Toward a Common Earth Data Publication Framework

Justin Rice¹, Daine Wright², Deborah Smith³, Amanda Leon⁴, Ajinkya Kulkarni³

Abstract

Data publication is an essential activity for all data archives. Each of NASA’s twelve Distributed Active Archive Centers (DAACs) have established publication workflows which account for the heterogeneous suite of missions, instruments, data providers, and datasets managed within the Earth Observation System Data and Information System (EOSDIS) program. Some aspects of data publication vary across DAACs: workflows range from manual to automatic, terms used to describe publication elements differ, and systems used to publish and manage data vary. Despite these differences, the DAAC data publication processes are generally the same: obtain the data and related information from data providers, describe the data with metadata and documentation, release the data for access by the user community, archive and perform routine tests on the data, and provide support for the data and user community.

In order to improve consistency and reduce the time to publish data, we have developed a cross-DAAC initiative called the Common Earth Publication Framework (Earthdata Pub). Earthdata Pub seeks to: standardize communications and interactions with data providers; identify and standardize common workflows and steps in the data publication process; and design/implement a front-end system with features that include a common web interface, email & status tracking, and common application programming interfaces (APIs) to communicate with various DAAC-specific software components (services and applications) on the back-end. This poster presents the latest updates on this effort’s progress and future plans.

Motivations

All DAACs follow data publication workflows to ingest and describe data for public distribution.

<table>
<thead>
<tr>
<th>DAAC</th>
<th>Data Publication Workflow Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAAC1</td>
<td>Receive data, metadata, documentation, release to public</td>
</tr>
<tr>
<td>DAAC2</td>
<td>Verify data, verify/create metadata, release to public</td>
</tr>
<tr>
<td>DAAC3</td>
<td>Curate metadata, release to public</td>
</tr>
</tbody>
</table>

Workflows vary from DAAC to DAAC. Some differences are inherent to the different data sources, datasets, and data users that the DAACs support or the data systems they operate.

There are opportunities to improve consistency and efficiency.

DAAC Survey Results

Earthdata Pub Goals

- **Provide resources**: tools are unique to each DAAC
- **Develop software framework**: consistency and efficiency
- **Assess and document**: workflow and publication processes

Division of Labor

- **Information Group**: information requirements, implementation of software framework, verification & validation
- **Development Group**: implementation of software framework, verification & validation

Group Goals

- **Information Group**: identify and select common publication elements and tools
- **Development Group**: create software project plan

Progress

- **Information Group**: identified and selected publication requirements, implementation of software framework
- **Development Group**: developed software framework and backbone

Future Work

- **Information Group**: create software project plan
- **Development Group**: begin development of software framework and modules

Authors

1 NASA Goddard Space Flight Center
2 Oak Ridge National Laboratory
3 University of Alabama in Huntsville
4 National Snow and Ice Data Center DAAC