Standardizing Algorithm Documentation for Improved Scientific Communication and Data Understanding

Aaron Kaulfus[1], Kaylin Bugbee[1], Alyssa Harris[2], Rahul Ramachandran[3], Sean Harkins[2], Sean Bailey[4], Aimee Barciauskas[4]

(1) University of Alabama in Huntsville
(2) Development Seed
(3) NASA Marshall Space Flight Center
(4) NASA Goddard Space Flight Center

2019 American Geophysical Union Annual Meeting. December 9-13, 2019. San Francisco, CA

Current Data Infrastructure

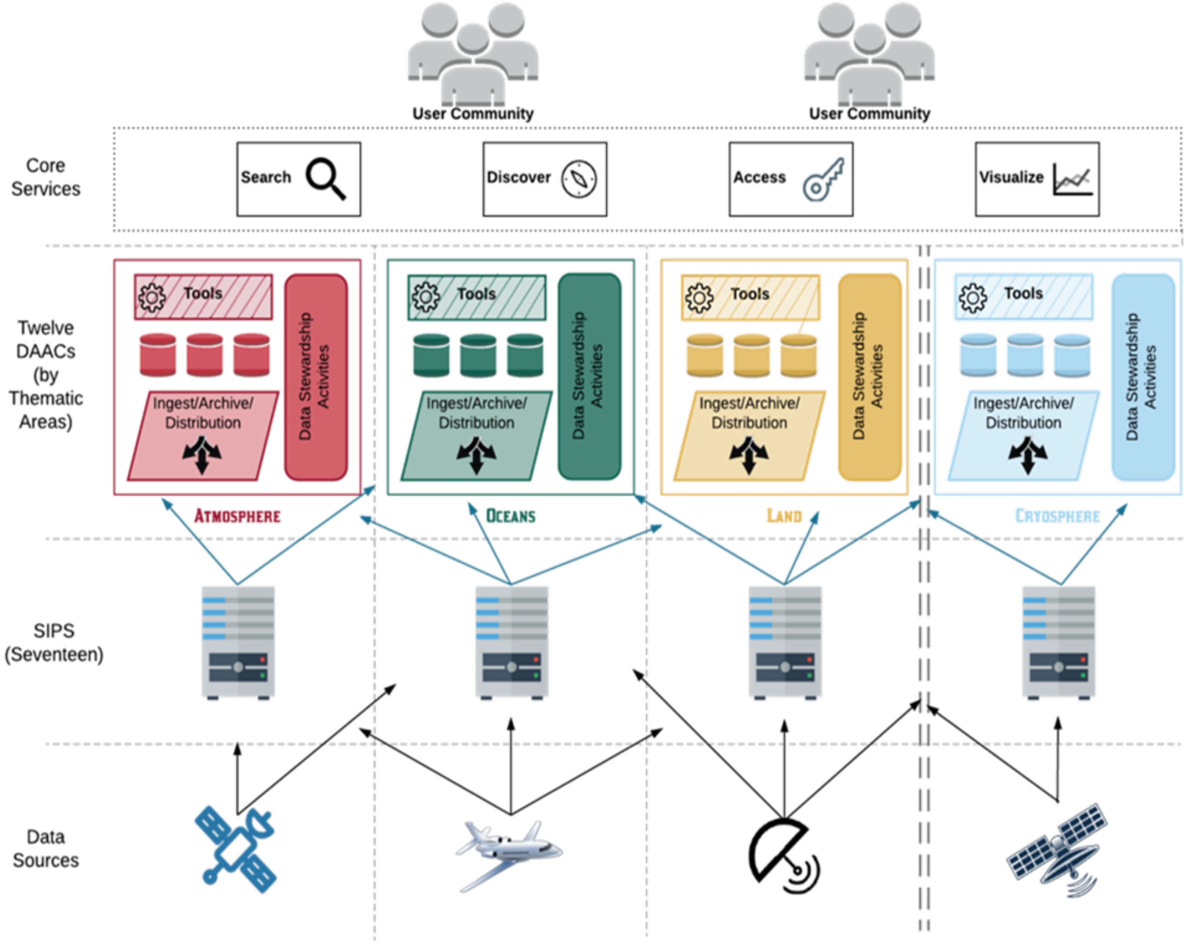
Collect data by Earth observing remote sensing instruments

Process at Science Investigator-led Processing Systems (SIPS)

Archive at Distributed Active Archive Centers (DAACs)

Distribute to user community through key services

Detailed data documentation and curation is a service that aids in distribution of data by enhancing search and discovery while promoting transparency and scientific reproducibility



Current NASA EOSDIS Architecture

Algorithm Theoretical Basis Documents

Algorithm Theoretical Basis Documents, or ATBDs, provide data users the physical theory, mathematical procedures and assumptions made for developing algorithms which convert radiances received by remote sensing instruments into geophysical quantities

ATBDs are *required* for every NASA Earth Observing System (EOS) instrument product

The Documentation Problem

No standard template or content requirements

- Creates confusion and uncertainty for scientists writing ATBDs
- Large volume of instruments, products and associated science teams
- An ATBD may address multiple products or a product may be addressed by multiple ATBDs

No central repository for search and discovery

- Documents are delivered to archival centers for preservation and distribution
- Important for data distribution velocity and science reproducibility that discovery is consistent and efficient

Difficult to update or maintain

- Data and associated algorithms may change rapidly
- Documents must be readily updated for advancements in data processing or when corrections are identified

Limited ability to efficiently parse content

- Often available as PDFs; not search engine optimized

What is APT?

The Algorithm Publication Tool (APT) is a cloud-based publication tool for standardizing ATBDs and streamlining the authoring process

APT has the goal of moving from a static to dynamic model of documentation with intelligent connections to software, data and other supporting resources to improve transparency and promote scientific reproducibility

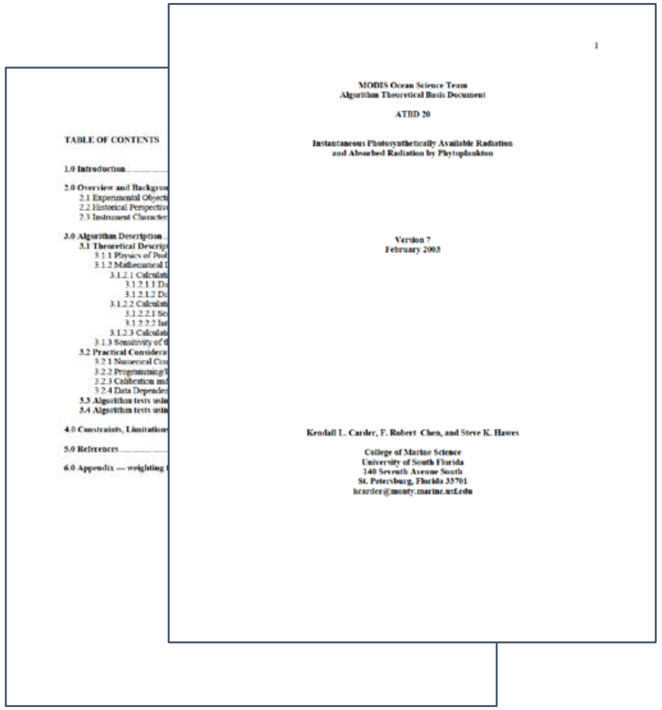


TABLE OF CONTENTS

1

MODIS Ocean Science Team
Algorithm Theoretical Basis Document

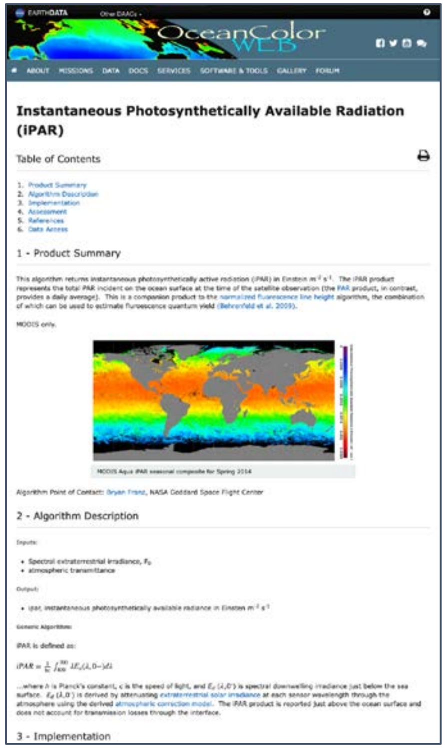
ATBD 20

Instantaneous Photosynthetically Available Radiation
and Absorbed Radiation by Phytoplankton

Version 7
February 2003

Kendall L. Carder, F. Robert Chen, and Steve K. Hawes

College of Marine Science
University of South Florida
140 Seventh Avenue South
St. Petersburg, Florida 33701
kcarder@monty.marine.usf.edu



Instantaneous Photosynthetically Available Radiation (iPAR)

Table of Contents

1. Product Summary
2. Algorithm Description
3. Implementation
4. Assessment
5. References
6. Data Access

1 - Product Summary

This algorithm returns instantaneous photosynthetically active radiation (iPAR) in Einstein $m^{-2} s^{-1}$. The iPAR product represents the total PAR incident on the ocean surface at the time of the satellite observation (the PAR product, in contrast, provides a daily average). This is a companion product to the normalized fluorescence line height algorithm, the combination of which can be used to estimate fluorescence quantum yield (Behrenfeld et al. 2009).

MODIS only.

MODIS Aqua iPAR seasonal composite for Spring 2014

Algorithm Point of Contact: Bryan Franz, NASA Goddard Space Flight Center

2 - Algorithm Description

Inputs:

- Spectral extraterrestrial irradiance, F_0
- atmospheric transmittance

Output:

- ipar, instantaneous photosynthetically available radiance in Einstein $m^{-2} s^{-1}$

Generic Algorithm:

iPAR is defined as:

$iPAR = \frac{1}{hc}\int_{400}^{700} \lambda E_d(\lambda, 0-) d\lambda$

...where h is Planck's constant, c is the speed of light, and $E_d(\lambda, 0^-)$ is spectral downwelling irradiance just below the sea surface. $E_d(\lambda, 0^-)$ is derived by attenuating extraterrestrial solar irradiance at each sensor wavelength through the atmosphere using the derived atmospheric correction model. The iPAR product is reported just above the ocean surface and does not account for transmission losses through the interface.

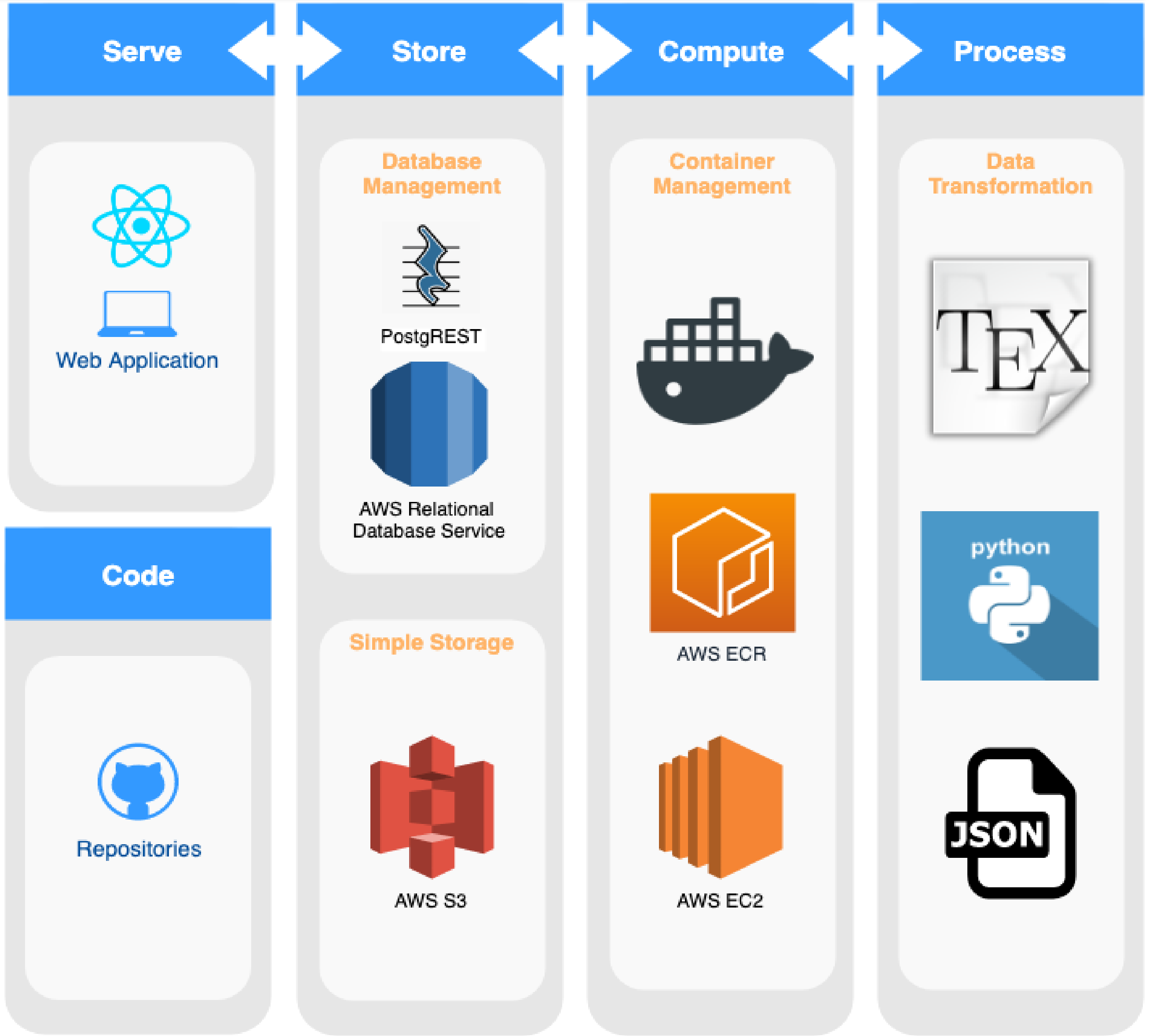
3 - Implementation

High Level Architecture

Web application front-end which serves, and is served by, a content model schema implemented as a database

Latex backend supports rich content required for scientific writing

All components implemented using AWS cloud resources

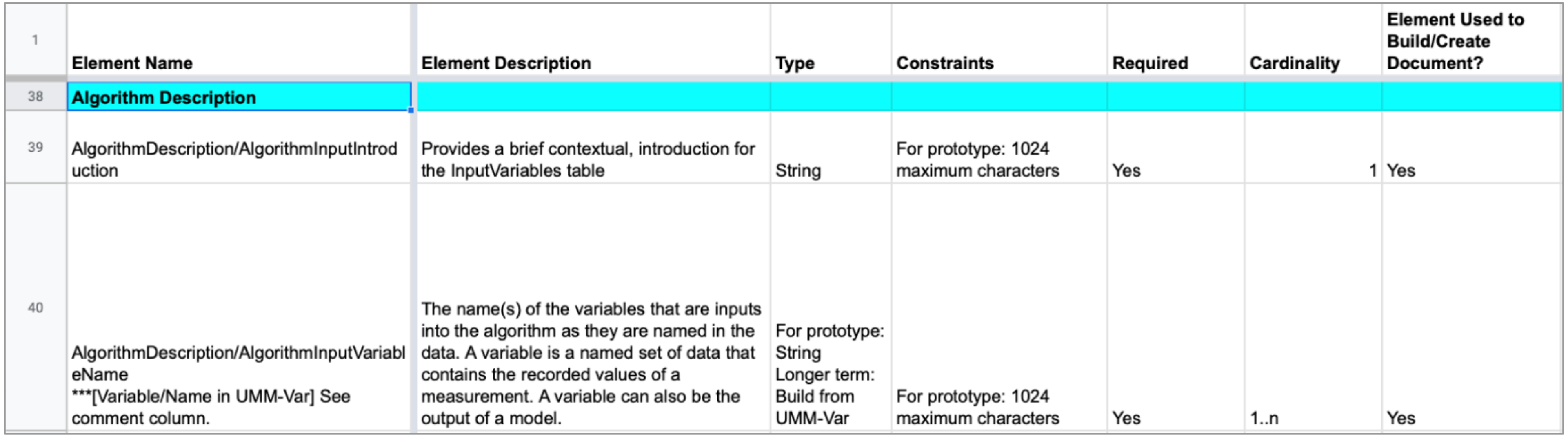


Algorithm Metadata Model

Basis of the tool are rich models for storing ATBD content

- Traditional metadata - information about the document
- ATBD content as metadata - standardize and simplify content

ATBD information model is based on review of Earth observations community algorithm documentation



1	Element Name	Element Description	Type	Constraints	Required	Cardinality	Element Used to Build/Create Document?
38	**Algorithm Description**						
39	AlgorithmDescription/AlgorithmInputIntroduction	Provides a brief contextual, introduction for the InputVariables table	String	For prototype: 1024 maximum characters	Yes	1	Yes
40	AlgorithmDescription/AlgorithmInputVariableName ***[Variable/Name in UMM-Var] See comment column.	The name(s) of the variables that are inputs into the algorithm as they are named in the data. A variable is a named set of data that contains the recorded values of a measurement. A variable can also be the output of a model.	For prototype: String Longer term: Build from UMM-Var	For prototype: 1024 maximum characters	Yes	1..n	Yes

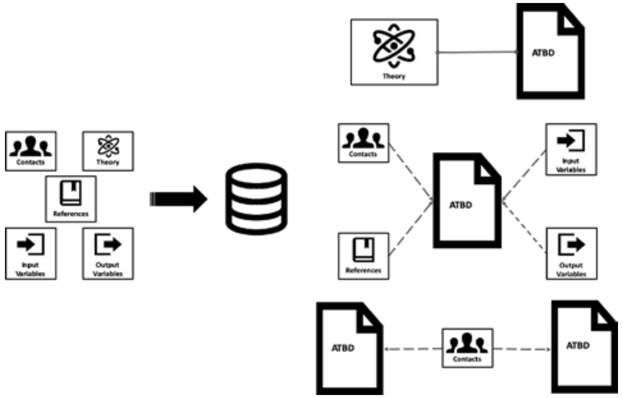
ATBD Publication

APT initial development focused on the publication web interface

- Ease document publication burden

Simplify embedding and generation of rich content

- Equation building with in-line validation
- UI for table construction; inserting figures
- Automatic formatting of tables, figures and equations within the document
- Database implementation promotes content reuse and consistency

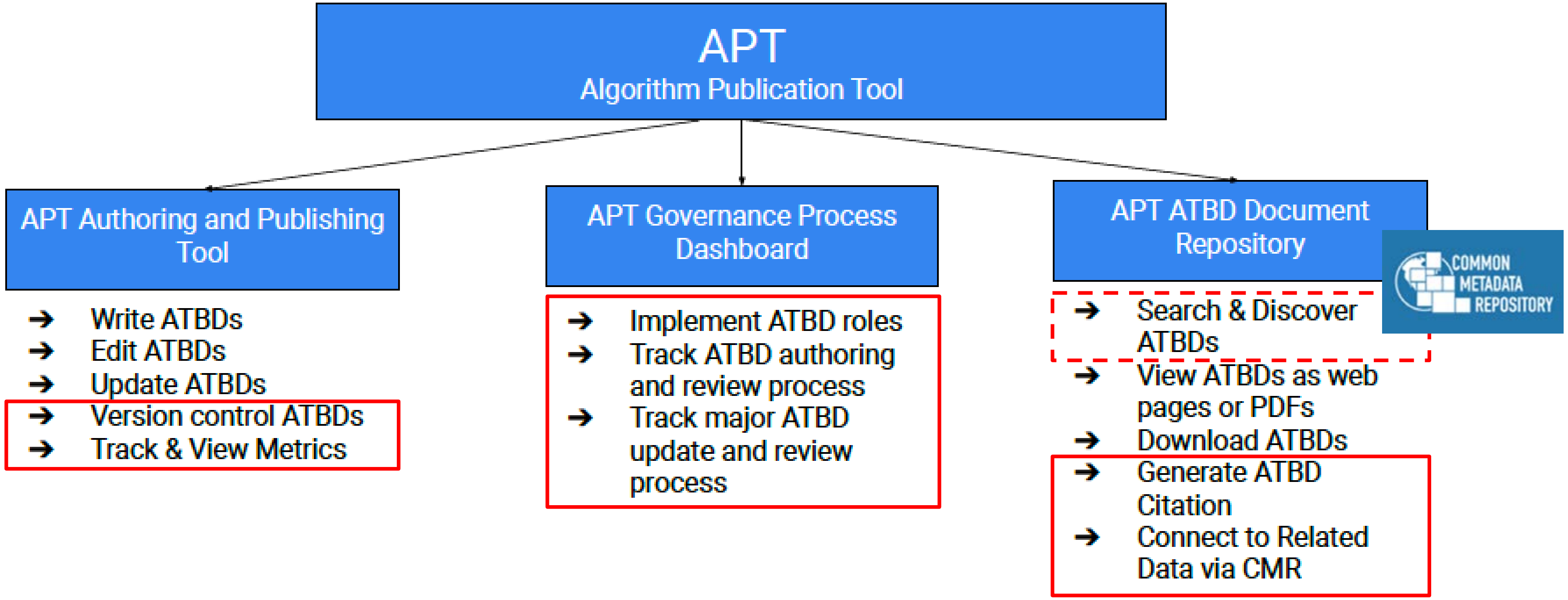


Long-term Vision

Integrate into NASA EOSDIS data curation and preservation system

- Interconnected with *input* and *output data* through CMR metadata
- Leverage future metadata schemas including that for *variables* and *software*

Adoption of APT for future mission documentation requirements



APT

Algorithm Publication Tool

APT Authoring and Publishing Tool

- ➔ Write ATBDs
- ➔ Edit ATBDs
- ➔ Update ATBDs
- ➔ Version control ATBDs
- ➔ Track & View Metrics

APT Governance Process Dashboard

- ➔ Implement ATBD roles
- ➔ Track ATBD authoring and review process
- ➔ Track major ATBD update and review process

APT ATBD Document Repository

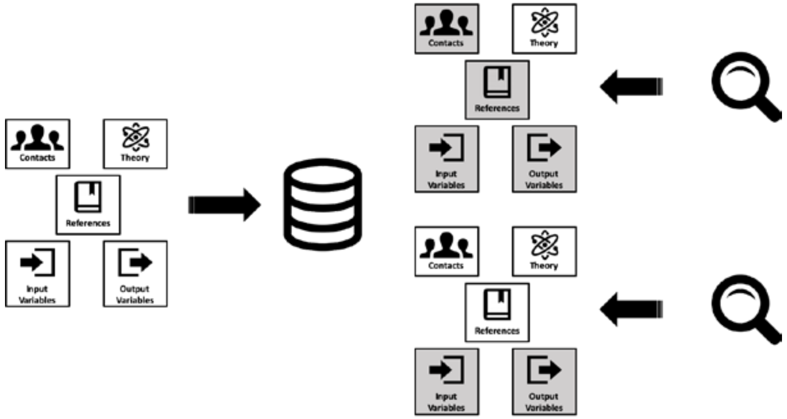
- ➔ Search & Discover ATBDs
- ➔ View ATBDs as web pages or PDFs
- ➔ Download ATBDs
- ➔ Generate ATBD Citation
- ➔ Connect to Related Data via CMR

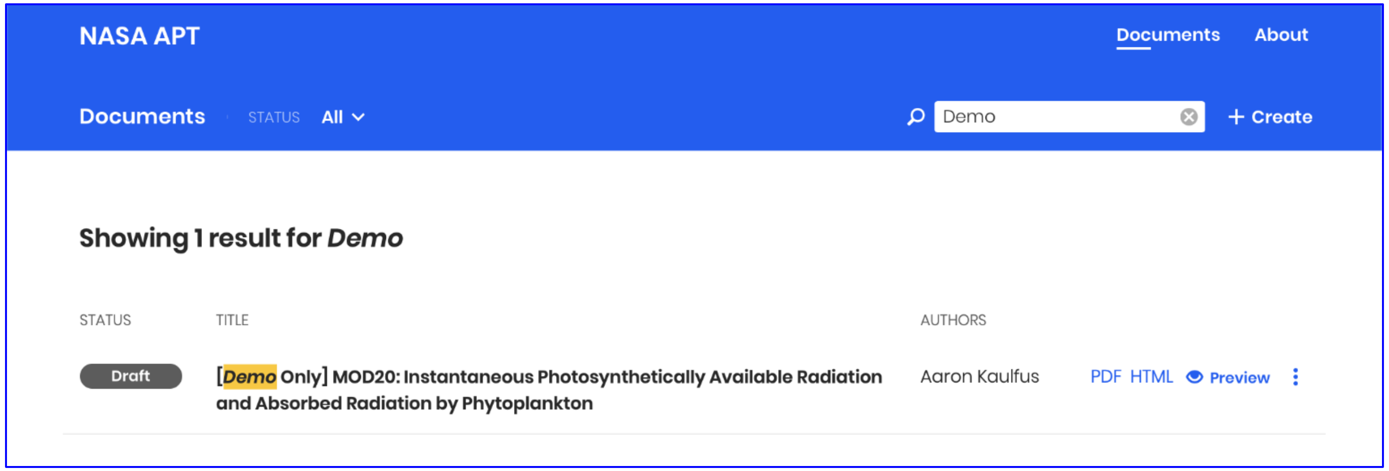
Long-term Vision: Centralized Repository

Implement interface and capabilities for search and discovery of documents by

- Identifying metadata, such as citation information and keywords *and*
- Document content, such as equations or scientific concepts

Integration of existing NASA ATBDs into the document repository





Showing 1 result for *Demo*

STATUS	TITLE	AUTHORS	
Draft	[*Demo* Only] MOD20: Instantaneous Photosynthetically Available Radiation and Absorbed Radiation by Phytoplankton	Aaron Kaulfus	PDF HTML Preview

Long-term Vision: Governance

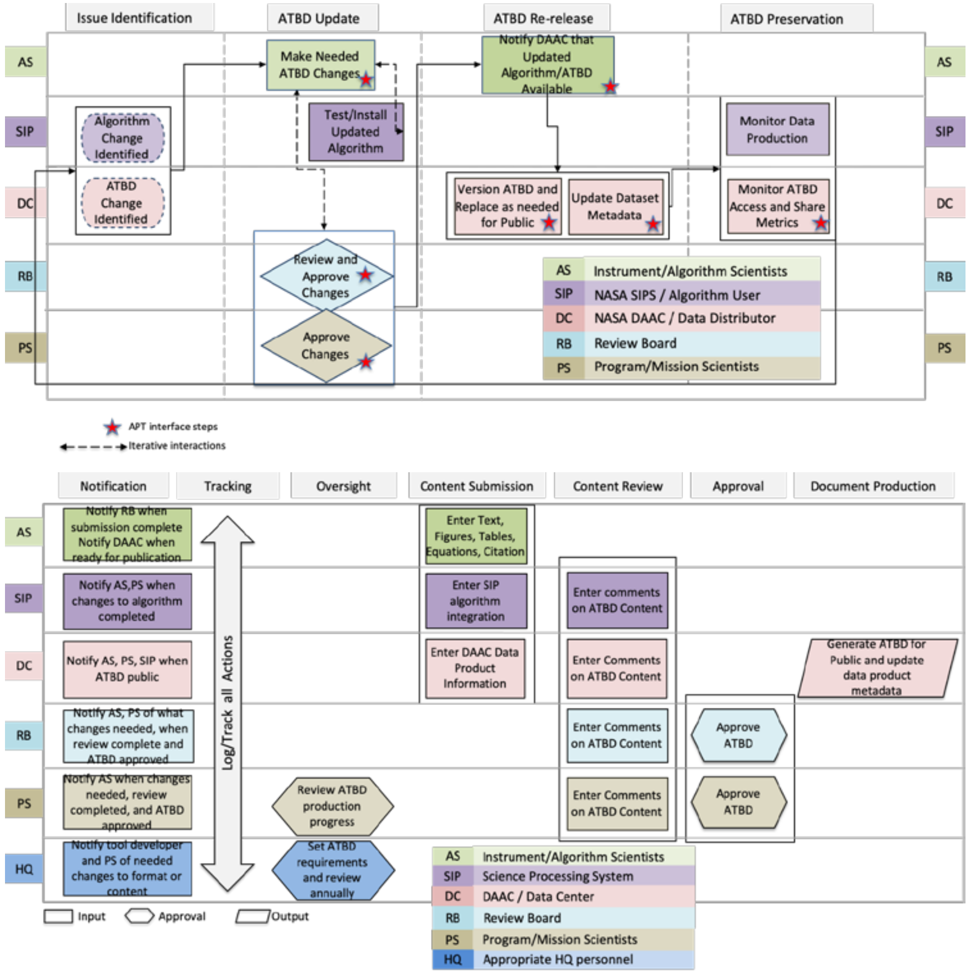
APT will reinforce NASA EOSDIS and data provider responsibilities throughout the data lifecycle

Provides independent workflows for

1) Initial creation and publication of an ATBD (formulation phase)
2) Updates to a ATBD (operations phase)

Dashboard for tracking/managing workflows

- APT serves as a centralized location for interactions between all parties involved in the data stewardship process


Issue Identification
ATBD Update
ATBD Re-release
ATBD Preservation
AS
SIP
DC
RB
PS
Algorithm Change Identified
ATBD Change Identified
Make Needed ATBD Changes
Test/Install Updated Algorithm
Review and Approve Changes
Approve Changes
Notify DAAC that Updated Algorithm/ATBD Available
Version ATBD and Replace as needed for Public
Update Dataset Metadata
Monitor Data Production
Monitor ATBD Access and Share Metrics
AS Instrument/Algorithm Scientists
SIP NASA SIPS / Algorithm User
DC NASA DAAC / Data Distributor
RB Review Board
PS Program/Mission Scientists
APT interface steps
Iterative interactions
Notification
Tracking
Oversight
Content Submission
Content Review
Approval
Document Production
HQ
Notify RB when submission complete Notify DAAC when ready for publication
Notify AS,PS when changes to algorithm completed
Notify AS, PS, SIP when ATBD public
Notify AS, PS of what changes needed, when review complete and ATBD approved
Notify AS when changes needed, review completed, and ATBD approved
Notify tool developer and PS of needed changes to format or content
Log/Track all Actions
Review ATBD production progress
Set ATBD requirements and review annually
Enter Text, Figures, Tables, Equations, Citation
Enter SIP algorithm integration
Enter DAAC Data Product Information
Enter comments on ATBD Content
Enter Comments on ATBD Content
Enter Comments on ATBD Content
Enter Comments on ATBD Content
Approve ATBD
Approve ATBD
Generate ATBD for Public and update data product metadata
AS Instrument/Algorithm Scientists
SIP Science Processing System
DC DAAC / Data Center
RB Review Board
PS Program/Mission Scientists
HQ Appropriate HQ personnel
Input
Approval
Output

Sample governance workflow and required interactions

Discussion

Versioning for dynamic documents is difficult

- Must decide what changes constitutes a new version of an ATBD
- How are persistent identifiers (document DOI) impacted?
- All versions must be archived and made available for transparency

For repository completeness, existing archived ATBDs should be incorporated

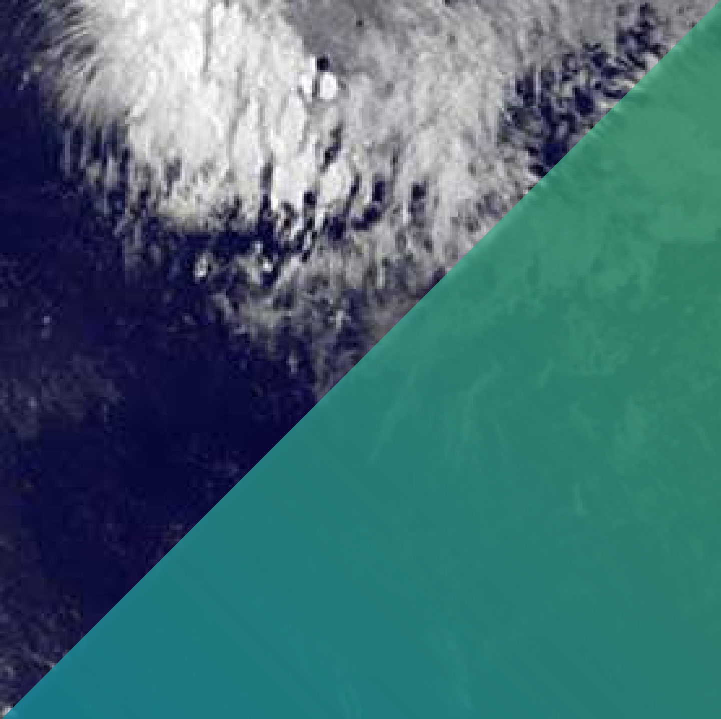
- Significant effort to rewrite all of these documents
- Can all ATBDs and their versions be accounted for?

Community support is required for success

- The tool should minimize resistance to adoption of a single typesetting service
- The tool must accommodate all scientific writing needs
- A standard governance plan should not overly/unnecessarily burdensome

Maximum value is achieved when integrated into EOSDIS existing metadata resources

- Need a flexible, future-proof tool to fit an evolving documentation vision and supporting suite of tools



Questions?

ak0033@uah.edu