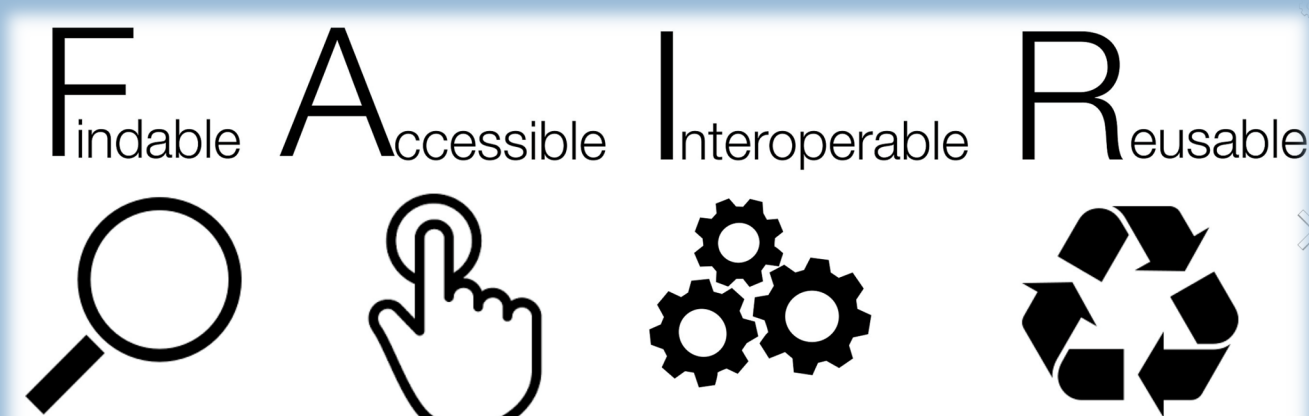# USING METADATA STANDARDS TO SIMPLIFY DISCOVERY AND NAVIGATION OF RESULTS

The Atmospheric Radiation Measurement (U.S. DOE: ARM) Data Center handles processing, submission, archiving, and distribution of data collected from instruments, value-added products, and field campaigns to the scientific community.

Data producers and repositories are currently faced with the challenges of making ever larger, more numerous, and increasingly heterogeneous collections of data available to users. General searches are typically met with an overwhelming number of results which are difficult to evaluate for relevance. This poster discusses ARM Data Center's approach to simplifying the discovery and navigation of relevant search results.
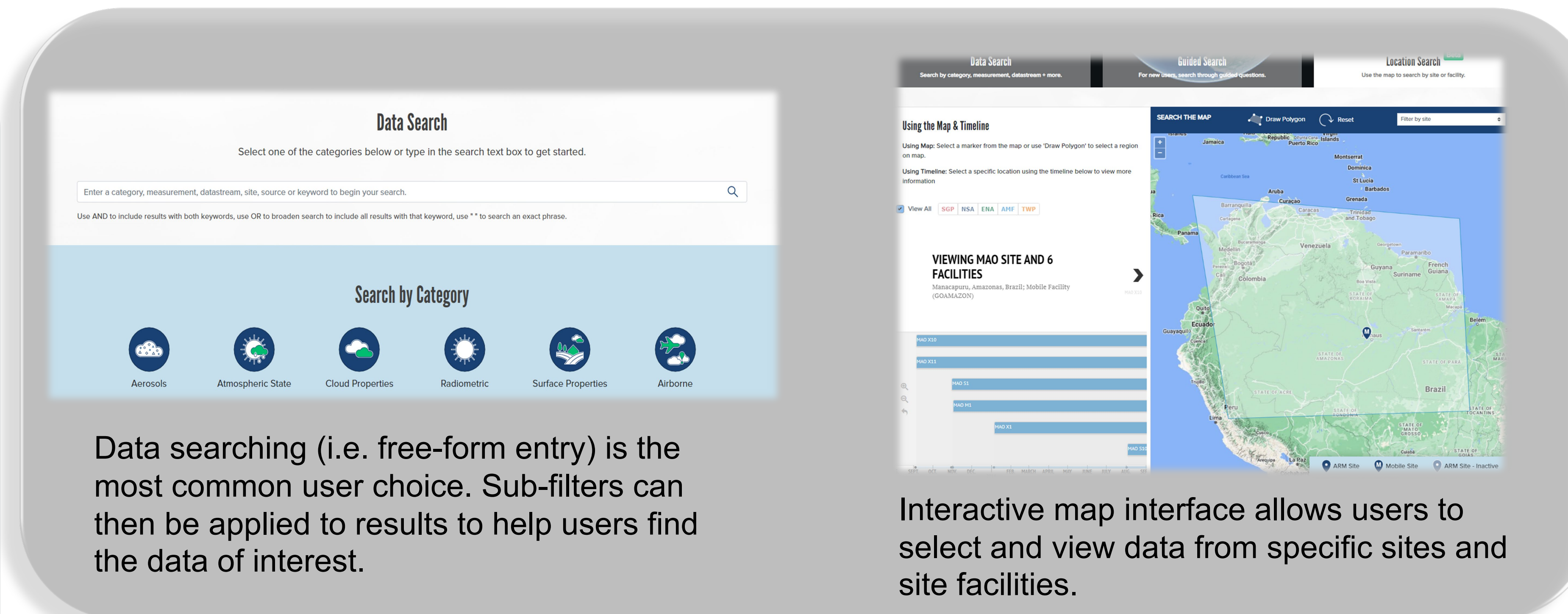
The ARM Data Discovery Tool (https://adc.arm.gov/discovery) helps scientists find and access datasets and ensure FAIR data practices.

Findable **A**ccessible **I**nteroperable **R**eusable

mao: (ARM site identifier) Manacapuru, Amazonas, Brazil; Mobile Facility (GOAMAZON)

S10: (ARM Facility designation) Manacapuru, Amazonas, Brazil; Supplemental site

20141014: (date) - YYYYMMDD

maoraintbS10.b1.20141014.212000.cdf

212000: (time UTC) - hhmmss

b1: (Data level) quality-control checks applied to at least one measurement.

raintb: (ARM instrument abbreviation) Tipping bucket rain gauge data

For the ARM data center, a **standardized datastream** name includes some of this metadata in a standardized format: Site: All characters are lowercase, only "a-z" and limited to 3 characters. Instrument: a code consisting of "A-Z", "0-9", and "." characters are allowed. ARM has excluded any proprietary names of instruments. Facility: a two letter code, following standards. Data level: level of data processing E.g., 00 for raw data. Other key metadata includes: Investigators, Data Format, File Naming Convention, Stratum Keywords, Data Type, Access Restriction, Use Restriction, Start Date and End Date, among others.

Data searching (i.e. free-form entry) is the most common user choice. Sub-filters can then be applied to results to help users find the data of interest.

Interactive map interface allows users to select and view data from specific sites and site facilities.

Recommendations:
Our key practice to "bin" ARM data by various means, along with data quality information allows recommendations for various user types e.g., core measurements address the ARM scientific mission, and have facilitated recommendations across measurements criteria.
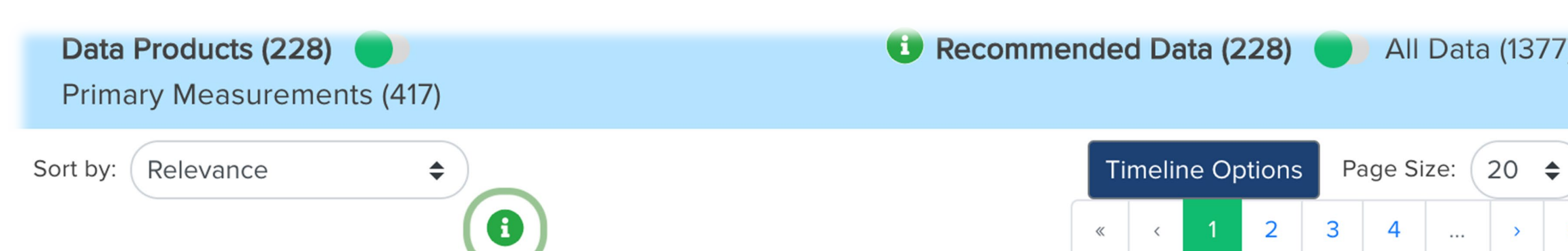
Recent improvements:
- **Justifications** of recommended data will be released to the user community as "Tweets"
- **Continual updates** of recommendations will save time and effort
- **Tagged descriptors** provide users another way to search on data, and in some cases facilitate new groupings of data (e.g., for data epochs)

Back-end work on metadata facilitates better user experience and better citations of ARM data:
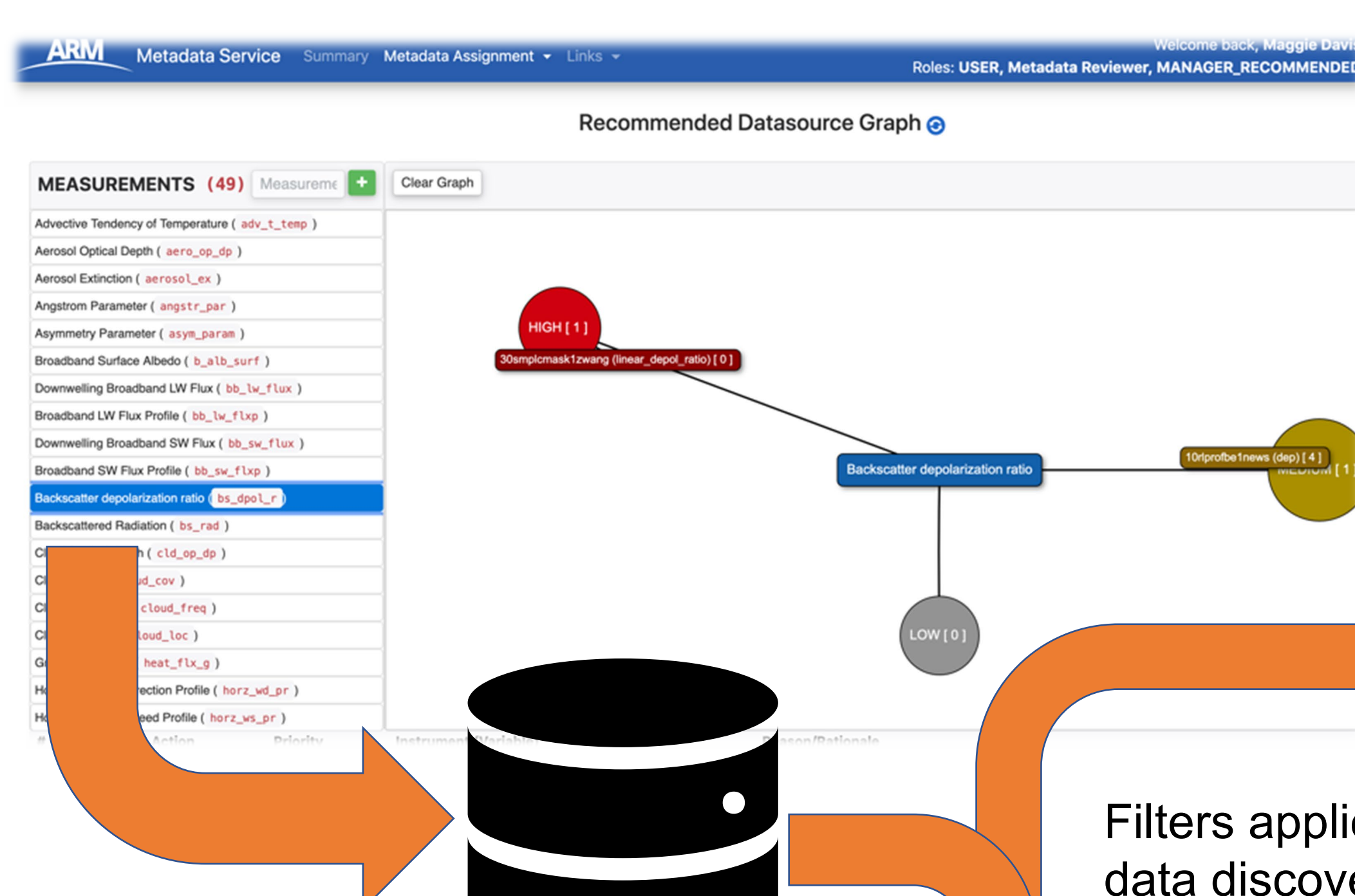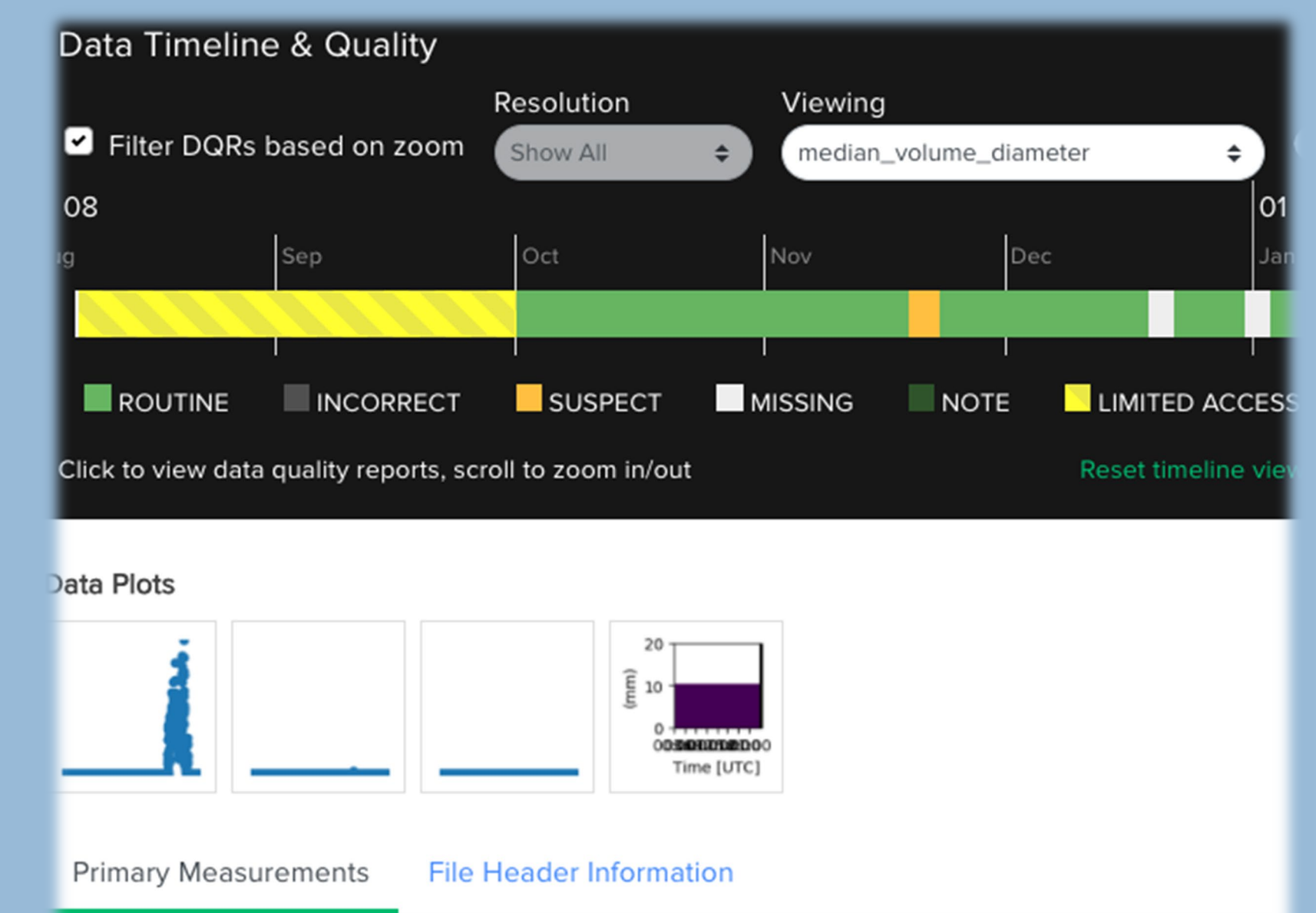
Recommendations & DOIs by datastream (i.e. code + data level) stored in the database

Filters applied in data discovery for users; user preferences allow changes

Other stored metadata facilitate a guided search for users of data discovery, including a new epoch experience. The guided search assess the user's needs, and returns limited datasets, decreasing the need to manually filter through results.

Data Quality Reports (DQRs)—visible while browsing for data, received with data orders, and accessible through the DQR web service. User contributions "inform ARM of any issue"
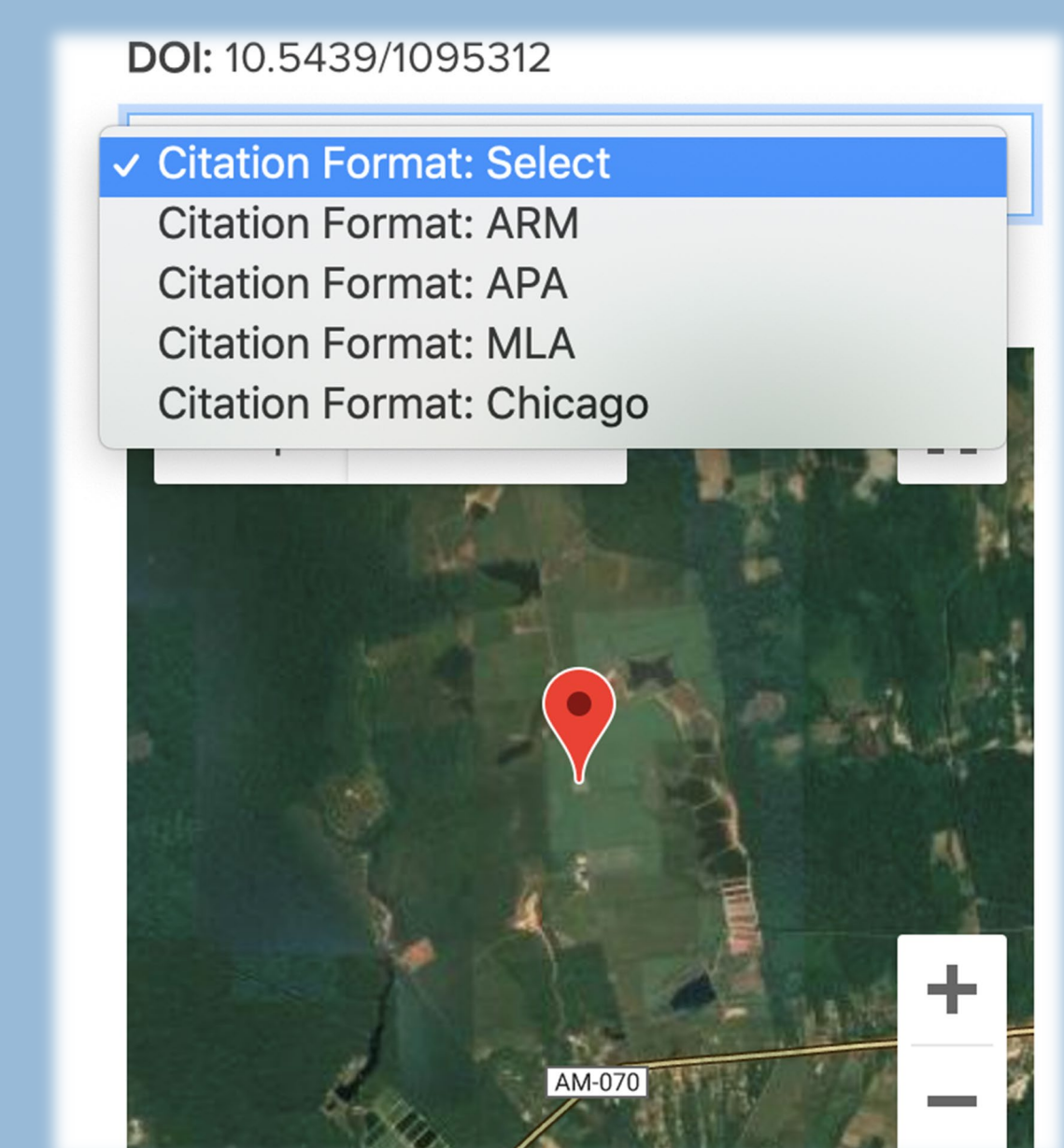
Automated data citation facilitates proper attribution, allowing ARM to track user metrics across publications as well. Prefix is ARM-specific, automatically reserved (OSTI)