete pre-capture preparation of the collection and capture of basic information - species name, author, year and preparation – of Species Inventory of moths (Sphingidae, Saturniidae and North American Geometridae), beetles (bark beetles) and Collembola (springtails). Year 1: 33%. Year 2: 33%. Year 3: 34%
- Complete capture of basic records for main collection across all taxonomic groups. Not possible to estimate the number of records at this time. Highly dependent of extra resources not available yet. Year 2: 20%. Year 3: 80%

*Invertebrate Zoology targets:*

- Create new descriptive records for 1,600 drawers of the Land Snails dry collection in preparation for a move from MSC to NHB. Project funded. Year 1: 80 %. Year 2: 20%

- Create new descriptive records for 869 cases of the Mollusk dry collection in preparation for an internal move within NHB. Assistance requested from NMNH-Collection Program Technicians. Year 2: 60%. Year 3: 40%
- Cataloging Department of Interior (BOEM) collections (on-going contract from DOI). Year 1: 8,000 records, Year 2: 8,000 records, Year 3: 8,000 records
- Create 24,000 new descriptive records of the voucher collections from benchmark surveys of various off-shore sites, targeted for oil drilling. Funding for cataloguing available through Department of Interior. Year 1: 33.3 %. Year 2: 33.3%. Year 3: 33.4%

*Mineral Sciences targets:*

- Create 20,000 descriptive records of rock specimens from the Princeton Collection. Year 1: 50 %. Year 2: 25%. Year 3: 25%

*Paleobiology targets:*

- Create 9,000 descriptive records for the Brown-Rose collection. Currently there are 195 records of the estimated 20,000 records needed to complete this project. Year 1: 33%. Year 2: 33%. Year 3: 34%
- Create 5,000 descriptive records from Paleobotany collection: Cleared leaves (2,500) and Pre Camb-Devonian (3,500). This project will complete 25% of the cleared leaves and 100% of the Pre Camb-Devonian collections. Year 1: prepare data to be migrated. Year 2: migrate 50%. Year 3: migrate 50%
- Create 2,000 new descriptive records of the Ostracoda Biological collection, which currently has 50% of the expected records already in EMu. Year 1: 33%. Year 2: 33%. Year 3: 34%

*Vertebrate Zoology targets:*

- Create 100,000 new descriptive records from backlog processing and data capture from ledgers of fishes (25,000), amphibians and reptiles (25,000), birds (25,000), and mammals (25,000). These are on-going projects that rely on volunteers, support staff and extra resources. Year 1: 33 %. Year 2: 33%. Year 3: 34%

**b) Create new surrogate records in RCIS and SIRIS.** Surrogate records include images and audio files. Creation of new records might be accomplished by migration of existing image or audio files into the system, and capture of images of specimen/objects prior to uploading into the system. The following table details the targets by department.

*Anthropology targets:*

- Create 10,000 digital surrogates for slides of before and after treatments for the ACL descriptive records. Year 1:50 %. Year 2: 50%.
- Capture 4,975 digital images of objects in the Catlin (175), Harrington (600), Wilkes (3,000), and Southeastern Ethnology (1,200) collections. Partial funding requested. Year 1: 45%. Year 2: 30%. Year 3: 25%
- Capture 3-D images of 35 Tlingit objects prior to repatriation. CCPF Funding Requested. Year 1: 100 %.

- Creation of digital surrogates for 122 hours of deteriorating sound recordings for Human Studies Film Archives (HSFA), including creating metadata and ingesting into DAMS for accessibility into SIRIS. Funded through CCPF12. Year 1: 100 %.
- Create digital surrogates and enhancement of records for 3,000 sound recordings and 35,000 pages of manuscript materials. Funding for \$1 million from the Arcadia Fund for a 2-year project received. Year 1: 25%. Year 2: 75%
- Rapid Capture Capacity Pilot- Two-Dimensional Materials-Capture of 120,000 surrogates of catalog ledgers, historical manuscripts and oversized materials as a rapid capture project to be uploaded into SIRIS and EMu/DAMS/EDAN. Funded from Walter Scott Foundation. Year 1: 25%. Year 2: 75%
- Create digital images for 500 photographs in the Mooney collection. This project is in collaboration with an outside researcher for collection processing, scanning and addition to records of identifications provided by Cheyenne, Arapaho, and Kiowa tribal members. A proposal to NEH currently in development. Year 1: 100%.
- Digitization for preservation of 2,200 HSFA legacy video tapes. OP&A (Office of Policy and Analysis) and DPO (Digitization Program Office) SI wide Audiovisual Digitization Study to be presented to Smithsonian administration for seeking outside support. HSFA is contributing to this study and is recommending that the most urgent need is magnetic media with a focus on legacy video tape because we are now in a critical window of opportunity to digitally preserve before lack of equipment and media degradation makes duplication impossible. Along with video we are including all deteriorating analog audio recordings that are also in critical need of digital preservation. Year 1: 34%. Year 2: 33%. Year 3: 33%

*Botany targets:*

- Capture images to enhance 50% of the new descriptive records to be generated. Year 1: 33 %. Year 2: 33%. Year 3: 34%
- Complete migration of live plant images to EMu, including live plant images of specimens and plant images donated by non-SI photographers. Year 1: 90%. Year 2: 10%.
- Complete migration of botanical illustrations to EMu, with associated plant specimen data. Year 1: 50%. Year 2: 50%.

*Education & Outreach Targets:*

- Create images for 20,000 catalog records (40-60,000 images) in the Q?RIUS collection. Year 1: 33 %. Year 2: 33%. Year 3: 33%
- Create at least 800 narratives for collections in Q?RIUS that describe multiple objects with at least one photo or video each Year 1: 33 %. Year 2: 33%. Year 3: 33%
- Create images for 3,000 EMu specimen records for E&O collections in the Discovery Room, Exhibit Cart Program, and Forensic Laboratory. Year 1: 33%. Year 2: 33%. Year 3: 33%

*Entomology targets:*

- Create approximately 40,000 surrogate records for images of type specimens and data labels ready to be migrated into RCIS (ants and parasitic wasps) or funded for capture during 2013 (moths, beetles, flies and true-bugs). Capture of digital images is ongoing across the collection with 34,754 surrogate records already in RCIS. Year 1: 50%. Year 2: 50%

- Migration of 7,500 digital images of one specimen of each sex when available for moths and butterflies to enhance descriptive records of specimens and Species Inventory. Year 1: 50%. Year 2: 50%

*Invertebrate Zoology targets:*

- Create 1,350 surrogate records for the dry mollusks collection, multiple images will be combined in plates for each record. Capture of digital images is ongoing with close to 30% completed. Year 1: 33.3 %. Year 2: 33.3%. Year 3: 33.4%
- Photograph Mollusk Primary types (30% done, on-going work by volunteer) Year 1: create 450 multimedia records, Year 2: create 450 multimedia records, Year 3: create 450 multimedia records

*Mineral Sciences targets:*

- Capture 22,500 digital images of exhibit specimens and high value minerals in the Blue Room to enhance descriptive records. Ongoing project with a backlog of at least 8,000 images in need to be uploaded into EMu. Year 1: 33 %. Year 2: 33%. Year 3: 34%
- Migrate into EMu 8,000 images of gem and mineral specimens to enhance descriptive records. Year 1: 100 %
- Capture 4,000 new images and migrate into EMu ~12,000 existing digital images of polished thin sections from Antarctic meteorites to enhance descriptive records. Year 1: 50%. Year 2: 50% of existing, plus 100% new. Year 3: 100% new images going forward.

*Paleobiology targets:*

- Capture 10,220 digital images of type specimens for foraminifers (5,700), trilobites (1,000), paleobotany (3,000), dinosaurs (417) and HVCI (103). Ongoing project with 4,081 images already in the system. Year 1: 33%. Year 2: 33%. Year 3: 34%
- Establish a 3-D specimen collection scanning program for the marine mammal collection. Currently scanning specimens for research and preliminary outreach. Year 1: 33%. Year 2: 33%. Year 3: 34%

*Vertebrate Zoology targets:*

- Create 32,000 surrogate records for images of fishes (12,000), birds (8,000) and mammals (12,000). Capture of digital images is ongoing across the collection with 33,724 surrogate records already in RCIS. Year 1: 33 %. Year 2: 33%. Year 3: 34%
- Capture 3D models for skull and skeletal elements of each primate species. Year 1: 50 %. Year 2: 50%

- c) Maintenance, standardization and data clean-up.** These tasks are an important aspect of our digitization activities as they either, create new records, facilitate future generation of records, archive images in DAMS, or move records from a substandard level (2) to an acceptable level (3).

*Anthropology targets:*

- Enhancement of descriptive records and data clean-up of the post crania material in the Terry collection. Year 1: 10%. Year 2: 40%. Year 3: 50%



- Migrate images of plaster busts into EMu and enhance information of descriptive records in the non-skeletal series of Physical Anthropology. Year 1: 15%. Year 2: 25%. Year 3: 60%
- Migration of legacy digital surrogates from NAA into DAMS for use in SIRIS and EMu. Funding requested from CIS-IRM Pool 2013. Year 1: 30 %. Year 2: 40%. Year 3: 20%
- Standardization of Mexican ethnonyms for SIRIS records. Funding pending. Year 1: 5%. Year 2: 40%. Year 3: 55%
- Enhancement of catalog information in Archaeobiology. Funding requested from CCPF13. Year 1: 50%. Year 2: 50%
- Processing and migration of NAA Born-Digital materials determination of files for DAMS ingest and establishment of accessioning guidelines and workflow. Ongoing project partially funded through CCPF. Year 1: 10 %. Year 2: 20%. Year 3: 70%
- Development of systematic plan for guidelines and integration into DAMS of audio and video materials generated as part of Recovering Voices initiatives. Year 1: 20 %. Year 2: 50%. Year 3: 30%
- Enhancement and standardization of NAA's collection-level catalog records for photograph collections (CIS and other grants-Ongoing project through a CIS-IRM Pool 2012 for the first year, funding from CIS-IRM Pool 2013 pending.). Year 1: 34 %. Year 2: 33.3%. Year 3: 33.3%
- Migration of NAA and Anthropology Collections images into DAMS that relate to treatments within the ACL database. Year 1: 40%. Year 2: 30%. Year 3: 20%

*Botany targets:*

- Cataloguing of new type specimens is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year.
- Cataloguing of new specimens from inventoried plant families is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year.
- Migrate specimen records from Cuba, Passifloraceae, Zingiberales, Marquesas, South America and Theaceae from non-standard databases in to EMu. Approximately 81,000 records. Year 1: 50%. Year 2: 50%.
- Enhance locality data for Cuban Collection Event records. Year 1: 100%.
- Enhance Collection Event records from the Guiana Shield with georeferencing codes. Year 1: 50%. Year 2: 50%.

*Education & Outreach Targets:*

- Standardization, cleanup and review of 20,000 Q?RIUS specimen records in EMu including data in Catalog, Taxonomy, Collections Events, Locations, Transactions and Parties Modules. Year 1: 33 %. Year 2: 33%. Year 3: 33%
- Cataloging of new specimens in E&O collection – an ongoing activity; count estimates per year not determinable.
- Complete data reconciliation of E&O Invertebrate Zoology specimens and update information in EMu. Year 1: 10%. Year 2: 90%

*Entomology targets:*

- Estimate the number of catalog records that will represent the close to 33.5 million specimens in the collection. Year 1: 100%
- Cataloguing of new dragonfly specimens is an ongoing activity performed as part of maintenance as they are received. The complete collection of pinned and papered dragonflies and damselflies is the only group that has been cataloged at the specimen level with 104,031 descriptive records. It is not possible to estimate an accurate number per year
- Cataloguing of new type specimens is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year
- Data clean-up and update information for 59,000 records and capture new descriptive records for the Diptera species inventory legacy dataset. Funding requested from CIS-IRM Pool 2013. Year 1: 10%. Year 2: 90%
- Establish standards and determine scope of work to standardize descriptive records available in EMu. Year 1: Define standards and determine scope of work. Year 2: standardize basic fields needed for reporting. Year 3: continue standardization of other fields

*Invertebrate Zoology targets:*

- Cataloguing of new lots is an ongoing activity performed as part of maintenance as they are retrieved from the collection. Collections to be moved will be targeted
- Cataloguing of new type specimens is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year
- Inventory Tissue Collections (dependent on the addition of “tissue” fields in EMu) Year 1: 10%, Year 2: 90%

*Mineral Sciences targets:*

- Standardize, append and clean-up data for existing ~14,000 descriptive records of Volcanic Glass. Year 1: 50 %. Year 2: 50%

*Paleobiology targets:*

- Cataloguing of new lots is an ongoing activity performed as part of maintenance as they are retrieved from the collection
- Cataloguing of new type specimens is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year
- Migrate Paleobotany Localities database to EMu. Year 1: prepare data for migration. Year 2: 50%. Year 3: 50%
- Migrate Green River database to EMu. Year 1: prepare data for migration. Year 2: 50%. Year 3: 50%
- Update 2,000 records in EMu from specimens in Exhibits, including new images. Year 1: establish workflow and 25%. Year 2: 37%. Year 3: 38%
- Establish standards and determine scope of work to standardize descriptive records available in EMu, standardization of records and establish collaborations for data sharing. Year 1: define standards and identify scope of work. Year 2: establish procedures, standardize 50% of records. Year 3: standardize 50% of records, begin collaboration for data sharing

*Vertebrate Zoology targets:*

- Cataloguing of new specimens is an ongoing activity performed as part of maintenance as they are retrieved from the collection
- Cataloguing of new type specimens is an ongoing activity performed as part of maintenance as they are received. It is not possible to estimate an accurate number per year

***Non-SD-600 Collections:***

- Anthropology: Migration of new CT scan data to MeSA for centralized back-up and distribution to departmental content owners. Planned migration of new CT scan data of human remains or other Anthropology collections into DAMS for preservation. For MeSA migration: Year 1: 100%. Following years 100% as data are generated. For DAMS migration: workflow under development
- Botany: Complete migration of live plant images to EMu, including live plant images of specimens and plant images donated by non-SI photographers. Year 1: 90%. Year 2: 10%.
- Botany: Complete migration of botanical illustrations to EMu, with associated plant specimen data. Year 1: 50%. Year 2: 50%.
- Vertebrate Zoology: Convert TIFF images into 74,000 PDF files of documentation materials (field books, journals and notes) to be uploaded as surrogate records attached to existing descriptive records in EMu for fishes (67,000), amphibians and reptiles (1,000), birds (3,000), and mammals (3,000). Year 1: 33 %. Year 2: 33%. Year 3: 34%
- Vertebrate Zoology: Digitize important ledger catalogs, radiographs and photographs to be attached to descriptive records in RCIS for fishes (5,000 Kodachrome slides), amphibians and reptiles (1,000 radiographs and photographs), and mammals (1,000 radiographs and 150 ledger catalogs). Year 2: 50%. Year 3: 50%

**Goal 2: Provide Museum collections information and data to the general public in novel and informative ways.**

We have been developing a number of innovative methods for the public to access our digital collections:

- **Q?RIUS:** In the new Q?RIUS education center, which will open to the public in the Fall of 2013, the public will be able to access the digitized education collection as well as any public records in EMu, through touchscreens and on a mobile-compatible website. The education collections records are being enhanced through extensive use of the Emu narratives and multimedia modules, providing public-friendly contextual information, including videos, associated researchers and links to more information for each object. The onsite and online interfaces provide many ways by which users can search and browse for objects, including location and time.

- **Ocean Hall:** Historically, there is usually not public access to digital versions to all the collections objects in an exhibit. For the Ocean Hall, each of the objects in the exhibit was entered into a simple spreadsheet so that docents and staff can easily find metadata about the objects. Sometimes, this data is accessed by staff via tablets in order to help visitors find objects. This effort was a pilot to test how exhibit objects might be accessed in the future.
- **Birds of DC** is a mobile-compatible website that allows museum visitors to access descriptive information, catalogue data, audio of bird calls, images and video related to the specimens in the Birds of DC exhibit. Prompts for users to access this information on their mobile devices will be integrated into the physical exhibit itself.
- **Smithsonian WILD** is a multi-organizational initiative to create a central repository for curating and safeguarding camera-trapped images (both legacy and citizen scientist contributed) and essentially treat them as specimens. Images are on both Flickr, and Smithsonian WILD. Later in the year, the eMammal website will be launched with data visualization tools for the public.
- **The Human Origins website** allows the public access to hundreds of 3D scanned artifacts and specimens.

**Goal 3: Create necessary infrastructure and documentation involved in the action of creating, maintaining and disseminating digital content.**

NMNH periodically needs to assess, identify, and integrate NMNH collections management systems that support on-going and new initiatives. When appropriate, new functionality should be added to existing systems to meet the needs of strategic goals in support of the NMNH Digitization Plan and Big Ideas - with primary emphasis in support of the target number of collection objects to represent the collection based on Goals 1 and 2 above, and to help realize NMNH Big Ideas, such as the Global Genome Initiative or outreach needs such as Q?RIUS. NMNH information resources will grow to meet the needs of its stakeholders, guided by NMNH and SI strategic goals. NMNH stakeholders include, but are not limited to, the scientific research community, collections, and Education and Outreach.

The following are known NMNH development needs based on feedback from the NMNH community. The development needs are divided into four sections: Ongoing Informatics Development Initiatives, Development Initiatives to be completed in Calendar Year 2013, Development Initiatives Planned for 2014 to 2018, and Completed Informatics Development Initiatives.

**Ongoing Informatics Development Initiatives**

- Database Migration
- Data Import Templates and Bulk Import

- Media Staging Area (MeSA)

#### **Development Initiatives to be completed in Calendar Year 2013**

- EMu-FreezerPro Integration
- RCIS Training Videos
- Transaction Management Integration Phase II
- Collections Level Index
- NMNH Collections Search Design – Intranet
- Implementation of the EMu Conservation, Events, and Narrative modules
- Use of EMu for Q&RIUS
- Integrated Publishing Toolkit (IPT)
- Connecting NMNH to Global Genome Biodiversity Network (GGBN)
- NMNH Collections Digital Record Assessment

#### **Development Initiatives Planned for 2014 to 2018**

- Develop and Implement MSNGR Aggregator
- Phase 2: Connection of EMu to the Smithsonian Digital Asset Management Systems (Audio/Video Assets)
- EMu Georeferencing
- Collection Level Description (CLD)
- Field Information Management System (FIMS)
- Report and Completion of Rapid Capture Capacity Pilot

#### **Completed Informatics Development Initiatives**

- Transaction Management Integration
- NMNH Collections Search Redesign – Internet(Phase I)
- Unified EMu Environment
- EDAN Cross-search
- Connection of EMu to the Smithsonian Digital Asset Management System
- Integrated Pest Management (IPM)
- EMu LDAP Integration via Centrify

From 2012 CDRS Annual Assessment Reporting (RCIS)<sup>5</sup>:

| DEPARTMENT                  | COLLECTIONS ITEMS COUNT per 2012 ACSR | ANTICIPATED RCIS RECORDS (or DB of record) | RECORDS COMPLETED as of Q1 FY 2013 | PERCENT COMPLETED OF RECORDS GOAL | RECORDS STILL NEEDED to REPRESENT COLLECTIONS | NUMBER OF SPECIMENS REPRESENTED BY RECORDS AS OF Q1 FY 2013 | PERCENT COMPLETED OF SPECIMEN GOAL | NUMBER OF DIGITAL SURROGATES | NUMBER OF RCIS RECORDS ON WEB |
|-----------------------------|---------------------------------------|--|------------------------------------|-----------------------------------|---|---|------------------------------------|------------------------------|-------------------------------|
| <b>Anthropology</b>         | 2,363,703                             | 577,589                                    | 540,389                            | 93.56%                            | 37,200  | 2,326,503   | 98.43%                             | 185,731                      | 439,738                       |
| <b>Botany</b>               | 4,904,071                             | 4,904,071                                  | 1,216,473                          | 24.81%                            | 3,687,598                                     | 1,216,826   | 24.81%                             | 152,041                      | 1,175,089                     |
| <b>Entomology</b>           | 33,385,565                            | 920,000*                                   | 234,961                            | 25.54%                            | 685,039                                       | 241,010   | 0.72%                              | 33,968                       | 228,516                       |
| <b>Invertebrate Zoology</b> | 33,897,667                            | 5,080,816                                  | 982,947                            | 19.35%                            | 4,097,869                                     | 8,497,482   | 25.07%                             | 21,764                       | 930,641                       |
| <b>Mineral Sciences</b>     | 687,066                               | 463,822                                    | 393,322                            | 84.80%                            | 70,500  | 596,123   | 86.76%                             | 4,045                        | 388,244                       |
| <b>Paleobiology</b>         | 42,727,932                            | 1,888,857                                  | 599,771                            | 31.75%                            | 1,289,086                                     | 7,814,358   | 18.29%                             | 18,776                       | 599,714                       |
| <b>Vertebrate Zoology</b>   | 9,608,405                             | 2,745,552                                  | 1,944,013                          | 70.81%                            | 801,539                                       | 3,988,866   | 41.51%                             | 33,724                       | 1,924,544                     |
|                             |                                       |  |                                    |                                   |   |   |                                    |                              |                               |
| <b>TOTAL</b>                | 127,574,409                           | 16,580,707*                                | 5,911,876                          | 35.66%                            | 10,668,831                                    | 24,681,168  | 19.35%                             | 450,049                      | 5,686,486                     |

\* This figure was based on an incremental goal; in 2013 an extensive analysis will be done to improve the estimate of anticipated records, an outside number for this total is ~33 million.

<sup>5</sup> Education & Outreach collections were classified as SD600 in April 2012 and therefore not reported in FY2012 CDRS. In future years it will be included.

From 2012 CDRS Annual Assessment Reporting (SIRIS):

| DEPARTMENT                    | NUMBER OF COLLECTIONS | ANTICIPATED COLLECTION LEVEL SIRIS RECORDS BY END OF FY 2013 | COLLECTION LEVEL RECORDS COMPLETED AS OF Q1 FY 2013 | PERCENT COMPLETED OF RECORDS GOAL FOR FY 2013 | RECORDS STILL NEEDED TO REPRESENT COLLECTIONS | NUMBER OF COLLECTIONS REPRESENTED BY RECORDS AS OF Q1 FY 2013 | PERCENT COMPLETED OF COLLECTIONS GOAL | NUMBER OF DIGITAL SURROGATES AT "STANDARD" AS DEFINED IN CDRS ASSESSMENT | NUMBER OF COLLECTION LEVEL IN SIRIS ON THE WEB<br><small>(DOES NOT INCLUDE ITEM AND SERIES LEVEL RECORDS)</small> |
|-------------------------------|-----------------------|--|---|---|---|---|---------------------------------------|--|---|
| <b>Anthropology: NAA/NSFA</b> | 13,394                | 12,022   | 100   | 63%   | 1432  | 11,962  | 89%                                   | 239,215  | 11,962  |

#### 4.2 Target Groups (MANDATORY)

Audiences that will benefit from digital access to our collections are varied depending on the collection characteristics. We have identified the three major external groups that drive NMNH digitization efforts as: Researchers and Scholars from the scientific, historical and academic communities, the General Public, and K-12; in addition there is a fourth group, internal to the Institution, the Collections Management staff.

**Researchers and Scholars:** The mobilization of information content of our collections will support a wide range of research in many disciplines. Examples:

- *NMNH Invertebrate and Fish Collection records were used to generate a map of Gulf of Mexico specimens (<http://ocean.si.edu/ocean-photos/map-gulf-mexico-collection>).*

This map shows the localities represented by the Gulf of Mexico collection of the Smithsonian Institution’s National Museum of Natural History. The Smithsonian has been collaborating with the Minerals Management Service (now Bureau of Ocean Energy Management, BOEM) since the 1970s to archive collections from marine surveys, resulting in more than 300,000 lots. It is estimated that the Smithsonian holds about 58% of all the publicly available specimen-based records from the Gulf of Mexico. This provides an important baseline for comparative surveys. The Museum’s Dr. Jonathan Coddington on June 15, 2010, provided testimony to a U.S. House Subcommittee that details the scope and significance of the collection. [Credit: the Smithsonian National Museum of Natural History Ocean Portal]

Invertebrate Zoology data from the BOEM collections were used to generate a map which may be seen on-line at <http://ocean.si.edu/ocean-photos/map-gulf-mexico-collection>

Station data from the BOEM collections have been digitized and made available to researchers at <http://invertebrates.si.edu/boem/boem.htm>

- By making our materials accessible to researchers it allows for new discoveries within the materials and across the collections, not only within NMNH but Smithsonian as a whole. An object held within the Anthropology collections might relate to a manuscript material held in the National Anthropological Archives. For example, a researcher discovered a particular image from a manuscript online, which facilitated his research on an archeological artifact. Based on this, he later visited the archives and also the Department of Anthropology's collections area where the three dimensional object resided. In the end, it provided new evidence for his research.

**General Public:**

- As an extension of NMNH's partnership with the Encyclopedia of Life (EOL), the Office of Education & Outreach created a new mobile experience for visitors wandering through the Birds of D.C. exhibit. Content is streamed from the EOL website, with specimen catalog information supplied by NMNH. The mobile website (<http://dcbirds.si.edu/>) contains audio, video, images, maps and text, so users can listen to bird vocalizations, watch their behaviors, see details of living birds in their environments, view their migratory and distribution patterns, and learn specimen information about the birds on display. The new content will take the visitor experience to a deeper level of appreciation for the birds in Washington, D.C. For the optimal user experience, visitors can connect their mobile devices to the new SI-Public wireless zone available in the Baird Ambulatory. This is the first ever instance of public Wi-Fi being available at NMNH for a dedicated public use case and it greatly enhances site access and responsiveness, which is especially critical to the mobile experience.

**K-12:**

- Middle and high school teachers and students will be able to access all of the digitized education collections onsite, through the Q?RIUS website, and through a mobile compatible website. Each object will have associated high-resolution images, descriptive public-friendly text, information about associated NMNH researchers, and multimedia. Students can use online fieldbooks to curate and annotate their own collections. Distance learning programs and online teacher guides will help teachers further use the digitized collections in classroom activities.



**Collections Management staff:** Cataloguing has been a collections management tool for many years. Starting with card catalogs and ledgers, we have recorded entries of specimens, collecting events, expeditions, list of species holdings, photographs and negatives. During the '70s and '80s some of those catalogs were continued in electronic form. Today, the NMNH RCIS is called EMu with a vast array of functionalities that allows us to manage the collections information at: collection, taxonomic, specimen and lots levels in the Catalog module; geographical levels in the Collection Events module; documentation about origins, images, audio and video in the Multimedia module; the names of collectors, donors and authors in Parties; bibliographical references of important specimens such as type specimens in Bibliography. Having access to all this information in a single system allows for automated reports and faster searches to answer many questions regarding holdings, preservation status, pest management, and legal status among others. Digitized collections data supports analysis of collections management needs and problems using contemporary tools.

The following case of particularly valuable specimens for scientific research exemplifies some of the benefits for collection management staff charged with their care.

- *Digitization of the Botanical Type Specimens*

Botanical type specimens form the descriptive basis of newly published plant names and thus are critical for conducting systematic and taxonomic research. Historically, access to these specimens was cumbersome for researchers (who incurred significant time and costs in identifying the repositories where particular type specimens resided) and often imposed a level of risk for the specimen (which frequently was sent out on loan to a researcher who could not travel to a repository to study the type specimen on site).

In 1970, the Department of Botany began developing the world's first digital catalog of type specimens, and augmented the online catalog data with high quality images starting in 2000. Today, over 110,000 type records and over 90,000 high quality images comprise our online Botany Type Register.

As researchers began to discover this resource, we experienced an 80% reduction in requests for loans of type specimens whose images were available. Researchers now find and study our type specimens online, which better serves them and removes the risk that loans incur on specimens. Thus digitization of specimen data and images offers a powerful solution to our conflicting missions of providing broad access to collections while ensuring their preservation in perpetuity. In addition, we are facilitating research for thousands of scientists worldwide and developing a growing public recognition of the value of our collections and our research.

## 5. CAPACITY

The Capacity section identifies the resources you're already bringing to bear on the digitization challenge, shows how far those resources take you, clarifies what resources are required to realize the ambitions of this UDP, and outlines a strategy for finding additional funding. This information will be crucial in assessing whether the pan-Institutional investment in digitization is appropriate, and provides arguments for increased resources (if warranted).

---

USE THE EXCEL WORKBOOK ENTITLED "UDP SECTION 5 CAPACITY WORKBOOK" TO COMPLETE THIS SECTION.

*Each of the questions in subsections 5.1 and 5.2 (with the exception of Question 5.2.3) has a corresponding worksheet in an Excel Workbook that can be found at <https://collab.si.edu/sites/SI/Digi/UDP/Template%20and%20Instructions/Forms/AllItems.aspx>. Enter your data into the appropriate worksheets in this Workbook. When you are finished with each worksheet, select it and then copy and paste it into the relevant subsections below so that all the information pertaining to this plan is as self-contained as possible.*

### **5.1 Snapshot of FY2013 Digitization Investment and Digitization Capacity**

*In this subsection, use the following guidelines when entering the number of FTE's (Federal and Trust) working on each of the digitization activities listed in the worksheets:*

- *Only count individuals who contribute at least a quarter of their time ( $\geq 0.25$ FTE) to any of the tasks listed. Note that 1 FTE might be two individuals spending half of their work hours on digitization, or three people spending approximately one third of their work hours on digitization.*
- *Use the following definition from SD610 to determine which activities are in scope for "digitization":*  
*"Digitization (is) a set of processes that converts physical resources to a digital form, or that creates materials in a digital form (born digital). These processes include:*
  - *Identification, selection and prioritization of materials to be digitized;*
  - *Digital asset creation or conversion;*
  - *Creation of descriptive and technical metadata sufficient to allow retrieval and management of the digital assets and to provide basic contextual information for the user;*

- Quality control of digital assets and metadata;
- Enhancement of the digital assets and metadata.”

Do NOT include staff whose activities are involved in the delivery of data assets (i.e., Web design, mobile app development, etc.)

### 5.1.1 Current Staff (MANDATORY)

Enter your response into the Excel Worksheet tab entitled, “5.1.1 Current Staff”.

| CURRENT Staff Investment   | FEDERAL          | FEDERAL      | TRUST            | TRUST       | CONTRACTORS      | AFFILIATED*<br>AGENCIES | VOLUNTEERS/<br>INTERNS**       | TOTAL CURRENT \$\$    |
|--|------------------|--------------|------------------|-------------|------------------|-------------------------|--------------------------------|-----------------------|
|  | # of individuals | FTEs         | # of individuals | FTEs        | FTE (equivalent) | FTE (equivalent)        | # of<br>Volunteers/<br>Interns |                       |
| <i>Note that service contracts should be reflected in 5.1.2</i>  |                  |              |                  |             |                  |                         |                                |                       |
| Total Unit Staff (includes staff who are <b>NOT</b> involved with digitization tasks/functions).                         | 341.00           | 293.00       | 128.00           | 109.00      | 50.00            | 42.00                   | 600.00                         | \$46,836,528.00       |
| Number of Total Unit Staff dedicated to creating, maintaining and/or enhancing digital descriptive records (cataloguing) | 37.00            | 18.25        | 4.00             | 2.05        | 8.40             | 2.04                    | 11.00                          | \$2,467,351.00        |
| Number of Total Unit Staff dedicated to creating, maintaining and/or enhancing digital surrogates                        | 25.00            | 6.02         | 3.00             | 0.54        | 6.25             | 0.87                    | 10.00                          | \$820,609.00          |
| Number of Total Unit Staff dedicated to managing digital assets  | 28.00            | 7.11         | 1.00             | 0.33        | 1.35             | 0.56                    | 3.00                           | \$883,563.00          |
| Number of Total Unit Staff dedicated to other digitization tasks/functions   | 30.00            | 7.25         | 3.00             | 1.54        | 2.15             | 0.30                    | 4.00                           | \$1,003,949.00        |
| <b>TOTAL</b>   | <b>120.00</b>    | <b>38.63</b> | <b>11.00</b>     | <b>4.46</b> | <b>18.15</b>     | <b>3.77</b>             | <b>28.00</b>                   | <b>\$5,175,472.00</b> |
| *Affiliated agency costs are not included in totals as these costs are borne by agencies                                 |                  |              |                  |             |                  |                         |                                |                       |
| **This number is an estimate as it is a constantly changing figure.  |                  |              |                  |             |                  |                         |                                |                       |

**5.1.2 Current Outsourcing for Service Contracts (RECOMMENDED)**

Enter your response into the Excel Worksheet tab entitled, "5.1.2 Current Outsource".

| <b>CONTRACTS/OUTSOURCING Investment-- Describe below. Note that contractors working alongside Smithsonian Staff should be reflected in 5.1.1</b>           | <b>FEDERAL funds</b> | <b>TRUST funds</b> | <b>TOTAL \$\$</b>   |
|--|----------------------|--------------------|---------------------|
| ENTOMOLOGY: Capture digital images of type specimens (CIS)   | \$22,008.00          |                    | \$22,008.00         |
| ENTOMOLOGY: Capture digital images of type specimens (CIS)   | \$31,114.80          |                    | \$31,114.80         |
| ENTOMOLOGY: Pre-capture preparation of collection and capture of partial descriptive records (CCPF)  | \$10,230.00          |                    | \$10,230.00         |
| ENTOMOLOGY: Pre-capture preparation of collection and capture of partial descriptive records (CCPF)  | \$13,432.00          |                    | \$13,432.00         |
| ENTOMOLOGY: Pre-capture preparation of collection and capture of partial descriptive records (CCPF)  | \$14,000.00          |                    | \$14,000.00         |
| ENTOMOLOGY: Capture of digital images of valued butterflies (EOL) [NOTE: this contract was active January 2012-January 2013, eliminate if not appropriate] | \$37,116.00          |                    | \$37,116.00         |
| ENTOMOLOGY: Capture digital images of types (USFS)   | \$25,000.00          |                    | \$25,000.00         |
| ENTOMOLOGY: Pre-capture preparation of collection and capture of partial descriptive records (CCPF)  | \$33,000.00          |                    | \$33,000.00         |
| ENTOMOLOGY: Capture digital images of type specimens (Diptera Fund) two contracts (one (4,000) completed recently and the other (1920) on-going)           |                      | \$5,920.00         | \$5,920.00          |
| IZ: Professional Photographer  |                      | \$8,000.00         | \$8,000.00          |
| MINERAL SCIENCES: Slide-scanning project with ScanCafe (10,000 slides)   | \$7,000.00           |                    | \$7,000.00          |
| NAA: Audio Digitization (per FY)   |                      | \$5,000.00         | \$5,000.00          |
| NAA: Microfilm Digitization (per FY) includes that currently out so some overlap   |                      | \$29,747.00        | \$29,747.00         |
| PALEONTOLOGY: Specimen imaging and EMu record enhancement  | \$10,000.00          |                    | \$10,000.00         |
| <b>TOTAL</b>   | <b>\$202,900.80</b>  | <b>\$48,667.00</b> | <b>\$251,567.80</b> |

**5.1.3 Current Hardware/Software (RECOMMENDED)**

Enter your response into the Excel Worksheet tab entitled, "5.1.3 Current HW SW".

| Digitization hardware and software (equipment, maintenance, parts, post-processing software) expenditures. | Federal funds         | Trust funds        | Total \$\$            |
|--|-----------------------|--------------------|-----------------------|
| Purchased equipment  | \$1,048,620.52        | \$44,251.16        | \$1,092,871.68        |
| Donated equipment*   |                       |                    | \$1,500.00            |
| <b>TOTAL</b>   | <b>\$1,048,620.52</b> | <b>\$44,251.16</b> | <b>\$1,094,371.68</b> |

\*Total Investment by unit (2000 - 2013) = \$3,408,409

**5.1.4 Current Storage (RECOMMENDED)**

Enter your response into the Excel Worksheet tab entitled "5.1.4 Current Storage".

| Local and centralized digital file storage expenditures. N.B. "Local storage" is storage managed by the unit; "Centralized storage" is storage managed by OCIO. | Total TB committed | Total TB Used | Total \$\$          |
|---|--------------------|---------------|---------------------|
| Local (Unit managed)  | 100.00             | 48.00         | \$131,000.00        |
| Centralized (OCIO managed) (if known)   |                    |               |                     |
| <b>TOTAL</b>  | <b>100.00</b>      | <b>48.00</b>  | <b>\$131,000.00</b> |

**5.1.5 Current Other Expenditures (RECOMMENDED)**

Enter your response into the Excel Worksheet tab entitled, "5.1.5 Current Other Exp".

| OTHER expenditures-- describe below | FEDERAL funds     | TRUST funds       | Total \$\$        |
|-------------------------------------|-------------------|-------------------|-------------------|
| Botany: Barcode labels              | \$1,200           |                   | \$1,200.00        |
| NAA: Paid internships (2 per year)  |                   | \$5,200           | \$5,200.00        |
| <b>TOTAL</b>                        | <b>\$1,200.00</b> | <b>\$5,200.00</b> | <b>\$6,400.00</b> |

**5.1.6 Space (RECOMMENDED)**

Enter your response into the Excel Worksheet tab entitled, "5.1.6 Space".

|   |                          |
|---|--------------------------|
| Space available to operate digitization hardware (digital capture). Please describe below                         | <b>Total square feet</b> |
| Digitization workstations, photography labs, scanning stations, imaging stations, barcoding and printing stations | 11395.00                 |

**5.1.7 Rapid Capture Capacity (MANDATORY IF APPLICABLE)**

If you have deployed any rapid capture workstations, enter the object type you are capturing in this manner, and the rate of throughput you are achieving, into the Excel Worksheet tab entitled, "5.1.7 Rapid Capture".

| <b>Rapid capture station: Object type</b>  | <b>Rate of throughput (Avg rate/day)</b>  |
|--|---|
| BOTANY: GPI - Historical Collections   | 80 records/day/person   |
| BOTANY: Mobile Data Entry  | 50 records/day/person   |
| BOTANY: Australia Living Atlas   | 15 records/day  |
| ENTOMOLOGY: Digital images of Entomology drawers of specimens. The resulting stitched image is stored locally as both a Gigapan Stitch file and a tiff image; the image file is also uploaded to the Gigapan website where the individual unit trays of a drawer are annotated for species names. Currently, there is no one assigned to the unit; the department is advertising for interns to run it on certain projects for Hymenoptera; not tried this with contractors. | 15 fully annotated drawers can be generated per day. Capturing and stitching amounts to 3.5 min/drawer; the annotation of the drawer contents is more time consuming, and there is no room for error. So, for a drawer of one species of Morpho, the whole drawer would be done in ca. 3.5 min; a drawer of gall wasps, which may contain 20 species in 40 unit trays, will take up to 10-15 min. |
| NAA: Two-dimensional (manuscripts)   | 300-400 day capture not to include post- processing   |

**5.1.8 Current Digitization Capacity (MANDATORY)**

Enter your response into the Excel Worksheet tab entitled, "5.1.8 Current Digitization Capacity".

Consult your SD600 Unit Digitization Assessment (formerly called Digitization Statistics) to help you fill out this worksheet. If your UDP includes digitization for non-SD600 assets (see Section 3 Scope), include those assets in the information you enter into this worksheet.

| DIGITAL SURROGATES (type of output) | Current Production Avg rate/ year |
|-------------------------------------|-----------------------------------|
| Image files (#)                     | 80,000/year*                      |
| Audio files (hrs)                   |                                   |
| Video files (hrs)                   |                                   |
| Other                               |                                   |
|                                     |                                   |
| DIGITAL DESCRIPTIVE RECORDS         | Current Production Avg rate/ year |
| Record Creation                     | 123,600/year**                    |
| Record Enhancement                  |                                   |

\*Includes NMNH + NAA/NSFA [NMNH=55,000/year (averaged over 2004-2012), NAA/NSFA=25,000/year]

\*\* Includes NMNH + NAA/NSFA [NMNH=123,500/year (averaged over 2004-2012), NAA/NSFA=100/year]

### 5.1.9 Digitization Rate (MANDATORY)

Enter your response into the Excel Worksheet tab entitled, "5.1.9 Digi Rate".

|   |   |
|---|---|
| Is your rate of digitization keeping up with your rate of collections growth? |   |
| Yes   |   |
| No  | X |

If you answer "no" in this worksheet, briefly explain below why digitization is not keeping up with your rate of collections growth:

Staffing levels and available budget do not accelerate at the same rate as the growth of our collections. The staff and budget we currently have for digitization are put towards the reduction of the backlog.

**5.2 Required Digitization Investment (FY14-16) and Capacity for Meeting UDP Goals (as articulated in Section 4.1)**

**5.2.1 Required Digitization Investment (FY14-16) (MANDATORY)**

*To deliver on the unit goals outlined in Section 4, and align with pan-Institutional plans and goals, enter the capacity increases that will be required into the Excel Worksheet tab entitled, "5.2.1 Projected Investment".*

| Complete all that apply                           | FY14     |                | FY15     |                | FY16     |                |
|---|----------|----------------|----------|----------------|----------|----------------|
|   | #        | \$\$           | #        | \$\$           | #        | \$\$           |
| STAFF   | 52.00    | \$2,877,904.00 | 52.00    | \$2,877,904.00 | 52.00    | \$2,877,904.00 |
| AFFILIATED AGENCIES STAFF                         |          |                |          |                |          |                |
| CONTRACTS   |          |                |          |                |          |                |
| EQUIPMENT<br>(Hardware/Software)                  |          | \$62,000.00    |          |                |          |                |
| DIGITAL FILE STORAGE<br>additional (GB/TB) needed |          |                |          |                |          |                |
| OTHER Expenditures                                |          |                |          |                |          |                |
| SPACE square feet needed                          | 10400.00 |                | 10400.00 |                | 10400.00 |                |

|   |                |
|---|----------------|
| TOTAL FTE needed FY14-16:                                       | 156.00         |
| TOTAL \$ needed FY14-16:  | \$8,695,712.00 |
| TOTAL \$ needed FY14-16<br>(without Affiliated Agencies Staff): | \$8,695,712.00 |
| TOTAL Space (Square Feet) needed FY14-16:                       | 31200.00       |



**5.2.2 Digitization Capacity (if the investment identified in 5.2.1 is fully funded) (MANDATORY)**

Enter your response into the Excel Worksheet tab entitled, “5.2.2 Projected Digi Capacity”.

If your UDP includes digitization for non-SD600 assets (see Section 3 Scope), include those assets in the information you provide in this worksheet.

| DIGITAL SURROGATES (type of output) | Projected Production Avg rate/ year |
|-------------------------------------|-------------------------------------|
| Image files (#)                     | 100,000.00                          |
| Audio files (hrs)                   |                                     |
| Video files (hrs)                   |                                     |
| Other                               |                                     |

| DIGITAL DESCRIPTIVE RECORDS | Projected Production Avg rate/ year |
|-----------------------------|-------------------------------------|
| Record Creation             | 550,000.00                          |
| Record Enhancement          | 150,000.00                          |

**5.2.3 Projected Impact on SD600 Digitization Assessment (formerly called *Digitization Statistics*) (RECOMMENDED)**

What impact will successful execution of this UDP, and funding at the required capacity level (see 5.2.1 Required Digitization Investment (FY14-16)) have on your unit’s SD600 Digitization Assessment? What estimated percentage of throughput can realistically be achieved specifically for SD600 collections?

*Some background: The Smithsonian has made a strong commitment to digitization, as exemplified in the pan-Institutional strategic plan “Inspiring Generations through Knowledge and Discovery” through the goal (pg. 24) of “doubling the current rate of digitization” over its 5 year run-time. (This goal corresponds to a 20% increase of digitization throughput every year.) This question is not meant to hold you accountable for a 20% increase, but to provide data that allows us to better understand what kind of increase might be possible.*

Development of this plan has given the NMNH Digitization Steering Committee a better framework for examining resources and goals. We realize that there are many people spending small amounts of time on digitization and it has been challenging to capture

their data for inclusion in this plan, which presents a challenge in accurately representing our full digitization output given that this “small” area includes a large number of people. We also realize that we must change our digitization paradigm if we are to reach our goals: more emphasis must be placed on developing alternative methods of connecting information (structured and unstructured) and on rapid digitization methods, and we must look at ways to focus digitization labor.

If we gain funding at the required capacity level (see table 5.2.1), we should be able to reach the goals stated in this plan. We currently have 6.1 million records of an anticipated 16.5 million records, making our new target baseline 10.4 million records. With future completion of the FY14-16 tasks NMNH records would increase by 1.65 million descriptions and 300,000 surrogates, moving NMNH 16% closer to the new baseline. It is important to recognize one of NMNH’s challenges: the large and scientifically important entomological collections can be described at species levels (*Morpho hercules*) and at specimen levels (the individual butterfly). For CDRS reporting, NMNH focused on the species level target yielding 920,000 anticipated records. We must also plan for the ultimate goal of specimen level records, and the analysis of the collection to allow that planning is part of the museum’s Digitization Plan. The ultimate number of descriptive records could be as many as 33 million – quite the challenge – so NMNH’s goals and numbers will change as FY 13 and FY 14 unfold.

### **5.3 Strategies for Digitization Funding (MANDATORY)**

*What strategies are you currently pursuing to increase funding for digitization, for example through the federal process, philanthropic avenues, or public/private partnerships?*

NMNH has used its FY 2010 Collections Assessment to focus attention on the need to digitize collections, and to explain how projects are prioritized. Several presentations about the Assessment have been made to OMB examiners as part of the plan to build understanding of NMNH’s collections challenges and opportunities. The museum seeks digitization funding in federal budget requests, includes digitization projects in calls for Federal Stimulus Funds projects, submits CIS Pool Fund proposals, and requests funds from the NMNH Director’s Discretionary Funds. Internal collections funds are applied to some projects.

The museum has not sought private funds for digitization, per se. Rather, funding is sought for initiatives such as Recovering Voices and applied to meeting a programmatic goal. In some cases the discrete tasks may include digitization.

Currently, the NMNH Digitization Committee is experimenting with an Australian partner on crowd sourcing. While the experiment does not include transfers of funds, it does share resources.

Since its establishment of the digital lab in 1999 the NAA has continually worked on relationships with tribal communities, researchers and scholars, and private foundations to support digitization of its collections. Over the years, this process has evolved from an ad hoc funding approach to a more strategic approach based on the selection and prioritization of collections and the ability to coordinate digitization and access simultaneously. In addition to creating digital images, the NAA has also focused on the dissemination, descriptive information and the long term sustainability of the digital asset. This approach has enhanced the potential to secure funding of the NAA's needs.

The NAA also has collaborated on grant applications to NSF, to secure project funding to ensure that a digital asset management plan is in place.

## 6. PRIORITIES

The Priorities section links high-level priorities that are articulated broadly in narrative form with the actionable digitization projects that are executing on these priorities. Both narrative priorities and actionable projects are crucial components for crafting unit and pan-Institutional funding pitches.

---

REFER TO THE COLLECTIONS & DIGITIZATION REPORTING SYSTEM (CDRS) TO COMPLETE THIS SECTION

*Beginning in the Fall of 2012, units will report their NCP Collections Statistics, NCP Collections Assessment and DPO Digitization Assessment in the new Collections & Digitization Reporting Systems (CDRS). CDRS also includes a Digitization Priorities component that SD600 units must complete before they can answer the questions in this section of the UDP. (For guidance on how to complete the Digitization Priorities component of CDRS, consult CDRS and [Annual Collections Statistics/Collections and Digitization Assessment Guidance and Instructions](#).)*

*CDRS currently compiles data for SD600 collections only. If you have included non-SD600 collections in the scope (Section 3) of this UDP, you will need to cite priorities and prioritization methodologies for these materials using whatever lists and methodologies you have established in your unit.*

### **6.1 Narrative Digitization Priorities Summary (MANDATORY)**

*In a narrative statement, broadly summarize your unit's digitization priorities. This summation should agree with the objectives you set forth in Section 3 (Scope) of this UDP, and should include the priorities you identified in the Digitization Priorities component of CDRS (e.g., in the section entitled, "Enter Unit Digitization Priorities"). When summarizing your priorities, identify both your greatest priorities (i.e., those you weighted as a "3" or "4" in CDRS), as well as your lowest priorities (i.e., those you weighted a "0" or "1").*

The overriding priority for NMNH digitization is to provide descriptive records and surrogates for researchers. In general, NMNH aims to have digital representation of the collections including all species (for recent biology), and all specimens and objects for Anthropology and Mineral Sciences, and representation for fossil collections. Accessibility, preservation, availability of external experts to identify and curate collections, and other collections management activities such as accessioning or cataloging, relocation, or clearing of a backlog are considered when initiating projects.

## HIGHEST PRIORITIES

The highest priority projects at the NMNH are tied to the following list of priorities:

- Creation of digital descriptive records
- Make collection materials more readily available for collaboration or distant research
- Preserve and protect collection materials that are at risk and cannot be used
- Make collection materials available for wider educational purposes
- Digitization of collection and associated metadata to improve access to materials that are hidden or hard to find.
- Collection materials selected for exhibition digitized to facilitate exhibition activities.
- Digitization is completed in conjunction with moving or rehousing collections.
- Digitization is completed in conjunction with improving efficiency of administering digital assets.
- Digitization is completed in support of a Grand Challenges initiative.
- Creation of 3-D objects of all sizes and complexity.

## LOWEST PRIORITIES

The lowest priority projects at the NMNH are tied to the following list of low priorities as defined by the NMNH Digitization Priority Statement:

- Collection materials selected for publication being made more broadly available.
- Assets selected for web presentation for use to publicize the type of resources available for use.
- Digitization to make it possible to share some collection materials in multiple venues as a means to generate revenue.
- Digitization is done in support of a special event.
- Creation of mixed media.
- Digitization of live collections.
- Digitization of other materials that include books, journals, maps, etc., in and of themselves. Digitization of these materials to enhance descriptive records or for preservation purposes (i.e. ledgers) may be a higher priority

## **6.2 Overall Digitization Challenge and Ambition (MANDATORY IF APPLICABLE)**

*For SD600 collections only. Once your unit's Digitization Assessment is complete, copy the total number of collections items that are listed as "not digitized" below. (This total should encompass items that you intend to digitize, and those that have been digitized at a substandard level. Archival collections should report this number in cubic feet.) This number is a snapshot of your unit's overall digitization challenge.*

### **SD 600 COLLECTION ITEMS NOT YET DIGITIZED AND DIGITAL DISCRIPTIVE RECORDS LABELED AS SUBSTANDARD**

The total number of collections items that are not yet digitized or are digitized at a substandard level totals 108,626,937 collections items. The total cubic footage of archival materials that are not yet digitized or are digitized at a substandard level is 9,262 ft<sup>3</sup>.

## **6.3 Digitization Priorities Assessment (MANDATORY IF APPLICABLE)**

*For SD600 collections only. In the Digitization Priorities section of CDRS (in the section entitled "Enter Collection Level Assessments—Digitization Priorities"), you assessed your collections against your digitization priorities by identifying the percentage of each collections' subset that "does not meet/meets/exceeds" these priorities. [You also had the option to state that you cannot make this determination ("undetermined") at this time.]*

*This assessment in CDRS gives you a snapshot of where your collections subsets rank in terms of the digitization priorities you summarized in 6.1 above. Copy and paste the information from the "Digitization Subsets – Priorities" section of CDRS below. (For large units: if your snapshot is very long, consider moving it to the end of this document and referring to it as an appendix.)*

This year there is not a match between the nearly 100 digitization priority projects and the seven subsets of collections that were reported in the Collection Level Assessments. NMNH is making a concerted effort to create metadata at a more refined subset level that will allow us to generate this information in future.<sup>6</sup>

---

<sup>6</sup> This statement was approved by Jessica Beauchamp of DPO due to the lack of data export from CDRS

**6.4 Digitization Priority Projects (MANDATORY)**

After you have created your list of digitization projects in CDRS, enter into the following table the list of prioritized digitization projects currently underway in your unit, as well as those you would like to undertake (e.g., aspirational projects) if you had the extra capacity stated in Section 5.2 (Required Digitization Investment (FY14-16) and Capacity Required for Meeting UDP Goals). We expect the number of projects you list here – both current and aspirational - to be commensurate with the size of your unit, i.e., small units will list fewer projects than large units. Large units with many projects may consider moving this table to the end of this document and referring to it as an appendix.

If applicable, add any non-SD600 projects to this table and indicate which projects are SD600/non-SD600 by checking off the appropriate column.

Due to the inability of CDRS to report our nearly 100 digitization projects we are unable to complete the table below at this time.

| Digitization Projects<br>(In priority order) | SD600 | Non-SD600 | Currently underway* or aspirational?** | Total # of Collection Items |
|--|-------|-----------|--|-----------------------------|
|  |       |           |  |                             |
|  |       |           |  |                             |
|  |       |           |  |                             |
|  |       |           |  |                             |
|  |       |           |  |                             |

\*Projects *currently underway* are those projects you are pursuing within your current digitization capacity as articulated in section 5.1 (Snapshot of FY2013 Digitization Investment and Digitization Capacity).

\*\**Aspirational* projects are projects you *would* pursue if your unit had the digital investment and capacity stated in Section 5.2 (Required Digitization Investment (FY14-16) and Capacity for Meeting UDP Goals).

## **6.5 Prioritization Methods (RECOMMENDED)**

*In the previous sections (6.1 – 6.4) you identified digitization priorities using a pan-Institutional methodology (via CDRS). Do you have a unit-specific methodology for choosing collections for digitization? If yes, briefly describe your prioritization methodology (or cite the URL of any documentation that describes this methodology).*

### **UNIT SPECIFIC METHODOLOGY FOR CHOOSING COLLECTIONS FOR DIGITIZATION**

Each of the scientific departments within NMNH is responsible for setting their own priorities for digitization projects within the overall goal of representation and focus on species. The Collections Program conducted a collection assessment in 2006 that was updated in 2010 that has also been used on a museum scale to help identify those collections in most need of support for digitization. The Assessment uses 13 criteria that could be combined so that a collection could be scored on Importance, Condition or Information. Those collections that score highly on Importance but less favorably for Information become targets for digitization, especially if there are other tasks or initiatives that contribute to a need for information, such as a pending collections move or a research project. This comparison can be done at museum, department, and sub unit scales so that very particular decisions can be made.



## 7. RELATIONSHIPS AND DEPENDENCIES

The Relationships and Dependencies section allows your unit to identify the many internal and external partnerships that may be vital to your digitization efforts.

### 7.1 Dependencies (RECOMMENDED)

Identify the resources, infrastructure, systems, collaborators, funders, etc. that this plan is dependent upon. (“Dependent” means that if the relationship with the resource, funder, collaborator, etc. should cease, your unit’s ability to continue some aspect of digitization articulated in this UDP would be diminished.) You may add as many rows as you need to these tables to accommodate your response.

| Internal SI Systems and infrastructure<br>(e.g., CIS, DAMs, storage areas) | External Systems, infrastructure |
|--|----------------------------------|
| EMu  |                                  |
| EDAN   |                                  |
| Islandora  |                                  |
| SIRIS  |                                  |
| DAMS   |                                  |
| Tivoli   |                                  |
| MeSA (Media Staging Area)  |                                  |

| Internal Collaborators (individuals, units, funders) | External Collaborators                 |
|--|--|
| CIS IRM Pool Fund                                    | Sloan Foundation                       |
| CCPF IRM Pool  | Harvard University                     |
| DPO  | Marine Biological Laboratory           |
| Smithsonian.com [Ocean Portal]                       | CONABIO [funder of EOL]                |
| Smithsonian Institution Libraries                    | Cosmos Prize [funder of Ocean Portal ] |
| Latino Pool Fund                                     | Walter Scott Foundation                |
| Smithsonian’s Women’s Committee                      | Encyclopedia of Life                   |
| Recovering Voices                                    | GBIF                                   |
| NMNH Web Advisory Group (WAG)                        | Mellon Foundation                      |
| NMNH Operating Fund                                  | Arcadia Fund                           |

## 7.2 Aggregators and Institutional Users of Unit Digital Assets (RECOMMENDED)

Identify other projects, units, organizations or institutions that are using/reusing/repurposing your digital assets **or** systems, or any projects or units that depend upon your digital assets.

| Internal Aggregators/Institutional Users | External Aggregators/Institutional Users      |
|--|---|
| EDAN and Collections Search Center       | Encyclopedia of Life                          |
| DAMS                                     | Global Plant Initiative                       |
| Q!RIUS                                   | Biodiversity Heritage Library                 |
| Ocean Portal                             | Digital Public Library of America             |
| NMNH Species of the Day Widget           | Map of Life                                   |
| Birds of DC exhibit app                  | Catalogue of Life                             |
| Temporary exhibits such as X-ray fishes  | TerraMar                                      |
| SIRIS                                    | Ecosystem Explorer                            |
|  | EOL Field Guide tool                          |
|  | Biodiversity Library Explorer                 |
|  | GBIF (Global Biodiversity Information System) |
|  | Reciprocal Research network (RRN)             |
|  | OBIS  |
|  | VertNet                                       |
|  | Flickr  |
|  | Barcode of Life                               |
|  | Gale Cengage Library Research Tool            |
|  | Bison   |
|  | JSTOR   |
|  | GGBN  |
|  | Internet Archive                              |
|  | EarthCube (in development)                    |

## 8. POLICIES AND STANDARDS

The Policies and Standards section identifies the availability of your assets and the impact your unit's choices have on data sharing. Clarifying how your unit applies policies, and which standards it promotes, are foundational to data integration and access in the future.

### 8.1 Legal and Policy Issues (MANDATORY)

*The Smithsonian has many directives that address policy issues affecting digital access. For example, SDs 205, 600, and 806 offer guidelines on various copyright issues; SDs 600 and 950 address privacy issues; SD807 addresses confidentiality issues, etc.*

*SD609 Digital Asset Access and Use provides the most recent and direct guidance on digital asset access. It states that all SI assets are to be considered accessible to the public for all noncommercial uses unless these assets fall under certain categories of restrictions (see 8.2 below).*

*For unit collections that do not fall under these restrictions, what is your unit's plan for making these digital assets available? Identify when, where, and how you will make these collections available.*

### 8.2 Restricted Digital Assets (MANDATORY)

*Place a checkmark next to the following SD609 restrictions that **are most applicable** to the digital assets in your unit. [For definitions and further details of these restrictions, refer to SD609.]*

#### SD609 Allowable Restrictions

| Legal Restrictions – Intellectual Property |   |
|--|---|
| Copyright                                  | X |
| Moral Rights                               | X |
| Trademark                                  | X |
| Patent                                     | X |
| Privacy Rights                             | X |
| Publicity Rights                           | X |
| Personal Identifiable Information (PII)    | X |

| <b>Legal Restrictions – Other</b>   |   |
|---|---|
| Contractual restrictions  | X |
| Native American & Hawaiian human remains & sacred objects                                     | X |
| Cultural object repatriation (due to illegal acquisition)                                     | X |
| Endangered species  | X |
| Asset reveals location of archeological, paleontological, geological, sacred or historic site | X |
| Uncertain provenance and export records (e.g., Holocaust era)                                 | X |
| Conservation, management, inventory, valuation, other business records not covered by SD609   | X |

| <b>Policy Restrictions</b>          |   |
|-------------------------------------|---|
| Sensitive content                   | X |
| Unpublished research data/resources | X |
| Resource limitations in unit        | X |
| Commercial use                      | X |

|   |           |
|---|-----------|
| <b>In your estimation, what percentage of your total collection falls under SD609 restrictions?</b> | <b>4%</b> |
|---|-----------|

### 8.3 Life Cycle Management (RECOMMENDED)

*Life cycle management (LCM) is an approach to managing digital assets from creation through maintenance and use or disposition. (See “Best Practices of Life Cycle Management for Digital Assets” in [Life Cycle Management Best Practices for the Smithsonian Institution’s Digital Asset Categories](#) for a summary of LCM practices). Does your unit have a documented LCM approach? If yes, provide a URL to the relevant document(s). If this information is not online, briefly describe the approach below.*

### 8.4 Digitization Standards

#### 8.4.1. List of Standards (MANDATORY)

*Do you have a documented list of unit-preferred digitization standards? If yes, provide a URL for this list. If this list is not online, identify the standards below. [N.B., DPO has started a list of standards in use at the Smithsonian. This list, which can be found at [Standards in Use at SI](#) on the DPO Sharepoint site, is not exhaustive. However, you may wish to consult it and should feel free to add to it any standards used by your unit that are not represented there].*

#### EMu Catalog Standards:

- Darwin Core (DwC)
- Biodiversity Information Standards (TDWG)
- Distributed Generic Information Retrieval (DiGIR)
- Unicode Standard (Unicode 5.2.0)

#### EMu Multimedia Standards:

- [Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images](#) (Embedded Metadata Working Group, Smithsonian Institution)
- [File Nomenclature Schemas](#) (NMNH Informatics)
- [Image Capture Format Quality Standards](#) (NMNH Informatics)
- Adobe Digital Negative Format Standard (DNG)
- TIF Image Format standard
- JPEG Format Standard,
- IPTC Metadata Format Standard
- Exif Metadata Standard

#### National Anthropological Archives Digital Imaging Guidelines

- NAA\_DAMS\_Working\_Guide
- NAA\_Digital\_Filenaming\_Guidelines
- NAA\_PDF\_Creation\_and\_Workflow
- NAA\_DAMS\_SD600tag\_guideline

In addition, each Research and Collections department within NMNH has discipline specific data and multimedia standards for their records.

Also in use at NMNH:

FADGI ([Federal Agencies Digitization Guidelines Initiative](#))

- Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files (2010)

#### **8.4.2 Process for Adopting Standards (RECOMMENDED)**

*What is your unit's process for adopting standards? If you also use internally developed standards, identify the process for how you create and maintain these standards.*

## 9. UNIT DIGITIZATION PROGRAM ADMINISTRATION AND UDP PLAN MANAGEMENT

The Unit Digitization Program Administration and UDP Plan Management section clarifies who is responsible for digitization in your unit, and provides DPO with a single point of contact as it coordinates various pan-Institutional and unit digitization pursuits.

### 9.1 Coordinating Office/Person (MANDATORY)

*Identify the unit's office/person and title/committee tasked with coordinating digitization efforts across the unit.*

The NMNH Digitization Steering Committee: Co-Chairs Tom Orrell (Informatics Branch Chief) and Carol Butler (Assistant Director for Collections), and Vice Chair Rebecca Snyder.

### 9.2 Responsible Officer (MANDATORY)

*To whom does the person or entity listed in Section 9.1 report to?*

Orrell reports to Associate Director for Operations, Butler reports to Associate Director for Science. Ultimate oversight is by Director.

### 9.3 Process for Coordinating Digitization in the Unit

#### 9.3.1 Coordinating Digitization (RECOMMENDED)

*Briefly identify how digitization is managed across your unit. For example, who are the principals? Does the unit set performance measures for its digitization projects and activities? How does it compile and review information on all its digitization activities?*

### **9.3.2 UDP Development Process (MANDATORY)**

*Briefly (in no more than a few sentences) identify the unit's process for developing this UDP.*

Responsibility for the creation of the NMNH Unit Digitization Plan was given to the NMNH Digitization Steering Committee, which includes members from all of the scientific collecting departments, as well as education, exhibits, archives, and libraries. The committee collaborated on and contributed to the draft, before it was submitted to the Associate Director of Science and the NMNH Unit Director for comment and review, prior to final submission.

### **9.4 Plan Administrator (MANDATORY)**

*Provide the name of person responsible for management of this UDP (i.e., the plan administrator).*

*The plan administrator is responsible for ensuring this plan is completed and submitted to the Digitization Program Office (DPO) by its due date. The administrator also coordinates and tracks revisions of the plan and serves as primary contact between the unit and DPO in all matters that involve the plan.*

Carol Butler, NMNH Digitization Steering Committee Co-Chair

### **9.5 Amending the UDP (RECOMMENDED)**

*Briefly describe your unit's procedures for amending this UDP.*

Annual review of UDP by NMNH Digitization Steering Committee to determine if corrections are warranted. If no revisions required during annual mid-plan review sessions, UDP will be updated and reviewed according to the prescribed 3-year cycle.

### **9.6 Next UDP Review Date (MANDATORY)**

*SD610 requires UDPs to be reviewed and updated at least once every three years (significant interim changes are to be reported to DPO annually). Identify the date of your unit's next mandatory UDP review (month, day and year).*

April 2016.

## UDP/DAMP GLOSSARY OF TERMS

| Term   | Definition   |
|--|--|
| Collections & Digitization Reporting System (CDRS) | A newly developed online system designed to facilitate the collection and reporting of information provided by units in response to various SI collections and digitization reporting requirements (i.e., the National Collections Program (NCP) Collections Statistics and Collections Assessment, and the Digitization Program Office's (DPO) Digitization Assessment). Units completing a UDP will use a worksheet in the CDRS system to identify their digitization priorities.  |
| Digital asset                                      | Content that is recorded and transferred in a digital format. It may include text, still images, moving images and sound recordings, collections that are digital (i.e., digital art), research datasets and other types of media originally created in digital format or digitized from another format or state (i.e., a digital surrogate) that are created, stored, or maintained by the Smithsonian. Digital assets also include the metadata used to describe the digital asset and its content.  |
| Digital Asset Management Plan (DAMP)               | A written plan that defines the roles, responsibilities and processes needed to ensure the systematic attention to a digital asset throughout its lifecycle, from creation or collection, through use, preservation and, if appropriate, disposition.  |
| Digitization                                       | A set of processes that converts physical resources to a digital form, or that creates materials in a digital form (born digital). These processes include: identification, selection and prioritization of materials to be digitized; digital asset creation or conversion; creation of descriptive and technical metadata sufficient to allow retrieval and management of the digital assets and to provide basic contextual information for the user; quality control of digital assets and metadata; and enhancement of the digital assets and metadata. |
| Fidelity   | The degree to which a system output accurately reproduces the sound or image of an input. Fidelity is a measure of quality for a reproduction, i.e., the greater the fidelity, the better the quality.   |
| General DAMP                                       | A type of digital asset management plan (DAMP) that covers all the digital assets in a unit that are created/collected as part of a set of <i>ongoing</i> activities in the unit rather than from a distinctive, finite project. For example, units routinely create digital assets on their collections as part of their daily stewardship activities. These digital assets need to be managed with as much care as the digital assets created within the context of a formal project. Similarly, a unit may treat a class of                               |



|                                |  |
|--------------------------------|--|
|                                | objects dispersed across a unit, such as GIS data, in a similar manner, and thus create a General DAMP that describes its management for this particular group of digital objects.   |
| Life Cycle Management (LCM)    | A comprehensive approach to managing digital assets that addresses these assets through all the stages of their “life.” (Also see “Project DAMP” below.) It begins with planning for the creation or acquisition of a digital asset, continues through the maintenance and use of the asset, and ends only when the asset is legally transferred to another entity or disposed. Life cycle management functions are sequential but a digital asset may go through certain stages of the life cycle multiple times as it is used by different groups or for different purposes. |
| Non-centrally supported system | Hardware and/or software used by a unit for digitization or with digital assets but that is not supported by OCIO.   |
| Project DAMP                   | A digital asset management plan (DAMP) that covers the digital assets created/collected within the context of a specific, finite project undertaken in the unit.   |
| SD600 collections              | Smithsonian Directive (SD) 600, <i>Collections Management</i> , designates certain units as “SD600 units” (meaning they must comply with the policies stated in SD600.) The collections held by SD600 units are referred to as “SD600 collections”.  |
| SD610                          | The Smithsonian directive entitled, “ <i>Digitization and Digital Asset Management Policy</i> ” that establishes the importance of digitization to the Institution and sets forth requirements for plans (Unit Digitization Plans and Digital Asset Management Plans) that will help the organization better care for and manage digitization and the digital assets that result from digitization activities.   |
| Unit Digitization Plan (UDP)   | A written plan that defines a unit’s digitization program. The plan addresses aspects such as unit objectives and priorities for digitization, responsible parties for unit-based digitization activities, performance metrics and digitization funding sources.   |
| Use case                       | A brief scenario that demonstrates how a unit’s efforts at digitization will serve its users/audiences.  |