June 2012

## U.S. FISH & WILDLIFE SERVICE

# SERVICE CATALOG PILOT PROJECT SUMMARY

Deliverable for: MBS Contract, Service Catalog Pilot

Written by: Sarah Shultz
Sara McLaughlin
Rande Ramsey-Cross

## Natural Resource Program Center

# Table of Contents

# List of Tables

# Executive Summary

In 2011, the U.S. Fish & Wildlife Service (Service) Inventory and Monitoring (I&M) Initiative began the process of digitizing refuge documents and uploading them into a centralized database.  This database, known as the U.S. Fish & Wildlife Service Catalog (ServCat), was developed through collaboration with the National Parks Service Inventory and Monitoring Program.  ServCat is a repository for compiling, organizing, and making accessible recent and historic information that is important to refuge resource management.   The information entered in the database is discoverable and retrievable using text and geospatial search tools.  ServCat will complement the NCTC Conservation Library and will be an important collaboration tool for the Service and its partners.

The ServCat pilot project was initiated as the first step in the data mining process.  Two data technicians, contracted through Managed Business Solutions, were hired to execute the pilot.  The technicians visited at least one refuge per Service region and spent a week scanning documents, entering the associated metadata in the database, learning how long it takes to create a ServCat record, and gathering ideas about how the full-scale implementation of the database should proceed.  Five types of resource management documents were targeted during the pilot: annual narrative reports, management plans, refuge specific reports, surveys, and maps.

During the pilot project a total of 2,473 documents from 10 different refuges were entered into the ServCat database.  A document can take anywhere from 5 to 60 minutes to scan.  Maps and plans can generally be scanned quickly; annual narrative reports and refuge specific reports require more time.  These reports are often fragile, contain photographs, and are difficult to unbind.  All of these factors increase scanning time.  Metadata entry requires significantly less time than scanning; 5 to 10 minutes per document is standard.  Overall, a general approximation is that it takes 30 minutes to scan a document and enter it as a record in ServCat.

The data technicians developed techniques to make the metadata entry process more efficient.  Templates were created to make abstracts for annual narrative reports and plans more consistent and easier to write.  A keyword master list was produced to standardize the keywords used in the database.  These tools will make it easier to search for a document in the database and they will help users create records for documents outside of their area of expertise.

Several lessons were learned throughout the course of the pilot project that will help guide the implementation of the data mining effort at the full-scale.  One of the most important lessons was the importance of the refuge staff in the data mining process.  Staff members who participated in the pilot helped prioritize documents, answered questions, and became familiar with ServCat.  This ensured that the refuge's most crucial documents were entered in the database and it increased the efficiency of the process and the quality of the records created.  A second lesson learned was that having a centrally located team that travels to the refuges is inefficient.  It took between one and two days of travel each way to reach most of the refuges.  Weather conditions increased travel time to four of the refuges.  Reducing travel time would provide more time to create ServCat records.  Finally, a great deal of variability was found from refuge to refuge.  This variability applied to document and binding types, number of documents, number of refuge staff, organization of refuge files, staff assistance, refuge accessibility, and working space.

These lessons learned were taken into account as alternatives were considered for the full-scale implementation.  The following alternatives were considered:

> *Alternative 1*:  Regions hire and support a team of data technicians to travel to refuges within the region.

> *Alternative 2*:  Refuges scan and enter documents on their own using equipment supplied by the regions or the refuge's scanner/copier.

> *Alternative 3*:  Refuges send documents to a professional scanning company to be digitized and then send the digital files to the Natural Resource Program Center (Center) or the regional office to be entered.

> *Alternative 4*:  Refuges send boxes of documents to the Center or to the regional offices to be scanned and entered.

> *Alternative 5*:  Regional I&M staff and/or refuge staff prioritize and scan refuge documents and post the digital files to an FTP site.  The Center will create the metadata records, and the regional/refuge staff will perform a quality control check.

> *Alternative 6 (Preferred alternative)*:  A combination of alternatives 1-5.  Regions hire data technicians and refuges have the choice of having the technicians come to them or sending their documents to the technicians.  The Center provides a ServCat Coordinator to deliver training, assist the regions, and oversee the development of the database.

Alternative 6 is recommended because it provides flexibility for every refuge and region.  This alternative will provide both regional and national support to the refuges, and database population should proceed quickly with this alternative.  Traveling time will be minimized because data technicians will be traveling within a single region.  The ServCat Coordinator will ensure that effective trainings are provided to the regions and refuges which should lead to quality and consistency in the records that are created.

Full-scale implementation of the ServCat database will require collaboration between refuges, regions, and the Center.  Populating the database will be an ongoing effort; dealing with the legacy data will likely take years.  It is important to focus on capturing a refuge's priority documents in a systematic and coordinated manner.  Emphasis should also be placed on entering new documents into ServCat as they are produced.  Standardizing the metadata entered in ServCat will make it easier to find a document in the database and faster to create a record.  Templates, the master keyword list, and the ServCat Guidance should be used whenever possible to expedite data entry.

# Introduction

In 2011, the U.S. Fish & Wildlife Service (Service) Inventory and Monitoring (I&M) initiative began the process of digitizing refuge documents and uploading them into a centralized database.  This database, known as the U.S. Fish & Wildlife Service Catalog (ServCat), was developed through collaboration with the National Parks Service Inventory and Monitoring Program.  ServCat is a repository for compiling, organizing, and making accessible recent and historic information that is important to refuge resource management.   The information entered in the database is discoverable and retrievable using text and geospatial search tools.

ServCat will complement the NCTC Conservation Library, which catalogs curriculum-related materials and published documents.  ServCat is unique because it is centralized, anyone in the Service can upload documents to the database, and it encompasses a wide variety of document types.

Once ServCat goes live, it will be a valuable tool for a number of reasons.  First of all, because documents can be retrieved quickly from the database, it will reduce the amount of time that Service personnel spend looking for paper files.  Secondly, information in the database will not be lost if original documents are unavailable or damaged.  Finally, ServCat will promote collaboration efficiencies by making past and current work available.


## The ServCat Pilot Project

A pilot project was initiated in 2011 as the first step in the data mining process.  The complete guidance for the pilot can be found in Appendix E.  There were five major goals for the project:

(1)   To enable a refuge to capture important resource management information into permanent records that can readily be accessed by refuge staff and by others.
(2)   To learn how long it takes to create a record in ServCat (both the electronic form of the document and the necessary metadata) for the different major types of resource management documents.
(3)   To provide information to refuges, regions, and the national office through the consistent collection and storage of resource information across the National Wildlife Refuge System (Refuge System).
(4)   To begin populating the ServCat database with resource management information.
(5)   To develop a strategy for the full-scale implementation of the data mining process.

Two data technicians, contracted through Managed Business Solutions, were hired to execute the pilot.  The technicians spent a week at a refuge in each Service region (two in Region 8), to create records for the database.  The following refuges volunteered to participate: Stillwater National Wildlife Refuge (NWR), Don Edwards San Francisco Bay NWR, Aransas NWR, Malheur NWR, Neal Smith NWR, Parker River NWR, Kenai NWR, St. Vincent NWR, and Tewaukon NWR.  The technicians also created records of documents that were sent to the Natural Resource Program Center (Center) from Ash Meadows NWR for a total of 10 refuges.

At each refuge the data technicians visited, they scanned priority documents and copied digital files onto their laptops.  They also entered the metadata that was associated with each document into the ServCat database.  The technicians focused on capturing a specific type of document each day of the week, as listed in Table 1.  It was important to gather a variety of information so that amount of time

required to create records for various types of documents could be estimated.  On average, there was one week between refuge visits and the technicians spent this time in the Center completing the metadata entry.

**Table 1: Types of documents targeted during the ServCat pilot**

| Day | Document Type |
|---|---|
| Monday | Annual narrative reports |
| Tuesday | Management plans |
| Wednesday | Refuge specific reports |
| Thursday | Surveys and inventories |
| Friday | Maps, aerial photos, and GIS data |

The refuge staff, with guidance from regional and national staffs, was responsible for organizing and prioritizing documents prior to the technicians' visit.  This was done to ensure that the refuge's most valuable documents were uploaded to ServCat and to maximize the time the technicians spent creating records.

During the pilot project, records were entered into a Microsoft Access database because ServCat was still under development.  The records in the Access database were batch uploaded to ServCat.  Once ServCat is live, the Access database will no longer be used.


## Continuation of the ServCat Project

Since the completion of the site visits, the data technicians have been creating records for digital files sent to the Center by regional offices.  Regions 3, 4, and 6 have sent a total of 3,129 files to be entered into the ServCat database.  Thus far over 850 annual narrative reports, 260 plans, and 7 miscellaneous documents have been made into records in addition to those reported as part of the pilot.

Currently the Center is investigating the efficiency of batching groups of similar documents, such as annual narrative reports, and entering them into ServCat.  A spreadsheet is being compiled that lists all of the known annual narratives for every refuge in the Service, more than 8,000.  Information about where these documents are located and whether or not they've been uploaded to ServCat is provided on the spreadsheet.   As mentioned above, many of the narratives have already been entered into ServCat.  At this writing the teams are investigating several possible solutions for creating records for the remaining annual narratives.

It has also been found that some refuges have catalogued documents in their own databases.  These databases are typically in Excel, Access, or EndNote.  It is worth investigating the metadata provided in these databases.  In some cases, this information could be imported directly to ServCat with only minor QA/QC.  Doing so would save time and reduce duplication of work.

# Application of the Pilot Guidance

## The Data Mining Process

The two main components of the pilot included travel and the creation of ServCat records.  Time was also spent communicating with refuge and regional staff prior to site visits, and on average these staff members devoted a week prior to the data mining trip to organize and prioritize documents.   The following sections summarize how the data mining effort proceeded during the ServCat pilot project.

### *Travel logistics*

Traveling to the refuges was an expensive and time-consuming element of the pilot project; one to two days of travel was required on either end of a visit.  Flight lengths ranged from two hours to ten hours, and driving times ranged from half an hour to five hours.  Additionally, several hours were spent making travel arrangements, packing, waiting for flights, and driving to and from the airport.  Weather conditions increased travel time on four occasions.

The scanners and other equipment were packed into two Pelican cases and shipped from one refuge to the next.  This arrangement worked well with one exception.  The data technicians visited two refuges back-to-back and the shipping company failed to pick-up the scanners from the first refuge.  As a result, the scanners never showed up at the second refuge.  The repeated shipping may have taken a toll on the scanners; two of the three had to be replaced before the end of the project.

### *Scanning equipment*

Two scanner models were used during the ServCat pilot project.  Each data technician had an Epson Workforce Pro GT-S50 auto-feed scanner.  This model is capable of scanning double-sided pages, pages up to 8.5" by 36", and at a resolution of up to 600 dpi.  Carrier sheets are available so that fragile pages can be safely fed through the scanner.  In an ideal situation, 75 sheets can be loaded into these scanners at a time and they can scan 25 letter-sized sheets in one minute (Epson America Inc., 2012b).  A variety of factors- including the condition of the paper, the weight of the paper, and the resolution- can affect these statistics.

In addition to the auto-feed scanners, the technicians shared an Epson GT-20000 flatbed scanner that was able to scan pages up to 11.7" by 17" at a resolution of up to 600 dpi (Epson America Inc., 2012a).  This scanner was not as efficient as the GT-S50's, but it was a good option for pages that were too large or too brittle to run through the auto-feed.

The Epson Scan software provided with the scanners was used to control the scanners.  ABBYY FineReader Sprint Plus OCR was used in conjunction with Epson Scan to make the text in the scanned documents searchable (ABBYY, 2012).

### *The scanning process*

A detailed description of the scanning guidelines for the pilot project can be found in Appendix E under the "Scanning" heading.  All documents were scanned as Adobe PDFs, and optical character recognition (OCR) was run so that the words in each document would be searchable.  Documents were scanned as one PDF instead of as individual pages that would later need to be reassembled.  The recommended settings listed in Appendix E  were used when possible; these settings were typically satisfactory for

new, clean documents.  Adjustments were made for older documents- those on onion skin paper, with faded or blurred ink, or with dirty pages.  These documents needed to be scanned in color, at 300 dpi or higher, in order to be legible.  For optimum results, the first page of an older document was scanned using various combinations of settings, such as brightness, contrast, and resolution.   The results- including the visual quality and the accuracy of the OCR- were compared to determine the best combination of settings for a particular document.

### The metadata entry process
The data technicians entered some metadata on site and entered the rest at the Center.  Information about how to create a record in ServCat is provided in Appendix E, "Creating a Record."  Each record was linked to the corresponding PDF document.  Groups of related documents were bundled together into a project, which means that they will be discoverable as a group in the database.

### Staff training
When refuge and regional staff had time, the technicians trained them to create ServCat records.  The training sessions were brief and informal, but they were important because they provided an opportunity for the technicians to explain the value and function of the database.  These staff members may require additional training when the online application is available; however, they will have a general understanding of how to create a complete, searchable ServCat record.

## Results of the Pilot Project
The data technicians entered 2,473 individual documents from the ten refuges into the ServCat database.  An average of 247 documents was entered per refuge.  The smallest number of documents entered was 92, and the largest number of documents was 420.  Table 2 shows the number of each document type entered in the database, and a more detailed account is provided in Appendix A.  The number of documents uploaded per refuge varied greatly.

Table 2:  Total number of documents entered into ServCat

| Type of Document | Total # of Documents | Avg. # of Documents Per Refuge |
|---|---|---|
| Plans | 244 | 24 |
| Reports/Documents | 1,360 | 136 |
| Data/Surveys | 309 | 31 |
| Maps | 57 | 6 |
| Photographs | 160 | 16 |
| Presentation/Poster | 6 | 0.5 |
| Letters/Articles | 337 | 34 |
| **Total Documents** | **2,473** | **247** |
| **% Already Digital** | **17%** | **15%** |

Four of the refuges that participated in the pilot provided digital files for the technicians to upload.  A total of 396 documents that were already digitized were entered into ServCat.  Because the need for scanning was eliminated, these documents could be entered very quickly.  With one exception, document totals were highest for the four refuges that provided digital files.

## Time Required to Create a Record

Making a record includes both digitizing the document and entering it in ServCat. Table 3 summarizes the approximate length of time required to create ServCat records for various types of documents.

Scanning a document is almost always the most time-consuming part of the process. Scanning time is highly variable; factors that increase this time include:

- A large number of pictures in a document
- The way a document is bound
- An aged document (i.e. frail)
- A document in color
- A lengthy document
- A document printed on fragile paper
- Presence of staples or paper clips

**Table 3: Approximate time required to create a ServCat record**

|  | Avg. # of Pages | Scanning Time (min) | Data Entry Time (min) |
|---|---|---|---|
| Annual Narrative | 46 | 15-30 | 6 |
| Plan | 38 | 5-10 | 10 |
| Refuge Report | 45 | 5-25 | 10 |
| Survey | 26 | 5-20 | 6 |
| Map/Map Series | 5 | ≤5 | 5 |

Annual narratives can take anywhere from 15 minutes to 30 minutes to scan. Even narratives from the last 20 years require a great deal of time because of the large number of pictures usually present. Refuge-specific reports and surveys can be scanned more quickly; except for the occasional outlier, 5 to 20 minutes is typical for a survey and 5 to 25 minutes is standard for a report. Plans were the quickest multi-page document to scan; 5 to 10 minutes was the normal range. Maps and aerial photographs are often a single page, but because they are scanned in color and often at high resolution, they can take up to 5 minutes to scan. Sometimes trial-and-error is required to find the settings that capture these documents with the best quality.

Though not targeted in the pilot, tract records were encountered at one of the refuges. They are classified as "generic documents" in Appendix A. The tract records were of high importance to the refuge; the staff said that they referenced them on a daily basis. These files were more challenging to scan than any of the other document types. They were lengthy and contained correspondence, legal papers, maps, planning documents, and court proceedings. Due to the variety of subdocuments contained within the tract records and the significant range in the ages of these subdocuments, different settings had to be used to scan almost every page. As a result, 30 to 60 minutes was generally required to scan a single tract record.

Metadata entry requires significantly less time than scanning. Five to ten minutes per document is standard. Use of the templates described on page 12 can greatly improve data entry efficiency. The only hurdle that was frequently encountered when entering data was the absence of important metadata (such as a date or author) on a document. When this data was unavailable, the data technicians consulted with the refuge staff to try and recover the missing information.

When budgeting time for a data mining project, a general approximation is that it takes 30 minutes to scan a document and enter it in ServCat. This approximation can be used to calculate how many documents can be scanned during a data mining trip, which will help the refuge staff determine how many documents to prioritize. Additional time should be factored in when a set of documents is known to contain many of the elements listed as time-consuming to scan. Alternatively, if the documents are known to be short, black-and-white, and straight-forward, 30 minutes to create a record will be an overestimate.

## Lessons Learned and Recommendations

In general, the pilot project was a success at every level of the Refuge System, and it showcased the power and value of ServCat. It has furnished a valuable foundation to direct the future of the ServCat database. Many lessons were learned throughout the pilot and these were applied to the subsequent recommendations.

### Variability

The most significant finding was the extreme variability from refuge to refuge. This variability applied to document and binding types, number of documents, number of refuge staff, organization of refuge files, staff assistance, refuge accessibility, and working space. All of these factors played into how successful and efficient the data mining effort was at a particular refuge. Ultimately, a one-size fits all approach is not likely to be successful when implementing this project across the Refuge System. Things that were found to be effective at one refuge were not as useful at another. The recommendations section includes suggestions on how to tailor this process for each region and/or refuge.

### Preparation

#### *Prioritization*
Prioritization was the biggest underlying factor in how many records were created during the week for two reasons. The first was it allowed the data technicians to focus on scanning and entering metadata into ServCat rather than spending time sorting through refuge files and documents. At one refuge, very little prioritization was completed before the technicians arrived. A tremendous amount of time was spent looking for paper documents in multiple locations within the refuge office. Approximately 15 documents were later found to be duplicates or even triplicates in the database. Through the mining process it was discovered that some reports had been copied over the years and were stored in several places. Without refuge staff first pulling and sorting through files, this duplication was inevitable. After this experience, the data technicians increased their communication with refuges prior to data mining trips and instituted a pre-visit consultation. As a result, the refuges better understood how to prioritize their documents, and, on average, the number of ServCat records created increased by 40 records per refuge with little or no duplication.

The second reason prioritization was key to increased efficiency was that it ensured refuge specific documents and those of high importance to the refuge staff were entered into the database. Priority was given to documents that were irreplaceable or valuable to the refuge. One refuge utilized the data

mining project by making surveys regarding an endangered species found on their land high priority. Having these documents in the ServCat, they stated, will help with ongoing litigation and public relations.

It is recommended that refuge staff, with support from the regional and national staffs, continue to lead the prioritization process. The amount of legacy data in a refuge office is daunting; it is not practical to expect that every document will be uploaded to ServCat in the immediate future. Therefore, it is important that the documents that are used the most frequently by the refuge staff are the first to be entered in the database. The staff is familiar with the refuge's filing system and is best qualified to determine which documents are valuable to the refuge. Their help in prioritization will help the full-scale implementation proceed efficiently.

### *Refuge support*

The support of refuge staff, more specifically the refuge point of contact (POC), impacted not only the efficiency of the data technicians but the quality of the metadata entered. The POC was the main contact at the refuge for the ServCat project. They worked closely with the technicians to ensure high priority documents were entered and were also available to assist with any questions or concerns. Ultimately the POC becomes the "owner" of the records entered into ServCat, meaning they have the ability to edit those records. At one refuge the POC, who had prioritized the documents prior to the technicians' arrival, was unavailable for most of the week. Much of the data set out to scan was of a new subject matter and organized in a method not familiar to the technicians. Over the course of the week, decisions were made and some guess work took place to make the documents set out by the POC into records. Without the POC present the methods of organization and document grouping used when entering the documents couldn't be confirmed. More in-depth QA/QC may be required for this refuge.

It was also important to have refuge staff involved in the data mining process because it allowed them to become familiar with the purpose of the pilot project and the function of the database. The data technicians provided demonstrations or training whenever possible. The staff was able to get feedback and recommendations from the data technicians on how to proceed with the data mining effort. Many staff members became excited about the project and made plans to continue scanning documents and creating records after the technicians left.

The role that refuge staff and volunteers played also proved to be invaluable for some of the more tedious tasks involved with the data mining process. Many of the document bindings, for example, were complicated or time-consuming to take apart and then put together again, especially annual narratives. Volunteers were often assigned the task of disassembling and reassembling these reports. This contribution either before the arrival of the technicians or during their visit was monumental in increasing efficiency.

The support of the entire refuge staff will prove to be an important factor in the success of the full-scale implementation of ServCat. It is suggested that refuge staff members and volunteers be involved in the continuation of the data mining effort as much as possible.

### *Unhealthy conditions*

Most refuge documents are stored in filing cabinets and older documents may not have been handled in many years.  Some of these documents can be very dusty and some are in close contact with mouse feces.  These files should not be handled without taking safety measures.

### *Binding types*

Binding and unbinding documents was an unexpected hurdle during the archiving process.  Though most bindings were fairly easy to bind and unbind- such as staples, paperclips, comb bindings, and prong fasteners- there were a few outliers.  Some documents (usually reports) were permanently bound with bolts or a heat sealed strip called velobinding.  To scan these documents would have been very time consuming or would have required that the bindings be broken.  A method will need to be developed for dealing with these types of bindings in the future.  Velobindings can be broken without damaging the document, but without access to a velobind machine, the documents need to be rebound using an alternative method.  High priority documents that are impossible to disassemble will need to be scanned page-by-page on a flatbed scanner.

### *Outdated digital formats*

Several of the refuges that participated in the pilot had old digital files stored in outdated formats.  A number of floppy disks, both 5.25-inch and 3.5-inch, were found.  Some of the disks were labeled with the files they contained, but others were unmarked.  In addition, there were files stored on refuge computers that were unreadable with standard software.  WordPerfect and Quattro Pro files were the most common.  Hardcopies of some of these files may exist, but without being able to open the digital copies, it was not possible to confirm this.

Files in outdated digital formats were not dealt with during the pilot, but these files could be captured as the data mining process continues.  A converter can be downloaded that allows Microsoft Excel to open Quattro Pro files.  Open Office supports a wide range of file formats, including WordPerfect.  It is recommended that data technicians have access to software that can read these file formats.  It is also suggested that external floppy disk drives are made available so that information from disks can be collected.

## Travel

### *Accessibility*

Accessibility became a concern at several of the refuges.  As mentioned previously, a large amount of time was spent traveling to and from the sites, taking time away from capturing data.  Refuges in urban areas were easy to reach, but others required up to 15 hours of travel one way.  In some cases weather conditions or poor directions increased this time.  The Refuge System contains over 500 refuges, a significant number of which are in remote locations.  This hurdle will need to be taken into consideration for future data mining visits.

Weather conditions not only impacted travel time, they also prevented the data technicians from reaching one refuge.  A major snowstorm in Alaska made Tetlin NWR completely inaccessible.  After arriving in Anchorage and finding the roads to Tetlin impassible, the technicians changed plans and spent their Region 7 data mining trip at Kenai NWR.  This situation was a challenge for a number of people.  The staff at Tetlin had spent a lot of time getting ready for the technicians and was

disappointed that their documents weren't digitized.  Clearly, the Kenai staff did not have time to prioritize documents for the technicians.  The technicians lost nearly two full days of scanning time due to the road conditions and were therefore unable to create as many ServCat records as they normally would have.

The time required to travel to refuges needs to be considered as the data mining project proceeds.  Refuge visits should be planned to try to avoid periods when major weather events may occur.  While refuge staff may be more available during winter, refuge access in that season must be considered.  It is recommended that the full-scale implementation minimize travel time as much as possible.


### *Time between visits*

In addition to accessibility, the importance of a time gap between trips was realized after two trips were scheduled back to back.  Some confusion concerning the shipping of the equipment from the first refuge to the second resulted in the scanners never arriving at the second refuge.  Fortunately, many documents at the second refuge were already scanned and the data technicians spent most of the time entering the metadata into ServCat.  The documents that needed to be scanned were done so using the office copier which was a slow process.  The network speed at this refuge was slow and not all of the documents scanned were retrieved from the drives.

When a longer gap was scheduled between refuge visits, it allowed plenty of time for the equipment to get from refuge to refuge, which often took up to five days for remote locations.  If equipment is being sent to a refuge, it is recommended that it be sent well in advance and that the data technicians confirm that it arrived before traveling to the refuge.

The time between trips also allowed the technicians to catch up on data entry.  To maximize the number of items scanned, priority was given to scanning documents during the week over entering metadata.  The week following the visit was focused on cleaning up scanned documents and entering associated data into ServCat.  Without the time to enter metadata later, the technicians had to spend more time at the refuge entering data.  As a result, fewer documents were scanned at these refuges.


## Equipment

The scanning equipment purchased for the project proved to be very efficient and relatively durable; however, there were a few inevitable difficulties.  Part of the challenge was with the scanner drivers, which occasionally became corrupt.  When this happened, the technicians reinstalled the drivers, which quickly fixed the problem.  Towards the end of the pilot project, there were also hardware issues.  Both of the Epson GT-S50 scanners quit working and had to be replaced.  It took several days to get replacement equipment, and the number of documents the technicians were able to scan was greatly reduced.  Having a back-up scanner available for data mining trips is advisable.  It is also recommended that data technicians have local administrative access to their computers so they can reinstall the scanner drivers when necessary.

Despite the challenges that arose, the scanners used in the pilot project are recommended for future data mining efforts.  The hardware issues that the technicians encountered were likely due to wear-and-tear on the equipment due to constant use and being shipped around the country.  Overall, the efficiency of the scanners outweighed the challenges associated with them.

### Miscellaneous supplies

Additional supplies were purchased to accommodate the diversity in bindings and document types. The following is a list of the archivist's basic equipment brought to each refuge:

- Stapler
- Staple remover (large and small)
- Double sided tape
- Acid-free glue stick
- Carrier sheets (10)
- Comb binding machine
- Dust mask
- Clam Clips
- Static-free cloth to clean scanner

## Data Entry Process

The data technicians developed techniques that made the metadata entry process more efficient, and they are discussed below. A quick reference guide for creating a ServCat record is located in Appendix B.

### Abstract templates

The most time consuming element of the data entry process is the creation of abstracts for reports and plans. Often, for documents such as annual narrative reports and management plans, an abstract may only vary by a few words, dates, or ideas. A template handbook, intended to speed up and bring consistency to the data entry process, can be found in Appendix B. These templates contain a basic outline of what a specific plan or narrative report contains. The dates, refuge names, and a few specific details are the primary things needed to make them satisfactory. Supplementary information can be added to these templates when necessary.

### Keyword master list

Listing the keywords in a document is another challenging part of metadata entry. It can be difficult to choose important, descriptive words from a report, particularly if the subject matter is outside the reader's expertise. Using a master list of keywords is recommended because it will help users pick out relevant keywords from a document and it will standardize the keywords used in the database. A proposed keyword master list is provided in Appendix D. This list will be updated as new document types are encountered.

Standardizing keywords will make it easier to select words to use when searching for a document. For example, the words "controlled burn", "prescribed burn", and "planned burn" can be used interchangeably. If a document is entered into the database using the term "prescribed burn," and a user does a search for "controlled burn", that user may have difficulty finding the "prescribed burn" document. If a keyword list is used, and the same term is always used as a keyword for documents of a particular subject matter, then the user can reference the keyword master list and search the database for the terms that will provide them with the most comprehensive list of results.

The first step in the metadata entry process is determining what type of reference a document should be classified as.  In most cases, this is very straightforward; however, there was some confusion at the beginning of the pilot project regarding the difference between a survey and a report that covers a survey.  The following clarification was made:

- A **survey** is a document composed primarily of raw data
- A **report** gives details about the survey process and presents the results of the survey along with analysis.  This type of document should typically be classified as an unpublished report.

It was also common to encounter plans, surveys, and reports that were contained within a memorandum.  These documents should be classified as plans, surveys, or unpublished reports- not memorandums; however, memorandum should be included as a keyword.

# Full-Scale Implementations and Recommendations

The following alternatives, outlined in Table 4, were developed by the data technicians based on several criteria.  The first addresses the demands and time restraints already faced by refuges.  These have driven the focus on ease and a minimal time requirement of record creation.  A second motive was the cost effectiveness of the project on such a broad scale.  Many aspects of the pilot may not roll over to full scale implementation smoothly.  The final motive was efficiency.  Entering the data from over 500 refuges could take many years.  Directing the project over multiple levels within the Refuge System may ease this challenge and allow for better efficiency and cost effectiveness.

## Alternative 1

The technicians recommend providing ServCat scanning and data entry support at the regional level.  This option seems to be the most time efficient and will provide the most support to refuges.  Time and time again refuges have expressed their interest in the project and their willingness to participate but do not have the staff or time to scan and enter such large amounts of data.  Some refuges the technicians visited had as few as three full-time staff members, each person juggling many responsibilities.

A solution to this problem would be to hire two ServCat data technicians in each region that would assist refuges with the data entry process.  The regional technicians could be stationed in the regional offices or in the zones.  The benefit to having them in the zones is that they would be closer to the refuges, which would cut down on travel time and expense.   On the other hand, this would increase the travel time when the data technicians needed to visit the regional offices or the Center.

Having regional support as opposed to strictly national support will allow the refuges to receive more direct contact, assistance, and training.  This option, however, may not be the most cost effective strategy.  Besides hiring the technicians, each region would need to purchase equipment and pay for any travel expense.

**Table 4: Comparison of alternatives for the continuation of the ServCat data mining effort**

| Alternative 1: Regions buy their own scanning equipment and hire two data technicians to travel to refuges within the region. | |
|---|---|
| Pros | Cons |
| • More people will be working on the data mining effort leading to increased efficiency<br>• Traveling will be more localized<br>• More support will be provided to the refuges | • Some refuges won't want technicians going through files<br>• Many refuges are far away from regional offices, so there will be a significant amount of travel for the data techs<br>• Regional techs may not receive the same support as the technicians from the Center<br>• There will be more inconsistencies in record creation<br>• Challenges with travel logistics (weather delays, equipment difficulties, finding lodging, etc.) still exist<br>• Data technicians will need to spend longer than a week at most refuges to capture all priority documents |

| Alternative 2: Refuges scan and enter documents on their own. | |
|---|---|
| Pros | Cons |
| • Records will be created by people who are most familiar with the refuges and documents<br>• Need for travel is eliminated<br>• Regions will not need to hire technicians | • Availability of scanning equipment<br>• Refuge staff most likely will not have time<br>• Inconsistencies in how records are created<br>• A large number of people will require ServCat training<br>• Some refuges will not participate<br>• Database population will take a long time<br>• Difficult to provide adequate support |

| Alternative 3: Refuges send documents to a professional scanning company to be digitized and then send the digital files to the Center to be entered. | |
|---|---|
| Pros | Cons |
| • Consistency of records entered<br>• Need for travel is eliminated<br>• Reduces responsibility and time demands of refuges<br>• Regions will not need to hire technicians | • Quality of scanned documents may not be satisfactory<br>• Database population will take a long time<br>• Cost of having documents scanned<br>• Documents may be disorganized<br>• Some refuges will hesitate to part with their documents |

| Alternative 4: Refuges send boxes of documents to the Center to be scanned and entered. | |
|---|---|
| Pros | Cons |
| • Reduces responsibility and time demands of refuges<br>• Ensures consistency with scanning and record creation<br>• Regions will not need to hire techs | • Data entry will take much longer – backlog<br>• Large amount of physical space required to house and scan documents<br>• Some refuges will hesitate to part with their documents |

| Alternative 5: Regional I&M staff and/or refuge staff prioritize and scan refuge documents and post the digital files on FTP site. The national office will create the metadata records. | |
|---|---|
| Pros | Cons |
| • Consistency of records entered<br>• Crews focus on one aspect of database population (scanning or metadata entry) and become efficient at that aspect<br>• Need for travel is reduced<br>• Formal quality control process for records | • Data techs could get burned out on task repetition<br>• Upload/download times of FTP or data transfer site may be slow<br>• Techs from the Center would need to interact with refuge staff over the phone if they had questions |

| Alternative 6 (Recommended alternative): A combination of alternatives 1-5 | |
|---|---|
| Pros | Cons |
| • Allows regions and refuges to use a method that works specifically for them<br>• Reduces need for travel<br>• Refuges receive regional and national support<br>• Database population will move quickly | • Inconsistencies with record creation<br>• Regional data technicians will need to be hired<br>• Regions will need to buy scanning equipment |

## Alternative 2

An alternative to hiring regional technicians is to provide training to the refuges so they may ultimately create the records.  Apart from training, this option eliminates the need for travel, but it seems to be reasonable only for refuges with a large staff.  Equipment would still need to be purchased at the regional level, but if each staff member participated in the effort, additional support may not be needed.  Refuge staff would need to be very disciplined, entering documents into ServCat as they are produced.  Limited time and multiple responsibilities; however, make this process complicated.  This alternative might be feasible for current documents; refuge staff would be able to create a record for a document they've written more quickly than a data technician could.  Given the large amount of documents that exist at most refuges, this alternative is not likely to be effective without additional assistance for legacy data.

## Alternative 3

Send documents to a professional scanning service to be digitized and then send the digital files to the Center to be entered into ServCat.  This alternative would put the least demand on the refuges; a refuge would need only to sort through documents, prioritize them, and ship them to be scanned and entered.  A disadvantage would be the loss of interaction between refuge and regional I&M staff in determining priority documents.  Also, additional data techs would need to be hired at the Center to enter the data, but even with an increased staff, populating the database would likely take a long time.  Another concern is that the quality of the scanned documents might not be satisfactory; the text in older documents may not be legible.

## Alternative 4

Sending the hardcopy documents directly to the Center to be both scanned and entered would carry many of the same benefits as alternative 3.  Additionally, this option would ensure that the quality of the scanned documents is high and that the documents are handled with care.  The Center would need to hire additional technicians to support this alternative.  The primary concern is that the process would move even more slowly than it would for alternative three.

## Alternative 5

Regional I&M staff and/or field staff prioritize and scan refuge documents and post the digital files on an FTP site.  The Center will create the metadata records and notify the regions and refuges when the records are complete.  The regional I&M staff and/or refuge staff can then perform a quality control check of the metadata records.  They will edit the records if necessary and then activate them.  This alternative would allow for consistency and control of the metadata records by the Center and it would inject a formal quality control process for verification and activation of the records.  It would also increase efficiency because it would allow data technicians to focus exclusively on one aspect of the process- either scanning or metadata entry.  Conversely, this task repetition might make data technicians burn out on the job more quickly, leading to a high turnover.  Another concern is that the upload/download times of the FTP site might be slow, especially on the refuge's end.  Finally, technicians from the Center would have to contact refuge staff by phone if they had questions; there would be no face-to-face interaction.

## Alternative 6 (recommended alternative)

Ultimately, a combination of these alternatives seems to be the best option to address the variability within the Refuge System as well as time and money restrictions.  Hiring two regional technicians is recommended and refuges would have the option of mailing documents to the regional office or having the technicians visit the refuges to scan documents.  Professional scanning services could also be used to

digitize the documents. Scanning documents off site would allow refuges to take the time they need to thoroughly sort and prioritize their files. Some refuges will prefer this option, as it ensures that all priority documents are entered but leaves the time consuming scanning to the regional office. For the refuges that need the extra support of the regional technicians, the second option allows for relatively easy travel and more face time. Data entry could be done at the regional level or the Center, depending on the amount of resources the region is able to dedicate to hiring and training data technicians. This combination alternative allows flexibility for every refuge and region and ultimately will provide the most support for the refuges.

In addition to the regional support, it is advised to hire a "ServCat Coordinator" based out of the Center. The coordinator will guide efforts by the regional technicians, provide support to the regions, and lead efforts to batch and enter groups of related documents. The coordinator will also oversee development and QA of the database and arrange trainings. Part of the training can include webinars for refuge staff so that they can learn to use ServCat and begin creating records for documents as the refuge produces them.

## Conclusions

The Service Catalog pilot project was a great first step in the U.S. Fish & Wildlife Service's data mining effort. During the pilot, 2,473 ServCat records were created from ten wildlife refuges. These records include annual narrative reports, management plans, refuge reports, surveys, maps, and other resource management documents. A number of lessons were learned throughout the course of the pilot, and the following conclusions can be drawn from the information that was gathered.

1. Full-scale implementation of the ServCat database will require collaboration between refuges, regions, and the Center. There is extreme variability in documents from one refuge to another, and a single implementation strategy will not be useful for every refuge in the System. Flexibility should exist so that refuges can select the data mining approach that will best suit their individual needs.

2. Populating ServCat will be an ongoing effort. Dealing with the legacy data will likely take years and will require a tiered approach. It takes approximately half an hour to create a single record; therefore, it is not realistic to assume that all of a refuge's legacy data can be entered in ServCat without it being a refuge priority requiring additional funds. Instead it is important to focus on capturing priority documents and entering new documents into ServCat as they are produced.

3. Standardizing the metadata entered in ServCat will make it easier to find a document in the database and faster to create a record. Templates, the master keyword list, and the ServCat Guidance should be used whenever possible to expedite data entry.

# References

ABBYY.  (2012).  ABBYY FineReader Sprint Plus OCR [Software].  Available from
    http://finereader.abbyy.com/

Epson America, Inc.  (2012a).  *Epson GT-20000 Document Scanner*.  Retrieved April 23, 2012 from
    http://www.epson.com/cgi-bin/Store/jsp/Product.do?BV_UseBVCookie=yes&sku=B11B195011

Epson America, Inc.  (2012b).  *Epson WorkForce Pro GT-S50 Document Scanner*.  Retrieved April 23, 2012
    from http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=B11B194011

# Appendix A: Number of documents entered for each refuge by document type

| | Aransas | Ash Meadows | Don Edwards | Kenai | Malheur | Neal Smith | Parker River | St. Vincent | Stillwater | Tewaukon | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **PLANS** | | | | | | | | | | | |
| **Plan** | 4 | 3 | 20 | 5 | 31 | 18 | 49 | 37 | 4 | 19 | **229** |
| **REPORTS/DOCUMENTS** | | | | | | | | | | | |
| **Book** | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | **8** |
| **Book Chapter** | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | **9** |
| **Brochure** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | **1** |
| **Dissertation** | 0 | 0 | 3 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | **8** |
| **Generic Document** | 11 | 1 | 0 | 8 | 1 | 55 | 14 | 11 | 4 | 68 | **173** |
| **Journal Article** | 0 | 7 | 0 | 0 | 3 | 1 | 3 | 7 | 4 | 0 | **25** |
| **Published Report** | 0 | 2 | 5 | 0 | 5 | 0 | 0 | 1 | 3 | 0 | **16** |
| **Thesis** | 0 | 1 | 4 | 3 | 1 | 13 | 4 | 1 | 3 | 1 | **31** |
| **Unpublished Report** | 195 | 38 | 113 | 63 | 113 | 76 | 202 | 115 | 135 | 39 | **1,089** |
| **DATA/SURVEYS** | | | | | | | | | | | |
| **Generic Dataset** | 0 | 0 | 0 | 0 | 0 | 8 | 2 | 0 | 0 | 0 | **10** |
| **Protocol** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | **3** |
| **Survey** | 29 | 0 | 16 | 8 | 4 | 17 | 32 | 82 | 51 | 20 | **259** |
| **Tabular Dataset** | 22 | 0 | 0 | 0 | 2 | 0 | 7 | 5 | 0 | 0 | **36** |
| **Vector Dataset** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | **1** |
| **MAPS** | | | | | | | | | | | |
| **Map** | 3 | 0 | 1 | 0 | 33 | 0 | 1 | 4 | 1 | 0 | **43** |
| **Map Series** | 0 | 0 | 0 | 0 | 13 | 1 | 0 | 0 | 0 | 0 | **14** |
| **PHOTOGRAPHS** | | | | | | | | | | | |
| **Aerial Photograph** | 0 | 0 | 79 | 0 | 0 | 0 | 72 | 0 | 2 | 0 | **153** |
| **Photograph** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | **7** |
| **PRESENTATION/POSTER** | | | | | | | | | | | |
| **Presentation** | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | **3** |
| **Poster** | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | **3** |
| **LETTERS/ARTICLES** | | | | | | | | | | | |
| **E-mail** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | **1** |
| **Letter** | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | **4** |
| **Memo** | 0 | 22 | 1 | 1 | 3 | 3 | 15 | 14 | 4 | 25 | **88** |
| **Newsletter Article** | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **5** |
| **Newspaper Article** | 2 | 178 | 0 | 1 | 0 | 0 | 0 | 15 | 0 | 3 | **199** |
| **Notes** | 0 | 14 | 2 | 0 | 0 | 0 | 0 | 9 | 2 | 0 | **27** |
| **Permit** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 12 | **13** |
| **TOTAL** | | | | | | | | | | | |
| **Total** | **266** | **281** | **248** | **100** | **217** | **198** | **405** | **318** | **253** | **187** | **2,473** |

# Appendix B: ServCat Guidelines for Creating a Reference

ServCat contains 69 reference types. Each reference type requires specific metadata, so data entry steps will vary. Before creating a Reference, it is important that you confirm whether a Reference already exists in ServCat. This will prevent the creation of a duplicate Reference in ServCat.

- Select Create from the Menu items > References > Create a Reference Option
- Select the reference type from the dropdown list (*Refer to ServCat help manual if uncertain about the reference type)*
- Select "Next"

## Core Information Tab

The Core tab contains important information needed in a reference record. The number of fields and labels changes for each reference type. An asterisk indicates that this field is required. It is recommended that you populate as many fields as possible. This increases the discoverability of the Information Resource.

### *Title*
- Enter the title that appears on the "title page" (*Note: information on the cover page is not always complete)*
- Do not omit or abbreviate words in the title
- Capitalize the first letter of every word except articles, conjunctions and prepositions
- If no title is known, construct a title, derive it from the Information Resource itself, and place brackets around it; e.g. [Baseline Inventory Data for Pea Island National Wildlife Refuge]
- Indicate in the "Reference Notes" that the title was constructed because an original title was not available
- If "Draft" is stamped on the cover, add this to the title and enclose in brackets
- If draft was not on the cover but mentioned in the document, again, add draft to the title and enclose it in brackets; e.g. Conservation Agreement for the Tricolored Heron. [Draft]

### *Brief description:* Provide a brief account of the Information Resource; enter up to 400 characters

### *Date of publication:* Use the date the Information Resource was issued
- Select date format from dropdown list
- Enter data

### *Contacts (Author(s))*
- Click "Add New Contact"
- Select contact type (Author, Editor, Producer) from the dropdown list
- Choose Individual or Business
- Enter data
- Click on "Save Individual Contact" button

*Abstract:* Provide a summary of the Information Resource's content; enter up to 4,000 characters

## *Size definition*
- Click on "Add New Size" button
- Click on "Label", select from the dropdown list
- Click on "Value", enter number of pages
- Click on "Units", select from the dropdown list

# Information Resource Evaluation Tab

The Information Resource Evaluation tab evaluates key qualities of the information resource. This evaluation is strictly for the Information Resource and not for the Reference. DO NOT embed sensitive or proprietary information into the reference itself. This evaluation determines who can see the Reference and download the associated digital file. The tab defaults to non-sensitive & non-proprietary; if this is not the case, enter appropriate information

## *Sensitivity evaluation*
There are three sensitivity options for an Information Resource; non-sensitive, sensitive, and unknown.
- A non-sensitive information resource contains no sensitive information.
- If sensitive is selected, specify why the Information Resource is sensitive. An Information Resource may have a number of reasons it is sensitive. Select all that apply.
- Select "Unknown" if the Information Resource has not been evaluated and may contain sensitive information.

## *Information evaluation*
There are also three proprietary options under the Information Evaluation tab; non-proprietary, proprietary and unknown.
- Non-proprietary:  not protected by trademark, patent, or copyright.
- Proprietary:  protected by trademark, patent, or copyright; made or produced/distributed by one having exclusive rights (e.g. journal article, thesis, proceedings paper).
- Select "Unknown" if the Information Resource has not been evaluated and may contain proprietary information.

## *Quality evaluation*
There are five options to choose from under the Information Resource Quality Evaluation section; High, Medium, Low, Poor, and Unknown. This provides you an opportunity to select the quality of the Information Resource, if known.

## Organizations and Geographical Tab

The Organization and Geographical tab contains tools to a link a reference record to a Fish and Wildlife organization and geographic area. An Information Resource is linked when it contains specific information about the Organization. By defining a link, you can easily search and discover information specific to an Organization. An Information Resource can be linked to one or more Organizations.

- Keep the default box checked to "Automatically Add Bounding Box"
- Choose the Organization(s) that relate to your Information Resource
- Click "Add Organization*" (Note: If you delete an Organization, you also have to delete the bounding box)*
- Custom bounding boxes can be added:
    - Select "Add Bounding Box" button
    - Select the "Draw Bounding Box" tool on the mapper
    - Sketch the bounding area on the map; in the name section about the mapper, assign the new bounding area
    - Click Save

## Keywords Tab

ServCat supports two types of keywords; free-form keywords and biological taxa. Consider the following when assigning keywords;

- User needs; assign keywords that assist users in identifying the Information Resource
- Current usage; user's needs are best met if keywords reflect current usage in regard to terminology
- Consistency; when selecting a keyword maintain consistency; consider using a pre-defined list of keywords
- The more keywords assigned, the greater the probability the user is erroneously led to an Information Resource
- Too few keywords can have an equally adverse effect; the user may miss an item because a keyword was not assigned
- A multiple- word phrase can be used (e.g. Habitat Selection or Wildlife Monitoring)

### Entering keywords

- Assign 5-10 keywords that best describe the overall content of the item
- Keywords should be comma-delimited *(spaces are not required after a comma)*
- Keywords should be Proper Case, meaning that the first letter of each word should be capitalized (e.g. Wildlife Refuges)
- Common words, including 'and' and 'or' are generally lower case (e.g. Boats and Boating)
- Species names are represented in lower case (e.g. Quercus gambellii)
- Do not use Organization name or reference type as a keyword
- Acronyms should be capitalized (e.g. CCP)

### Biological taxa

ServCat treats Biological Taxa like it handles keywords. It is another way both to communicate the content of the Information Resource and also facilitate discovery. Biological Taxa should be linked to a Reference if the taxon is mentioned somewhere in the body of the text or if the reference contains a record for the specific taxon. *(Note: When searching for your species records, you must use the scientific name)*

- Select the "Add Biological Taxa" button
- Type in the taxon, select it
- Click next
- Check the Organization link
- Select "Finish and Close"

## Permissions Tab

The last tab in the form is the Permissions tab.  By default, the record creator is the owner. In this tab, you can add and remove Reference Owners and Viewers. Reference Owners have permission to edit the Reference and download all digital attachments for the Reference.

- Select the "Add Owner/Viewer"  button *(Note: "viewer" is the default when selecting owners)*
- Type in "your active directory name" (e.g. Todd_Sutherland@fws.gov)
- Add additional owners/viewers as needed

## Saving a Record

After each tab has been completed, you will be presented with three options; Save as Draft, Active, or Cancel.

*Save as Draft:* If the record is incomplete (e.g. fields not validated or metadata is missing) save the Reference as a Draft. A Draft reference will only be visible to the record owner.

*Active:*  If all the fields have been completed and validated, you can activate the record.

*Cancel:* If you select "Cancel", this closes the form and data entered will not be saved.

## Adding a Holding Location

It is important that an Information Resource includes the associated digital file. This will promote information sharing and collaboration between Organizations.

### Uploading PDF's
- Select "Add Digital File"
- In the "Upload New Files" panel click on the "Select Files"  button

- Browse to your PDF
- Click the "Upload" button
- Type a brief description of the digital document (e.g. Scanned PDF of paper document)
- Assign additional owners of the digital holding as needed

## Create a Reference by Cloning

If the existing Reference has attributes so similar to the new one you can use the "Clone" tool. When you clone an existing Reference, the new Reference is identical to the original except for the following:

- You are the Owner
- The Reference lifecycle state is "Draft"
- Any Approved Reference Unit links will be set to Pending
- There are no Holding Locations
- There are no comments
- There is no Source XML
- Most related References are not preserved
    - Project bundles are ignored (i.e., Was-Created-By link)
    - Child References are ignored (i.e., Is-Part-Of link) For example, cloning a Book does not automatically create new book chapters. However, when cloning a book, the link to the book series will be preserved.

*To clone a reference:*
- Select "Clone" from the Options Banner
- Click "OK" to continue
- Modify metadata as needed (e.g. contacts, date fields, pagination)
- Upload PDF

## Create a Reference by Uploading XML Metadata

One or more local XML metadata records can be uploaded to create new References in lieu of creating one in the online form. The benefit of using this process is that you don't have to recreate attributes that are common to the two formats.

- Select Create from the Menu items >References>Create a Reference from XML option
- Select a Reference Type option *(Note: "Geospatial Dataset" is the default Reference Type, and currently the only one supported)*
- Hit the "Select Files" button and navigate to the XML metadata
- Select the "Upload" button
- Select the "Edit" option to evaluate the attributes imported from XML *(Note: Each tab should be viewed to confirm that all critical fields are complete)*
- Select the "Active" button *(Note: A copy of the original XML can be viewed by selecting "View Source XML" in the Options Banner; also, an External Holding Location is added to the record. This location reflects the path to where the digital file can be downloaded.*)

# Appendix C: Template Guidebook

This guidebook is intended to provide information about converting annual narrative reports and management plans into ServCat records. Recommendations on entering titles, brief descriptions, and keywords are provided for annual narrative reports. Additionally, fill-in-the-blank templates are provided to help create abstracts for annual narratives and management plans.

## I.  Annual Narrative Reports

Narrative reports summarize the activities of a refuge during a specific period of time; if the report covers a full year (calendar or fiscal) it is referred to as an annual narrative report. The required material in the reports changed over time; however, the narratives generally followed a similar format within a given time period, as summarized in Table 5. Until 1942, narrative reports were typically issued to cover a fiscal year or a fiscal quarter. The fiscal year began on July 1 of the prior year and ended on June 30 of the year in question until 1976. Between 1942 and 1963, narrative reports were issued three times a year in four month segments- January to April, May to August, and September to December. In 1964 the requirements changed, and narrative reports were written to cover an entire calendar year. This was continued until 1973. In 1973, a planning and accomplishment narrative was written for the fiscal year. Narrative reports were written for the 1974 and 1975 fiscal years, and in 1976 the reports covered calendar years once again. A narrative report for July 1 to December 31, 1975 was written to cover the gap between fiscal year 1975 and calendar year 1976. Narrative reports continued to be written for calendar years from 1976 until 2005 when they were discontinued.

**Table 5: Time periods covered by narrative reports**

| Time Period | Period of Time Covered by Narrative Report |
|---|---|
| <1942 | Fiscal year (July 1 to June 30) or quarterly |
| 1942-1963 | Tri-annually (Jan-April, May-Aug, Sept-Dec) |
| 1964-1972 | Calendar year |
| 1973-1975 | Fiscal year (July 1 to June 30) |
| 1975 | July 1 to Dec 31 |
| 1976-2004 | Calendar year (with a few exceptions) |

When creating a ServCat entry, narrative reports should be classified as unpublished reports. Each narrative report for a refuge should be entered into ServCat as a separate record. If there were three narratives for one year, then there should be three entries in ServCat for that year. If a single narrative report was written for an entire refuge complex, then it should be scanned as a single PDF and entered as a single record; however, the metadata for that entry must include information about all refuges in the complex.

### Title

The title should be entered as it appears on the document. Sometimes the phrase "Reviews and Approvals" or the location of the refuge is printed on the title page. This information does not need to be included in the title. The range of dates that the report covers should be included in the title, even if it is not part of the title of the original document. An example is provided below.

[Insert refuge name] : Narrative Report : [Insert dates]

## *Brief Description*
The brief description of a narrative report is a single sentence that repeats the information provided in the title.

> This narrative report for [insert refuge name] summarizes Refuge activities from [insert dates].

## *Keywords*
The keywords for a narrative report will be different for each refuge.  It is not necessary to repeat words from the title or abstract in the keywords list.  If the phrase "annual narrative report" does not appear in the report's title, it should be listed as a keyword.  If the narrative includes a "highlights" section, this is a good place to find keywords.  The following should also be included in the keywords list:
- Any major weather events that occurred on the refuge (droughts, hurricanes, tornadoes, floods, etc.)
- Major fires
- Specific wildlife species that are discussed at length
- Projects or programs that took place at the refuge
- Refuge research projects

## *Abstract*
These abstract templates were developed using the table of contents from a variety of narrative reports.  Specific details do not need to be included here, only broad concepts.  The bold text represents major headings from the report.  Subheadings listed below the main headings should be inserted into the following templates.  Some sections may not be included in a narrative, and the corresponding information should be deleted from the abstract.

### *Tri-Annual Narrative Reports (1942 to 1963)*
This narrative report for [insert refuge name] summarizes Refuge activities from [insert dates].  The report begins with a summary of the **weather and habitat conditions** during the period.  **Water conditions and food and cover** are described in this summary.  **Wildlife**- including [insert wildlife]- is also covered.  The report includes a **Refuge development and maintenance** section which discusses [insert development and maintenance activities].  **Resource management** is outlined; topics include [insert resource management topics].  A progress report on **field investigations and applied research** is also provided.  The **public relations** section of the report discusses [insert public relations topics].  **Items of interest, N-R forms, and photographs** are provided at the end of the report.

### *Planning and Accomplishment Narrative (FY 1973)*
This narrative report for [insert refuge name] summarizes Refuge **highlights and accomplishments** for the 1973 fiscal year.  [Insert highlights] are discussed in the highlight statement.  The **Refuge accomplishments** section covers [insert refuge accomplishments].

*Fiscal Year Narrative Reports (FY 1974-FY1975)*

This narrative report for [insert refuge name] summarizes Refuge activities during the [insert year] fiscal year.  The report begins with a summary of the **weather** and **habitat conditions** during the period.   **Water conditions and food and cover** are described in this summary.  **Wildlife**- including [insert wildlife]- is also covered.  The report includes a **Refuge development and maintenance** section which discusses [insert development and maintenance activities].  **Resource management** is outlined; topics include [insert resource management topics].  A progress report on **field investigations and applied research** is also provided.  The **public relations** section of the report discusses [insert public relations topics]*.*  **Items of interest** are provided at the end of the report.

*Calendar Year Narrative Reports (1976 to 1980)*

This annual narrative report for [insert refuge name] summarizes Refuge activities during [insert year].  The report begins with an **introduction to the Refuge** and a summary of the **year's climatic and habitat conditions.**   **Land acquisition** is also covered in this section, as is the **system status**.  A **construction and maintenance section** is provided next.  **Habitat management** is also covered; [insert habitat management topics] are discussed.  The **wildlife** section of the report discusses [insert wildlife topics].  **Interpretation and recreation** are also described; topics include [insert recreation topics].  **Field investigations, cooperative programs, items of interest, and safety information** are provided at the end.

*Calendar Year Narrative Reports (1981 to ~2000)*

This annual narrative report for [insert refuge name] summarizes Refuge activities during the [insert year] calendar year.  The report begins with a summary of the year's **highlights and climatic conditions.**   **Land acquisition**-including [insert land acquisition topics]- is also covered.  The report includes a **planning section** which discusses [insert planning activities].  **Refuge administration** is outlined; information about [insert administration topics] is given.  **Habitat management** is also covered.  Subjects include [insert habitat management subjects].  The **wildlife** section of the report discusses [insert wildlife].  The **public uses** of the Refuge described in this report include [insert public uses].  The **equipment and facilities** section of the report provides information about [insert equipment and facilities topics].  **Items of interest** are provided at the end.

*Calendar Year Narrative Reports (~2000-2004)*

This annual narrative report for [insert refuge name] summarizes Refuge activities during the [insert year] calendar year.  The report begins with an **introduction to the Refuge** and a summary of the year's **highlights and climatic conditions**.  Information about **monitoring and studies**- including [insert specific studies]- is provided next.  **Habitat restoration and management projects** are also covered; activities include [insert habitat management projects].  **Fish and wildlife management** is discussed next with emphasis on [insert wildlife topics].  **Coordination activities**, such as [insert activities] are outlined.  The **resource protection** section provides information about [insert topics].  Information about **public education and recreation**

is given including [insert public education topics].  Finally, **Refuge planning and administration** are discussed.

## II. Management Plans

Refuge management plans guide a variety of actions and activities on a refuge.  These plans vary from visitor services plans to contaminant contingency plans.  Like the annual narrative reports, their objectives are fairly standard and change little from refuge to refuge.  The following abstract templates are for common management plans.  Supplementary information can be added.

### Cropland Management Plan

The [insert refuge name] Cropland Management Plan focuses to the production of supplemental grain and browse foods to maintain wildlife populations at approved objective levels.  In addition, emphasis is given on the prevention of undesirable brush or trees.  Division and management strategies are discussed.

### Disease Contingency Plan

The DVE Contingency Plan for [insert refuge name] is intended to serve as a ready reference for background information; an inventory of refuge personnel, equipment, and resources; and the outline of a general plan for resolving a potentially hazardous situation.

### Fire Management Plan

This plan considers fire on [insert refuge name] both from the standpoint as a tool for management and as a potential problem to be dealt with.  This document discusses environmental impacts and alternatives covering implementation of fire management on the refuge.  The use of prescribed burning for manipulating vegetation to benefit wildlife or to maintain desirable plant forms is also presented.  Included are burn unit charts and miscellaneous management forms.

### Forest Management Plan

The [insert refuge name] Forest Management Plan is a general plan which outlines the refuge management objectives, forest description, forest management objectives, silvicultural techniques, and problem areas relating to forest management on the refuge.

### Furbearer Management Plan

The [insert refuge name] Furbearer Management Plan directs the management and regulation of trapping.  The furbearer management program directly supports the environmental health, wildlife and habitat, and other recreational use goals of the Refuge.  Furbearers included in this plan are [insert furbearers here].

### Hunting Plan

This plan provides guidelines for administration of hunting activity and for development, maintenance, and enforcement of regulations and guidelines on [insert refuge name].  It specifically

addresses the areas of information, access, use patterns, regulations, area restrictions, permits, and enforcement.  Target species include [insert species here].


### *Law Enforcement Plan*
The [insert refuge name] Law Enforcement Plan clarifies Fish and Wildlife enforcement policies as they apply to the refuge.  It provides information about problems, responsibilities, procedures, and situations.  Specific topics include: [insert topics].


### *Refuge Master Plan*
The [insert refuge name] Master Plan guides the long-range development of the Refuge by identifying and integrating appropriate habitats, management strategies, program elements, and facilities which support the goals and objectives for which the Refuge was established.  This plan is divided into the following chapters: [insert chapters].


### *Safety Plan*
The [insert refuge name] Safety Plan discusses policies for the safety of the station employees, volunteers, and public.  This plan seeks to identify and correct unsafe conditions and preclude unsafe acts through education and on-the-job training.  Contained in this plan are refuge safety objectives, an emergency contact listing, details on equipment and training, and a list of general hazardous conditions.


### *Trapping Plan*
The [insert refuge name] trapping plan outlines trapping areas, species, regulations, equipment, and seasons.  This plan will allow harvest of a renewable natural resource without adverse impacts on furbearer populations or other refuge resources.  Trapping operations will not interfere with non-target species.


### *Water Management Plan*
The [insert refuge name] Water Management Plan has been developed to meet the station objectives set forth in the Master Plan.  The purpose of this plan is to establish a schedule of operations for the manipulation of managed waters.

# Appendix D: Keyword Master List

**Bolded/CAPS**=Primary keyword
**Bolded/lowercase**=Secondary keyword
Keyword following a dash=additional keyword to use if appropriate

## ANIMAL STUDIES
Aquatic Animals
Breeding
Carnivore
Duck Plague (DVE)
Ducks
Furbearers
Habitat Suitability
Predation
Predator
Predator Control
Reproductive Success
Scat
Scent Station
Tagging
Waterfowl Breeding
Waterfowl Diseases
Wildlife Habitat Improvement
Wildlife Monitoring
Wildlife Refuges

**Amphibians**

**Birds**
Aquatic Birds
Bird Banding
Bird Refuges
Game Birds
Migratory Birds
Migratory Waterfowl
Nest and Nesting
Perching Birds
Songbirds
Wading Birds
Waterfowl

**Endangered Animals**
**Exotic Species**

**Fish**

Anadromous Species
Carp and Minnow
Fish Habitat Improvement
Fish – Habitat
Fish Hatcheries
Fish Kill
Freshwater Fishes

**Insects**
**Invasive Species**
**Invertebrates**

**Mammals**
Deer - Hunting
Rodents
Ungulates

**Reptiles**
Alligator
Snake
Sea Turtle
Turtle

**Sensitive animals**
**Threatened animals**

## ANTHROPOGENIC IMPACTS
Cooperative Farming
Dam Effect
Dams
Grazing
Land Development
Land Use
Mining
Ranching

**Contaminant**
Pollution
Pollution Cleanup
Pollution Control

**Hazardous Material**
Chemical Spill
Mercury Contamination
Oil Activity
Oil Spill
Pesticides

**Recreation**
Concessions
Fishing
Picnic
Trapping

# AQUATIC SCIENCES
Water Resource
**Coastal Systems**
**Hydrology**
Flow
Gaging Station
Rivers and Streams
Watersheds

**Limnology**
Lakes
Waterway

**Water Quality**
Dissolved Oxygen

**Water Quantity**
Water Level Management

**Water Rights**

**Wetland - Riparian**
Marsh
Wetland Restoration

**Marine Systems/Oceanography**
Barrier Islands
Current
Coastal Environments
Coastal Marsh
Coastline
Convergence

Current
Dune
Erosion
Estuarine Environments
Estuary
Hurricane
Marine
Marine Birds
Marine Environments
Marine Flora
Marine Mammals
Marine Species
Saltwater Fishes
Sediment Transport
Shore
Tide
Watershed

# ATMOSPHERIC SCIENCES
Acid Rain

**Air Quality**
Air pollution
Climate Change
Climate Effects
**Climatology**
Drought
Moisture
Precipitation
Snow
Storm
Temperature
Tornado
Weather
**Light Pollution**
**Night Sky**

# ECOLOGY
# HABITAT
Alpine Environments
Aquatic Environments
Arid Environments
Chaparral
Grasslands
Habitat Fragmentation

Habitat Restoration
Forest Ecology
Prairies
Rangeland
Tundra
Wildlife Habitat Matrix

## FIRE
Fire Cycle
Fire Ecology
Fire Effects
Fire Prevention
Fire Protection
Fire Regime
Fire Suppression
Forest Management
Prescribed Burning
Prescribed Burn
Wildfire
Wildland Fire

## FUNGI

## MANAGEMENT
CCP
Coastal Zone Management
Comprehensive Conservation Plan
Culling
Ecosystem Management
Environmental Impact Statement
Fire Management
Hunting Surveys
Poaching
Resource Management
Resource Protection
Soil Management

**Fisheries Management**
**Forestry**
Forest Conservation
Forest Management
Logging
Timber Management
**Integrated Pest Management**
**Management**

**Recreation - Aesthetics**
Bird Watching
Fishing
Outdoor Recreation
Rivers – Recreational Use
Sport Fisheries
Sport Fishing
Trapping
Visitor Services
Waterfowl Shooting
Wilderness Area
Wildlife Viewing

**Restoration**
Preservation
**Visitor Impacts**
Public Access
Public Use
**Watershed Management**
Water Supply
**Wildlife Management**
Furbearer Resources
Game Management
Goose Transplant
Pest Control
Pest Management
Predator Control
Predator Management
Upland Game Bird Management
Waterfowl Production

## PALEONTOLOGY
Fossils

## PLANT STUDIES
Boreal Forests
Conifers
Deciduous Trees
Grasses
Ground Cover
Hardwoods
Old Growth Forest

Plant Ecology
Plant Conservation
Prairie Planting
Riparian Environments
Riparian Plant
Riparian Vegetation
Silviculture
Softwoods
Timber Stand Improvement
Vegetation Survey
Wetland Plants
Wildflowers
Woody Vegetation
**Endangered Plant**
**Exotic Species**
Exotic Vegetation

**Invasive Species**
Invasive Plant – Control
**Plant Communities**
**Non-vascular Plants**

## PHYSICAL SCIENCES
**Erosion**
**Geology**
**Sedimentation**
**Sedimentology**
**Soil Science**
Clay
Seismic Survey
Silt
Soil
Soil Conservation
Soil Moisture
Soil Stability
Soil Survey

## [OTHER [not used as a keyword]]
Adjoining Landowners
Buildings, Facilities and Structures
Education Outreach
Emergency Procedures

Environmental Education
Environmental Hazards
Haying
History
Landscape Conservation
Law Enforcement
Preventive Search and Rescue
Public Involvement
Public Lands
Public Participation
Refuge Establishment
Regulation
Rescues
Safety
Signs
Signage
Survey - Opinion
Tourism

**Agriculture**
**GIS**
Geographic Information System
**History**
**Inventory**
**Sociology**
**Monitoring**
Natural History
Social Behavior
Socioeconomic

**ARCHAEOLOGY**
Archeology
Artifacts
Ceramic Archaeology
Historic Buildings – Conservation
Historic Buildings – Restoration
Historic Site – Conservation
Historic Site – Restoration
Human Remains
Pottery
Shard

## Appendix E: Pilot Guidance

National Wildlife Refuge System

# GRAS[1] Pilot Guidance

This document provides guidance to wildlife refuge staff and to regional and national I&M staff, on the pilot program for crating records for the Geospatially Reference Archive Service.   Once a refuge document becomes a GRAS record, it is protected and available for reference by the refuge, the FWS, and by other resource management partners.

**This Guidance was developed by the I&M Legacy Data Working Group**

Mike Artmann
Richard Easterbrook
Kaylene Keller
Brian Loges
Brian Van Druten
Peter Dratch

---

1. GRAS was the working name for ServCat

# Contents

# Introduction to the Natural Resource Data Mining Pilot

This is the Guidance for the Natural Resource Data Mining Pilot that will take place in 2011. The primary purpose of the Guidance is to provide clear instructions so data input can be completed consistently at eight diverse refuges, representing the FWS regions. The pilot data mining project has three goals. The first is to enable a refuge to capture important resource management information into permanent records that can readily be accessed by refuge staff and by others. The second is to learn how long it takes to create a record in GRAS (both the electronic form of the document and the necessary metadata) for the different major types of resource management documents. The third is to provide information to refuges, regions and the national office on the consistent collection and storage of resource information across the National Wildlife Refuge System (NWRS). Once the pilot is completed in 2012, a report will be prepared and regional I&M Coordinators and Data Managers will be able to thoroughly evaluate the pilot project results. The I&M Coordinators can then provide recommendations on how to move forward before data mining is conducted more widely on refuges.

The Inventory and Monitoring Program is collaborating with the National Park Service (NPS) to create a centralized repository for compiling, organizing and making accessible recent and historic information that is important to refuge resource management. This repository will mirror the Reference Application of the NPS Natural Resource Information Portal. This partnership will integrate systems across government agencies, minimize duplication of effort and leverage initial NPS investment into their application to ensure sound stewardship of public funds. The Refuge System application is called the Geospatially Referenced Archive Service (GRAS) and compiles documents and organizes datasets, such as reports, surveys, databases, geospatial data and images. The information is then easily discoverable and retrievable in GRAS using text and geospatial search tools.

It is intended that GRAS will become the electronic reference library of the NWRS for information both historical and current that has relevance to resource management. For it to achieve that purpose, refuges will need to provide their priority information, and they will need support to get those materials documented and uploaded. This pilot will enable us to work with the refuges that have volunteered to participate and learn how to best retrieve their valuable resource information most efficiently.

The data technicians will have two types of scanners to electronically capture paper documents that are stored at the refuge. If there are electronic versions of these documents available it will be a significant time-savings to locate them in that form. Then the document can be uploaded directly and only the addition of the metadata is required to complete the record. We estimate that it could take about forty hours of staff time by the refuge staff and zone biologists and eighty hours of data technician time at each location to complete the pilot project. If there are major problems we will adjust the data gathering protocol, but we see this as an experiment in data gathering with eight replications at diverse sites, represented by the size and age of the refuges. If we follow the following implementation plan, we should be in a better position to evaluate the effort and improve it in the future.

## The Refuge Role

The data mining effort is a partnership between the I&M Program at the national and regional levels and eight refuges that are participating in the pilot. The data technicians are well trained and equipped for

their week of intensive work at the refuge.  The crucial role of the refuge staff, working with regional I&M staff when they are available, is to prioritize the information so that documents of the highest resource management utility become the first GRAS records during the time that the data technicians are at the refuge.

The Legacy Data Working Group identified five major information categories to capture in GRAS.  They represent both important documents that occur across the NWRS (i.e. annual narratives and management plans) and documents that are unique to each refuge (generally handled later in the week so refuge and regional staff have time to discuss them).  The value of gathering diverse types of information for GRAS is to keep a refuge from only targeting one type of information and also to estimate how long it takes to complete a record for different types of documents.  The refuge and regional professionals determine the priorities for each day, and discuss these priorities with the data technicians.  The technicians will record how far they get with the refuge priorities on each day. The five major areas to be addressed in the pilot project are:

- Annual Refuge Narratives
- Management Plans (all types from Hunt Plans to Water Management Plans)
- Reports of Specific Refuge Relevance (as determined by refuge and regional staff)
- Biological Inventories and Surveys
- Maps and GIS Datasets Unique to the Refuge.

A thumb drive will be provided to the Refuge to facilitate copying of the available digital documents to be included in GRAS.  The desktop GRAS application requires a refuge point of contact that would be selected by the refuge as the technician's primary contact at the refuge during the pilot project.  This individual would likely also represent the refuge as a Unit Steward/Unit Point of Contact (UPOC) for the centralized web application of GRAS.

Ideally, the refuge staff will have the documents for the day that are to be turned into GRAS records ready in priority order for the technicians each morning of their week at the refuge.  The technicians will work through those documents, asking the staff if questions arise that is necessary for capture and GRAS documentation.  At the end of the day, the technicians will note on the Document Inventory Worksheet what information was not included because of time constraints, and estimate the percentage of priority documents that were processed.

## Data Technicians Responsibilities

Data technicians will work with refuge and regional staff to identify available documents in each of the major areas to be included in GRAS.  Each data record will be entered into the desktop version of GRAS by the data technicians. Creating a record entails scanning the documents to put them in digital form when necessary, uploading the documents in digital form, and entering the appropriate documentation or metadata.

Each day, the data technicians will focus on one of the 5 major areas to insure that some documents of that type become GRAS records.  After they have uploaded the priority information selected by the refuge for that category, they can move on to other documents.  This outlines the daily routine for creating GRAS records:

1. Ask participating refuge and regional staff for the prioritized documents for that day (e.g., specific years of annual narratives that are most valuable to the refuge).
2. Create the digitized file for the document
    a. Hardcopy – Scan the document following the scanning instructions and enter the reference into GRAS.
    b. Digital Documents – Convert to an Adobe PDF format.  If the document is in a format (e.g., WordPerfect) unreadable by available software, make a note of this format in the provided spreadsheet.
3. Enter metadata into GRAS to complete the record for the document
4. Write a brief daily summary that accounts for the records created and estimates the percentage of priority documents that are now GRAS records and those that remain.  Address any issues (e.g., hardware and software problems, quality and quantity of documents) that arise during the work day.  These five daily summaries should be compiled at the end of the week, discussed with the refuge and regional staff and presented to I&M staff once returning to Fort Collins.

If the data technicians complete entry of all of the material in a particular subject area, they can use the remaining time to move to the subject from another day requiring more time.

## Guidance for all data types

### Creating a Record

To create a record in GRAS, you should first be familiar with the FWS GRAS Documentation and Help Manual.  This manual contains an introduction, process overview and all the background information for the desktop GRAS application.  In general, the user is required to create a reference record for each document.  The following questions should generally be answered to properly cite a document:

- **Where** – NWR that the information resource covers? (Units)
- **What** – Is the title? Can the information content be described? (Title, Brief Description)
- **When** – Is there a date on the document? (Content Begin Date, Content End Date)
- **Why** – Purpose of the document? (Brief Description, / Abstract / Full Description)
- **Who** – Who produced the document? (Contacts)


### Scanning

Three image scanners will be mailed to each refuge in advance of the technicians.  There will be two duplex, high-speed, auto feed document scanners and one large format (11" x 17") image scanner.  This pilot project will not scan documents larger than 11" x 17".  If the refuge determines that these large format documents are a priority, then it will be the refuge's responsibility to digitize them.

All documents will be scanned to multi-page Adobe PDF format (black & white, grayscale or color) with optical character recognition (OCR) enabled.  These documents will then be searchable once they become GRAS records.  Below are the recommended settings based on the type of information being scanned:

| Information Type | Resolution (dpi) | Color Specification |
|---|---|---|
| Text | 300 | Monochrome |
| General B&W | 300 | 8 bit grayscale |
| Detailed B&W | 600 | 8 bit grayscale |
| General Color | 300 | 8 bit color |
| Detailed Color | 600 | 24 bit color |

## File Naming Convention

A standardized convention for the digital files will be used.  The default naming conventions is as follows:  RefugeLITCode_FirstAuthorLastName_YearOfPublication_Title.  Digital files will be placed on portable hard drive until they are uploaded.  Refuge LIT Codes can be obtained from: http://www.fws.gov/stand/standards/IFWS_LIT_LRS_012406.txt

## Folder Structure

Scanned documents will be placed on portal hard drives.  These hard drives will be pre-formatted with a folder structure that mirrors the five major areas addressed by this pilot project.

## Points of Contact

The Main Form in the desktop GRAS application requires that a refuge point of contact be selected.  The refuge point of contact should be the technician's primary contact at the local refuge during the pilot project.  This individual would also be selected as one of the refuge's Unit Steward/Unit Point of Contact (UPOC) for the centralized web application of GRAS (see section below).  Regional I&M Data Managers and Biologist will also be UPOC's for the refuges in their regions.  UPOC's are trusted data stewards who ensure the quality, integrity, and relevance of information about their refuge in GRAS. They use personal judgment and assistance from other subject matter experts to approve and activate references containing information specific and relevant to their refuge.

## Information Resource Evaluation

Each information resource will be evaluated by the UPOC for:

- Legal and Non-legal Sensitivity (Information Resource Sensitivity)
- Copyright (Information Resource Proprietary Rank)
- Quality (Information Resource Quality Rank)

These criteria ultimately are important not only for communicating the type of information being managed, but also who should have access to that information (via Visibility).  If the sensitivity, copyright and/or quality status of the information resource is unknown or unclear, the refuge point of contact should be consulted.   Refer to Section 5.1 of the GRAS Documentation and Help Manual for more information.

## Centralized Web Application

The technicians will be utilizing a desktop version of GRAS, which was created in Microsoft Access, to reference information resources.  Digital data and their references created in this pilot project will be imported from this desktop application into the centralized web application of GRAS when it is operational in early 2012.  Once this import is complete, information resources will be easily discoverable and retrievable in GRAS using text and geospatial search tools.  GRAS will eventually be integrated into the Environmental Conservation Online System (ECOS) and NRPC staff will provide training and technical support for NWRS users.

# Annual Narrative Reports – Day 1

Annual narrative reports offer a rich source of historical and contemporary information about wildlife refuges. In the early years, narrative reports were commonly written as monthly or quarterly summaries. In the early 1960s it was common to find a single document covering the entire calendar year. Often written by a refuge biologist or manager, these reports provide a detailed look at how refuges operated, what species were present, when infrastructure was constructed, weather information, historical photographs, and many other items regarding activities on refuges. Most refuges contain narrative reports from refuge establishment through the early 2000s.

Technicians have two primary goals for narrative reports.  First is to clearly document and inventory the range (number of years) and condition of narrative reports for a refuge. This information will assist future efforts if further narrative scanning is necessary after the initial pilot, with the assistance of the I&M Program. This inventory of annual narratives should be captured in the provided Document Inventory Worksheet. The Inventory should be sure to note any missing, damaged, or incomplete narratives in the description field.  Some stations have already scanned and catalogued many of their narrative reports. With preplanning before the technicians arrive at the refuge the technicians should learn from refuge staff about the location and status of narrative reports, and the order in which they should be scanned. Given the number of narrative reports in existence at every refuge (in some cases 60-80 years of documents) and the limited amount of time during the pilot study to focus on narratives, technicians should first determine from  refuge staff the priority order in which narratives are made into GRAS records.  If annual narratives have been previously scanned, only the metadata needs to be entered with the digitized document for them to become records.   The second goal of the technicians is to scan narratives according to the refuges priority for use of the documents. The scanning process will create a high resolution digital record of the document. The digital record can be safely stored, organized, and uploaded into GRAS.

## Scanning Narratives

### Document Quality

If the highest priority narratives are determined to be the oldest ones, special consideration and care will be required to scan these documents. Many of the early narratives were typed on carbon copy paper. While this medium is stable, it is very fragile and requires special handling during the scanning process. The Service recognizes old narrative reports as museum property, and thus they must be handled with great care. Narrative reports on carbon copy paper will be placed in special carrier sheets to prevent damage to the documents during the scanning process.   In addition, care should be taken during the unbinding process.

### Bindings

Narratives should be disassembled prior to scanning, if possible. If the binding does not allow it to be disassembled without causing damage, then the individual pages should be captured page by page as would be the case if scanning a bound book taking care to make sure pages are properly aligned, centered, etc.  Refuge staff should consider the large amount of time required for page-by-page scanning when prioritizing narratives.   If the document's condition is so poor that it cannot be scanned without causing damage, it should still be catalogued and its location referenced in GRAS, so that its contents can be captured at a later time without putting the document at risk.

### Scanning Process

Individual disassembled pages should be scanned at 300 DPI. They should be scanned as a multi-page PDF file. If a page contains an actual photograph taped to it, the photograph should not be removed. Scan the entire page as usual. Scan settings for text and black and white photos should be set to gray scale (for black and white photos) or to color (for color photographs). For those narratives that exist as four quarterly reports or monthly, they should be scanned and then combined into one document reflecting the calendar year. If a narrative is mostly black and white with only a few color pages, scan the sections separately (black and white sections in grayscale and color sections in color) and then recombine them into one document.

## Management Plans – Day 2

As land managers at National Wildlife Refuges for over 100 years, refuge personnel have long been guided by a multitude of documents, including resource management plans. The range of management plans at individual field stations can be extensive, so use the following list of typical planning documents to determine if such plans exist and are available.

1. Fill out the provided Document Inventory Worksheet.  Add more rows if they are needed.
2. Note the date of the plan or range of dates if multiple plans exist (e.g., 1964-1989).
3. If the plan is already digital, copy the complete document to the provided thumb/hard-drive.
4. Add the file name to the table if it has been copied to the thumb drive.

### Typical Planning Documents:

- Comprehensive Conservation Plan
- Habitat Management Plan
- Integrated Pest Management Plan
- Comprehensive Management Plan
- Endangered Species Act Recovery Plan
- Invasive Species Management Plan
- Fire Management Plan
- Visitor Services Plan
- Water Management Plan

- Disease Management Plan
- Hazard Communication Plan
- Spill Prevention Plan
- Land Protection Plan
- Refuge Master Plans
- Project Specific Plans and Records

Some of these documents may already be in digital format and available from an offsite location.  For example, many Comprehensive Conservation Plans have already been scanned and are available from the National Conservation Training Center.  Before scanning begins, and GRAS records are created, the refuge and regional staff should identify and prioritize the management plans that will become GRAS records.

Other types of management plans are historical in nature. Old Refuge Master Plans were typically created soon after refuge establishment.  Some information in these Master Plans may not be in Annual Narrative reports from that time period.  From a historical perspective, it is valuable to scan and document historical management plans. Consult refuge personnel for priority in scanning management plans no longer in use.

## GRAS Document Type

Using the desktop GRAS application, plans should be referenced as "Plan" reference types.   If there are other documents developed for a plan (e.g., Environmental Assessments used to develop CCPs or HMPS), a "Project" reference type should be used for these documents.  Both the plan and any associated documents should then be bundled together into this project (see Section 4.3 of the FWS GRAS Documentation and Help Manual).

## File Formats

Many of the current management plans should be in digital form in various formats. Convert digital files from their native format to the PDF file format before loading them into GRAS.  It is unlikely that all management plans will be scanned or converted to PDF files during the site visit.  Therefore, be sure to complete an inventory of remaining management plans using the provided Document Inventory Worksheet.

# Refuges Specific Reports – Day 3

The focus of Refuge Specific Reports day is to insure that the variety of information that is specific to the refuge in accomplishing its mission is documented.  Target documents will contain information on refuge natural resources and may include internal reports, publications, popular articles, theses and dissertations. In particular, the inventory captures references that may not be easily located via commercial on-line reference services (e.g., BIOSIS, JStor) or via public or academic libraries.

## Consult with Refuge Staff

Talk with refuge and regional staff to determine if there are high priority reports specific to the Refuge that need to be scanned.

1.  Ask if they have an on-going project (e.g., HGM, WRIA) that could benefit from the scanning of certain reports so they can be referenced as GRAS records.
2.  Also ask if there is a need to stratify by date or type of report.  For example, one station may want all documents from an era (1950s) where another station may want every report related to water resources scanned.
3.  Examples of target documents for GRAS input are described in the table below.

## Search Refuge Files

Start looking through the files (hardcopy and digital).  Examples of target documents for GRAS are described in the table below.

If there are different drafts of a refuge document before it is finalized, only the final report should go into GRAS.  If a refuge document is changed in later years, those versions should be entered into GRAS with their date, noting that they replaced the previous document.

| Source | GRAS Document Type |
|---|---|
| **Environmental Documents written for the NWR.  For example NEPA Environmental Impact Statements, Biological Opinions** | Unpublished Report |
| **Monitoring Reports** | Unpublished Report |
| **Inventories of biotic or abiotic resources** | Survey |
| **Cultural resource reports (Note:  Some cultural resources reports contain sensitive information and need to be documented as so in GRAS to restrict access (see Section 5.1 of the FWS GRAS Documentation and Help Manual)..  Refuge staff should be consulted prior to its inclusion** | Unpublished Report, Thesis, or Project |
| **Water resources reports** | Unpublished Report |
| **Documentation of management actions** | Project, Protocol, or Notes |
| **Research reports** | Published Report, Thesis, or Project |
| **Water management reports** | Unpublished Report |
| **Research reports on Refuge resources** | Published Report, Thesis, or Project |
| **Natural Disaster Occurrence Reports (hurricanes, floods, wildfires, etc.)** | Unpublished Report |
| **Herbicide Usage Reports** | Unpublished Report |
| **Contaminants Surveys** | Unpublished Report |
| **Cooperative Farming or Grazing Reports** | Unpublished Report |
| **Out of print reports, books, etc. that deal with important species or communities found on the refuge** | Book or Published Report |

# Biological Surveys and Inventories – Day 4

Many stations have a rich history of biological surveys, censuses, and inventories documented in a wide range of formats from final reports to unprocessed field data sheets.  It will often not be practical to reference all of these in GRAS. Refuge staff will need to review and provide guidance on which items should be captured in GRAS.  Refuge-specific reports will of course include biologically relevant data, and those were captured on Day 3 (Refuge Specific Reports).  Other common reference types housing biological data, GIS datasets, and maps, include: generic collections, generic datasets, notes, plans, posters, protocols, specimens, surveys, and tabular datasets.

## Overview of steps

1. Refuge staff find and list known surveys and inventories, prior to the refuge visit if possible.

2. Refuge and regional staff prioritize known surveys and inventories.

3. Data technicians begin capturing priority items and provide feedback to refuge staff on the feasibility of capturing all of the high priority items.

4. Refuge and regional staff can then refine the priority list as needed to meet time restrictions.

The following criteria of station relevance, quality, and permanence are a recommended approach to aid in the prioritization of biological survey and inventory data.  Each criterion could be evaluated equally, weighted, or as group to help develop a ranking of discovered items.  Once a list of all known surveys is developed, each can be rapidly assessed against the three parameters and prioritized. Following the assessment the list of inventories and surveys may be ranked or simply placed into categories (high priority, medium - capture if time allows, or low - no need to document).

- **Station Relevance:**  Data informing planned, past, or ongoing decisions on conservation actions would be considered most relevant and adhere to guidance outlined in the draft I&M policy. Data used to determine and prioritize resources of concern, set population or habitat objectives, or determine limiting factors (habitats) are a few examples of datasets with a direct relation to a station's effort to plan conservation delivery. A few specific examples of the above could include: inventory of refuge vegetation, refuge specific surveillance monitoring, invasive plant species inventories, cooperative surveys, and surveys of state or federally listed species. Although a data set can be important at larger spatial scales, those with a primary purpose of informing decisions at the landscape, regional, or national scale would rarely be unique to a station.  Questioning whether the dataset addresses a specific problem, a set of problems, or a plan objective will help establish its refuge-specific relevance.  Datasets that can address information needs identified in station specific structured problem solving techniques would be highly relevant in most cases.

- **Quality:**  Determining the quality of a dataset is not simple, but it is usually linked to a well-developed and complete protocol.  A well-developed protocol will completely address the following: objectives, sample design and methods, data management (including data standards),

analysis and reporting, personnel and training, operational requirements, and references.  A quick measure of dataset quality can be made by evaluating consistency in data formats across years or replicate samples, clearly defined fields for tabular datasets, and accuracy in transcribed versions (field sheets to spreadsheet files).  Many datasets in electronic file formats may require a high level of interpretation to assess quality, likely beyond the time commitments available for the GRAS entry.  In such situations it would be advisable to give the scientist the benefit of doubt and include it, enabling a more detailed review of the dataset at a later time.

- **Permanence:** Is the data at risk?  Data that have been submitted to a centralized database and remain accessible for queries, reports, and general use by station staff would be considered low risk.  Unprocessed paper data sheets housed in a station's library, an empty biotech's office, and a biologist's bookshelf would represent data at high risk.   Datasheets that have been transcribed and verified into an electronic file format would have reduced risk although attention should be given to the viability of the electronic file format and quality control measures applied during data entry.

# Refuge-specific Maps and GIS Layers – Day 5

Geospatial information is anything that can be rendered onto a map; that is, information that is associated with real-world coordinates that would specify their location somewhere on the Earth.  This information may be manifested as maps, Geographic Information System (GIS) layers (such as geodatabases, shapefiles or coverages), Computer-aided Design (CAD) files, (Global Positioning System (GPS) files or similar mechanisms (data models) that contain coordinate geometry.  Photographs, especially aerial photos, may be considered spatial if location information is included with the photos.  Cataloging, interpreting, and organizing this information may require experience in GIS, GPS, remote sensing, air photo interpretation, geography, and/or cartography.  At a minimum, experience with GIS is necessary for handling digital spatial data.

## GIS Datasets

GIS datasets should only be added to GRAS if they meet the following conditions: (1) it contains refuge-specific data; (2) it is a final version of a geospatial dataset; and (3) it has FGDC-compliant metadata.  Metadata must contain all of the 'required' elements to be considered complete and be in XML format.

GIS datasets do not need to have been created by the refuge to be referenced in GRAS.  However, GIS datasets produced by, and readily available from, other entities (such as USDA NAIP Imagery, NRCS SSURGO soils data, USGS DLG , DEM and DRG products) should not be uploaded unless they have been significantly modified by a refuge.  For example, clipping a USGS National Hydrologic Dataset (NHD) to the refuge boundary would not constitute a significant edit/modification.  However, if NHD stream networks were modified by refuge staff with GPS positioning, this would be considered substantial.

A dataset that is a work-in-progress should not be added to GRAS.  However, many GIS datasets would never be classified as 'final' because they are dynamic in nature.  For example, a shapefile of road kill

bears would be updated whenever there is a new mortality. If metadata exists for these dynamic resources they should be included in the GRAS record. The web-based GRAS application, available in early 2012, is scheduled to have versioning available to support these types of datasets.

All data sharing agreements and data restrictions should be carefully evaluated before data are incorporated into GRAS records. If a dataset is proprietary or legally sensitive (e.g., some Natural Heritage or archeological data) this must be documented in GRAS to restrict access (see Section 5.1 of the FWS GRAS Documentation and Help Manual).

If not contained within a single file, GIS datasets, CAD Files, GPS files or similar mechanisms should be compressed using WinZip before loading into GRAS.

The following list contains GIS datasets and maps that may be refuge priorities. If the dataset and corresponding metadata are complete:

1. Fill out the provided Document Inventory Worksheet. Add more rows if additional datasets are developed and maintained by the refuge.
2. Copy the complete files and metadata to the proper folder location on the provided thumb/hard-drive.
3. Add the file name to the spreadsheet if it has been copied to the thumb/hard-drive

**GIS Datasets and Maps (not in priority order)**
* Created/Contracted by Refuge

- Aerial Photography *
- Buildings/Structures
- Cultural Resources
- Culverts, Dams and Water Structures
- Fences and Gates
- Geology *
- Historic Maps
- Hydrography *
- Invasive Species
- LiDAR, Contours and Elevation *
- Monitoring Sites
- Oil, Gas and Mining
- Orthophotography *
- Refuge Management Units/Zones
- Roads, Railroads and Trails *
- Signs
- Soils *
- Species (Flora & Fauna) Observations
- Springs & Wells
- Survey Monuments / Data *
- Utilities
- Vegetation, Land Cover and Habitat *
- Wildfire & Prescribed Burning

The FWS GRAS Documentation and Help Manual should be consulted to properly reference GIS datasets. Using the desktop GRAS application, GIS datasets will be documented as one of the following reference types:

1. **Geospatial Dataset** – Any dataset that is inherently geospatial in nature. A geospatial dataset would not easily be categorized as either a raster or vector dataset. For example, ESRI geodatabases and Autocad Drawing files can contain both raster and vector data.
2. **Raster Dataset** – Storage format for managing the location of geospatial information as grids and their associated attribute information. Raster datasets are natively geospatial. Orthorectified aerial photography, USGS digital elevation models and national land cover datasets are just a few examples.
3. **Vector Dataset** - Storage format for managing the location of geospatial information (points, lines and polygons) and their associated attribute information. Vector datasets are natively geospatial. Trails, structures and soils GIS layers are common stored in vector format.

## Analog/Hardcopy Maps

Similar in nature to digital GIS datasets, hardcopy maps and aerial photos contain a wealth of geographic information pertinent to refuge management. This information should be added to GRAS if they meet the following conditions: (1) contains Refuge-specific data; and (2) is a "final" version.

These analog products do not need to have been created by the refuge to be referenced in GRAS. However, if they were produced by and are readily available from other entities (e.g., USGS 7.5-minute topographic maps, USFS Forest Maps and BLM District Maps) they should not be uploaded unless they have been annotated by the refuge.

Using the same thematic list (above) search the refuge files for hardcopy report maps, aerial photographs, or stand-alone maps with a legend and enter the information into GRAS.

Adequate information about each map/aerial photograph must be available to properly document these analog products in GRAS. The following questions should generally be answered:

- **Where** – NWR that the information resource covers? (Units)
- **What** – Is there a title? Can the information content be described? (Title, Brief Description)
- **When** – Is there a date on the map? (Content Begin Date, Content End Date)
- **How** – How was the data collected? (Brief Description, / Abstract / Full Description)
- **Why** – Purpose of the map? (Brief Description, / Abstract / Full Description)
- **Who** – Who produced the map? (Contacts)

Do NOT include maps that were generated using GIS unless the Refuge indicates that they do not have the GIS data used to create the map.

Scanning of maps and aerial photos larger than 11" x 17" is not possible during this pilot project as large format scanners are not being provided to the technical team. If a map or aerial photo is significant to the refuge, but is too large to scan, it should still be referenced in GRAS. In this situation the holding

location would be set to "Physical Holding" (see Section 4.2.3 of the FWS GRAS Documentation and Help Manual) and no digital file would be associated with the record.

Prior to the team's arrival, the refuge should consider having high-priority, large-format maps (larger than 11" x 17") scanned by private entities, such as architect firms, copy services, surveyors or by other nearby government offices if large format scanners are available. Maps and aerial photos should be scanned at 600 dpi. Monocolor maps and black and white aerial photos should be scanned in grayscale.

Using the desktop GRAS application, maps would be referenced as "Map" and aerial photos would be referenced as "Aerial Photograph" reference types.

## Finishing Day 5 and Wrapping Up

As with the creation of GRAS records in previous days, priority maps and GIS layers that are not captured in GRAS due to time constraints should be identified with location, in the Document Inventory Worksheet. This will enable them to be included at a later date, and allow the percentage of refuge materials documented in the data mining pilot to be estimated.

Allow at least an hour for Friday close out – for refuge and regional staff and the data technicians to meet and discuss the week. This includes going over the daily narrative notes and Document Inventory Worksheet and discussing anything that might have made the data mining effort go more smoothly. After they have been reviewed and edited in Fort Collins, copies of the digitized documents including examples of metadata documentation will be sent to the refuge and the region as tangible deliverables of the GRAS pilot project.