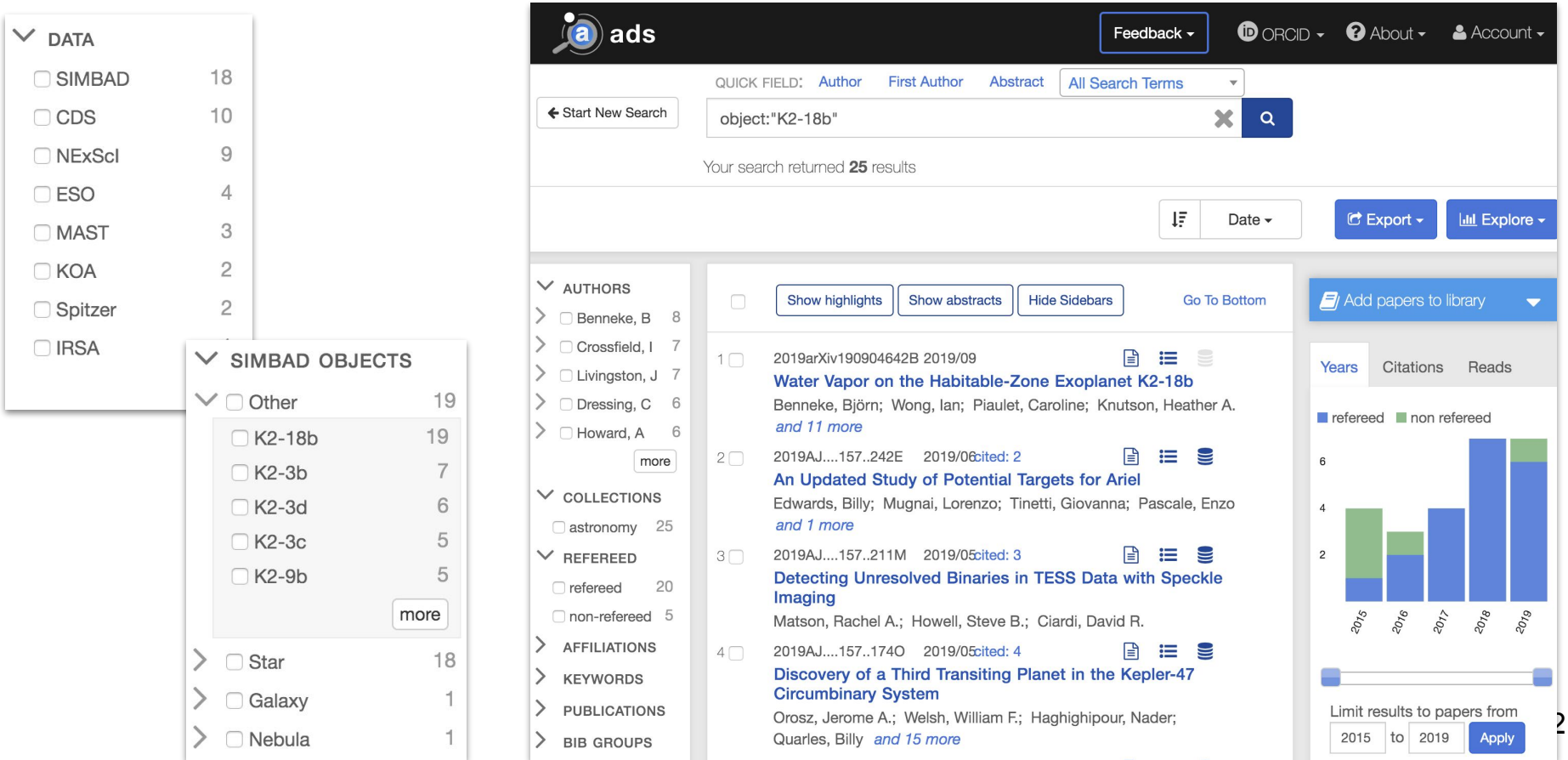
Data Citation and Linking

Alberto Accomazzi
aaccomazzi@cfa.harvard.edu | *@aaccomazzi*
NASA Astrophysics Data System | *@adsabs* | *https://ui.adsabs.harvard.edu*

IVOA Interop Meeting | 10/18/2022

Literature-Data Integration in ADS



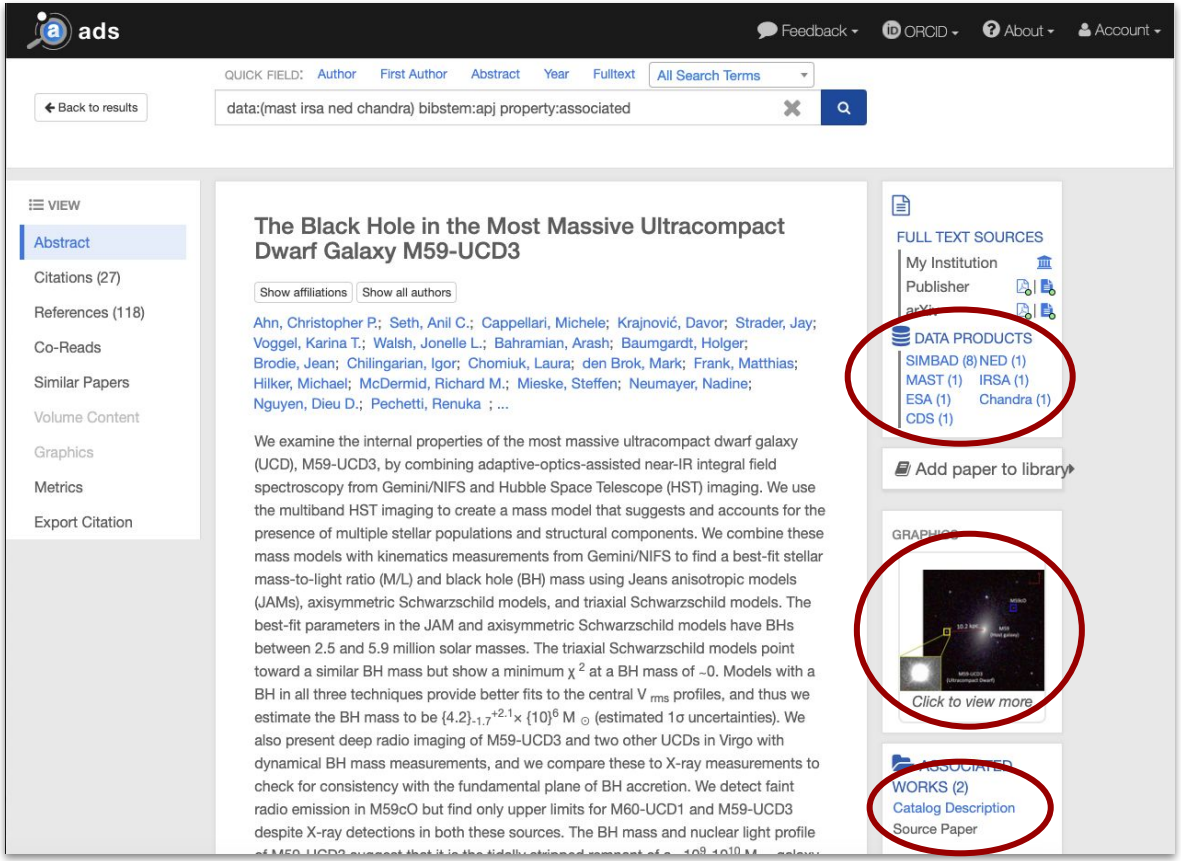
FAIRness via Archival Data Linking

Links to 5 archives with data!

Links to NED and SIMBAD!

Links to Vizier Tables!

Links to paper plots and images!



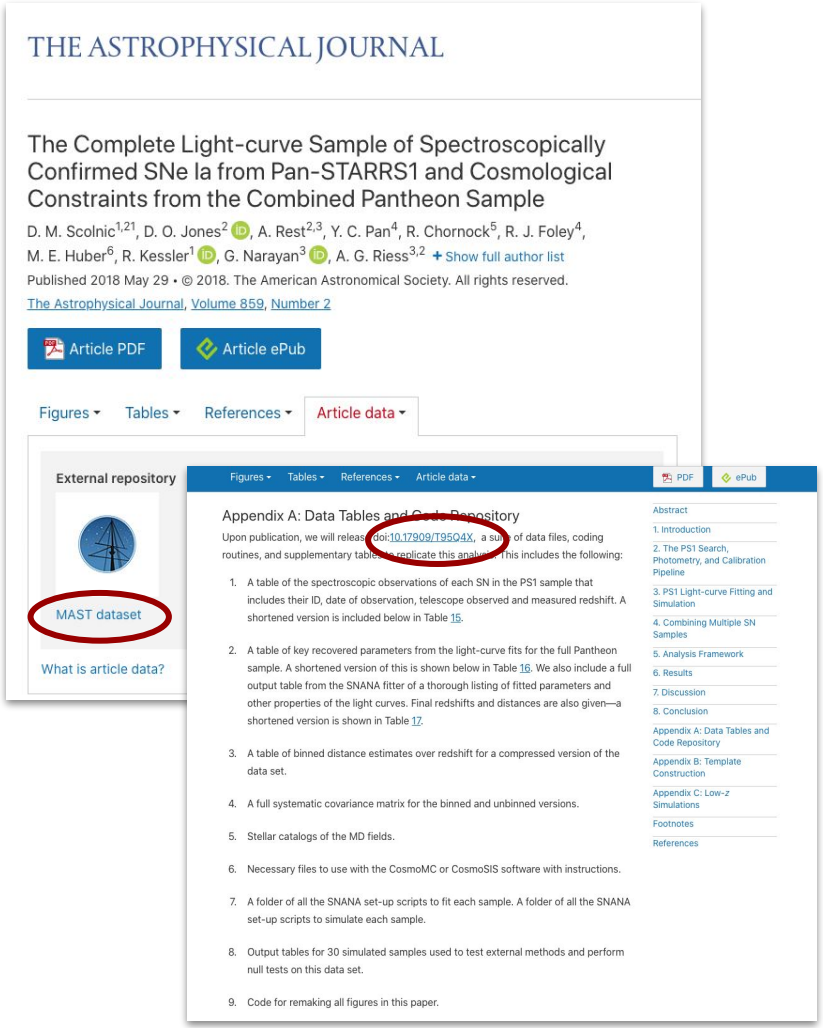
Example of Data Mention: Scolnic et al, ApJ 859, 101

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

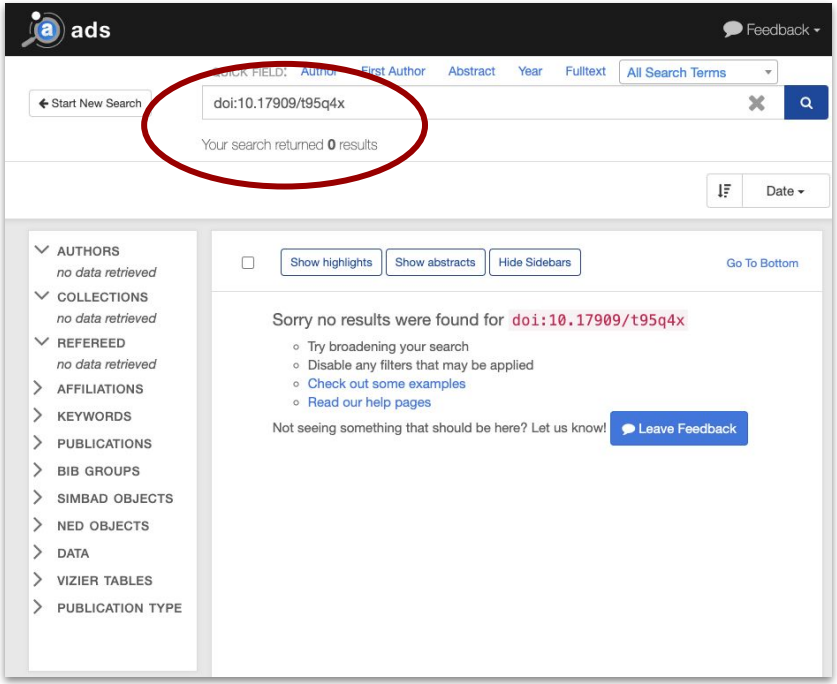
DOI 10.17909/t95q4x is mentioned 7 times in HTML and PDF document:

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

However, there is no citation for it

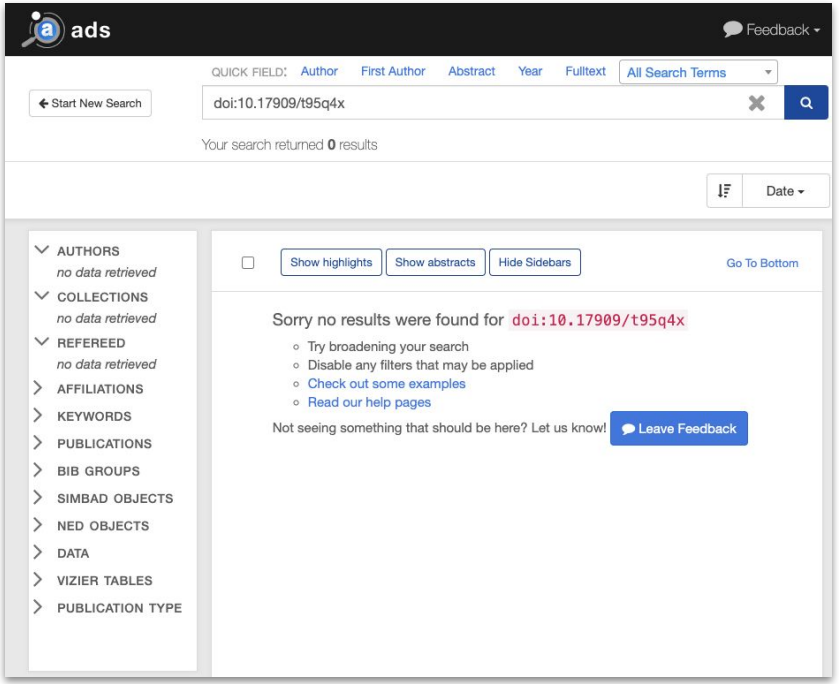


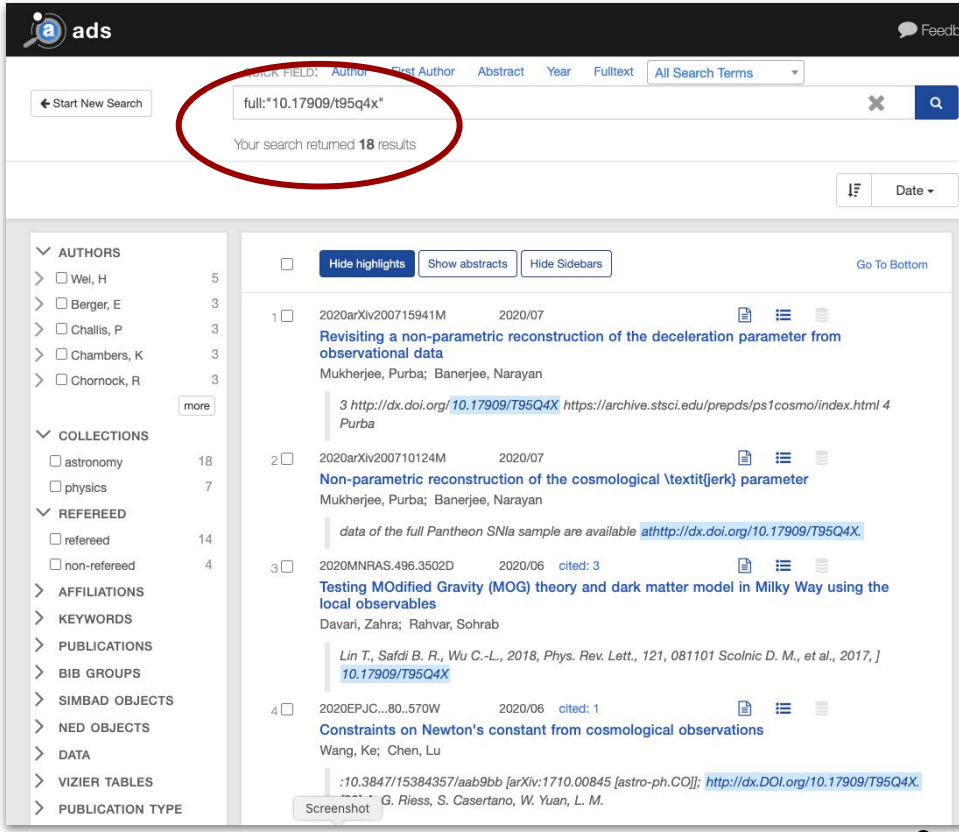
What does ADS know about 10.17909/t95q4x?



The record is not indexed in ADS...

What does ADS know about 10.17909/t95q4x?





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

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

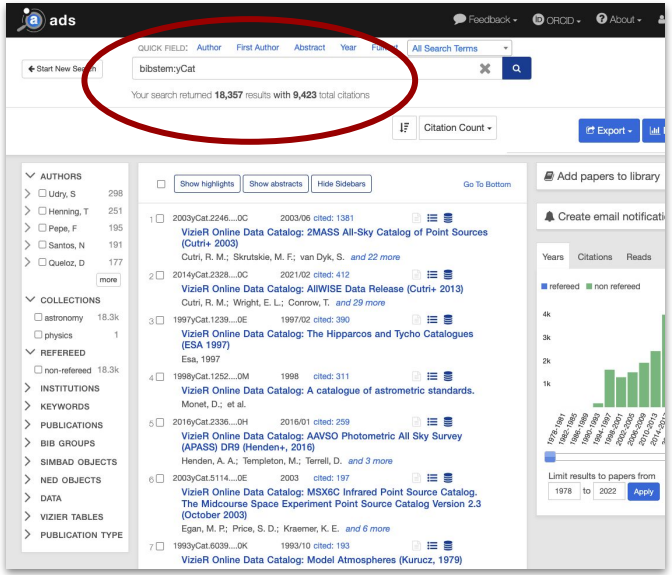
Data: Mention vs Citation, Ongoing efforts

	Mention		Citation	
	Free text or URL	DOI	Free text or URL	DOI
Use cases	in Data Avail. Section: 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 (http://vipers.inaf.it).	in Data Avail. Section: The HST data used in this Letter are available as part of the MAST archive[6] and can be accessed at doi:10.17909/t9-3tsk-qh26.	in references: GPy since, 2012, GPy: A Gaussian Process Framework in Python. http://github.com/SheffieldML/GPy	in references: Thyagarajan, N., Harish, S., Kolopanis, M., Murray, S., & Jacobs, D. 2020, Precision Radio Interferometry Simulator (PRISim), v2.2.1, Zenodo, doi:10.5281/zenodo.3892099
in ADS	Searchable in ADS fulltext / ack fields	Searchable in ADS fulltext / ack fields; Link to data product created in ADS	URL ignored, citation may not get counted unless matched via heuristics	Record for corresponding DOI created(*)/identified and citation counted
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



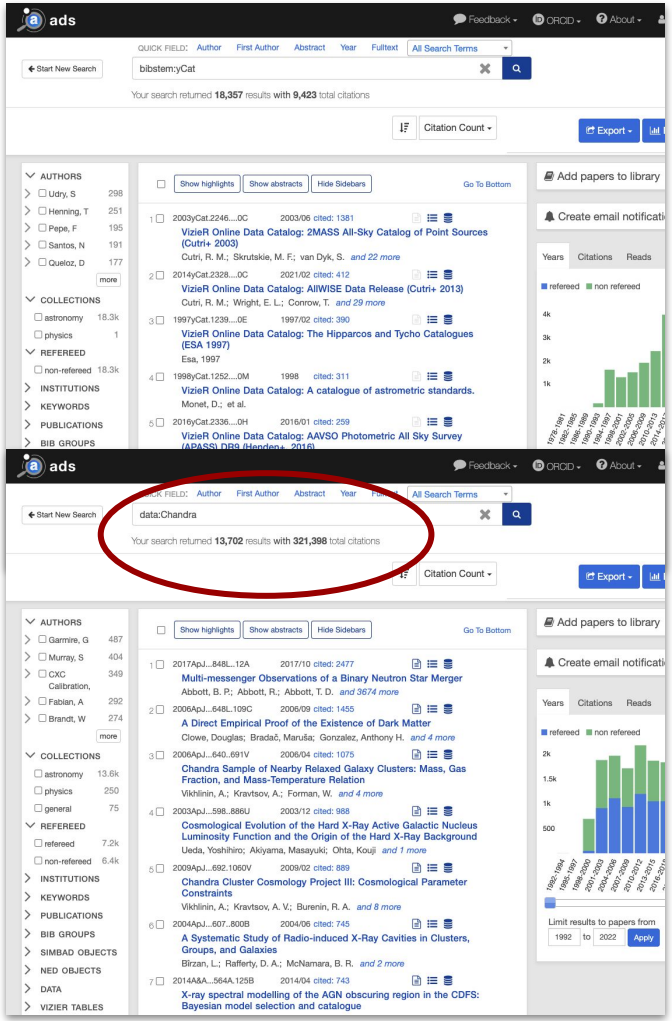
What's the difference

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



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