Linking to Data

—

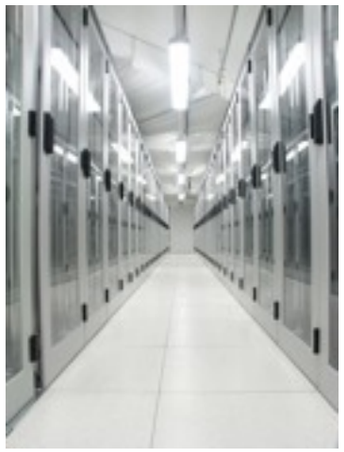
Effect on Citation Rates in Astronomy

Edwin Henneken
Smithsonian Astrophysical Observatory

Progress in Science: information/knowledge transfer



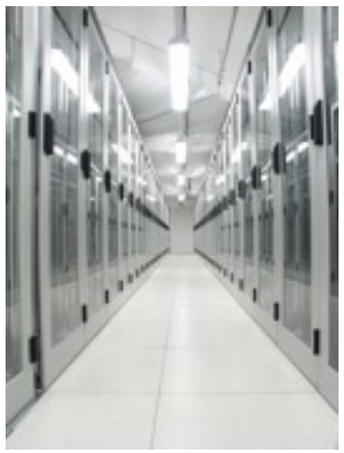




Progress in Science: information/knowledge transfer







Discoverability:
Publication & Underlying Data

Linking to Data

Effect on Citation Rates in Astronomy

305

The Literature-Data Connection: Meaning, Infrastructure and Impact

Special Session

Wednesday, May 25, 2011, 10:00 AM -11:30 AM

Staffordshire

This special session will provide a forum for data providers, publishers and scientists to review the current efforts in the curation and preservation of astronomical data products published with or analyzed in scholarly publications. Speakers will be addressing a range of issues related to the publication and preservation of data products, namely: Nomenclature: how to uniquely name and refer to all digital assets (archival and user-provided); Persistence: what underlying technology should be used to manage the resolution of identifiers (Handles, DOIs, IVOA identifiers); Preservation: where user-provided data and metadata is stored, how to implement validation, verification, migration and versioning; Policies: what are the responsibilities of all parties involved in the publication process: authors, editors, archives, projects; Dissemination: ensuring that the metadata for all digital assets collected and preserved as part of this effort are properly exposed to the VAO and IVOA community to facilitate harvesting, indexing and discovery. Invited speakers will be from leading institutions, societies and projects involved in the curation and archival of astronomical publications, data and metadata.

Linking to Data: Citation Advantage?

305

The Literature-Data Connection: Meaning, Infrastructure and Impact

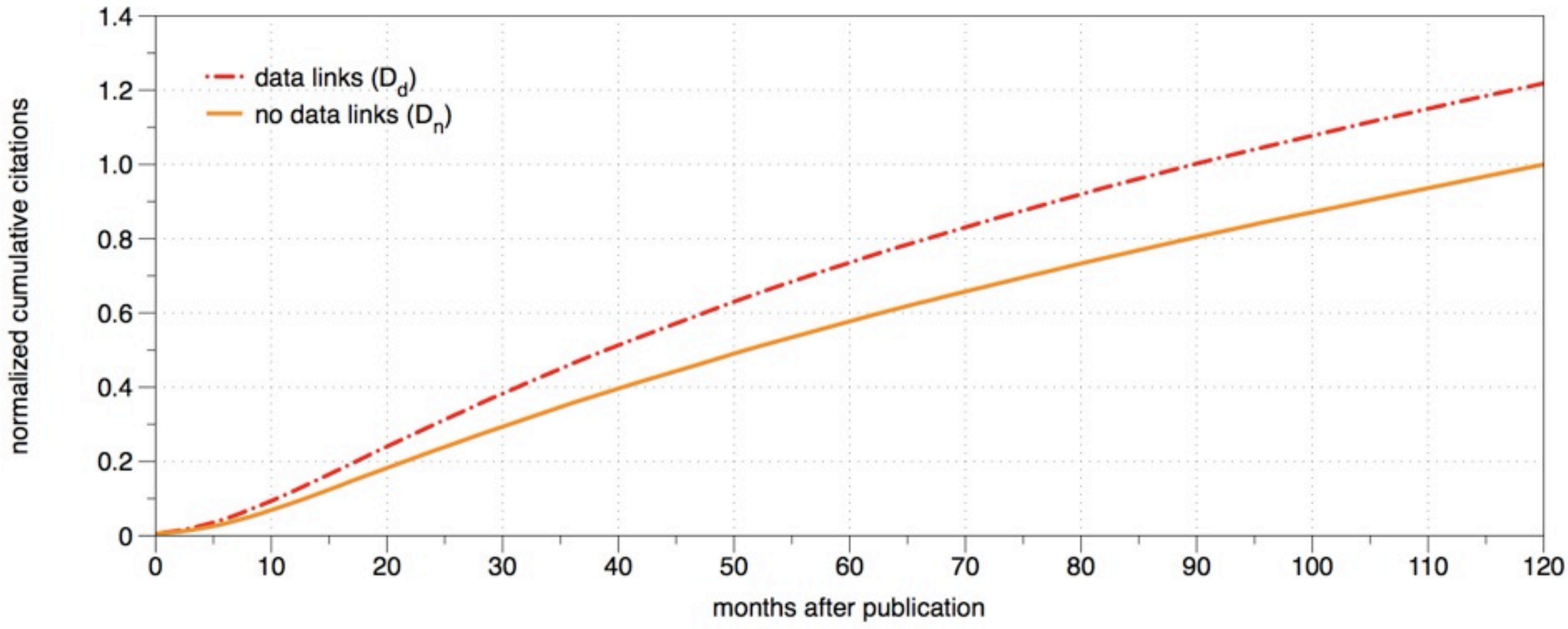
Special Session

Wednesday, May 25, 2011, 10:00 AM -11:30 AM

Staffordshire

This special session will provide a forum for data providers, publishers and scientists to review the current efforts in the curation and preservation of astronomical data products published with or analyzed in scholarly publications. Speakers will be addressing a range of issues related to the publication and preservation of data products, namely: Nomenclature: how to uniquely name and refer to all digital assets (archival and user-provided); Persistence: what underlying technology should be used to manage the resolution of identifiers (Handles, DOIs, IVOA identifiers); Preservation: where user-provided data and metadata is stored, how to implement validation, verification, migration and versioning; Policies: what are the responsibilities of all parties involved in the publication process: authors, editors, archives, projects; Dissemination: ensuring that the metadata for all digital assets collected and preserved as part of this effort are properly exposed to the VAO and IVOA community to facilitate harvesting, indexing and discovery. Invited speakers will be from leading institutions, societies and projects involved in the curation and archival of astronomical publications, data and metadata.

Linking to Data: Citation Advantage? YES



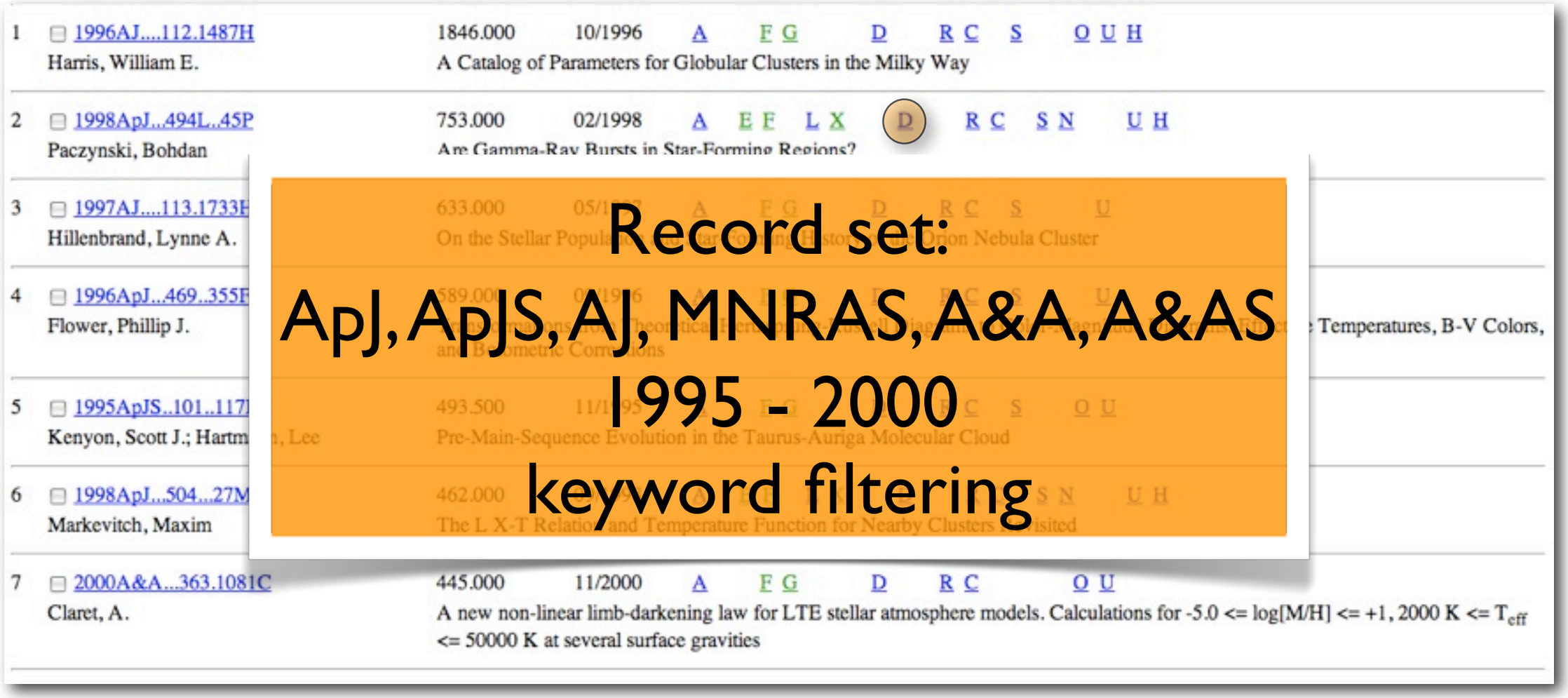
ADS holdings & citation data

1	1996AJ....112.1487H Harris, William E.	1846.000	10/1996	A F G D R C S O U H	A Catalog of Parameters for Globular Clusters in the Milky Way
2	1998ApJ...494L..45P Paczynski, Bohdan	753.000	02/1998	A E F L X D R C S N U H	Are Gamma-Ray Bursts in Star-Forming Regions?
3	1997AJ....113.1733H Hillenbrand, Lynne A.	633.000	05/1997	A F G D R C S U	On the Stellar Population and Star-Forming History of the Orion Nebula Cluster
4	1996ApJ...469..355F Flower, Phillip J.	589.000	09/1996	A F G D R C S U	Transformations from Theoretical Hertzsprung-Russell Diagrams to Color-Magnitude Diagrams: Effective Temperatures, B-V Colors, and Bolometric Corrections
5	1995ApJS..101..117K Kenyon, Scott J.; Hartmann, Lee	493.500	11/1995	A F G D R C S O U	Pre-Main-Sequence Evolution in the Taurus-Auriga Molecular Cloud
6	1998ApJ...504...27M Markevitch, Maxim	462.000	09/1998	A E F L X D R C S N U H	The L X-T Relation and Temperature Function for Nearby Clusters Revisited
7	2000A&A...363.1081C Claret, A.	445.000	11/2000	A F G D R C O U	A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for -5.0 <= log[M/H] <= +1, 2000 K <= T_{eff} <= 50000 K at several surface gravities

ADS holdings & citation data

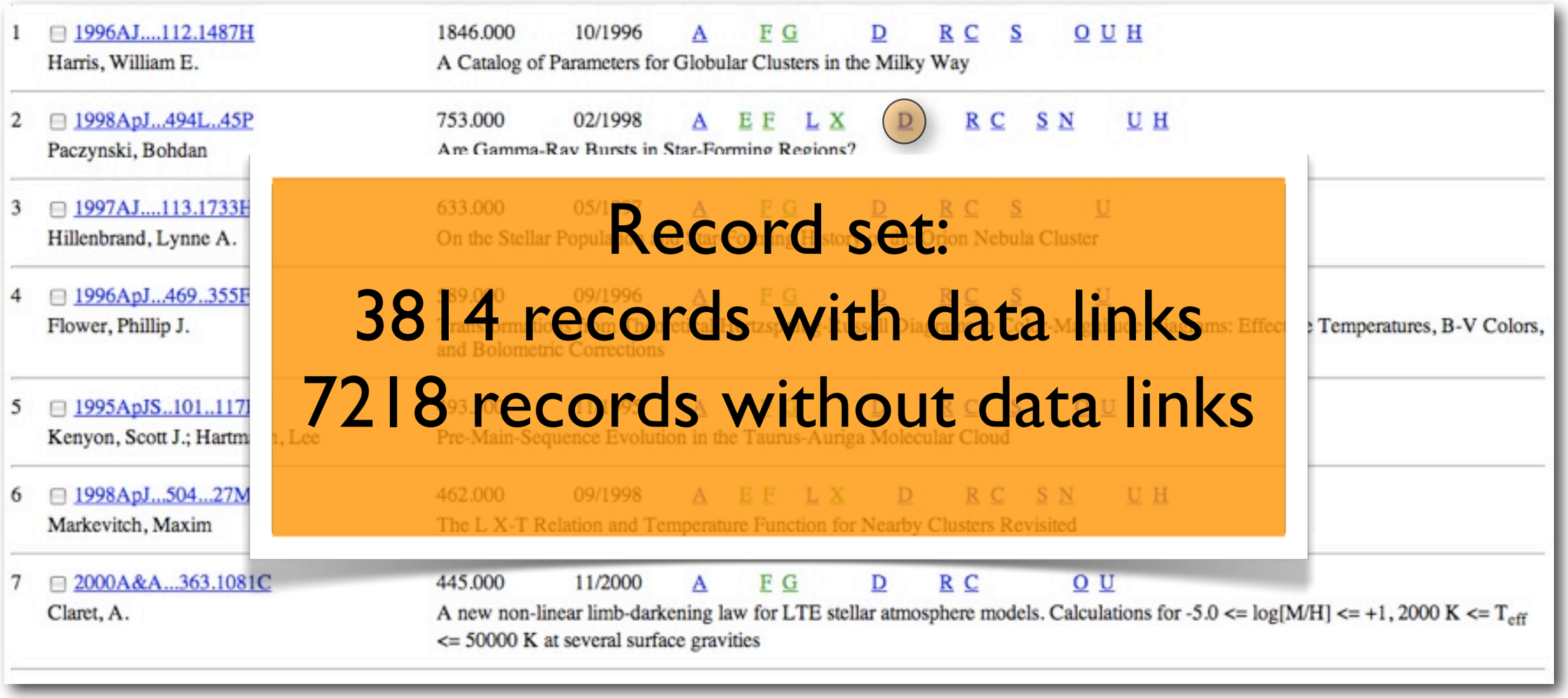
	Bibcode	Author	Score	Date	Links	Title
1	1996AJ....112.1487H	Harris, William E.	1846.000	10/1996	A F G D R C S O U H	A Catalog of Parameters for Globular Clusters in the Milky Way
2	1998ApJ...494L..45P	Paczynski, Bohdan	753.000	02/1998	A E F L X D R C S N U H	Are Gamma-Ray Bursts in Star-Forming Regions?
3	1997AJ....113.1733H	Hillenbrand, Lynne A.	633.000	05/1997	A F G D R C S U	On the Stellar Population and Star-Forming History of the Orion Nebula Cluster
4	1996ApJ...469..355F	Flower, Phillip J.	589.000	09/1996	A F G D R C S U	Transformations from Theoretical Hertzsprung-Russell Diagrams to Color-Magnitude Diagrams: Effective Temperatures, B-V Colors, and Bolometric Corrections
5	1995ApJS..101..117K	Kenyon, Scott J.; Hartmann, Lee	493.500	11/1995	A F G D R C S O U	Pre-Main-Sequence Evolution in the Taurus-Auriga Molecular Cloud
6	1998ApJ...504...27M	Markevitch, Maxim	462.000	09/1998	A E F L X D R C S N U H	The L X-T Relation and Temperature Function for Nearby Clusters Revisited
7	2000A&A...363.1081C	Claret, A.	445.000	11/2000	A F G D R C O U	A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for -5.0 <= log[M/H] <= +1, 2000 K <= T_{eff} <= 50000 K at several surface gravities

ADS holdings & citation data



Record set:
ApJ, ApJS, AJ, MNRAS, A&A, A&AS
1995 - 2000
keyword filtering

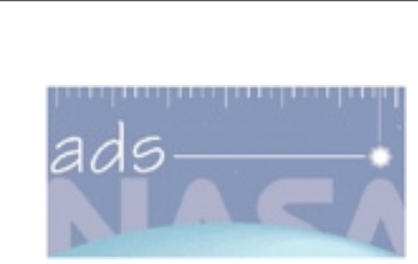
ADS holdings & citation data



Record set:

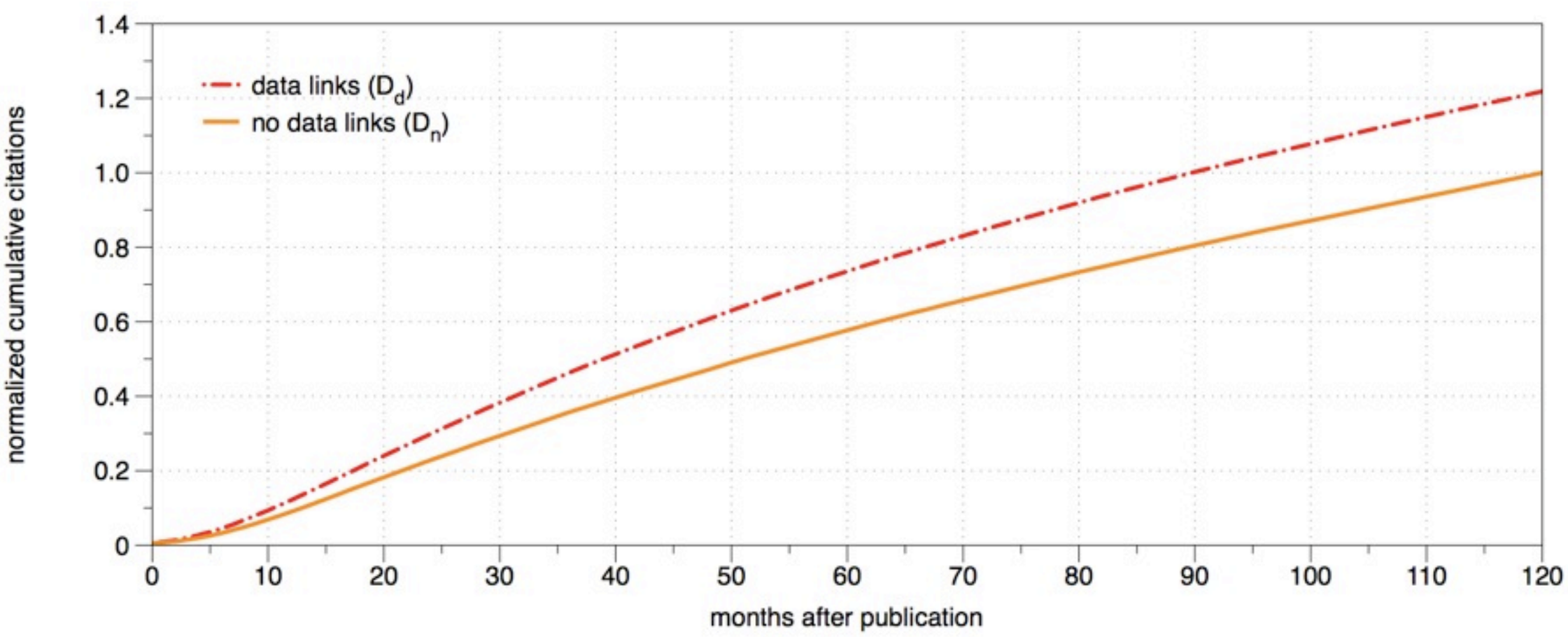
3814 records with data links

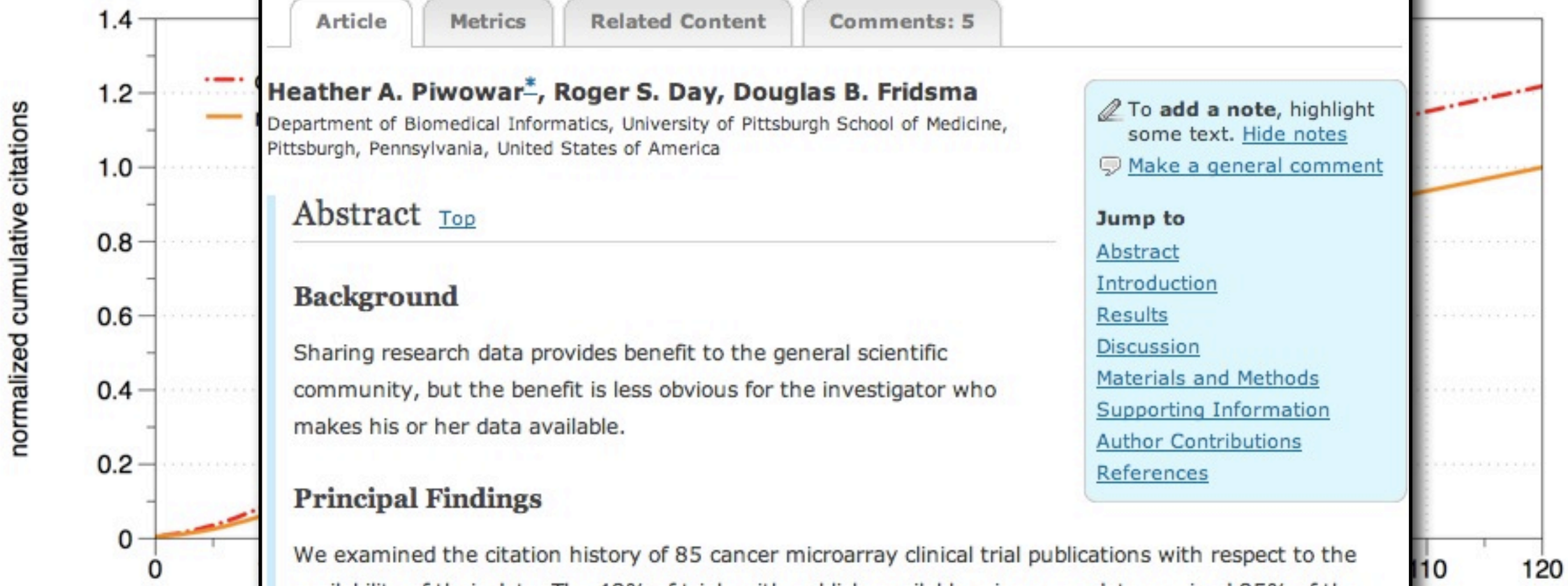
7218 records without data links



Linking to Data

Effect on Citation Rates in Astronomy





RESEARCH ARTICLE OPEN ACCESS

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Article | Metrics | Related Content | Comments: 5

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma
Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

To **add a note**, highlight some text. Hide notes
Make a general comment

Jump to
Abstract
Introduction
Results
Discussion
Materials and Methods
Supporting Information
Author Contributions
References

Abstract Top

Background

Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available.

Principal Findings

We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.

Received: December 13, 2006; **Accepted:** February 26, 2007; **Published:** March 21, 2007

Significance

This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

Could there be other systematic effects?

- E-printing, additional metadata
- "Cherry-picking"

Data sharing is a multi-discipline discussion

- "Scientific Journals can Advance Tropical Biology and Conservation by Requiring Data Archiving"
- "The Essential Nature of Sharing in Science"
- "Time for fair trade in research data"
- "Challenges and Opportunities of Open Data in Ecology"
- "Will a Data Registry Increase Professional Integrity?"
- "Reproducible Research in Signal Processing"
- "Data archiving in ecology and evolution: best practices"
- "Whose data are they anyway?"

Respectively published in:
Biotropica, *Science & Engineering Ethics*, *The Lancet*, *Science*, *Journal of Forestry*, *IEEE Signal Processing Magazine*, *Trends in Ecology & Evolution*, *British Medical Journal*

Avoid: "What's my data is mine and what's your data is also mine."

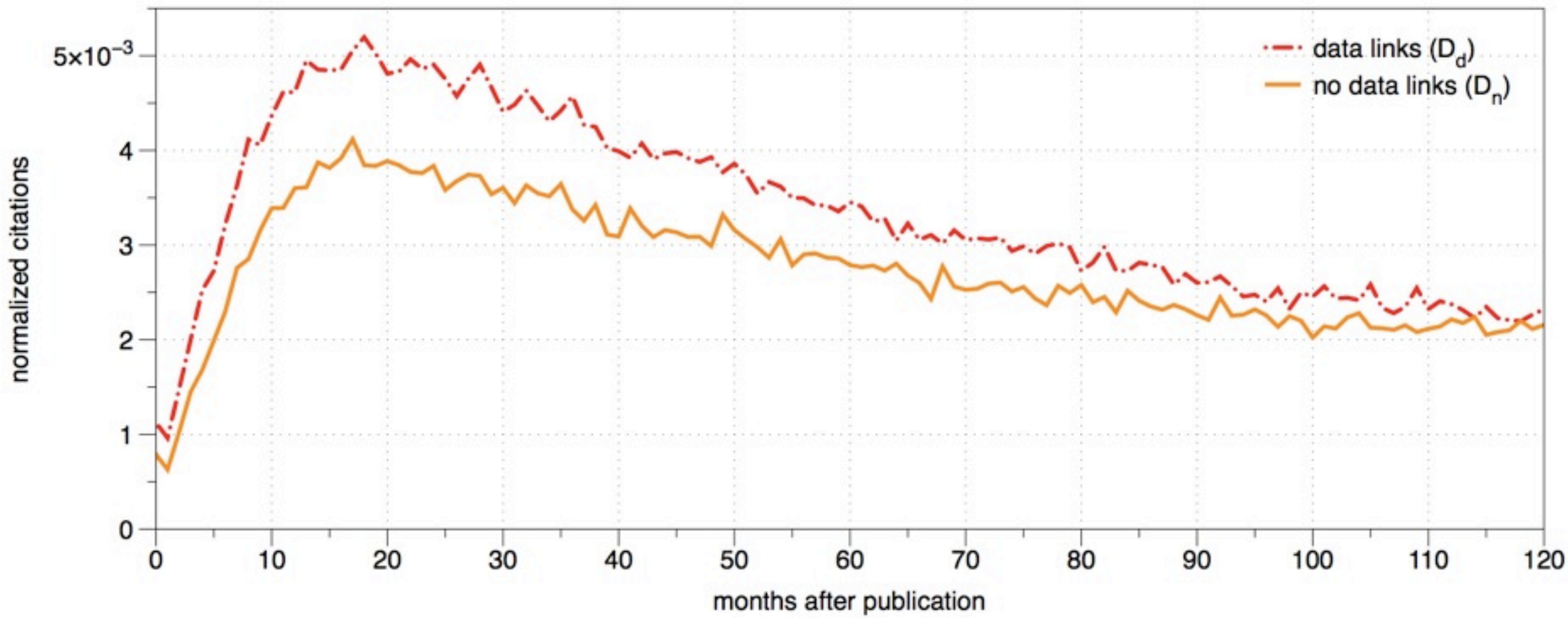
Sydney Brenner on data mining.

Thank you!

The ADS is Operated by the Smithsonian Astrophysical Observatory
under NASA Grant NNX09AB39G

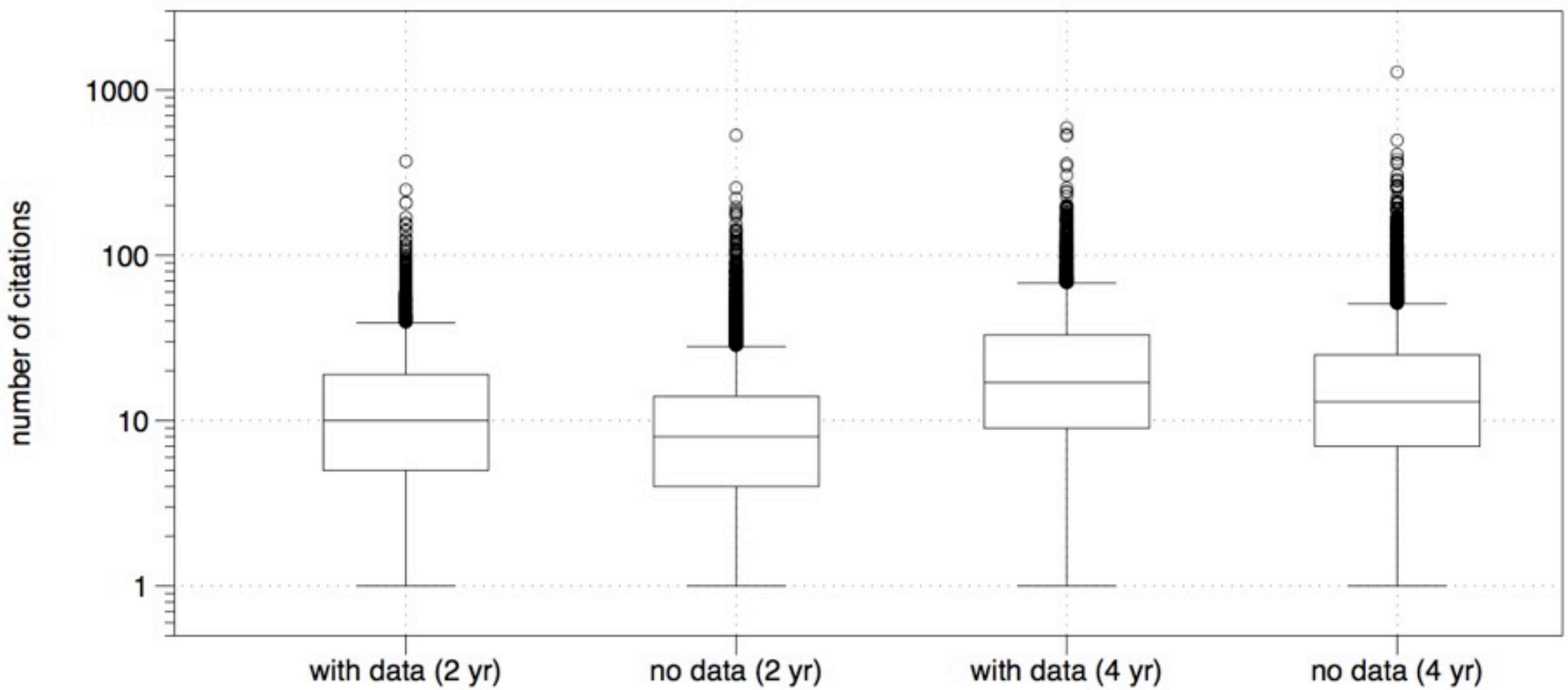
Linking to Data

Effect on Citation Rates in Astronomy



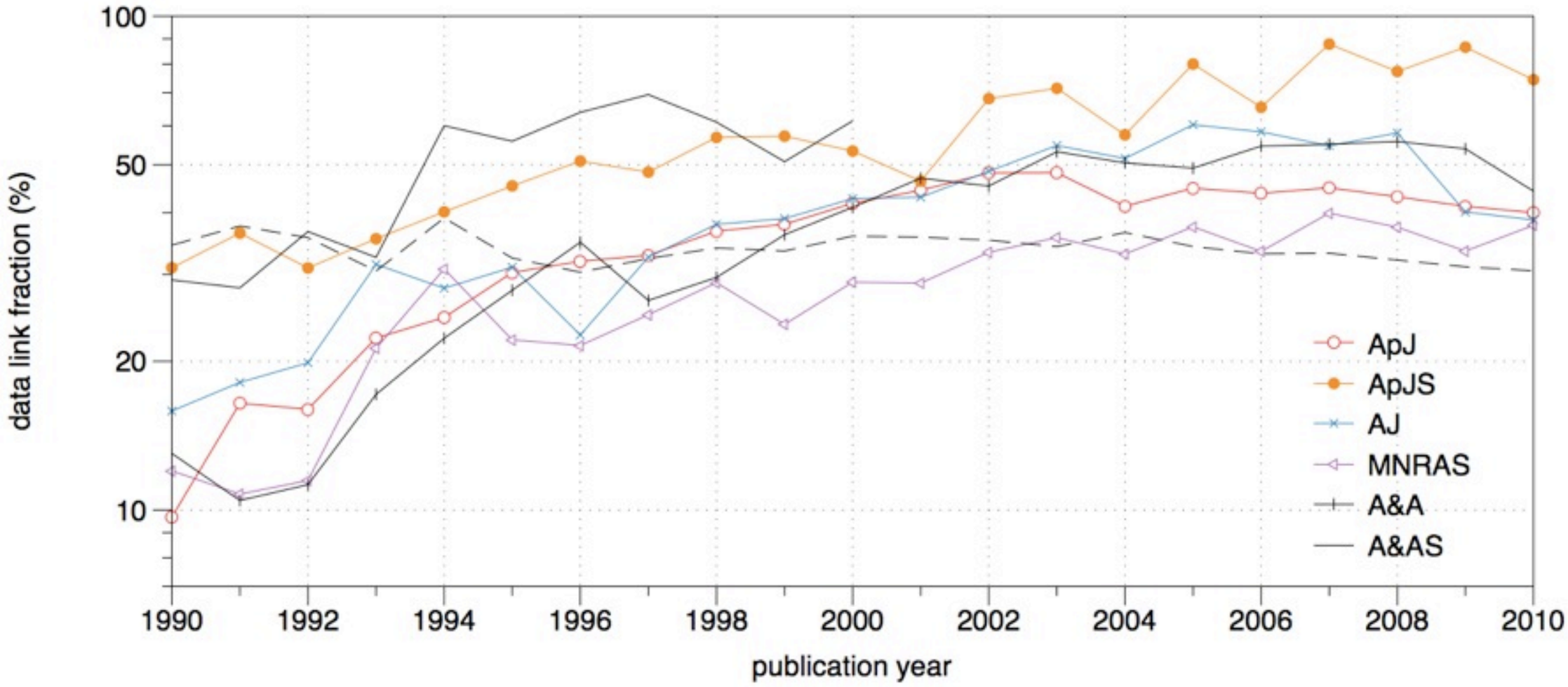
Linking to Data

Effect on Citation Rates in Astronomy



Linking to Data

Effect on Citation Rates in Astronomy



Linking to Data

Effect on Citation Rates in Astronomy