



# The utility of dataset DOIs in manuscript review and scientific publications

August (Gus) Muench, Data Editor  
10 May 2023

AMERICAN ASTRONOMICAL SOCIETY

State (2023) of Data Linking and Deposits to  
NASA & other Astronomy archives archives?

**Sometimes maybe good**

data. The normalized OVI absorption lines and relevant discussion are presented in Section 6.1.  
 All *MAST*, *FUSE*, and *GALEX* data used in this paper can be found in *MAST*: 10.17909/19FGGR.

The screenshot shows the MAST data archive interface. At the top, there are options to 'Select a collection' and 'and enter target'. Below that, there's a search bar and a table of data. The table has columns for ID, Access, Headers, Instruments, Project, Filter, Wavelength, and Target Name. An orange arrow points from the text above to the table.

## NASA Exoplanet Archive Digital Object Identifiers

A Digital Object Identifier (DOI) is a permanent and unique string used to identify a digital resource and link to it on the web. IPAC has created DOIs for many astronomical data sets and for its data or collection services, e.g., NED. If you are writing a manuscript that uses one of these data sets or services, we recommend that you reference the corresponding DOI in your manuscript's bibliography, in addition to the appropriate refereed journal article. If you are submitting to an AAS Journal, instructions for referencing IPAC DOIs in your manuscript are provided here: <https://journals.aas.org/aastexguide/#softwareandthirdparty>. Tools such as [doiZbib.org](https://doi.org/10.26434/chemrxiv-2019-08-29-2) are useful for quickly generating BibTeX markup for these DOIs.

35 DOIs Search for Author, Title, Abstract or DOI (3 characters minimum)

DOIs by project: All [IRSA](#) [NASA Exoplanet Archive](#) [NED](#) [ExoFOP](#) [KOA](#)

- ▶ Confirmed Planets Table
- ▶ Composite Planet Data Table
- ▶ K2 Candidates Table
- ▶ Kepler Objects of Interest Cumulative Table
- ▶ Kepler Objects of Interest DR 25 Table

## Widening adoption (data citation & deposit)

- [MAST](#): Mint DOIs for collections of MAST data used; cite HLSPPs directly (~210+ datasets);
- [Cite IRSA/NEA/IPAC resources/services](#) used in author's reference list (~600 dataset & service DOIs);
- Cite one of 30,000 [ESA](#) dataset DOIs!
- Every [Chandra](#) ObsID has a DOI (~28,000).
- [Vizier](#) has adopted dataset DOI for most of their holdings (though you can also cite them via yCats in ADS – that's not confusing, is it?).
- Adding data deposit repositories: CADK, China-VO

State (2023) of Data Linking and Deposits to  
NASA & other Astronomy archives archives?

**Sometimes maybe not so good**

data. The normalized OVI absorption lines and relevant discussion are presented in Section 6.1.

All *HST*, *FUSE*, and *GALEX* data used in this paper can be found in MAST: 10.17909/19F-GGR.

ObsID	Access	Hardware	Instruments	Project	Filter	Weekend	Target Name
1	***	HST	CSST/PCV	HST	01399	UV	PH1-029-026
2	***	HST	CSST/PCV	HST	01399	UV	PH1-027-026
3	***	HST	CSST/PCV	HST	01399	UV	PH1-027-026
4	***	HST	CSST/PCV	HST	01399	UV	PH1-027-044
5	***	HST	CSST/PCV	HST	01399	UV	PH1-027-044
6	***	HST	CSST/PCV	HST	01399	UV	PH1-028-067
7	***	HST	CSST/PCV	HST	01399	UV	PH1-028-067
8	***	HST	CSST/PCV	HST	01399	UV	PH1-028-067
9	***	HST	CSST/PCV	HST	01399	UV	PH1-028-024

## NASA Exoplanet Archive Digital Object Identifiers

A Digital Object Identifier (DOI) is a permanent and unique string used to identify a digital resource and link to it on the web. IPAC has created DOIs for many astronomical data sets and for its data or collection services, e.g., NED. If you are writing a manuscript that uses one of these data sets or services, we recommend that you reference the corresponding DOI in your manuscript's bibliography, in addition to the appropriate refereed journal article. If you are submitting to an AAS Journal, instructions for referencing IPAC DOIs in your manuscript are provided here: <https://journals.aas.org/aastexguide/#softwareandthirdparty>. Tools such as [doiZbib.org](https://doi.org/10.1002/doi) are useful for quickly generating BibTeX markup for these DOIs.

35 DOIs Search for Author, Title, Abstract or DOI (3 characters minimum)

DOIs by project: All IRSA **NASA Exoplanet Archive** NED ExoFOP KOA

## Confusing & Inconsistent Landscape

- **Variable practices:** **Only** MAST provides collection DOI services; use 10 Chandra ObsIDs? Cite 10 Chandra ObsIDs! Use a specific observation, e.g., Spitzer AOR, held at IRSA? Cite nothing!
- **Variable but improving metadata:** Discussions with IPAC, ESA, and Chandra about **inconsistent** DOI metadata in need of uniform repair to not confuse authors. Making progress!
- **Obscure data-deposit guidelines** : limited by archive “charters”; no one-stop deposit or FAQ. Send us an email!
- **There is no bit-level reproducibility:** None of the archives provides persistent, reproducible, **versioned** data citations.

- ▶ Confirmed Planets Table
- ▶ Composite Planet Data Table
- ▶ K2 Candidates Table
- ▶ Kepler Objects of Interest Cumulative Table
- ▶ Kepler Objects of Interest DR 25 Table

# Swamp? Well, sort of!

The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and links for 'Upload', 'Communities', 'Log in', and 'Sign up'. Below the header, the page is titled 'AAS Journals'. A 'Recent uploads' section is visible, featuring a search bar and a list of three items. Each item includes a date, a 'Dataset' or 'Software' label, an 'Open Access' button, and a 'View' button. The first item is 'The Radius of PSR J0740+6620 from NICER with NICER background estimates' from October 6, 2022. The second is 'A rapidly evolving high-amplitude  $\delta$  Scuti star crossing the Hertzsprung Gap' from September 4, 2022. The third is 'Evidence of Long-Term Period Variations in the Exoplanet Transit Database (ETD)' from September 21, 2022. To the right of the search results is a 'New upload' button and a 'Community' section for 'AAS PUBLISHING', which includes a description of the journal collection, contact information, and a 'Curated by' section.

zenodo Search Upload Communities Log in Sign up

## AAS Journals

### Recent uploads

Search AAS Journals

October 6, 2022 (v1.0.0) Dataset Open Access View

**The Radius of PSR J0740+6620 from NICER with NICER background estimates**

Salmi, Tuomo, Vinciguerra, Serena, Choudhury, Devarshi, Riley, Thomas E., Watts, Anna L., Remillard, Ronald A., Ray, Paul S., Bogdanov, Slavko, Guillot, Sebastien, Arzoumanian, Zaven, Chirenti, Cecilia, Dittmann, Alexander J., Gendreau, Keith C., Ho, Wynn C. G., Miller, M. Coleman, Morsink, Sharon M., Wadiasingh, Zorawar, Wolff, Michael T.

Posterior sample files associated with the preprint "The Radius of PSR J0740+6620 from NICER with NICER background estimates" by Salmi et al. (2022; arXiv:2209.12840 [astro-ph.HE], accepted for publication in ApJ). Also included are: the data products; the numeric model files including th

Uploaded on October 6, 2022

September 4, 2022 (v1.0.0) Dataset Open Access View

**A rapidly evolving high-amplitude  $\delta$  Scuti star crossing the Hertzsprung Gap**

Niu, Jia-Shu, Xue, Hui-Fang.

The datasets includes the work folder which we used in the paper to construct the evolutionary models based on MESA (r15140).

Uploaded on October 1, 2022

1 more version(s) exist for this record

September 21, 2022 (v1.0.0) Software Open Access View

**Evidence of Long-Term Period Variations in the Exoplanet Transit Database (ETD)**

Hagey, Simone.

New upload

Community

**AAS PUBLISHING**

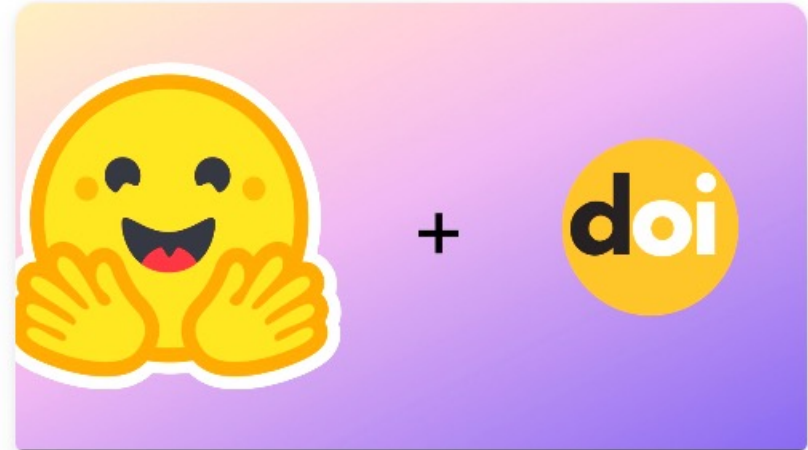
**AAS Journals**

Collection of records for replication or related materials for AAS Journal (ApJ, AJ, ApJL, ApJS, PSJ, RNAAS) articles.

More information at <https://journals.aas.org> and <https://github.com/AASJournals/Tutorials/tree/master/Repositories>

Curated by: augstmuench

Curator policy: Demonstration of relationship to submitted AAS Journal (ApJ, AJ, ApJL, ApJS, PSJ, RNAAS) manuscripts submitted to the AAS Journals via



## Introducing DOI: the Digital Object Identifier to Datasets and Models

By sylvestre • Oct 7, 2022

## What are the Authoring Experiences using dataset DOIs?

### 1. A FULL PROVENANCE EXAMPLE

*Yes, this example is to remind you of Data Origin!*

We obtained optical astrometric data of these sources from the Gaia (Gaia Collaboration et al. 2016) Early Data Release 3 (Gaia Collaboration et al. 2021) via the CDS/Vizier version (Gaia collaboration 2020a) of this catalog that was accessed using `astroquery` (Ginsburg et al. 2019) Version 0.4.5 (Ginsburg et al. 2021).

## 2. IRSA

*There are many copies of data, but I refuse to call them "mirrors" because provenance matters, and well, they are rarely pure mirrors:*

We obtained optical astrometric data of these sources from the Gaia (Gaia Collaboration et al. 2016) Early Data Release 3 (Gaia Collaboration et al. 2021) that was accessed using IRSA/Gator and the IRSA version (Gaia collaboration 2020b) of the EDR3 catalog.



### 3. ESA

*This is a little bit of a slam, but its critical to note that the quality of DOI metadata is a huge potential barrier for authors and well, everyone else:*

If I use the ESA data directly, and try cite the ESA dataset DOI for Gaia EDR3 then this is what an author would produce: (No Author 2020). This is because there is no author information in the ESA DOI Crossref metadata. There may be author information in the ESA HTML landing page (in HTML text or schema.org tags) but **we feel that these alternatives do not matter if that information is not in Crossref/Datacite!**

Adding a note from Christophe Arviset's talk this morning:

In a future where embargo periods are driven to zero, authors may be asked to cite proposals directly! It is worth knowing if citing a proposal abstract based on its current format in ADS, e.g., (Newman et al. 2015) (sans "DOI") or citing the ESA version of that same abstract, e.g., (No Author n.d.) that has been given a DOI is what authors should do – See Mark Parsons's talk in a few minutes.

#### 4. MAST (AS DATA LINKS OR CITATIONS)

*First we look at a collection dataset DOI; MAST has requested these not be cited and there's '(almost?) no reason they should ever be indexed.*

All the data obtained and used in this research can be found in MAST:  
[doi:10.17909/2w6t-bn31](https://doi.org/10.17909/2w6t-bn31).

*Next we have a direct citation of an HLSP dataset. Again, metadata matters: I'm not sure Pontoppidan & Gordon are the sole persons to receive attribution for these data but they are the only persons listed in the DataCite metadata.*

In addition we used the High-level Science Products available for the JWST Early Release Observations ([Pontoppidan & Gordon 2022](#)) available on MAST.

*MAST is also procuring dataset DOIs for HLSP datasets that will be released in the future and are derived from a particular article. This is what we ask authors to do with those:*

The CLEAR dataset created in this research will be made available as a High-Level Science Project on MAST [doi:10.17909/9cjs-wy94](https://doi.org/10.17909/9cjs-wy94) (Simons, Raymond et al. 2023).

## 6. NASA EXOPLANET ARCHIVE: SOMETIMES THINGS CHANGE

*Authors experience a lot of confusion when datasets disappear even if an archive tries to be clear about these transitions. Consider the "Confirmed Planets Table" at the NASA Exoplanet Archive:*

The number of confirmed planets in each system was additionally determined by searching the NASA Exoplanet Archive Confirmed Planets Table ([NASA Exoplanet Archive 2019](#)). 8 out of the 31 systems have at least one confirmed planet.

*Except that this reference did not exist at the date of submission! We can try to "fix" this, although this assertion really must have a date of access!*

The number of confirmed planets in each system was additionally determined by searching the NASA Exoplanet Archive Planetary Systems Table ([NASA Exoplanet Science Institute 2020](#)). 8 out of the 31 systems have at least one confirmed planet.

## Residual Questions

- Mirrors? Are they really mirrors? (no)
- Fix your metadata 🙏
- Versions mean lots and lots of different things!
- Science platforms: will they produce bit-wise reproducible datasets? With DOIs?

## REFERENCES

- Gaia collaboration. 2020a, Gaia EDR3, Centre de Donnees Strasbourg (CDS), doi: [10.26093/cds/vizier.1350](https://doi.org/10.26093/cds/vizier.1350)
- . 2020b, Gaia Source Catalogue eDR3, IPAC, doi: [10.26131/IRSA541](https://doi.org/10.26131/IRSA541)
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1, doi: [10.1051/0004-6361/201629272](https://doi.org/10.1051/0004-6361/201629272)
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, 649, A1, doi: [10.1051/0004-6361/202039657](https://doi.org/10.1051/0004-6361/202039657)
- Ginsburg, A., Sipócz, B. M., Brasseur, C. E., et al. 2019, *AJ*, 157, 98, doi: [10.3847/1538-3881/aafc33](https://doi.org/10.3847/1538-3881/aafc33)
- Ginsburg, A., Sipócz, B., Parikh, M., et al. 2021, *astropy/astroquery*, v0.4.5 [Software], Zenodo, doi: [10.5281/zenodo.5804082](https://doi.org/10.5281/zenodo.5804082)
- NASA Exoplanet Archive. 2019, Confirmed Planets Table, IPAC, doi: [10.26133/NEA1](https://doi.org/10.26133/NEA1)
- NASA Exoplanet Science Institute. 2020, Planetary Systems Table, IPAC, doi: [10.26133/NEA12](https://doi.org/10.26133/NEA12)
- Newman, A. B., Belli, S., & Ellis, R. S. 2015, Early Quiescent Galaxies Under the Magnifying Glass, HST Proposal. Cycle 23, ID. #14205
- No Author. 2020, Gaia EDR3, European Space Agency, doi: [10.5270/esa-1ugzkg7](https://doi.org/10.5270/esa-1ugzkg7)
- . n.d., \_14205, European Space Agency, doi: [10.5270/esa-kkv6wdg](https://doi.org/10.5270/esa-kkv6wdg)
- Pontoppidan, K., & Gordon, K. 2022, Data from the JWST-ERO program, STScI/MAST, doi: [10.17909/67FT-NB86](https://doi.org/10.17909/67FT-NB86)
- Simons, Raymond, Papovich, Casey, & Momcheva, Ivelina. 2023, Data from the CANDELS Lyman Alpha Emission at Reionization Survey (CLEAR), STScI/MAST, doi: [10.17909/9CJS-WY94](https://doi.org/10.17909/9CJS-WY94)

Extra slides

## 5. CHINA-VO (FROM BEGINNING TO END)

At submission (provided to peer-reviewer/data editors):

The data of this paper, including the two spectral catalogs and the spectra of the identified objects (GPQs/contaminants), are available on the China-VO PaperData Repository: <https://nadc.china-vo.org/res/r101095/>.

Post-acceptance (either the authors' original files or AAS-curated versions):

The data of this paper, including the two spectral catalogs and the spectra of the identified objects (GPQs/contaminants), are available on the China-VO PaperData Repository: [doi:10.12149/101096](https://doi.org/10.12149/101096) (Version 1).



# Data Editing

## Working with authors to standardize and document the data in their articles

- Three data editors (first in 2000; me in 2014; 3<sup>rd</sup> editor starts in 2 weeks)
- Review article submissions during peer review;
- Curate data & figures after acceptance;
- Support helpdesk; AASTeX.