

Improved Data Communication, Understanding and Discovery Using Algorithm Theoretical Basis Documents

Bradley Baker¹, Kaylin Bugbee², Aaron Kaulfus², Olaf Veerman³, Daniel da Silva³,
Rahul Ramachandran², Shawn Foley²

¹University of Alabama in Huntsville, ²NASA Marshall Space Flight Center, ³Development Seed

PRESENTED AT:



INTRODUCTION

Algorithm Theoretical Basis Documents (ATBDs)?

- ATBDs describe the physical theory, mathematical procedures and assumptions made for developing scientific algorithms that convert radiances received by remote sensing instruments into geophysical quantities.
- All NASA Earth Observing System instrument products require ATBDs.

Why are ATBDs important?

- Promote data transparency by ensuring reproducible science
 - Describe calibration and validation processes needed for data product development
 - Provide a mechanism for peer review of the algorithm
- Ensure algorithm and corresponding software are scientifically valid, efficient and forward looking or state of the art
- Promote understanding and proper use of Earth observation data
 - Serve as a reference for each data product for users
 - Describe scientific and mathematical assumptions in order to generate the data

What are the current challenges with ATBDs?

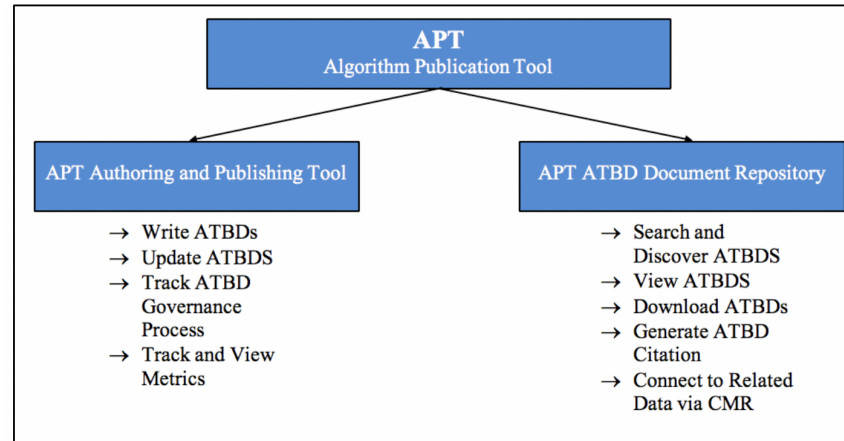
- A standardized ATBD template does not exist
- Difficult for scientists to know what content to provide
- Inconsistent structure and information across ATBDs provides a challenge for users
- ATBDs are difficult to discover
 - No centralized location to search and find ATBDs
 - Not consistently referenced in a dataset's metadata
 - Often stored as PDFs, making them difficult to discover via search engine
- Stewardship activities are variable and seen as less crucial to data stewardship activities

HOW DO WE MITIGATE CURRENT CHALLENGES ASSOCIATED WITH ATBDS?

The Algorithm Publication Tool (APT) is a cloud-based tool that accomplishes the following:

- Streamlines authoring, publishing and maintaining ATBDs
- Facilitates the ATBD governance process
- Provides a centralized discovery repository

APT has 3 component: (1) metadata model, (2) front-end tool and supporting content database and (3) centralized document repository.



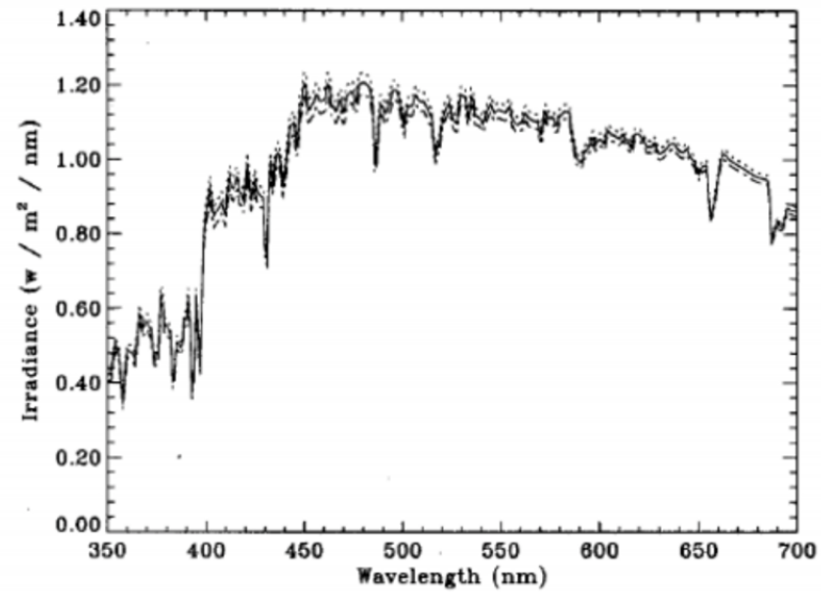
APT components

Key Features of APT's Front-end Tool

The APT uses a standardized ATBD template developed after reviewing ATBDs throughout the Earth observation community (e.g., NASA, NOAA, ESA, and NEON). The key features of the front-end tool include

- Guides users through the ATBD authoring process in 7 organized sections
 - Provides authors with required content
 - Reduces the need to organize and structure documents
 - Promotes standardization across ATBDs
- Uses:
 - LaTeX to properly render mathematical equations, tables and figures
 - BibTeX for citation management
 - State.js for writing dynamic rich text
- ATBD content is stored as metadata

$$IPAR = 1/hc \int_{400}^{700} \lambda E_d(\lambda, 0^-) d\lambda \quad (30)$$

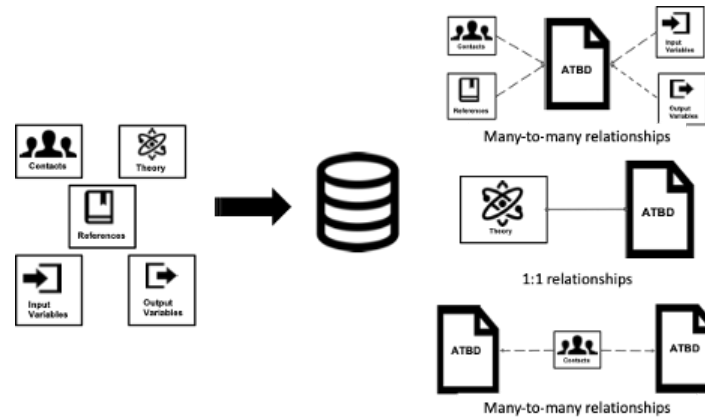


Equation and figure from ATBD drafted in APT

APT'S METADATA MODEL

The APT envisions ATBDs as structured metadata stored in a database that can be accessed and used by authors writing other ATBDs. Thus, APT

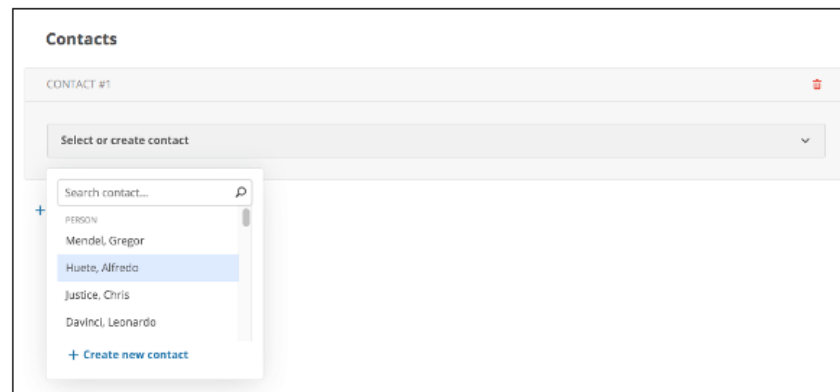
- Stores ATBD content as metadata
- Generates documents from metadata
- Provides component metadata as reusable across different ATBDs
 - Current reusable metadata includes citation and author information and scientific equations.



APT metadata model

The benefits of using a metadata model include the following:

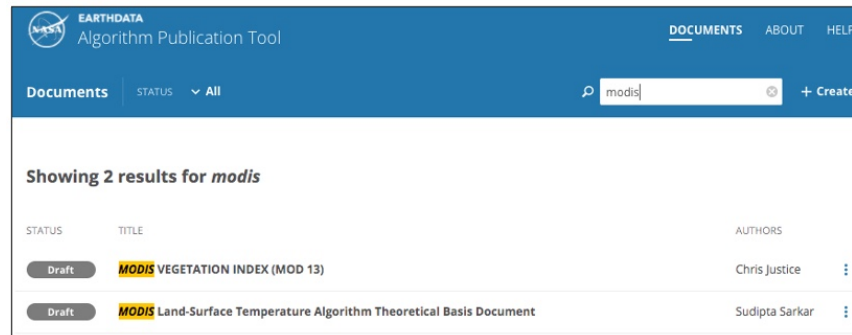
- Promote consistency and standardization while reducing human inputs
- Reduce manual input and therefore potential for errors
 - Prevents authors from repeatedly writing the same content
- Provide dynamic information that is easier to update
- Provide machine readable formats that are interoperable with other tools and services



DATA DISCOVERY WITH THE APT

The APT's centralized document repository [currently in work] will allow users to easily search and find ATBDs in a centralized location. The main features will include:

- Free and open ATBD discovery portal
- Search engine optimized
 - Machine readable
 - Use schema.org markup
- All ATBDs are given persistent identifiers for effective retrieval



ATBD elastic search

[Demo Only] MOD20: Instantaneous
Photosynthetically Available Radiation and
Absorbed Radiation by Phytoplankton

August 27, 2020

Contents

1	Introduction	2
2	Historical Perspective	3
3	Algorithm Description	4
3.1	Scientific Theory	4
3.1.1	Assumptions	5
3.2	Mathematical Theory	5
3.2.1	Assumptions	19
3.3	Algorithm Input Variables	19
3.4	Algorithm Output Variables	20
4	Algorithm Implementations	20
4.1		20
5	Algorithm Usage Constraints	20
6	Performance Assessment Validation Methods	21
7	Performance Assessment Validation Uncertainties	22
8	Performance Assessment Validation Errors	22

APT ATBD with dynamic table of contents

Conclusions

The APT addresses current challenges of traditional ATBDs by:

- Improving document authoring, publishing and updating
- Establishing standard ATBD content requirements and document organization
- Assisting in data discovery and promote users understanding
- Ensuring ATBDs are easily readable and accessible
- Promoting scientific reproducibility

ABSTRACT

Scientists and data repositories depend on the effective communication of scientific and physical theories used to derive Earth observation datasets from raw instrument data, which is important to understanding and properly using the data. The NASA Earth science data community communicates this information through Algorithm Theoretical Basis Documents (ATBDs). However, ATBDs lack a formal, standardized structure, which often results in ATBDs containing inadequate information to understand the algorithm or efficiently parse the document's content for the desired information. Additionally, science teams typically provide ATBDs in human, but not machine readable formats, which makes it difficult for modern information processing technologies to process the data. The Algorithm Publication Tool (APT) reconceptualizes ATBD content as metadata, or description about the products they represent, to streamline authoring and dynamic updating, encourage consistent information across ATBDs and promote human and machine parsing of information in order to simplify data understanding. The APT alleviates the aforementioned issues by envisioning ATBDs as metadata and providing a simplified cloud-based template for ATBD authoring, review, and publication. Data users can easily search and discover ATBDs using the APT's centralized repository. This presentation describes our effort to envision ATBDs as metadata and demonstrates how the tool supports data and information discovery.