Collaborative Data Curation to Support the Multi-Mission Algorithm and Analysis Platform (MAAP)

Kaylin Bugbee[1], Jeanne le Roux[1], Aimee Barciauskas[2], Betzy Hernandez[1], Marco Lavalle[3], Gary Geller[3], Amanda Whitehurst[4], Rahul Ramachandran[5]

1 – University of Alabama in Huntsville, 2 – Development Seed, 3 – Jet Propulsion Laboratory, California Institute of Technology, 4 – Arctic Slope Technical Services, Inc., 5 – NASA Marshall Space Flight Center

IMPACT
Interagency Implementation and Advanced Concepts Team

What is MAAP?

Upcoming space-borne missions will offer unprecedented data about Earth but will also feature exponentially high data volumes. These high data volumes will change the way the scientific community works with data and will also create a unique need for improved data sharing and collaboration. NASA and ESA are working together to address these issues by collaboratively developing the Multi-Mission Algorithm and Analysis Platform (MAAP) to improve the understanding of global aboveground terrestrial carbon dynamics. The MAAP will support ESA's BIOMASS mission, NASA's GEDI mission and NASA/ISRO's NISAR mission.

The MAAP will be developed in two phases: a pilot phase and a full production phase. The pilot phase will demonstrate collaboration and basic capabilities. The pilot phase will focus on biomass relevant airborne and field campaign data. Two NASA teams are supporting the development of the MAAP. The MAAP engineering team is responsible for the development, maintenance and operations of the MAAP system while the MAAP data team ensures the ongoing quality of the data, metadata and other information provided in the MAAP. The MAAP data team also supports the ingest and archive of identified data to the MAAP platform. This poster describes the use case development process for the pilot MAAP and the data curated in support of those use cases. Additionally, this presentation will outline the pilot MAAP data ingest process and metadata curation effort along with efforts to ensure interoperability between ESA and NASA data and metadata.

MAAP Use Cases

In order to better understand the capabilities needed for the MAAP, sixteen generalized use cases were developed from discussions with biomass scientists. A subset of these use cases were then identified in order to drive requirements and success criteria for the pilot MAAP. Tools and data needed to support the pilot MAAP were also identified and prioritized.

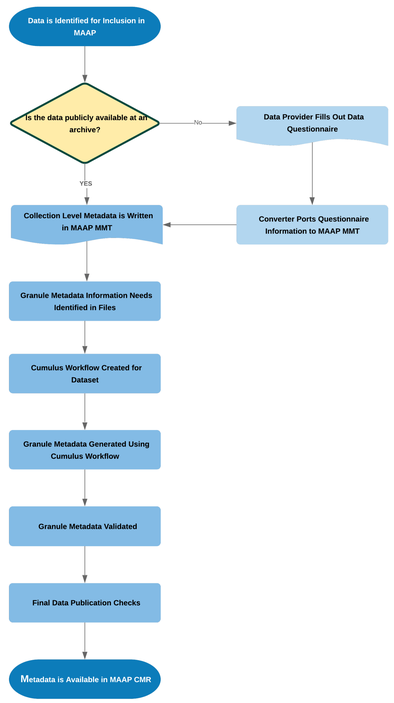
Example Use Cases	Use Case Purpose
User registration, orientation, and first search	Convey key capabilities to support a new user getting started with no background on the system.
Multi-sensor data order, visual comparison, and save	Demonstrate that user can find and work with data from multiple sensors and that user can save images on the MAAP or download them locally.
Sub-setting to different formats	Demonstrate how a user subsets data and saves the subset as images or a spreadsheet file.
Temporal comparison to look for change	Demonstrates temporal comparison capabilities using flicker or swipe technique.
Temporal trends	Demonstrate temporal trend capability using animation and trend graph; help the user review potential explanatory variables for those trends.
Algorithm calibration using in situ data	Demonstrate how an algorithm developer would test an algorithm by comparison with other data.
Information sharing among developers or scientists	Demonstrate how developers or scientists are able to share information and facilitate working as a team.
Code development and testing	Demonstrate how developers or scientists are able to import, develop and run new code in the MAAP.

Data needed to support the pilot use cases were identified and key metadata information was curated for these data. The data for the pilot phase primarily focuses on the AfriSAR field campaign and supporting ancillary data. Recognizing the limited time available during the pilot phase, data was prioritized to support the use case. Essential data for demonstrating the pilot MAAP was prioritized as a must have while supporting data were further prioritized as should and could have for the pilot. Data was also assessed for both ingest and curation complexity in order to better understand the effort required in providing these data in the pilot MAAP. Lastly, data was assessed for compliance to standard MAAP data formats.

MAAP Data & Metadata Curation

The data ingest process into the MAAP begins with creation of a collection level metadata record for the dataset using the MAAP Metadata Management Tool (MMT). Granule level metadata requirements for the dataset are also identified and documented. After both metadata steps are complete, a Cumulus workflow is created to ingest the data into the MAAP data store, to generate granule level metadata and to push the granule level metadata into the MAAP Common Metadata Repository (CMR). Cumulus is NASA's Earth observation cloud-based data pipeline. After this step is complete, both the granule level metadata records and the data files are checked for accuracy. Lastly, the collection level metadata is rechecked to ensure accessibility.

MAAP Data Team Workflow



Additional Metadata in MAAP

High quality metadata is essential to MAAP in order to support effective search and discovery. The use cases for pilot MAAP demonstrated a need for the development of additional metadata in order to support the unique search criteria of the biomass community. These search needs varied from platform and instrument type. Over twenty additional attributes have been identified to date. Examples of some of the requested search criteria can be found in the table below.

Platform/Instrument Type	Search Request	Example Values
Airborne & Satellite SAR	Polarization	HH, HV, VH, VV
Airborne SAR and Lidar	Flight Number	16008
Airborne SAR	Track Number	001
Airborne SAR and Lidar, Terrestrial Lidar, In situ Measurements	Site Name	Lope National Park Gabon

Additional information is being included in the metadata as additional attributes for both collection and granule level metadata. The additional attributes will continue to be refined as both NASA project scientists and ESA subject matter experts review the information and provide further feedback and guidance.

NASA and ESA Interoperability

With the goal of making biomass relevant data more discoverable and usable, NASA and ESA are collaborating together to make metadata and data more interoperable. For the pilot phase, ESA plans to use the MAAP Common Metadata Repository to ensure that all relevant metadata is provided in a centralized location. Additionally, ESA plans to use the MMT to curate collection level metadata to support metadata interoperability. Lastly, NASA and ESA are working together in identifying the additional metadata required to support the biomass research community needs.

Contact: Kaylin.m.Bugbee@nasa.gov

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

NASA