

AN INTRODUCTION TO THE ADVANCED TRACKING AND RESOURCE TOOL FOR ARCHIVE COLLECTIONS (ATRAC)

Philip R. Jones^{1*}, Ken Roberts¹ and Nancy Ritchey²

¹STG, Inc, Asheville, North Carolina

²NOAA's National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

Fundamental documentation and administrative responsibilities for long-term data stewardship demand resources of both the archive (the data center) and the data provider. The amount of resources necessary to do these activities has increased with recent archiving developments, including implementation of the "NOAA Procedure for Scientific Records Appraisal and Archive Approval", data provider to archive submission agreements, evolving metadata standards, inventory improvements, and archive requirements processes. The Advanced Tracking and Resource tool for Archive Collections (ATRAC) is a new, web-based tool developed at the NOAA National Climatic Data Center (NCDC) that provides an interface for users to enter, share and display information for a data archive project. This paper discusses how the capabilities of ATRAC can improve project communication and work flow. The ATRAC tool is accessible at <http://www.ncdc.noaa.gov/atrac/index.html>.

2. DATA ARCHIVE PROJECT

As part of its business model, NCDC has adopted the Open Archive Information System Reference Model (OAIS-RM). This ISO standard defines the interdependent functional entities of an archive organization and the kinds of information necessary to make an independently understandable data package for long-term data preservation. Complete and accurate metadata is the foundation of data preservation, and provision of this knowledge by the data provider is key to successful data support. Data rescue efforts at NCDC show that data stewardship relies heavily on active participation from the data provider, data center, and data users (Knapp et al 2007).

In order to address the challenges in data management with growing data volumes and complexities, NCDC follows an end-to-end process to support a data project, from the initial

archive request to data disposition (Table 1). The process is not implemented in isolation at the data center, but rather includes interaction with the data provider and the data user. The first phase of the process, the archive appraisal, includes assessing the data value and identifying and securing adequate funds for the anticipated long-term support. Some of these appraisal items may already be addressed in the initiator's enterprise-wide data management plan. For the appraisal steps, NCDC follows the existing "NOAA Procedure for Scientific Records Appraisal and Archive Approval". This NOAA-wide process was

Table 1: The steps of the data center archiving process with related participants and deliverables. ATRAC supports the collection and display of information throughout the project lifecycle.

Phase	Step	Actor	Deliverable
I. Appraisal	Request	Provider	<ul style="list-style-type: none"> • Assessment • Recommendation • Preliminary Agreement
	Evaluate	Data Center, Users	
	Negotiate	Data Center, Provider	
	Decide	Data Center	
II. Transfer	Prepare	Data Center, Provider, Users	<ul style="list-style-type: none"> • Requirements • Metadata • Submission Agreement
	Test	Data Center, Provider	
	Implement	Data Center, Provider	
III. Stewardship	Preserve	Data Center, Provider, Users	<ul style="list-style-type: none"> • Updates
	Maintain	Data Center	
IV. Disposition	Evaluate	Data Center, Users	<ul style="list-style-type: none"> • Assessment • Recommendation • Disposition Schedule
	Decide	Data Center	

developed using guidelines from NARA and from National Research Council reports on NOAA data management. A data center decision to support the data leads to the next phase – the data

* Corresponding author address: Philip Jones; NCDC/RSAD; 151 Patton Ave.; Asheville, NC, 28801; email: philip.jones@noaa.gov

transfer – when details of the project are confirmed and systems are developed to implement the operational management of the data. The archive metadata are also developed during this phase with the provider. For this effort, NCDL employs metadata formats compatible with international standards to facilitate data preservation, discovery, access, and information compatibility. The following long-term stewardship phase is the primary objective of the data archive project. Again, the stewardship responsibility must be shared by the data provider and participating user community in order to fully support the data. The decision process to end data center support is similar to the initial appraisal phase in that users are engaged in the process.

3. ATRAC OVERVIEW

3.1 Purpose

The ATRAC system tracks the progress of a data center project through all phases of the archiving process. The tool provides a common environment for project stakeholders to gather and share information on a project, and it is a common source for interested parties to find status

information on projects. This provided efficiency in communication reduces the cost, time and risks associated with data center projects. Use of ATRAC requires no training or prior experience with data archiving. It points data providers to archiving guidelines that establish project expectations and it facilitates the use of data standards for new users.

3.2 Design

There are three primary areas of ATRAC web pages: 1) the Home pages with introduction and guidelines; 2) the Project Display pages with project status and summary information; and 3) the Project Input pages that collect input to be published on the Project Display pages and in data center documents. Any user may acquire an ATRAC account through email in order to provide information on a project. Behind a login, a user has the ability to provide input on projects associated with his/her profile. A project is associated with a user's profile by either the user creating a new project or by the user being added to an existing project by a project stakeholder.

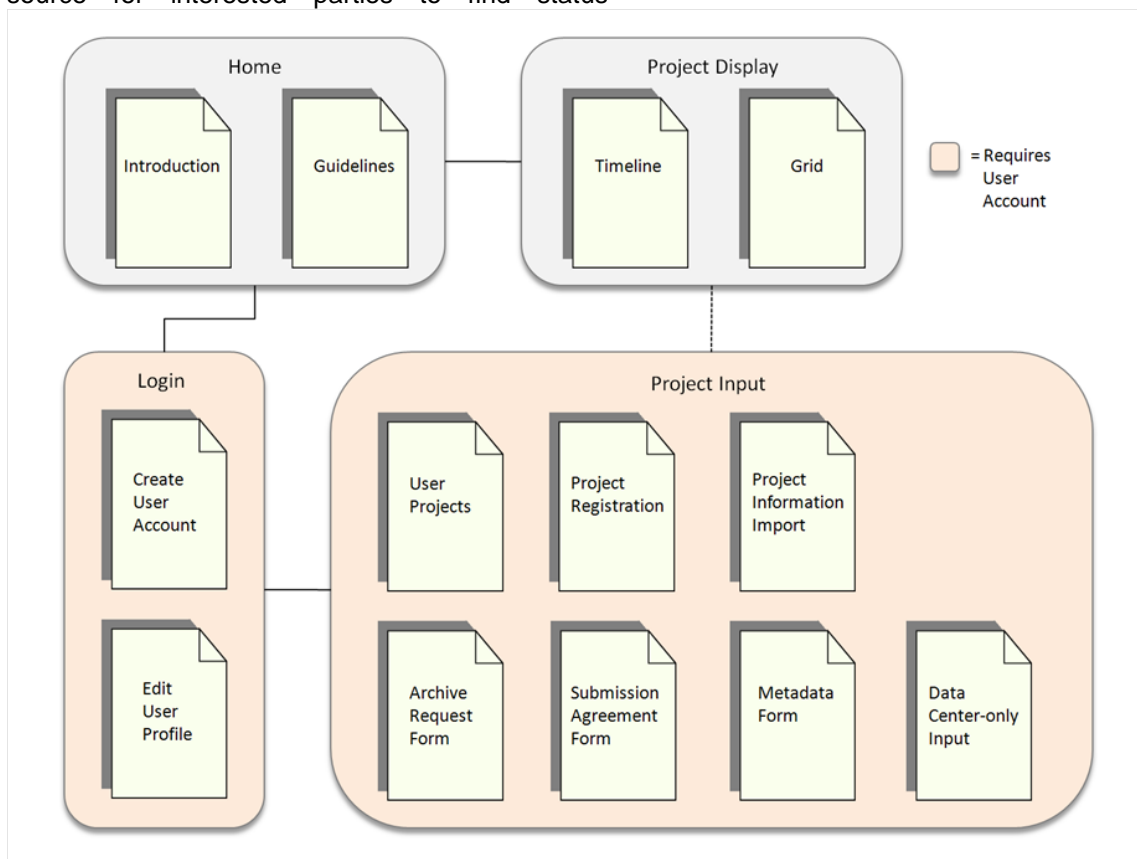


Figure 1: Web diagram of ATRAC, showing the pages in each primary functional area.

The input section of ATRAC is a set of web forms that collect information to feed the project display pages. The web forms are categorized by type of information with form content presented in a succinct layout for users. Submitted input must pass validation checks in order to be published in the project display. The input is also used by the data center to populate project documents.

2.3 Software Architecture

The Advanced Tracking and Resource tool for Archive Collections (ATRAC) features a technology stack centered around the Spring Framework, a popular source framework for enterprise Java applications. As one of the components of the framework, Spring Security provides a flexible and robust solution for authentication and access controls. ATRAC's data management and persistence requirements are met through the use of Hibernate, an object/relational mapping library that features nearly seamless integration with Spring. In terms of client-side architecture, ATRAC utilizes jQuery, an extensible JavaScript library, allowing for an interactive and dynamic user experience.

2.4 Features

ATRAC has several features that simplify its use. The input forms collect information on a project using a combination of menu selection, auto-population and free text entry. A user can save a partially completed form or submit a completed form. A validation check is run upon submission of a completed form and returns a message for success or errors. Form validation ensures compliance and returned error messages instruct users on how to correct an error. The validation routine checks form content against rules of completeness, conditionality and format. Validation rules are based on the intended use of the form content for display or in document templates. A submitted form that passes validation automatically triggers an email notification from ATRAC to the project stakeholders, including the data center representatives. Only validated form content is displayed on the project display pages or used in document templates.

The project status information maintained in ATRAC is presented on the project display pages. Projects are displayed in two compatible



Figure 2: Sample Timeline display of projects with milestones, tasks and status information.

ATRAC - Projects Overview									
		Home		Project Status		Project Input			
Timeline Overview Legend									
Search Project Title		Search							
Project Title	Provider	Data Center	CLASS	Ingest Status	Ingest Start	Planned Tasks	In Work Tasks	Complete Tasks	Modified
AVHRR Pathfinder Atmospheres - Extended (PATMOS-X)	TBD	NCDC	No	Static	9/15/2010	None	• Archive Recommendation	None	9/27/2010
Blended Total Precipitable Water (TPW) and TPW Anomalies	TBD	NCDC	TBD	Ongoing	11/1/2010	None	• Archive Recommendation	None	9/24/2010
CDMP Raw Keyed Lightship Observations	TBD	NCDC	No	Static	8/30/2010	None	• Archive Recommendation	None	9/24/2010
Continuously Operating Reference System (CORS)	TBD	NGDC	TBD	Ongoing	2/1/2005	None	• Other	• Submission Agreement • Archive Metadata	10/28/2010
Cooperative Summary of the Day - COOP Historical Integration Project (CHIP)	TBD	NCDC	No	Ongoing	4/9/2010	None	None	None	9/27/2010
ESRL 20th Century Reanalysis	TBD	NCDC	Yes	Ongoing	9/30/2010	None	None	None	9/27/2010
Extended Reconstruction Sea Surface Temperature (ERSST), Version 3b	TBD	NCDC	No	Static	9/15/2010	None	None	None	9/27/2010
Gridded Satellite (GridSat) Project for ISCCP B1 CDR	TBD	NCDC	No	Static	9/15/2010	None	None	None	9/27/2010
HIRS Upper Tropospheric Water Vapor (UTWV) CDR	TBD	NCDC	No	Static	9/1/2010	None	None	None	9/27/2010
Integrated Ocean Drilling Program (IODP)-U.S. Implementing Org.(USIO) Core Data and Images	TBD	NGDC	No	Ongoing	9/1/2010	None	None	• Submission Agreement	10/22/2010
International Best Track Archive for Climate Stewardship (IBTrACS), Version 3.0	TBD	NCDC	No	Static	8/25/2010	None	None	None	9/27/2010
International Comprehensive Ocean-Atmosphere Data Set Release 2.5	ESRL	NCDC	Yes	Static	9/16/2009	None	None	• Archive Recommendation • Submission Agreement • Archive Metadata	9/27/2010
NASA GEWEX SRB	TBD	NCDC	TBD	Ongoing	10/15/2010	None	None	None	9/27/2010
National Ice Center Sea Ice Charts	TBD	NCDC	TBD	Ongoing	11/1/2010	None	None	None	9/27/2010
NPOESS Data Exploitation (NDE) Microwave Integrated Retrieval System (MIRS)	TBD	NCDC	TBD	Pending	12/16/2011	None	None	None	9/27/2010
NPOESS Data Exploitation (NDE) NOAA-Unique CrIS/ATMS Processing System (NUCAPS)	TBD	NCDC	TBD	Pending	12/16/2011	None	None	None	9/27/2010
NPP Cal/Val Campaigns	TBD	NCDC	Yes	Pending	1/18/2012	None	None	None	9/27/2010
Rapid Update Cycle (RUC) Model	TBD	NCDC	TBD	Ongoing	9/15/2010	None	None	None	9/27/2010

Figure 3: Sample Grid display of projects with the capability to sort project details by column.

views: timeline and grid (Figures 2 and 3). The timeline view paginates the projects for a user-configurable time scale and period. The view displays the status of data ingest with the option to expand the view to include a project's tasks and events. Recorded details of task milestones and events are displayed when a user mouses over the respective icons. The ATRAC grid view displays characteristics of projects in sortable table columns. The data ingest and task columns have color-coded cells by status type to support the visual grouping of projects. Both the timeline and grid views share project search results to filter the project display. Project search criteria include project title, data producer, data center, ingest status and Comprehensive Large-Array [data] Stewardship System (CLASS) support.

Information in the forms is used by the data center for project assessments, planning, and to populate one or more standard document

templates. These templates include an archive recommendation letter, a data submission agreement and standards-based discovery metadata. While the document templates can be populated, formatted and exported by the system, the system itself is not a document management system or repository. The PDF or XML formatted documents produced by ATRAC are intended to be managed in another system as appropriate by data center and/or document type.

The previous ten saved or submitted (i.e., validated) versions of a form are accessible to users allowing retrieval of earlier form versions. A user accesses the previous versions from a history page for each form displaying the associated saved and submitted form date and author.

Input that's shared between one or more forms is automatically copied to the corresponding form thus requiring users enter like information in just any one form. Examples of this include the data

overview, data format and volume information. The system also shares components of information across all projects. Contact information components are one example, where upon typing a contact entry the system automatically queries stored contact components and presents the matching options for selection (e.g., see Figure 4). The system can store the entry of new components and save edits made to existing components.

a)

1. Identify a Provider contact who is knowledgeable about the data and whose contact information can be listed as the primary data point-of-contact.

Myself
 Other

Last Name:

First Name:

Organization:

Email:

Phone:

Contact Instructions:

Contact Hours:

b)

1. Identify a Provider contact who is knowledgeable about the data and whose contact information can be listed as the primary data point-of-contact.

Myself
 Other

Last Name:

First Name:

Organization:

Email:

Phone:

Contact Instructions:

Contact Hours:

Figure 4. Screen capture of the ATRAC interface showing (a) the presentation of possible matching contact names upon entering letters of a name, and (b) resulting auto-population of contact information after selection of the stored contact name.

Information from another comparable project can be imported to a user's project to facilitate data entry. A project import is achieved through a search and selection, where the selected project's information is copied to all respective project forms. The projects available for importing are limited by a user's read permission.

The majority of projects in ATRAC are publicly viewable throughout the system, though project stakeholders have the capability to "hide" the display of their project information behind a login viewable only to specified ATRAC users. Project stakeholders can also assign write permission for their project to a limited number of users (in addition to data center representatives). These permission features allow stakeholders to effectively manage the level of participation needed for their project.

4. PLANNED IMPROVEMENTS

Project information in ATRAC originates from and/or is already used in other systems, and so a natural progression for the tool is to develop interfaces with these related systems for efficient information management. The related systems include data and metadata inventory systems at NCDC and other data centers, and external product development tracking systems. Plans have been made for these new interfaces and some of the initial development is currently in work.

ACKNOWLEDGMENTS

The Authors thank and recognize Dan Kowal of the National Geophysical Data Center (NGDC), who significantly contributed his time and effort into testing, reviewing and providing recommendations on the tool.

REFERENCES

Consultative Committee on Space Data Systems (CCSDS), 2003: *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, adopted as ISO 14721.2003.

Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA, National Research Council, 2007: *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*, National Academies Press.

Knapp, K. R., J. J. Bates and B. Barkstrom, 2007: Scientific Data Stewardship: Lessons from a satellite data rescue effort, *Bulletin of the AMS*, 88 (9), 1359-1361 doi:10.1175/BAMS-88-9-1359.

NOAA Procedure for Scientific Records Appraisal and Archive Approval, August 2008.

