



Improving Bibliographic Records for the Alaska Region in the Integrated Resource Management Applications Data Store

Final Project Report

Natural Resource Report NPS/AKR/NRR—2014/752



ON THE COVER

Alaska Region Inventory and Monitoring Program bookshelf.

Improving Bibliographic Records for the Alaska Region in the Integrated Resource Management Applications Data Store

Final Project Report

Natural Resource Report NPS/AKR/NRR—2014/752

Steven P. Johnson
Julianna E. Braund-Allen

Alaska Resources Library & Information Services (ARLIS)
3211 Providence Drive, Suite 111
Anchorage, Alaska 99508

Cooperative Agreement H8W07110001: Improving the Natural Resources Information System (NRInfo) Records and Holdings for Alaska's Parks

January 2014

U.S. Department of the Interior
National Park Service
Natural Resource Stewardship and Science
Fort Collins, Colorado

The National Park Service, Natural Resource Stewardship and Science office in Fort Collins, Colorado, publishes a range of reports that address natural resource topics. These reports are of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

The Natural Resource Report Series is used to disseminate high-priority, current natural resource management information with managerial application. The series targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

This report received informal peer review by subject-matter experts who were not directly involved in the collection, analysis, or reporting of the data.

Views, statements, findings, conclusions, recommendations, and data in this report do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

This report is available in digital format from the Alaska Resources Library and Information Resources (ARLIS; www.arlis.org) and the Natural Resource Publications Management website (<http://www.nature.nps.gov/publications/nrpm/>). To receive this report in a format optimized for screen readers, please email irma@nps.gov.

Please cite this publication as:

Johnson, S. P. and J. E. Braund-Allen. 2014. Improving bibliographic records for the Alaska Region in the Integrated Resource Management Applications Data Store: Final project report. Natural Resource Report NPS/AKR/NRR—2014/752. National Park Service, Fort Collins, Colorado.

Contents

	Page
Contents	iii
Tables	iv
Acknowledgments.....	v
Project Team	1
Origin and Purpose of Project.....	2
Project Products	3
Milestones.....	3
Alaska Data Store Records	5
Distribution by Year	5
Distribution by Reference Type	5
Occurrence of Duplicates within Reference Types	6
Reference Type Differences: IRMA Data Store and EndNote.....	6
IRMA Data Store Reference Types as Exported to Endnote	7
Project Procedures	8
Retrieving Records by IRMA Data Store Reference Type in EndNote	8
EndNote Custom Fields.....	9
Duplicate Detection and Editing Procedures.....	9
Processing Records with No Year or Date of Publication	11
Data Management Practices	11
Lessons Learned and Recommendations	13
Future Deduplication/Record Editing Projects in IRMA Data Store	13
IRMA Data Store Functions and Documentation.....	13
Literature cited.....	15

Tables

	Page
Table 1. Distribution of records by publication year/date at conclusion of project.	5
Table 2. Distribution of records by IRMA Data Store reference type.	5
Table 3. Occurrence of duplicate records within IRMA Data Store reference type.	6
Table 4. IRMA Data Store reference types as exported to Endnote	7
Table 5. Endnote custom fields.....	9
Table 6. Format of uncompleted worksheet.....	10
Table 7. Model of completed worksheet.....	11

Acknowledgments

The project team thanks K. Dratch and B. Frakes for significant technical support, particularly for installing the project's recommended corrections into the production Data Store. W. Johnson and S. Wesser made helpful suggestions for the project reports. We thank M. Shephard for recognizing the need for and initiating the project.

Project Team

Agreement Technical Representative

Sara Wesser, National Park Service Alaska, Inventory & Monitoring

Technical Consultants, Inventory & Monitoring

Cuyler Smith, National Park Service Alaska, 2011

William Johnson, National Park Service Alaska, 2012–2013

Principal Investigators

Julianna E. Braund-Allen, ARLIS / University of Alaska Anchorage, 2011–2013

Steven P. Johnson, ARLIS / US Fish & Wildlife Service, 2011–2013

Record editors

Judy Alward, ARLIS / University of Alaska Anchorage, 2011–2013

Sonia Barajas, ARLIS / University of Alaska Anchorage, 2012–2013

Roxie M. Duckworth, ARLIS / University of Alaska Anchorage, 2013

Sara Evans, ARLIS / University of Alaska Anchorage, 2013

Adam Mehlhorn, ARLIS / University of Alaska Anchorage, 2013

Sharon G. Prien, ARLIS / Bureau of Land Management, 2011–2013

Mary L. Sonnichsen, ARLIS / University of Alaska Anchorage, 2011–2013

Network administrator

Adolf A. Pantoja, ARLIS / University of Alaska Anchorage, 2011–2013

Origin and Purpose of Project

In 2011, the National Park Service began a two-year project with the Alaska Resources Library and Information Services (ARLIS) to remove duplicate records and to improve record quality for approximately 30,000 bibliographic records in the Alaska Data Store portion of NPS's national database, the Natural Resources Information System (NRInfo; later renamed IRMA, the Integrated Resource Management Application). The goal was to have no duplicate records showing in the database and improvements made to the records that were retained.

NPS had recognized that the Alaska Data Store contained an unacceptable proportion of duplicate records. These records occurred when personnel had not successfully checked the database for an existing record for an item before creating a new record for that same item. Also, the records originated in de-centralized databases on microcomputers maintained at each national park in Alaska. Each park's copy of the database included records for items found at the park in which the database was created. Because many of these reports, journal articles, and other information sources were used in more than one park, separate records, often with differing information, had been created for the same items. When the databases were combined, and in the process of migrating the records first to the NatureBib platform and then to NRInfo/IRMA Datastore, duplicate references and their holdings were not condensed into single records.

In addition, record quality had suffered because the database records had been added at the different parks by individuals without background training in bibliographic work. They also had worked without explicit guidelines for selecting among the multiple reference types available and for entering the bibliographic information in a standardized fashion within and across parks.

During the fall of 2010, NPS and ARLIS began discussing how to identify duplicate records for removal from the Alaska Data Store, while also improving the quality of the remaining records. At that time, ARLIS estimated that 11 percent of the records were duplicates, based on its preliminary examination of every 100th record among the 8,800 records in the Southwest Alaska Network (SWAN) subset of the database. The project commenced in July 2011. ARLIS worked on exported records in Excel and EndNote formats, due to the non-availability of the multi-user software specified in the cooperative agreement. In autumn 2013, the database custodians at NPS Ft. Collins suppressed duplicate references based on reports supplied by ARLIS. NPS Ft. Collins then began updating Alaska's remaining records to match ARLIS' recommended content.

Project Products

In accord with the cooperative agreement, ARLIS furnished NPS with the following products in 2013:

- A delimited-text file containing reference codes for 4,110 duplicate records and corresponding master reference codes. These records represent 13.65 percent of the records held in the Alaska Data Store at the start of the project.
- An EndNote database containing 4,110 records for duplicate references in the Alaska Data Store.
- EndNote databases, one per IRMA Data Store document reference type, containing a total of 25,994 corrected records that correspond to non-duplicate references in the Alaska Data Store.
- A report, entitled *Best Practices for Creating Bibliographic References in the Integrated Resource Management Applications Data Store*, describing standard operating procedures relating to standardized database entry for NPS to follow in the future.
- This final report.

Milestones

2010

- **October.** NPS Alaska Inventory & Monitoring and ARLIS hold initial discussions.
- **November.** ARLIS reviews every 100th record of 8,800 SWAN records provided by NPS Alaska; identifies / counts 11 percent as duplicates of other SWAN records.

2011

- **July.** Cooperative agreement formally starts.
- **October.** NPS Ft. Collins provides first of several exports of Alaska Data Store in Excel and EndNote formats.
- **November.** ARLIS devises format for printing worksheets from Excel spreadsheet, showing IRMA Data Store reference code, IRMA Data Store reference type, a field for designating record as master or duplicate, a field for entering master reference code if duplicate record, display citation as formatted by IRMA Datastore, and comments by the Record Editor. Entries are arranged alphabetically by display citation.
- **December.** NPS Ft. Collins provides ARLIS with corrected export of Alaska Data Store in EndNote format.
- **December.** Due to remaining field mapping errors in export of records in formats for book chapters and for conference papers, ARLIS implements plan to start work with individual reference types, beginning with journal articles, with each editor taking on a certain span of years.

2012

- **September.** NPS Ft. Collins furnishes ARLIS with corrected records for conference papers and book chapters. As these are received, ARLIS rebuilds the EndNote database. As additional Record Editors are added to the project, ARLIS divides editing assignments by shorter time spans.

2013

- **July.** ARLIS sends duplicate reference codes and master reference codes to NPS Ft. Collins. ARLIS sends master record corrections to NPS Ft. Collins, first round.
- **July–December.** NPS Ft. Collins sends correction requirements for each Alaska Data Store format type to ARLIS.
ARLIS sends corrected master records for each format type to NPS Ft. Collins.
- **September.** ARLIS confirms accuracy of duplicate identification through test searches of IRMA Data Store at NPS Alaska Regional Office.
- **December.** ARLIS submits last sets of corrected records to NPS Ft. Collins.
- ARLIS submits *Best Practices for Creating Bibliographic References in the Integrated Resource Management Applications Data Store*.
- **December.** ARLIS submits this Final Project Report.
- **December.** Cooperative agreement for Alaska Data Store project formally ends.

Alaska Data Store Records

Distribution by Year

The Alaska Data Store records referenced items with publication years/dates starting in 1751 and ending in 2009. As shown in Table 1, roughly about 75 percent of the content was published from 1970 to 1999. About 12 percent of the records were dated from 1751 to 1969. Another 12 percent of the references were published from 2000 to 2009. At project's end, less than 1 percent of the records remained undated.

Year/date of record	Sub-total	Percentage of total
No year/date available	265	0.88%
1751–1899	140	0.47%
1900–1949	1048	3.48%
1950–1959	737	2.45%
1960–1969	1544	5.13%
1970–1979	4741	15.75%
1980–1989	8744	29.05%
1990–1999	9268	30.79%
2000–2009	3615	12.01%
Total	30,102	100.00%

Table 1. Distribution of records by publication year/date at conclusion of project.

Distribution by Reference Type

As shown in table 2, more than 68 percent of the records were for published reports or journal articles.

IRMA Data Store reference type (as contained in URL field in Endnote)	Sub-total	Percentage of total
Book	901	2.99%
Book chapter	2,559	8.50%
Conference proceeding paper	1,732	5.75%
Conference proceeding	53	0.18%
Dissertation	104	0.35%
Generic document	288	0.96%
Journal article	8,922	29.64%
Letter	124	0.41%
Map	533	1.77%
Memorandum	19	0.06%
Newspaper article	17	0.06%
Published report	11,722	38.94%
Thesis	889	2.95%
Unpublished report	2,239	7.44%
Total	30,102	100.00%

Table 2. Distribution of records by IRMA Data Store reference type.

Occurrence of Duplicates within Reference Types

Table 3 shows how the occurrence of duplicate records varied by reference type. As a percentage of records within each reference type, duplicate records occurred most often within maps, unpublished reports, published reports, and books.

Reference Type	Total	Master	Duplicate	Percentage duplicate
Book	901	762	139	15.43%
Book chapter	2,559	2,212	347	13.56%
Conference proceeding paper	1,732	1,669	63	3.64%
Conference proceeding	53	46	7	13.21%
Dissertation	104	99	5	4.81%
Generic document	288	261	27	9.38%
Journal article	8,922	8,219	703	7.88%
Letter	124	123	1	0.81%
Map	533	414	119	22.33%
Memorandum	19	17	2	10.53%
Newspaper article	17	17	0	0.00%
Published report	11,725	9,521	2,204	18.80%
Thesis	888	819	69	7.77%
Unpublished report	2,239	1,815	424	18.94%
Total	30,104	25,994	4,110	13.65%

Table 3. Occurrence of duplicate records within IRMA Data Store reference type.

Reference Type Differences: IRMA Data Store and EndNote

The relationship between reference types in the IRMA Data Store and EndNote is shown below. One-to-one relationships with EndNote record types do not exist for the following IRMA Data Store reference types: conference proceeding, dissertation, memorandum, thesis, and unpublished report. In addition, the IRMA Data Store parent reference type series appears in Endnote records only as a secondary title field within records for book chapters, proceeding papers, and other components of the series. These differences complicated the project.

Table 4 shows the relationship between IRMA Data Store reference types and the Endnote reference types to which these records were exported.

IRMA Data Store Reference Types as Exported to Endnote

IRMA Data Store reference type	Endnote reference type
Book	Book
Book chapter	Book section
Conference proceeding paper	Conference paper
Conference proceeding and generic document	Generic
Dissertation and thesis	Thesis
Journal article	Journal article
Letter	Personal communication
Map	Map
Memorandum and unpublished report	Unpublished work
Newspaper article	Newspaper article
Published report	Report

Table 4. IRMA Data Store reference types as exported to Endnote.

Project Procedures

Retrieving Records by IRMA Data Store Reference Type in EndNote

In EndNote records, the IRMA Data Store reference type was embedded at the end of the URL field in a string ending “reference type = <IRMA Data Store reference type>”. In EndNote, therefore, records can be selected by IRMA Data Store reference type by using Endnote’s “search for field ending with” function to identify records in which the URL field ends with the desired IRMA Data Store reference type. In order to distinguish between “published report” and “unpublished report” when using this technique, the search string for “published report” is “= published report”.

EndNote Custom Fields

EndNote custom fields were defined as follows for project purposes:

Number	Field label and use
1	–Master” or –Duplicate” designation, entered by ARLIS Record Editor.
2	–If duplicate, corresponding Master reference code,” entered by ARLIS Record Editor.
3	–Master edited? –Yes” entered if record modified by ARLIS Record Editor.
4	Preferred reference type, for records entered with incorrect or less appropriate reference type.
5	Flag for deletion. Used for unique –Master” but uniquely bad records, conveying no useful information. Recommend that record not be visible in IRMA Data Store.
6	IRMA Data Store reference code; custom export from IRMA Data Store.
7	IRMA Data Store physical location note; not modified by ARLIS Record Editor.
8	IRMA Data Store national park holding code; not modified by ARLIS Record Editor.

Table 5. Endnote custom fields

Duplicate Detection and Editing Procedures

ARLIS Record Editors reviewed the 30,000 plus IRMA Alaska Data Store records one at time, working from printed worksheets and from Endnote databases.

ARLIS provided each Record Editor with:

1. Copy of entire the EndNote database and Endnote software version X5.
2. Worksheets printed on tabloid ledger paper, from Excel spreadsheet of the Alaska Data Store, with references for a specific span of years, sorted alphabetically by display citation.

Printed worksheets were used to keep track of progress and for initial comparison of records sorted by display citation. The paper worksheets also made it easier to search the Endnote database for a match with a given reference

Record Editors identified duplicates by working through references for the assigned segment of years, using the following steps:

1. Browse through the worksheets laid out as in Table 6. –M” for Master, –D” for duplicate, and the Master reference code, if applicable, would be entered in pencil on the worksheet.

Reference code	Master / duplicate	If duplicate, master reference code	Reference type	Year	Display citation	Comments
599261			Journal article	2004	Stewart, I.J. and others. Site fidelity of spawning sockeye salmon in the presence and absence of olfactory cues. Ecology of Freshwater Fishes 13:104-110.	
585451			Journal article	2004	Stewart, I.J. and others. Site fidelity of spawning sockeye salmon in the presence and absence of olfactory cues. Ecology of Freshwater Fishes 13:104-110.	

Table 6. Format of uncompleted worksheet.

2. Search the entire EndNote database on keywords to identify possible duplicate citations lacking correct author or title or date of publication.
3. Select the best record as master when more than one record was present. Enter ~~M~~ or ~~D~~ in ~~Master/Duplicate~~ column of worksheet and in Custom Field 1 of the corresponding Endnote record. If duplicate, enter reference code of corresponding Master record in Endnote field 2.
4. Check library sources for authoritative information when a record contained anomalies, such as that presented by two versions of a record with differing author information or article pagination. If Master record was corrected, enter ~~Y~~ in Endnote custom field 3.

To obtain corrected and missing information for records in the Alaska Data Store, ARLIS Record Editors used the ARLIS online Library Catalog, OCLC WorldCat database, journal databases, other web sites, and the physical ARLIS Collection. Physical items were checked as a last resort when questions could not be resolved through other means.

Because Record Editors were not permitted to change IRMA Data Store reference types, sometimes the ~~best record~~ was the record with the correct reference type, even if other bibliographic details in that record required correction.

The completed worksheet would look like the example in Table 7.

Reference code	Master / duplicate	If duplicate, master reference code	Reference type	Year	Display citation	Comments
599261	M		Journal article	2004	Stewart, I.J. and others. Site fidelity of spawning sockeye salmon in the presence and absence of olfactory cues. Ecology of Freshwater Fishes 13:104-110.	
585451	D	599261	Journal article	2004	Stewart, I.J. and others. Site fidelity of spawning sockeye salmon in the presence and absence of olfactory cues. Ecology of Freshwater Fishes 13:104-110.	

Table 7. Model of completed worksheet.

Processing Records with No Year or Date of Publication

At the start of the project, 1,480 records lacked a year or date of publication. Without this information, the records could not be assigned to any particular segment until they were individually examined. By the conclusion of the project, year or date of publication had been obtained for all but 248 records. These records referenced 109 unpublished reports, 87 published reports, 19 generic documents, 12 books, 8 maps, 4 journal articles, 4 conference papers, 2 letters, and 1 conference proceeding.

Data Management Practices

EndNote databases were stored on a network drive at ARLIS, which was regularly backed up. In addition, on a daily basis, an ARLIS Principal Investigator would copy each Record Editor's individual database to a local drive. Each week, the Principal Investigator would copy the records from each Record Editor's segment to a new database. The new database was used to generate progress reports for the project as a whole.

When a new Record Editor joined the project, existing data editing assignments were divided by year. Another copy of the database was made from the segment with which the new Record Editor would be working.

In order to check for missing records, lists of reference codes from the starting and the concluding versions of the database were compared, using UNIX text comparison utilities. Any missing records were replaced using copies of the database retained from earlier phases of the project.

Throughout the project, ARLIS held weekly meetings of Principal Investigators and Record Editors to identify problems, refine processes, and devise next steps. In addition, ARLIS Team Leaders and the

NPS Agreement Technical Representative and Technical Consultants held regular project management teleconferences.

Lessons Learned and Recommendations

Future Deduplication/Record Editing Projects in IRMA Data Store

- Increase direct communication between Ft. Collins and project teams early in projects of this type.
- Provide access to IRMA Data Store software to identify duplicate records and edit non-duplicate records, rather than exporting records to different software / system and re-importing records after editing.
- If export to another system is required, select and use systems implemented on a one-to-one relationship of tables and fields between source and target databases, for each reference type, to facilitate reimport.
- If export / import model is used, perform dry run of import process early in project to provide opportunity for needed changes before extensive editing is carried out.
- If more than one editor will work simultaneously, employ multi-user software, rather than multiple copies of single-user software. The use of EndNote software complicated the project. To work around the limitations of EndNote software, the project worked with multiple copies of the database. Each editor worked on records for certain span of years. Periodically, a principal investigator would create a new copy of the EndNote database by copying each editor's segment from that editor's copy of the database. This design increased the number of file to back up on a daily basis. The need to combine segments from each editor's copy of the database increased the occasions on which errors could be made.

IRMA Data Store Functions and Documentation

- To prevent creation of duplicate records, a software function to identify possible and likely duplicate records should be added to the record creation process. After author, title, and year were entered in the IRMA record creation form, but before the record was saved, the IRMA Data Store would display possible duplicates selected on basis of comparisons against author, title, and other fields, across all IRMA Data Store reference types. The individual entering the record would have to examine the possible duplicates before saving a new record.
- Clarify standard operating procedures by incorporating or adapting instructions from the *Best Practices for Creating Bibliographic References in the Integrated Resource Management Application Data Store*, including standardized lists of titles, authors, and serial publications.
- On an annual basis, review *Best Practices* and revise to reflect changes in IRMA Data Store record data entry screens, reference types, and functions.
- Revise IRMA Data Store data entry screens and documentation to more clearly delineate how bibliographic information should be entered into each record field for each record type.
- Improve definitions for reference types in the IRMA Data Store, as follows:
 - Resolve conflict between IRMA Data Store reference types based on format and IRMA Data Store reference types based on content. Such conflicts lead to confusion in cases of documents that are published in multiple formats (e.g., conference proceedings that are published as a report or as a journal issue or as a book; or a thesis published as a report or as a book). These IRMA Data Store reference types defined by content include

dissertations, theses, conference proceedings, and maps. Potential confusion might be eliminated by explicitly assigning priority to content or to format, using real life examples, such as a thesis published in a NPS report series.

- Combine similar IRMA Data Store reference types. For example, theses and dissertations could be combined in a single category and differentiated in the bibliographic description, rather than the record format.
 - Revise the vague definition of “unpublished reports” in the IRMA Data Store. The distinction between “unpublished” and “published” has been eroded over time by widespread use of photocopiers, scanners, and the Internet. In this project, many documents described as “unpublished reports” were found in multiple library collections. The IRMA Data Store definition of “unpublished reports” is not helpful in assessing grey literature often described as “unpublished reports.” If “peer reviewed” versus “non-peer reviewed” is the major difference implied by “published” versus “unpublished” that distinction should be recorded in a check box field in the IRMA Data Store record, rather than being implied by record type. This practice is used in several commercial bibliographic databases which allow limiting of search results to peer-reviewed items.
 - Discontinue use of the generic document reference type. At time of writing, the IRMA Data Store uses the generic document reference type for multiple records imported by a batch process in XML format. That batch software process should be enhanced to differentiate reference types embedded in the imported XML file. Existing records with the generic document type should have the reference type changed to reflect the actual format of the reference described.
- The IRMA Data Store help file should describe how specifics of NPS preferred practice for entering bibliographic data in the Data Store depart from bibliographic guides for manuscript preparation recommended by NPS. For preparing manuscripts, NPS recommends bibliographic style guides issued by the Ecological Society of America, Council of Editors, and Chicago Manual of Style. IRMA Data Store practice departs from these authorities in important ways. The differences are most notable in preferred names of government agencies.
 - Involve librarians, or others well-trained in bibliographic style, in the database process and creation of instructions, as well as localized trainings

Literature cited

Braund-Allen, J.E. and S.P. Johnson. 2014. Best practices for creating bibliographic references in the Integrated Resource Management Applications Data Store: Developed for the Alaska National Park Service by the Alaska Resources Library and Information Services. Natural Resource Report NPS/AKR/NRR—2014/753. National Park Service, Fort Collins, Colorado.

National Park Service. 2012. The IRMA Data Store Help Manual. National Park Service, Fort Collins, Colorado. <https://irma.nps.gov/content/help/datastore/docs/ReferenceHelp.pdf>. Accessed 20131224.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 953/123447, January 2014

National Park Service
U.S. Department of the Interior



Natural Resource Stewardship and Science

1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov

EXPERIENCE YOUR AMERICA™