

# Data Citation and Linking

*Alberto Accomazzi*

[aaccomazzi@cfa.harvard.edu](mailto:aaccomazzi@cfa.harvard.edu) | [@aaccomazzi](https://twitter.com/aaccomazzi)

NASA Astrophysics Data System | [@adsabs](https://twitter.com/adsabs) | <https://ui.adsabs.harvard.edu>

IVOA Interop Meeting | 10/18/2022



# Literature-Data Integration in ADS

- DATA
  - SIMBAD 18
  - CDS 10
  - NExScI 9
  - ESO 4
  - MAST 3
  - KOA 2
  - Spitzer 2
  - IRSA
- SIMBAD OBJECTS
  - Other 19
    - K2-18b 19
    - K2-3b 7
    - K2-3d 6
    - K2-3c 5
    - K2-9b 5
- Star 18
- Galaxy 1
- Nebula 1

ads Feedback ORCID About Account

QUICK FIELD: [Author](#) [First Author](#) [Abstract](#) [All Search Terms](#)

[← Start New Search](#)

Your search returned **25** results

[Date](#) [Export](#) [Explore](#)

[Show highlights](#) [Show abstracts](#) [Hide Sidebars](#) [Go To Bottom](#) [Add papers to library](#)

**AUTHORS**

- Benneke, B 8
- Crossfield, I 7
- Livingston, J 7
- Dressing, C 6
- Howard, A 6

**COLLECTIONS**

- astronomy 25

**REFEREED**

- refereed 20
- non-refereed 5

**AFFILIATIONS**

**KEYWORDS**

**PUBLICATIONS**

**BIB GROUPS**

2019arXiv190904642B 2019/09  
**Water Vapor on the Habitable-Zone Exoplanet K2-18b**  
Benneke, Björn; Wong, Ian; Piaulet, Caroline; Knutson, Heather A.  
*and 11 more*

2019AJ....157..242E 2019/06:cited: 2  
**An Updated Study of Potential Targets for Ariel**  
Edwards, Billy; Mugnai, Lorenzo; Tinetti, Giovanna; Pascale, Enzo  
*and 1 more*

2019AJ....157..211M 2019/05:cited: 3  
**Detecting Unresolved Binaries in TESS Data with Speckle Imaging**  
Matson, Rachel A.; Howell, Steve B.; Ciardi, David R.

2019AJ....157..174O 2019/05:cited: 4  
**Discovery of a Third Transiting Planet in the Kepler-47 Circumbinary System**  
Orosz, Jerome A.; Welsh, William F.; Haghighipour, Nader; Quarles, Billy *and 15 more*

**Years** Citations Reads

■ refereed ■ non refereed

Year	referenced	non refereed	Total
2015	1	3	4
2016	2	1	3
2017	4	0	4
2018	7	0	7
2019	6	1	7

Limit results to papers from  to

# FAIRness via Archival Data Linking

Links to 5 archives with data!

Links to NED and SIMBAD!

Links to Vizier Tables!

Links to paper plots and images!

The screenshot shows the ADS interface for a search result. The search bar contains the query: `data:(mast irsa ned chandra) bibstem:apj property:associated`. The search results show the paper: "The Black Hole in the Most Massive Ultracompact Dwarf Galaxy M59-UCD3". The authors listed are Ahn, Christopher P.; Seth, Anil C.; Cappellari, Michele; Krajnović, Davor; Strader, Jay; Voggel, Karina T.; Walsh, Jonelle L.; Bahramian, Arash; Baumgardt, Holger; Brodie, Jean; Chilingarian, Igor; Chomiuk, Laura; den Brok, Mark; Frank, Matthias; Hilker, Michael; McDermid, Richard M.; Mieske, Steffen; Neumayer, Nadine; Nguyen, Dieu D.; Pechetti, Renuka ; ...

The abstract text reads: "We examine the internal properties of the most massive ultracompact dwarf galaxy (UCD), M59-UCD3, by combining adaptive-optics-assisted near-IR integral field spectroscopy from Gemini/NIFS and Hubble Space Telescope (HST) imaging. We use the multiband HST imaging to create a mass model that suggests and accounts for the presence of multiple stellar populations and structural components. We combine these mass models with kinematics measurements from Gemini/NIFS to find a best-fit stellar mass-to-light ratio (M/L) and black hole (BH) mass using Jeans anisotropic models (JAMs), axisymmetric Schwarzschild models, and triaxial Schwarzschild models. The best-fit parameters in the JAM and axisymmetric Schwarzschild models have BHs between 2.5 and 5.9 million solar masses. The triaxial Schwarzschild models point toward a similar BH mass but show a minimum  $\chi^2$  at a BH mass of  $-0.1$ . Models with a BH in all three techniques provide better fits to the central  $V_{ms}$  profiles, and thus we estimate the BH mass to be  $(4.2)_{-1.7}^{+2.1} \times 10^6 M_{\odot}$  (estimated  $1\sigma$  uncertainties). We also present deep radio imaging of M59-UCD3 and two other UCDS in Virgo with dynamical BH mass measurements, and we compare these to X-ray measurements to check for consistency with the fundamental plane of BH accretion. We detect faint radio emission in M59cO but find only upper limits for M60-UCD1 and M59-UCD3 despite X-ray detections in both these sources. The BH mass and nuclear light profile of M59-UCD3 suggest that it is the tidally stripped remnant of a  $10^8$  to  $10^9 M_{\odot}$  galaxy."

On the right side of the interface, there are several sections highlighted with red circles:

- DATA PRODUCTS**: A list of data products including SIMBAD (8), NED (1), MAST (1), IRSA (1), ESA (1), Chandra (1), and CDS (1).
- GRAPHICS**: A section containing a visualization of the galaxy M59-UCD3 with a red circle around it and the text "Click to view more".
- ASSOCIATED WORKS (2)**: A section containing links for "Catalog Description" and "Source Paper".

# Example of Data Mention: Scolnic et al, ApJ 859, 101

ApJ paper has DOI 10.17909/t95q4x  
linked under “Article data” tab.

DOI 10.17909/t95q4x is mentioned 7  
times in HTML and PDF document:

- Narrative (3 times)
- Table captions (3 times)
- Appendix A (data & code availability, 1 time)

However, there is no citation for it

doi:10.17909/t95q4x, a subset of data files, coding routines, and supplementary tables to replicate this analysis. This includes the following:'. A list of 9 items follows, detailing the data and code provided. A table of contents on the right side of the page lists sections from 'Abstract' to 'References'."/>

THE ASTROPHYSICAL JOURNAL

The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample

D. M. Scolnic<sup>1,21</sup>, D. O. Jones<sup>2</sup>, A. Rest<sup>2,3</sup>, Y. C. Pan<sup>4</sup>, R. Chornock<sup>5</sup>, R. J. Foley<sup>4</sup>, M. E. Huber<sup>6</sup>, R. Kessler<sup>1</sup>, G. Narayan<sup>3</sup>, A. G. Riess<sup>3,2</sup> + Show full author list

Published 2018 May 29 • © 2018. The American Astronomical Society. All rights reserved.

[The Astrophysical Journal, Volume 859, Number 2](#)

Article PDF Article ePub

Figures Tables References Article data

External repository

MAST dataset

What is article data?

Figures Tables References Article data PDF ePub

Appendix A: Data Tables and Code Repository

Upon publication, we will release [doi:10.17909/t95q4x](https://doi.org/10.17909/t95q4x), a subset of data files, coding routines, and supplementary tables to replicate this analysis. This includes the following:

1. A table of the spectroscopic observations of each SN in the PS1 sample that includes their ID, date of observation, telescope observed and measured redshift. A shortened version is included below in Table 16.
2. A table of key recovered parameters from the light-curve fits for the full Pantheon sample. A shortened version of this is shown below in Table 16. We also include a full output table from the SNANA fitter of a thorough listing of fitted parameters and other properties of the light curves. Final redshifts and distances are also given—a shortened version is shown in Table 17.
3. A table of binned distance estimates over redshift for a compressed version of the data set.
4. A full systematic covariance matrix for the binned and unbinned versions.
5. Stellar catalogs of the MD fields.
6. Necessary files to use with the CosmoMC or CosmoSIS software with instructions.
7. A folder of all the SNANA set-up scripts to fit each sample. A folder of all the SNANA set-up scripts to simulate each sample.
8. Output tables for 30 simulated samples used to test external methods and perform null tests on this data set.
9. Code for remaking all figures in this paper.

Abstract

1. Introduction
2. The PS1 Search, Photometry, and Calibration Pipeline
3. PS1 Light-curve Fitting and Simulation
4. Combining Multiple SN Samples
5. Analysis Framework
6. Results
7. Discussion
8. Conclusion

Appendix A: Data Tables and Code Repository

Appendix B: Template Construction

Appendix C: Low-z Simulations

Footnotes

References

# What does ADS know about 10.17909/t95q4x?

The screenshot shows the ADS search interface. At the top, there is a search bar with the text "doi:10.17909/t95q4x" entered. Below the search bar, it says "Your search returned 0 results". The search bar and the "0 results" text are circled in red. On the left side, there is a navigation menu with categories like AUTHORS, COLLECTIONS, REFEREED, AFFILIATIONS, KEYWORDS, PUBLICATIONS, BIB GROUPS, SIMBAD OBJECTS, NED OBJECTS, DATA, VIZIER TABLES, and PUBLICATION TYPE. On the right side, there are buttons for "Show highlights", "Show abstracts", and "Hide Sidebars", along with a "Go To Bottom" link. Below these buttons, a message states "Sorry no results were found for doi:10.17909/t95q4x" and provides suggestions: "Try broadening your search", "Disable any filters that may be applied", "Check out some examples", and "Read our help pages". At the bottom, there is a "Leave Feedback" button.

The record is not **indexed** in ADS...

# What does ADS know about 10.17909/t95q4x?

The screenshot shows the ADS search interface. The search bar contains the query 'doi:10.17909/t95q4x'. Below the search bar, it states 'Your search returned 0 results'. On the left sidebar, there are various filters such as 'AUTHORS', 'COLLECTIONS', 'REFEREED', 'AFFILIATIONS', 'KEYWORDS', 'PUBLICATIONS', 'BIB GROUPS', 'SIMBAD OBJECTS', 'NED OBJECTS', 'DATA', 'VIZIER TABLES', and 'PUBLICATION TYPE'. The main content area displays a message: 'Sorry no results were found for doi:10.17909/t95q4x'. Below this message, there are suggestions: 'Try broadening your search', 'Disable any filters that may be applied', 'Check out some examples', and 'Read our help pages'. At the bottom, there is a button that says 'Leave Feedback'.

The screenshot shows the ADS search interface with a search for 'full:10.17909/t95q4x'. The search bar is circled in red. Below the search bar, it states 'Your search returned 18 results'. The left sidebar shows filters for 'AUTHORS', 'COLLECTIONS', 'REFEREED', 'AFFILIATIONS', 'KEYWORDS', 'PUBLICATIONS', 'BIB GROUPS', 'SIMBAD OBJECTS', 'NED OBJECTS', 'DATA', 'VIZIER TABLES', and 'PUBLICATION TYPE'. The main content area displays a list of search results. The first result is 'Revisiting a non-parametric reconstruction of the deceleration parameter from observational data' by Mukherjee, Purba; Banerjee, Narayan. The second result is 'Non-parametric reconstruction of the cosmological  $w$  parameter' by Mukherjee, Purba; Banerjee, Narayan. The third result is 'Testing Modified Gravity (MOG) theory and dark matter model in Milky Way using the local observables' by Davari, Zahra; Rahvar, Sohrab. The fourth result is 'Constraints on Newton's constant from cosmological observations' by Wang, Ke; Chen, Lu. The search results are displayed in a list format with various icons and links.

The record is not indexed in ADS...  
...but is found as a **mention** in fulltext

# Example of Data Citation: Gonzales et al, ApJ 938, 56

ApJ paper cites two catalogs in ADS,  
[2012yCat.2311....0C](#) (Wise) and  
[2003tmc..book....C](#) (2MASS PSC)

Both cited data catalogs are indexed in  
ADS, and can be found in the  
reference list as linked records

Note: there are DOIs assigned to each one of  
them, although citations do not provide them

THE ASTROPHYSICAL JOURNAL

OPEN ACCESS  
A Comparative L-dwarf Sample Exploring the Interplay between Atmospheric Assumptions and Data Properties

Eileen C. Gonzales<sup>1,2</sup>, Ben Burningham<sup>3</sup>, Jacqueline K. Faherty<sup>2</sup>, Nikole K. Lewis<sup>1</sup>, Channon Visscher<sup>4,5</sup>, and Mark Marley<sup>6</sup>

Published 2022 October 13 • © 2022. The Author(s). Published by the American Astronomical Society.

[The Astrophysical Journal, Volume 938, Number 1](#)  
Citation Eileen C. Gonzales et al 2022 ApJ 938 56

Article PDF Article ePub

Figures Tables References

- ↑ Cushing M. C., Saumon D. and Marley M. S. 2010 *AJ* 140 1428  
Go to reference in article ADS Crossref Google Scholar
- ↑ Cutri R. M., Wright E. L., Conrow T. et al. 2012 *II/311 VizieR On-line Data Catalog*:  
Go to reference in article ADS Google Scholar
- ↑ Cutri R. M., Skrutskie M. F., van Dyk S. et al. 2003 *2MASS All Sky Catalog of Point Sources*  
Go to reference in article ADS Google Scholar
- ↑ ...

Abstract

1. Introduction
2. Literature Data on Sample
3. Fundamental Parameters and Overall SED Comparison of the Sample
4. The Brewster Retrieval Framework
5. Updating the J1416+1348A Retrieval
6. Model Selection
7. Retrieval Results
8. Discussion
9. Conclusions

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our [Privacy and Cookies](#) policy.

# What is indexed, linked in ADS

## Indexed (an actual database record, searchable)

- The scholarly literature of interest to Astronomers
- VizieR records, IVOA standards, observing and funding proposals
- Software products: ASCL records, software packages cited via DOI
- Soon: cited data products, other research objects such as notebooks

**Indexed records are scholarly research objects.**

**They are discoverable and citable via ADS, and their metrics are tracked**

## Linked (resource accessible from a record via a link)

- Data Products hosted by external collaborators (Archives, SIMBAD, NED)

**Linked data collections can be used as a filter in ADS,  
and to evaluate impact of linked data products**



# Data: Mention vs Citation, Ongoing efforts

## Mention

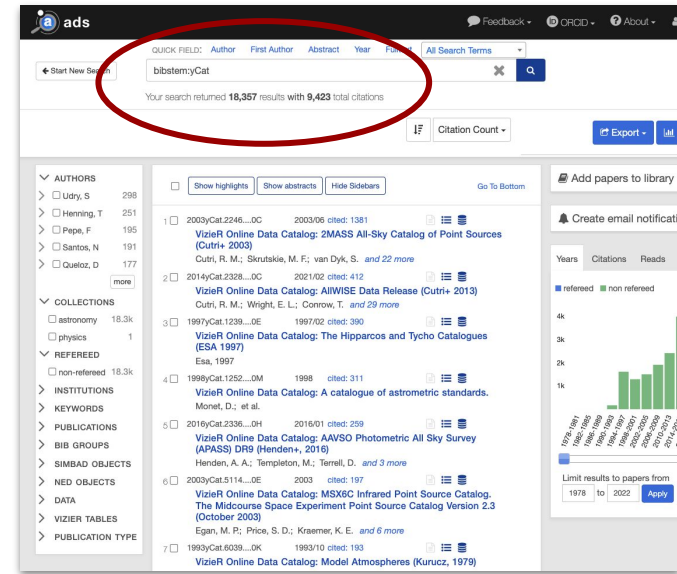
## Citation

	Free text or URL	DOI	Free text or URL	DOI
Use cases	<p>in Data Avail. Section:</p> <p>This paper uses data from the VIMOS Public Extragalactic Redshift Survey (VIPERS). VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at (<a href="http://vipers.inaf.it">http://vipers.inaf.it</a>).</p>	<p>in Data Avail. Section:</p> <p>The HST data used in this Letter are available as part of the MAST archive<sup>6</sup> and can be accessed at doi:<a href="https://doi.org/10.17909/t9-3tsk-qh26">10.17909/t9-3tsk-qh26</a>.</p>	<p>in references:</p> <p>GPy since, 2012, GPy: A Gaussian Process Framework in Python. <a href="http://github.com/SheffieldML/GPy">http://github.com/SheffieldML/GPy</a></p>	<p>in references:</p> <p>Thyagarajan, N., Harish, S., Kolopanis, M., Murray, S., &amp; Jacobs, D. 2020, Precision Radio Interferometry Simulator (PRISim), v2.2.1, Zenodo, doi:10.5281/zenodo.3892099</p>
in ADS	Searchable in ADS fulltext / ack fields	Searchable in ADS fulltext / ack fields; <a href="#">Link to data product created in ADS</a>	URL ignored, citation may not get counted unless matched via heuristics	<a href="#">Record for corresponding DOI created(*)/identified and citation counted</a>
outside of astro	Ignored so far	Ignored so far	URL ignored(?) citation not counted	Citation counted by CrossRef / DataCite

# What's the difference

## Indexed Dataset

- Data is accessible from paper via citation and data link
- Dataset has higher level of discoverability (retrieved by ADS search)
- Dataset has ADS metrics associated with it



The screenshot shows the ADS search interface. The search bar contains 'bibstemyCat' and is circled in red. Below the search bar, it states 'Your search returned 18,357 results with 9,423 total citations'. The results list includes several datasets with their respective citation counts and links to more information. A sidebar on the right shows a bar chart of citations over time and a table of results.

Item	Year	Citations
1	2003/06	1381
2	2021/02	412
3	1997/02	390
4	1998	311
5	2016/01	259
6	2003	197
7	1993/10	193

# What's the difference

## Indexed Dataset

- Data is accessible from paper via citation and data link
- Dataset has higher level of discoverability (retrieved by ADS search)
- Dataset has ADS metrics associated with it

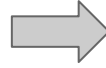
## Linked Dataset

- Data is accessible from paper via data link
- Associated papers typically part of a linked data collection (e.g. Chandra, IRSA, MAST)
- Metrics available via paper metrics

The image displays two screenshots of the ADS (Astrophysics Data System) search interface. The top screenshot shows a search for 'bibstemyCat' with 18,357 results and 9,423 total citations. The bottom screenshot shows a search for 'data:Chandra' with 13,702 results and 321,398 total citations, with the search results and citation count circled in red. Both screenshots show a list of search results with columns for citation count, year, and title, along with a sidebar for filters and a right-hand panel for additional actions like 'Add papers to library' and 'Create email notification'.

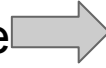
# What should happen in the (near) future

Data Archives register data products using DOIs and rich metadata which includes origin of data



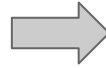
ADS (and others) can discover and index/link data products via text mining of DAS/reference sections

Data Archives provide clear instructions for acknowledgment (“cite as”/“acknowledge as”)



Authors know how to cite/mention the relevant data products when writing papers

ADS provides discovery capabilities and metrics for linked/indexed data products



Data Archives can use the ADS to discover mentions/citations of their data and related impact metrics

# Many details to still be worked out

## What should be indexed in ADS

- Curated, high-level datasets with good metadata and clear authorship information
- Research data collections that have shown reuse value (via eg. citations)

## What should not be indexed in (but linked from) ADS

- Every data product out there
- Every single version of a data product
- Temporary data collections created for convenience purpose (e.g. MAST user generated DOIs)

# For more Information

The Asclepias project, enabling software citation in Astronomy (AAS, ADS, Zenodo):

- Henneken et al 2022: [2022BAAS...54..046H](#)
- Muench et al 2020: [2020ASPC..522..711M](#)

FORCE11 Data Citation Principles:

- <https://www.force11.org/datacitationprinciples>

FAIR Data Principles:

- <https://www.force11.org/group/fairgroup/fairprinciples>