ARM Data Discovery – Planned Upgrades and Design Review

Giri Prakash, Ranjeet Devarakonda, Jim Mather, Kyle Dumas, Maggie Davies, Ric Cederwall + ~50 stakeholders

ARM Data Center
For ARM/ASR PI Meeting
06/12/2019

Scope

- Search providing the user the data/products they are looking for and other relevant data/products
    - Currently Too Many Results: Searches return too many products that require the user to sift through
    - Need to Provide 'Right Data' Results: Some searches do not yield datasets that are available or relevant to user

- Quick, intuitive searching ability
    - Finding data needs to be intuitive and serve new users as well as ARM knowledgeable users

- Search providing intuitive identification and access to the metadata and related information
    - Periods of high quality data, products, and event time periods
    - Quality and uncertainty
    - ARM related contacts understandable by new and existing users

Stakeholder discussions

Reaching out to each group to encourage individuals who might not be able to provide detailed experience and input in a large one meeting setup.

Radar Group

AAF Group

Data recommendation strategy

Instrument mentors

Translators and science products

Aerosol group

LASSO modeling

ASST

ADC user interactions
Communications group
UEC

DQ Office

Reprocessing and HPC

ASR Working groups:

- Aerosol Processes
- Warm Boundary Layer Processes
  Convective Processes
- High-Latitude Processes

SFA/site science leads

Recent campaign Pis

New(er) ASR Pis

Suggested users

Radar users

Additional recommended users from the above stakeholders

scientific users

Usability Review

- UT User Experience Lab conducted a day-long usability study and provided detailed report with specific recommendations

- Feedback from ARM Communications group

- Data tools and UI experts from the Labs, NASA and USGS

Study demographics represent relevant ARM user sciences

USER #	SEX	FIELD OF STUDY/DISCIPLINE	USE OF ARM DATA DISCOVERY IN PAST 60 DAYS
1	F	Climate Science	Never
2	M	Climate Science	Never
3	M	Biogeoscience	Never
4	F	Atmospheric Sciences	Never
5	M	Geospatial Information Management	Never
6	F	Water Resources	Never
7	M	Terrestrial Ecosystem Modeling	Never
8	F	Ecology	Never
9	F	Geospatial Information Management	Never
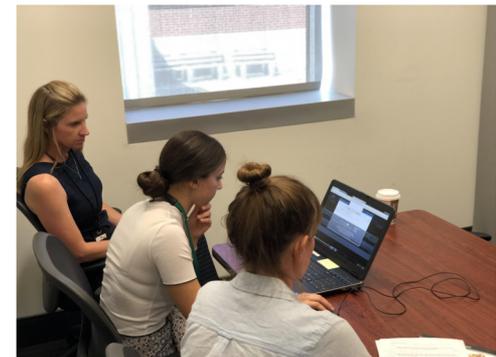
Usage Pattern Analysis

Heatmap analysis represents which functionality is being used most in the current Discovery Tool
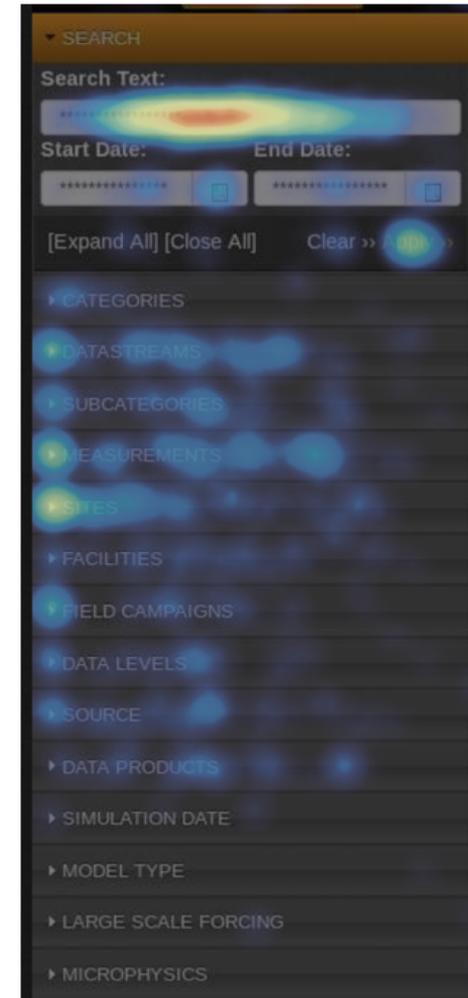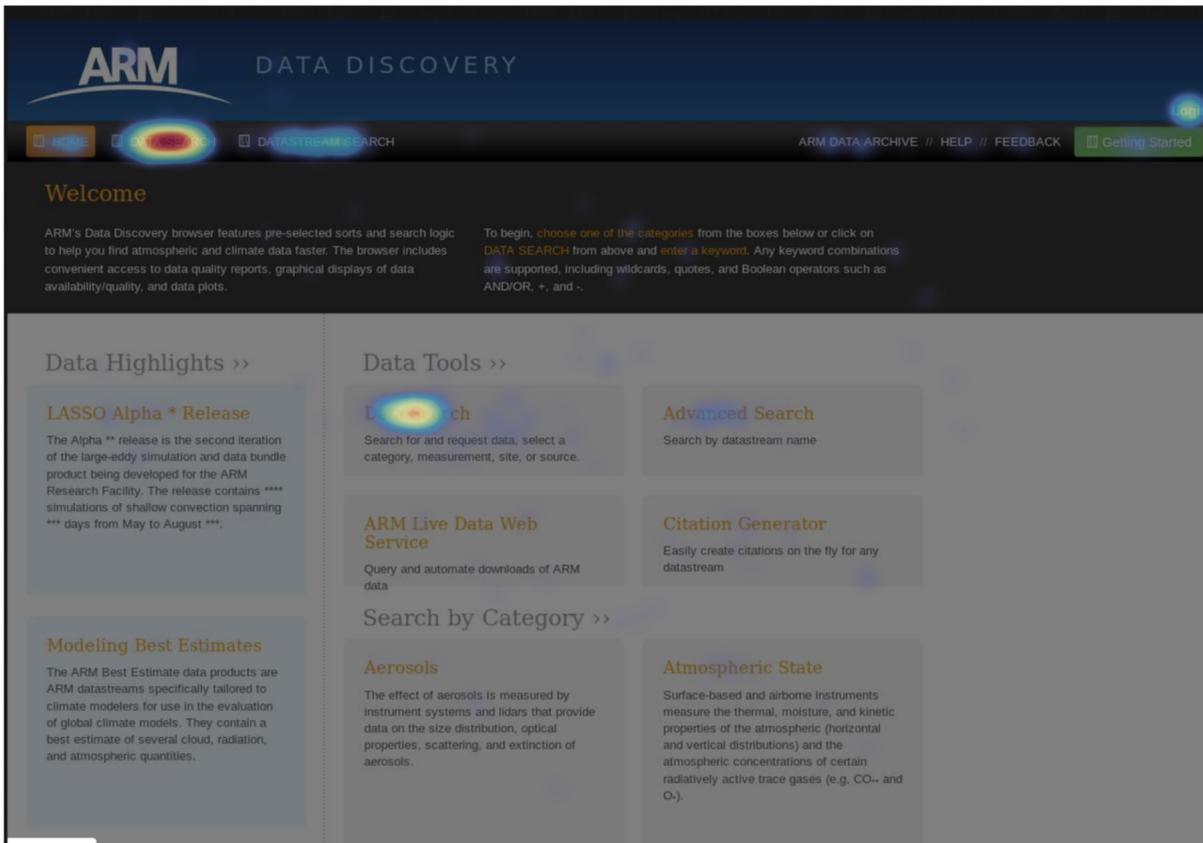
Stakeholder Feedback Summary

Data Discovery Compliments and Criticisms

Compliments:
- Powerful and user friendly tool
- By far the best data portal, easy to navigate
- Quick access to data (few clicks)
- Clear and one place to access the data
- Options to get multiple formats of data
- Advanced datastream search options
- Happy with continuous improvements (example: web services)

Criticisms:
- Too many
    - results and choices
    - facets that aren't necessary to all users
    - functionalities as default (The cluttered display distract users from important features)
- Difficulty undoing selected filters
- Datastream search doesn't show all datastreams (raw and processed)
- The search and filter options need to be intuitive for new users

U.S. DEPARTMENT OF
ENERGY

Key Recommendations

- Reduce the number of search results to new users by using recommended data sources with an option to show all results.

- Simplify data search results and the user interface

- Make searches more intelligent (by including recommended sources, broader ontologies etc..)

- Centralize metadata information to be cleaner and easier to find

- Provide guided and spatial search for new/novice users

- Ease the transition between arm.gov and data discovery

- My Accounts page with access to order history

Other Recommendations

- Homepage Redesign to dynamically update contents
- Add interactive data plotting capabilities that can be used to visualize larger datasets and browse data epics similar to LASSO Bundle Browser.
- Add tag support to the metadata process and on the UI to allow mentors and users to add keywords that can be searched on. These tags could be used to flag data "Data Epoch" or specific characteristics such as "storm" for specific time periods and be filtered on.
- Add a map view to help supplement location data and allow spatial filtering
- Add a view to provide detailed information about the data product
- Virtual Field campaigns (data bundling)
- Provide data provenance information (e.g: source datastreams)
- Add the ability to search on the value of another instrument (similar to LASSO)
- Standardize variable names across all sites, improve the semantics (sites vs facilities)

Planned Backend Upgrades

- Include DOD metadata records in the search index
- Better metadata mapping and auditing functionalities
- Deploy recommended data source workflow
- Interlink broader earth science ontologies (SWEET, GCMD, CF) and ARM metadata

Key Recent Upgrades
(Continuous Improvement Cycle)

- Whitepapers
  - recommended data sources (new prototype tool),
  - spatial data search
- Citation generator using DOIs
- Data extraction upgraded with:
  - ASCII-CSV format (request from AOS community-ENG0003662)
  - ASCII-ICARTT (request from AAF community-ENG0003856)
- Advanced data delivery using Globus (GridFTP), THREDDS/OpenDAP (ENG0003245)
- All evaluation data are grouped under one classification (ENG0003295)
- API/Webservice option for requesting data from ADC (ENG0003797)
- Event Driven Indexing (ENG0003907)
- Improved DQR displays (ENG0003662)
- Improved data retrieval and notification emails
- Enhanced the information bubble with links to instruments, handbooks and technical reports

Data Visibility with External Portals

ARM

ARM Data Center

Science metadata
- ISO 19115,CF
- DOI
- JSON/THREDDS

Data access
- Extraction
- Visualization
- OPENDAP

Data Gateways
Data.gov
DataCite
IASOA,
NGEE-Arctic

Google data search

AWI/Pangea

Metadata harvesting
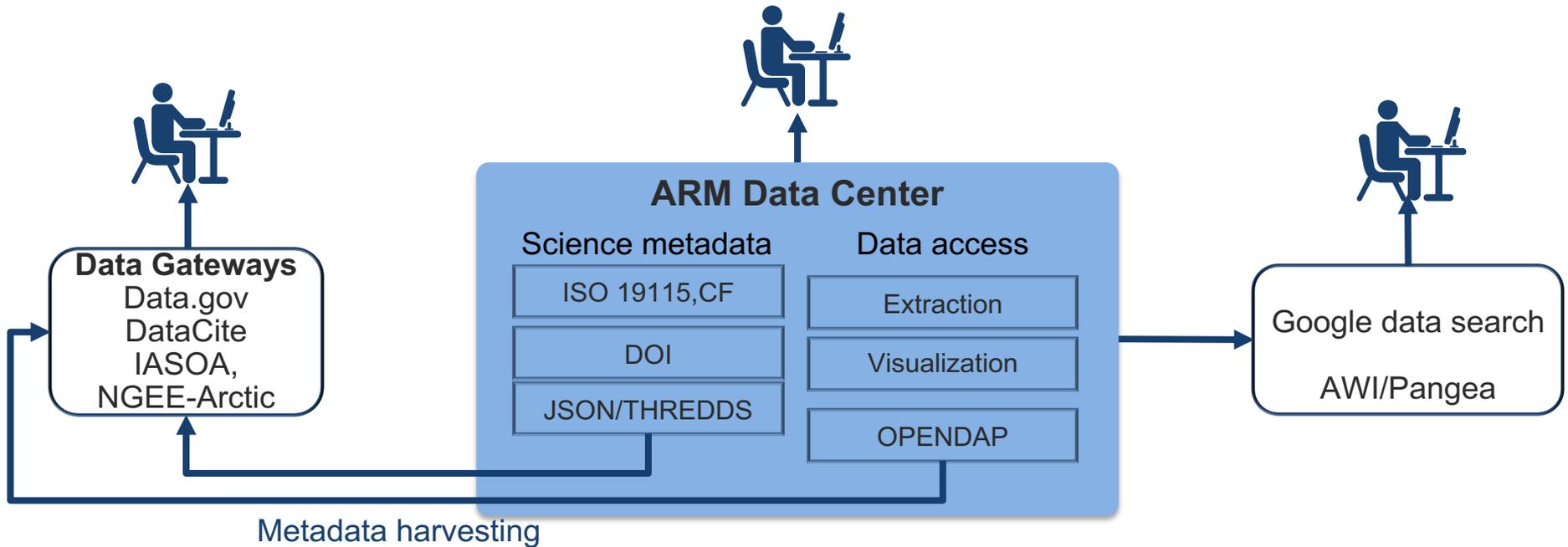
Download data and additional information

- Improve the visibility of ARM data products in external data clearinghouses and relevant scientific portals

ARM Standards:
- Data standard
- Data Quality Report
- Metadata ontology (hierarchical scientific keywords)

Status and Next Steps

- Continue the design review with various stakeholders

- Implement the design using continuous integration process (April 19)

- Present and collect feedback from scientists at the ASR/ARM PI meeting (June 19)

- Release of Beta version for initial feedback and usability (end of August 19)

- Release of UI - Phase I (Q1 2020) and Phase II (Q2 2020)