



# AMERICAN ASTRONOMICAL SOCIETY

American Astronomical Society public comment on “*DRAFT Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research*” (Document 2020–00689; Posted 2020-01-17)

Physical Sciences

16 March 2020

Authors:

August Muench, AAS Data Editor

Greg Schwarz, AAS Data Editor

Julie Steffen, AAS Director of Publishing

The American Astronomical Society (AAS) is the major organization of professional astronomers in North America. Its membership of over 8,000 individuals also includes physicists, mathematicians, geologists, engineers, and others within the broad spectrum of subjects comprising contemporary astronomy, planetary science, and heliophysics. The mission of the AAS is to enhance and share humanity’s scientific understanding of the universe.

As a 501(c)(3) nonprofit corporation, the AAS owns, operates, and publishes the most widely read and cited journals in the field, *The Astronomical Journal*, *The Astrophysical Journal*, *The ApJ Letters*, *The ApJ Supplements Series*, and *The Planetary Science Journal*. Since creating electronic editions starting in 1995, the AAS has encouraged researchers to submit data critical to their research result *along with their manuscript*. Machine-readable tables (MRT) and data-behind-figures (DbF) are examples of the research data integrated into and hence preserved for posterity in many thousands of research articles published in AAS journals. The AAS has employed trained astrophysicists as data editors and adopted publishing workflows that help researchers share their data for the past twenty years, which has led to the inclusion of a significant amount research data in the literature.

Additionally, the AAS has spearheaded efforts to link to important, related data sets in federally funded data repositories and will continue to develop and deepen these connections.

## **The proposed use and application of the desirable characteristics**

Astronomers share many of the problems experienced by other researchers in the physical sciences, although some of our common problems have been solved. Original data products resulting from NASA space missions are well-curated and stored in repositories that follow many (but not all) of the proposed characteristics. This culture of saving and sharing original data has made astronomy a leader among physical science disciplines in sharing research data. However, astronomy has a shortage of domain-specific repositories for most derived data products, for simulation and modeling results, and for data not resulting in a research publication (e.g., null result data). There are few existing repositories in astronomy that accept researcher data and satisfy all proposed characteristics. Requiring them for new repositories may narrow rather than expand the already limited options for federally funded investigators.

## **The appropriateness of the “Desirable Characteristics for All Data Repositories” (Section I)**

The AAS finds the proposed characteristics noteworthy and valuable and endorses them uniformly. Further commentary, informed by long experience working with researchers, is intended to highlight specific issues and to reflect on the current repository landscape for astronomy researchers.

In addition to endorsing the entire set of proposed characteristics, the AAS strongly endorses the need for curation and quality assurance mechanisms in data repositories (**Section I; Characteristic D**). Data submitted without curation to generalist repositories are of limited value and may be missing critical details, e.g., units, that are necessary for either human or machine reuse. Standard data review conducted at AAS uncovers errors in the tabulation of results that would otherwise not be detected, especially if those data are archived ex post facto in a generalist repository. Useful metrics should correlate successful compliance of an open data mandate with enhancement of the scientific record. Unreviewed data may distort these metrics, rendering them useless.

Adding open curation platforms to repositories would improve the quality and success of data sharing by researchers. Experience indicates that supporting and assisting researcher data submissions increases the likelihood of data sharing and improves the overall result. Generalist repositories tend to lack workflows for external review and improvement of submitted data and even domain-specific archives struggle with managing data review efficiently and expeditiously. Enabling curation by external teams, such as data scientists, or other stakeholders, such as data librarians, would make the sharing process more efficient and accurate.

The AAS also strongly endorses the need for domain-specific metadata (**Section I; Characteristic C**). It may be valuable to enable search and discovery across repositories using abstracted or “common” metadata; however, community-specific metadata schema, e.g, the standards of the International Virtual Observatory Alliance (IVOA), are even more vital for successful reuse and interoperability.

## **The ability of existing repositories to meet the desirable characteristics**

Astronomy is well positioned as a result of data archiving and release mandates put in place for NASA space missions and their data repositories (examples include the Infrared Science Archive [IRSA] and the Mikulski Archive for Space Telescopes [MAST]). The Astrophysics Data System (ADS) is a federally funded repository of bibliographic data that is fundamental for astronomy and astrophysics, providing links between the literature and data archives. As previously mentioned, however, more repositories are needed for derived data products not covered by the current scope of these NASA repositories, such as ground-based observations, model and simulation data, and laboratory astrophysics data.

The AAS is in active collaboration with NASA data repositories that are engaged in improving their functionality to support many of the proposed characteristics, including generating persistent identifiers for data. None of these NASA data repositories, however, are currently CoreTrustSeal certified. Some repositories do not accept data in advance of journal publication. Negative or “null” result data may go unarchived in the current repository landscape.

## **Summary Conclusions**

- The AAS has actively supported research data sharing for over twenty years by encouraging researchers to submit the data critical to their research result along with their manuscripts to our flagship research journals.
- The AAS believes that the most successful research data sharing involves curation and researcher support and has employed professional astronomers as data editors for this purpose.
- The AAS supports domain-specific repositories and metadata over more general solutions.
- The AAS actively collaborates with existing federally funded data repositories and would welcome further adoption of the OSTP proposed repository characteristics at these repositories.

## **END OF COMMENTS**

If you would like to follow up on any of the above comments, you may contact the AAS at [public.policy@aaas.org](mailto:public.policy@aaas.org).