

Implementation of Organizational IDs in NASA's ADS Abstract Service

Carolyn Grant, Matthew Templeton & the ADS Team
cgrant@cfa.harvard.edu

Pidapalooza Jan. 23, 2019



What is ADS?

- Digital library portal for researchers in astronomy and physics
- Operated by the Smithsonian Astrophysical Observatory (SAO) under a NASA grant
- Publications in astronomy and astrophysics, physics and geophysics, and the arXiv e-prints
- Journals, conference proceedings, theses, e-conferences, data catalogs, proposals, software citations, etc.

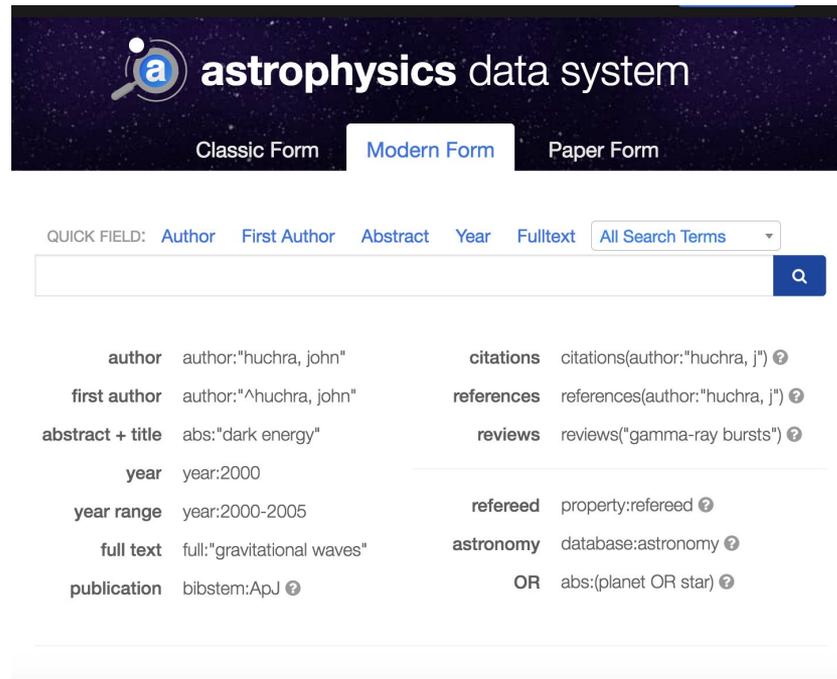
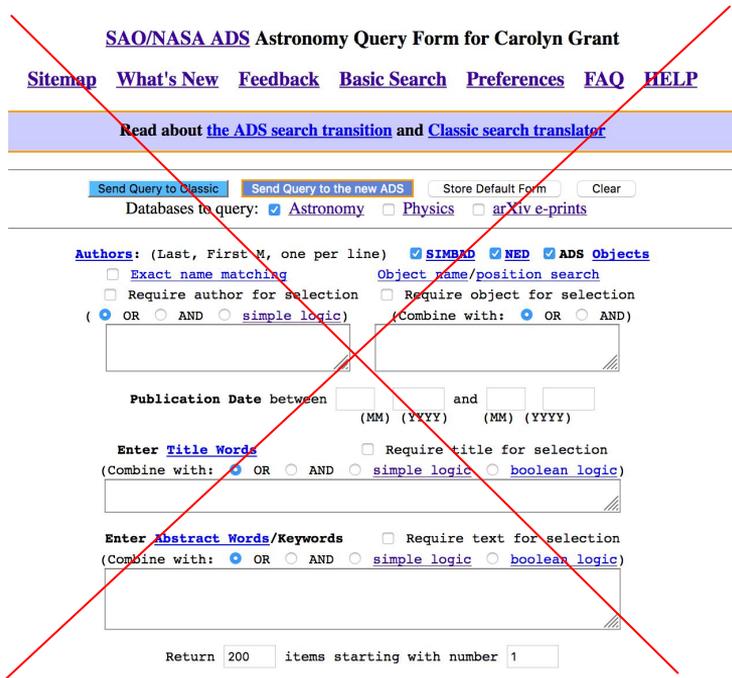


What is ADS?

- 14.3 million records
- 111 million citation pairs
- 5.2 million scanned pages (670,000 articles)
- 1.25 million hits a month ($\frac{1}{2}$ direct access; $\frac{1}{2}$ through google)
- Annual ingest of about 1 million records, 10 million citations



What is ADS?



Affiliations

The data problem:

- Affiliations are typically free-form (> 34 million unique strings)
- Although fairly complete for recent refereed journals, most are missing for the grey literature and older material
- Authors have multiple affiliations
- Inherent ambiguity (IoA)

Institute of Astronomy, Cambridge, UK
Institute of Astronomy, UNAM, Mexico
Institute of Astronomy, University of Tokyo
Institute of Astronomy, Moscow
Institute of Astronomy, Taiwan
Institute of Astronomy, University of Vienna
Institute of Astronomy, Bulgarian Academy of Sciences
Institute of Astronomy, University of Leuven



Affiliations

The usability problem:

- Searches by affiliation will inherently be incomplete - how to convey that to users?
- How best to integrate names, affiliations and ORCID ids in the user interface?
- Librarians and scientists potentially have different needs (level of detail)

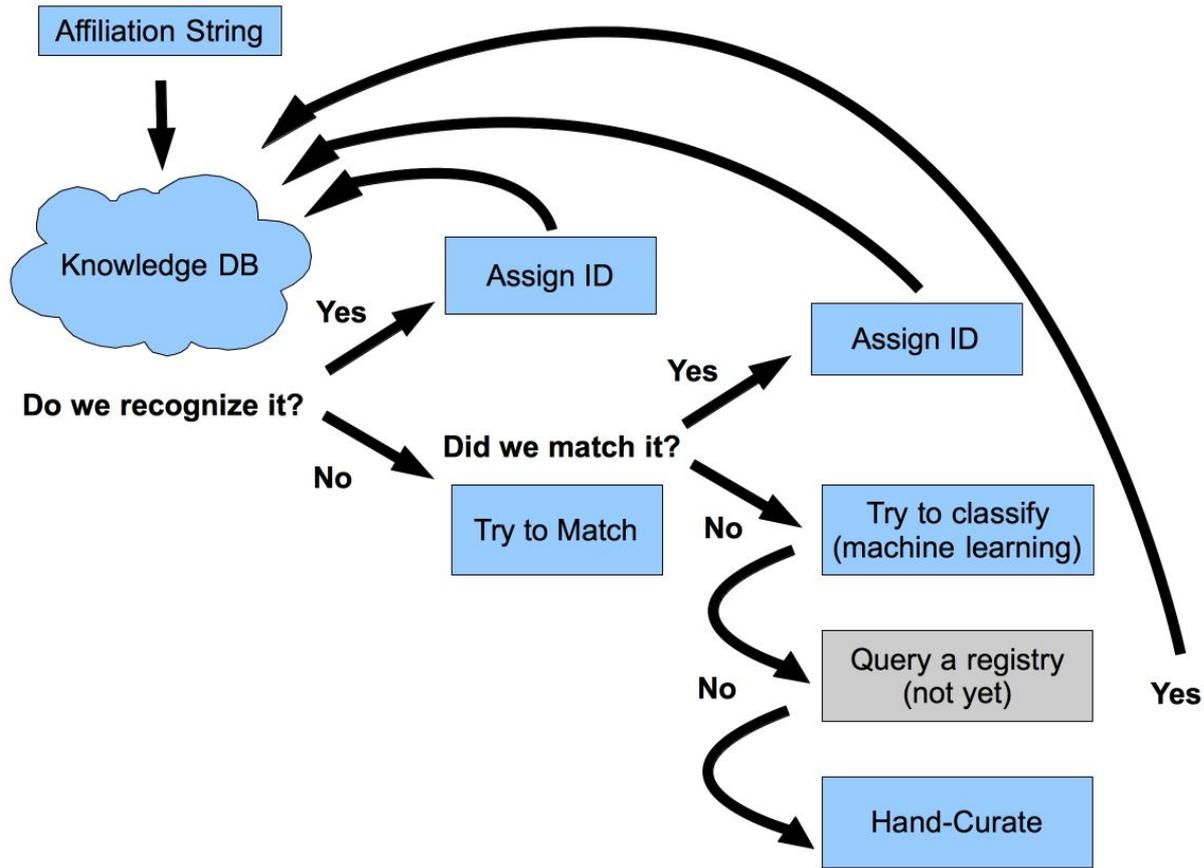


Affiliations

Current efforts:

- Creation of canonical institution names, ids and abbreviations for facet (3800 inst + 2800 divisions)
- Assignment of raw strings to institutional ids (93% for astronomy, 70% for physics)
- Development of python routine to match new and unmatched affiliations





Curation Challenges

- Sheer numbers:
 - Potential for huge amount of hand-work. **Have to automate**
 - Must to be able to bulk update, correct, etc.
 - Computer intensive
- Inconsistent naming:
 - **grid.473002.2**, University of the State of Paraná, Paranavaí,, Brazil
 - **grid.441795.a**, Universidade Estadual do Norte do Paraná, Jacarezinho,, Brazil
 - **grid.441662.3**, State University of West Paraná, Cascavel,, Brazil
- Incomplete or ambiguous affiliations
 - Dept Phys Univ Sezione INFN
 - UCB (Boulder, Colorado; Berkeley, California)



Curation Challenges

- Nothing systematic from publishers (only 2 give us institutional identifiers)
- Multiple and/or changing names:
 - State University of West Parana
 - UNIOESTE
 - Universidade Oeste do Paraná
 - Universidade Estadual do Oeste do Paraná
- No existing standard or registry (before last week)



- Attila (Jozsef) Univ Szeged (Hungary)
- Attila Jozsef University, Szeged
- Department of Experimental Physics University of Szeged Szeged Hungary
- Department of Experimental Physics, University of Szeged, Dómtér 9, H-6723 Szeged, Hungary
- Department of Experimental Physics, University of Szeged, Dómtér 9, Szeged, Hungary
- Department of Experimental Physics, University of Szeged, Szeged, Dómtér 9, 6720 Hungary
- Department of Optics, JATE University, Szeged, Hungary
- Department of Theoretical Physics, University of Szeged, Tisza Lajos krt 84-86, Szeged 6720, Hungary and
Department of Experimental Physics, University of Szeged, Dómtér 9, Szeged 6720, Hungary
- Departments of Theoretical and Experimental Physics, University of Szeged, Szeged, 6720 Dómtér 9., Hungary
- Dept Experimental Phys JATE Univ Szeged Dom ter H Szeged Hungary
- Dept Experimental Phys Univ Szeged Szeged Dom ter Hungary
- Dept Optics & Quantum Electronics Univ Szeged 406 Szeged Hungary
- Dept Optics Quantum Electronics Univ Szeged H Pf Szeged Hungary
- Dept Optics Quantum Electronics Univ Szeged PO Box: H Szeged Hungary
- Dept Phys Chem Materials Sciences Univ Szeged
- Dept Theoretical Phys Univ Szeged
- Dept Theoretical Phys Univ Szeged Hungary
- Dept Theoretical Phys Univ Szeged Tisza L krt Szeged Hungary H
- JATE Univ Research Group Laser Phys Hungary
- Jozsef Attila Tudományegyetem
- On leave from the Department of Optics and Quantum Electronics, University of Szeged, Dómtér 9, Szeged, Hungary



QUICK FIELD: [Author](#) [First Author](#) [Abstract](#) [Year](#) [Fulltext](#) [All Search Terms](#)[← Start New Search](#)Your search returned **1,178** resultsProperty
+property:refereed

Date ▾

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

- ✓ **AUTHORS**
- ✓ Schmidt, B 1.1k
- Schmidt, B 859
- Schmidt, B P 80
- Schmidt, Brian P 63
- Schmidt, B E 48
- Schmidt, B G 45
- [more](#)

- > Cardini, A 480
- > Veltri, M 480
- > Graziani, G 479
- > Soler, F 478
- [more](#)

COLLECTIONS

REFEREED

- ✓ **AFFILIATIONS**
- > CNRS 288
- > U Heidelberg 269
- > U Oxford 268
- > UCamb 263
- > ITEP 254
- [more](#)

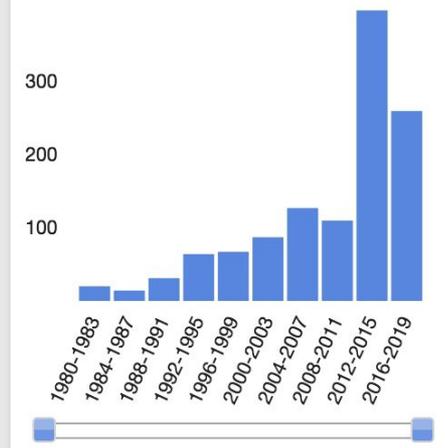
 [Show highlights](#) [Show abstracts](#) [Hide Sidebars](#)
[Go To Bottom](#)

1	<input type="checkbox"/>	2019PASP..131a4002H	2019/01	cited: 2			
		Carnegie Supernova Project-II: The Near-infrared Spectroscopy Program					
		Hsiao, E. Y.; Phillips, M. M.; Marion, G. H. <i>and 42 more</i>					
2	<input type="checkbox"/>	2019PASP..131a4001P	2019/01	cited: 3			
		Carnegie Supernova Project-II: Extending the Near-infrared Hubble Diagram for Type Ia Supernovae to z ~ 0.1					
		Phillips, M. M.; Contreras, Carlos; Hsiao, E. Y. <i>and 41 more</i>					
3	<input type="checkbox"/>	2018PhLB..787..124A	2018/12	cited: 2			
		Search for CP violation in $\Lambda_b^0 \rightarrow pK^-$ and $\Lambda_b^0 \rightarrow \pi\pi^-$ decays					
		Aaij, R.; Adeva, B.; Adinolfi, M. <i>and 836 more</i>					
4	<input type="checkbox"/>	2018JLTP..193..867D	2018/12				
		The CUORE Cryostat					
		D'Addabbo, A.; Alduino, C.; Bersani, A. <i>and 38 more</i>					
5	<input type="checkbox"/>	2018Icar..316...63H	2018/12	cited: 3			
		The Ac-5 (Fejokoo) quadrangle of Ceres: Geologic map and geomorphological evidence for ground ice mediated surface processes					
		Hughson, Kynan H. G.; Russell, C. T.; Williams, D. A. <i>and 19 more</i>					
6	<input type="checkbox"/>	2018Icar..316...46S	2018/12	cited: 2			
		Ceres' Ezinu quadrangle: a heavily cratered region with evidence for localized subsurface water ice and the context of Occator crater					
		Scully, Jennifer E. C.; Buczkowski, D. L.; Neesemann, A. <i>and 22 more</i>					
7	<input type="checkbox"/>	2018EPJC...78.1019A	2018/12				
		Evidence for an $\eta_c(1S) \pi^+ \pi^-$ resonance in $B^0 \rightarrow \eta_c(1S) K^+ \pi^-$ decays					
		Aaij, R.; Beteta, C.; Abellán, Adeva, B. <i>and 843 more</i>					

0 selected

[Add papers to library](#)

Years Citations Reads

 refereed non refereed


Limit results to papers from

QUICK FIELD: Author First Author Abstract Year Fulltext All Search Terms

Start New Search

{author:"Schmidt, Brian" AND year:1980-2019}

Your search returned 143 results

Author Schmidt, B P
Author Schmidt, Brian P
Property +property:refereed

Date

Export

Explore

- AUTHORS
Schmidt, B 143
Schmidt, B P 80
Schmidt, Brian P 63
Filippenko, A 42
Kirshner, R 42
Challis, P 35
Suntzeff, N 35
COLLECTIONS
REFEREED
AFFILIATIONS
ANU 112
UCB 48
Manoa 46
ESO 43
Harvard U 42
KEYWORDS
PUBLICATIONS
BIB GROUPS

Show highlights Show abstracts Hide Sidebars

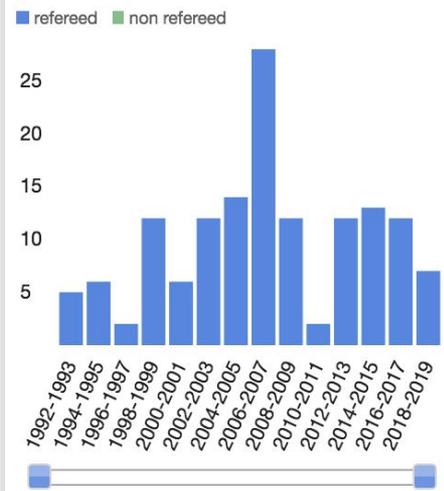
Go To Bottom

2019PASP.131a4002H 2019/01 cited: 2 Carnegie Supernova Project-II: The Near-infrared Spectroscopy Program
2019PASP.131a4001P 2019/01 cited: 3 Carnegie Supernova Project-II: Extending the Near-infrared Hubble Diagram for Type Ia Supernovae to z ~ 0.1
2018PASA...35...24W 2018/06 cited: 2 Discovery of the Most Ultra-Luminous QSO Using GAIA, SkyMapper, and WISE
2018NatSR...8.8561S 2018/06 cited: 5 Sensations from a single M-cone depend on the activity of surrounding S-cones
2018MNRAS.477..766D 2018/06 cited: 17 SMSS J130522.47-293113.0: a high-latitude stellar X-ray source with pc-scale outflow relics?
2018MNRAS.475.1046I 2018/03 cited: 46 On the nature of hydrogen-rich superluminous supernovae
2018PASA...35...10W 2018/02 cited: 41 SkyMapper Southern Survey: First Data Release (DR1)

0 selected

Add papers to library

Years Citations Reads



QUICK FIELD: [Author](#) [First Author](#) [Abstract](#) [Year](#) [Fulltext](#) [All Search Terms](#)[← Start New Search](#)

(author:"Schmidt, Brian" AND year:1980-2019)

Your search returned **112** results**Author**
Schmidt, B P**Author**
Schmidt, Brian P**Property**
+property:refereed**Affiliations**
+ANU

Date ▾

[Export](#)[Explore](#)▼ **AUTHORS**

- ▼ Schmidt, B 112
 - Schmidt, B P 68
 - Schmidt, Brian P 44
- > Filippenko, A 30
- > Kulkarni, S 27
- > Suntzeff, N 27
- > Challis, P 26

[more](#)> **COLLECTIONS**> **REFEREED**▼ **AFFILIATIONS**

- ▼ ANU 112
 - Res Sch Ast Astrop 97
 - CAASTRO 17
 - MSSSO 13
 - ANU 5
- > UCB 47
- > Manoa 46
- > ESO 42
- > CalTech 37

[more](#)[Show highlights](#)[Show abstracts](#)[Hide Sidebars](#)[Go To Bottom](#)

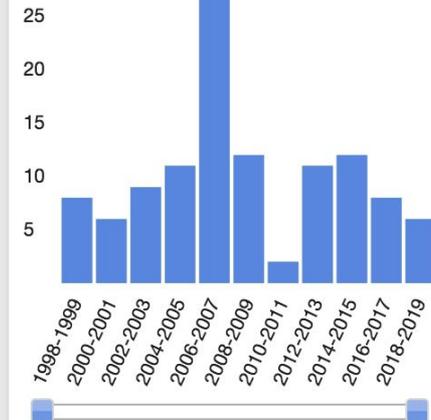
- 1 2019PASP.131a4002H 2019/01 cited: 2
Carnegie Supernova Project-II: The Near-infrared Spectroscopy Program
Hsiao, E. Y.; Phillips, M. M.; Marion, G. H. [and 42 more](#)
- 2 2019PASP.131a4001P 2019/01 cited: 3
Carnegie Supernova Project-II: Extending the Near-infrared Hubble Diagram for Type Ia Supernovae to $z \sim 0.1$
Phillips, M. M.; Contreras, Carlos; Hsiao, E. Y. [and 41 more](#)
- 3 2018PASA...35...24W 2018/06 cited: 2
Discovery of the Most Ultra-Luminous QSO Using GAIA, SkyMapper, and WISE
Wolf, Christian; Bian, Fuyan; Onken, Christopher A. [and 5 more](#)
- 4 2018MNRAS.477..766D 2018/06
SMSS J130522.47-293113.0: a high-latitude stellar X-ray source with pc-scale outflow relics?
Da Costa, G. S.; Soria, R.; Farrell, S. A. [and 17 more](#)
- 5 2018MNRAS.475.1046I 2018/03 cited: 46
On the nature of hydrogen-rich superluminous supernovae
Inserra, C.; Smartt, S. J.; Gall, E. E. E. [and 34 more](#)
- 6 2018PASA...35...10W 2018/02 cited: 41
SkyMapper Southern Survey: First Data Release (DR1)
Wolf, Christian; Onken, Christopher A.; Luvaul, Lance C. [and 14 more](#)
- 7 2017PASA...34...69A 2017/12 cited: 28
Follow Up of GW170817 and Its Electromagnetic Counterpart by Australian-Led Observing Programmes
Andreoni, L.; Ackley, K.; Cooke, J. [and 122 more](#)

0 selected

[Add papers to library](#)

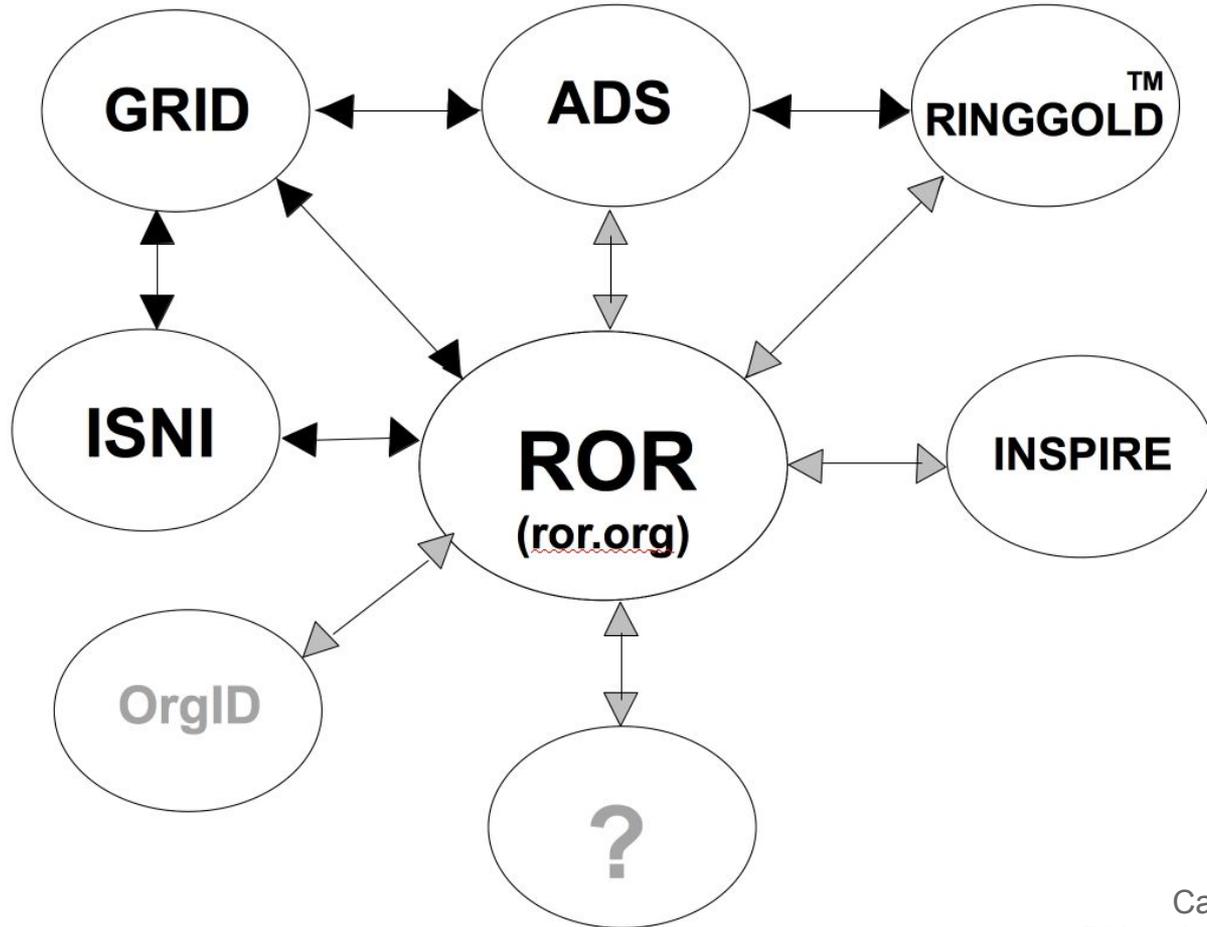
Years Citations Reads

■ refereed ■ non refereed



Limit results to papers from

A Registry!



Institutional Identifiers

- **Ringgold (proprietary)**
 - > 400,000 institutions (< 200,000 academic)
 - 165,000 parent (35,000 academic)
 - 500,000 correlated with ISNI
- **ISNI (open)**
 - >500,000 (lots of non-institutions)
- **GRID (open)**
 - 90,000 institutions (18,000 academic)
 - 14,000 relationships
- **OrgRef (defunct)**
 - 32,000 institutions
- **ADS (“open”)**
 - 6600 institutions
 - 3800 parents, 2800 children
 - ~1000 ADS-created (mostly department level)
- **ROR ! (open)**
 - 91,000 institutions



Questions

- What are use cases?
 - How many tiers?
 - To anglo-size or not? (University, Università, Université, Universität, Universiteit, Univerzita, etc.)
 - Standardize? (Polytechnical, Politechnical, Technical)



Questions

- What about acquisitions, mergers, name changes?
 - Current, recent past, distant past
- What about relationships?
 - Reality is complicated



ISNI Relationships:

- isMemberOf
- hasMember
- isUnitOf
- hasUnit
- isSupersededBy
- supersedes
- isAffiliatedWith
- isRelatedTo
- formerName
- laterName
- see also from
- no relation

Recommended additions*:

- acquired
- acquired by
- isGovernedBy
- Governs
- isHostedBy
- Hosts
- isPartneredWith



*Smith-Yoshimura, Karen et al. "Addressing the Challenges with Organizational Identifiers and ISNI", OCLC Research Report, 2016, <https://www.oclc.org/research/publications/2016/oclcresearch-organizational-identifiers-and-isni-2016.html>

Still to do

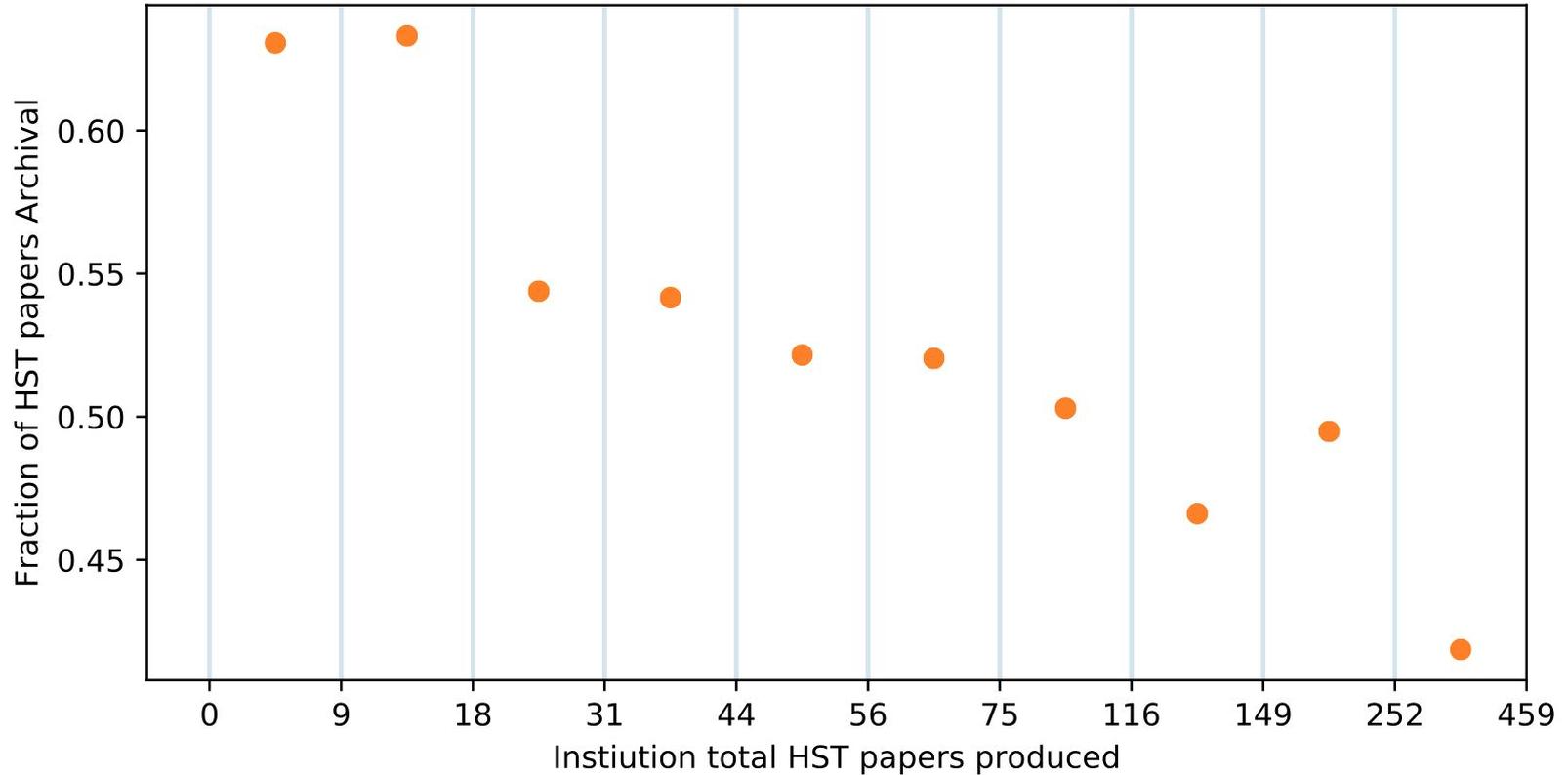
- Improvement of user interface (e.g. hover over, auto-complete)
- Development of curation tool
- Couple with authors
- Split joint affiliations better

- **ROR!**





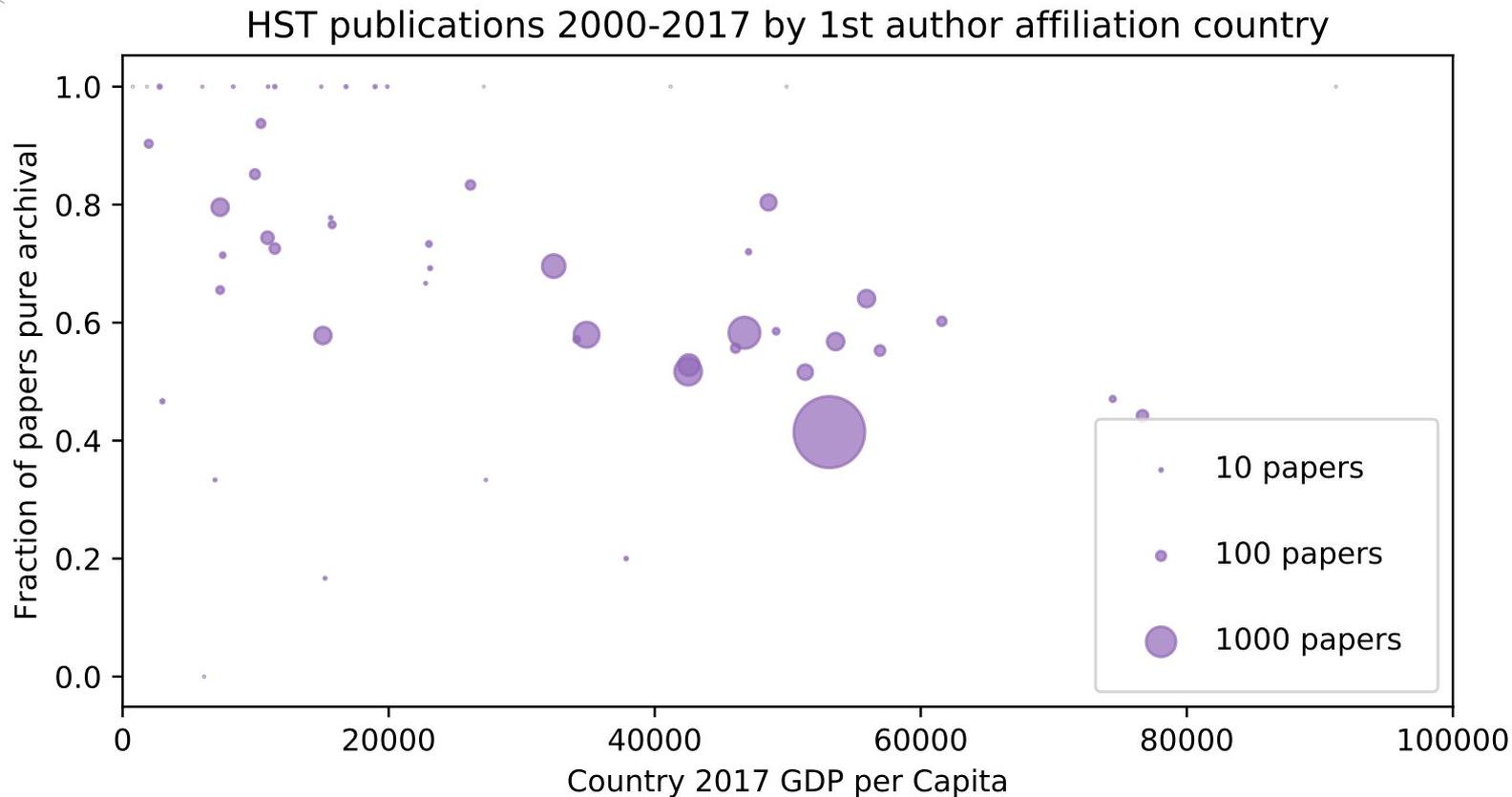
Institutions that produce fewer total papers are more archival



Slide courtesy of Josh Peek, STScI



Strong, similar trends hold by country



Thank you!

cgrant@cfa.harvard.edu



Why capture software citations?



"Research Life Cycle" image from UC Irvine Library Digital Scholarship Services (<https://www.lib.uci.edu/dss>)

- Research process: more and more digital in nature
 - research products: digital objects
- Citation system is centered around articles
 - “software papers”: bad proxy for software products
 - all research products represent aspects of research knowledge

Challenges:

- identify non-traditional objects in a paper
- identify non-traditional objects used within research process!