Title
The invisible knowledge infrastructure of astronomy: A sharper focus on blurry data

Author
Borgman, Christine L.

Publication Date
2018-10-03

The invisible knowledge infrastructure of astronomy: A sharper focus on blurry data

Christine L. Borgman

Distinguished Research Professor
Director, Center for Knowledge Infrastructures
University of California, Los Angeles
http://christineborgman.info
@scitechprof

Visiting Scholar, Harvard, October 2018
Center for Astrophysics (primary)
Data Science Initiative
Berkman Klein Center for Internet & Society

Harvard-Smithsonian Center for Astrophysics
3 October 2018


Christine Borgman


Bernie Boscoe


Peter Darch


Milena Golshan


Irene Pasquetto


Michael Scroggins


Cheryl Thompson


Morgan Wofford

UCLA Center for Knowledge Infrastructures

https://knowledgeinfrastructures.gseis.ucla.edu

Data sharing policies

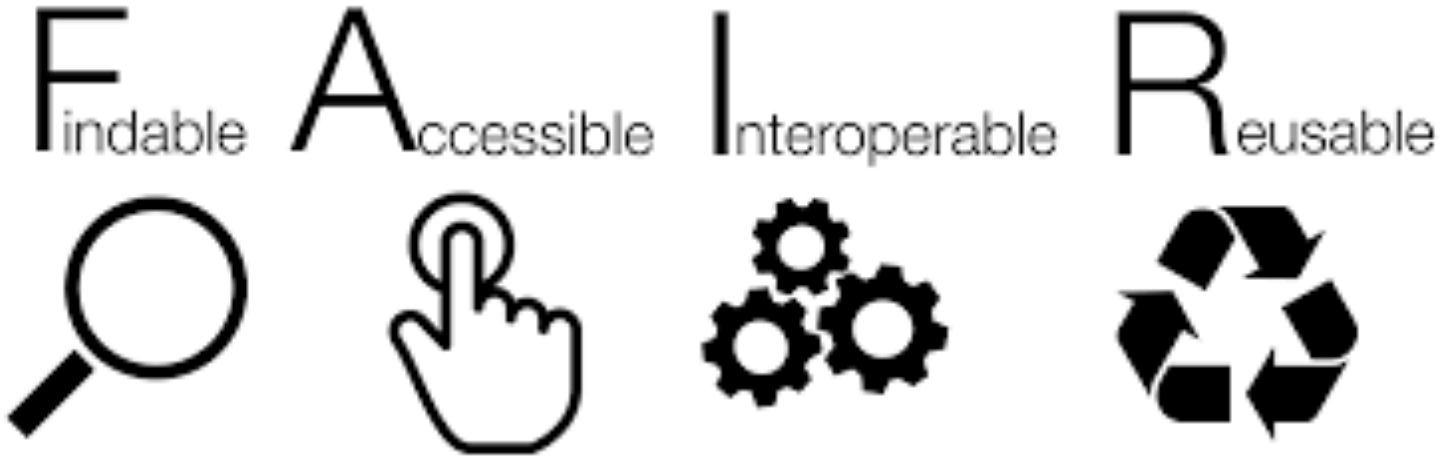
- European Union
- U.S. Federal research policy
- Research Councils of the UK
- Australian Research Council
- Individual countries, funding agencies, journals, universities

Supported by wellcome trust

National Institutes of Health
Turning Discovery Into Health

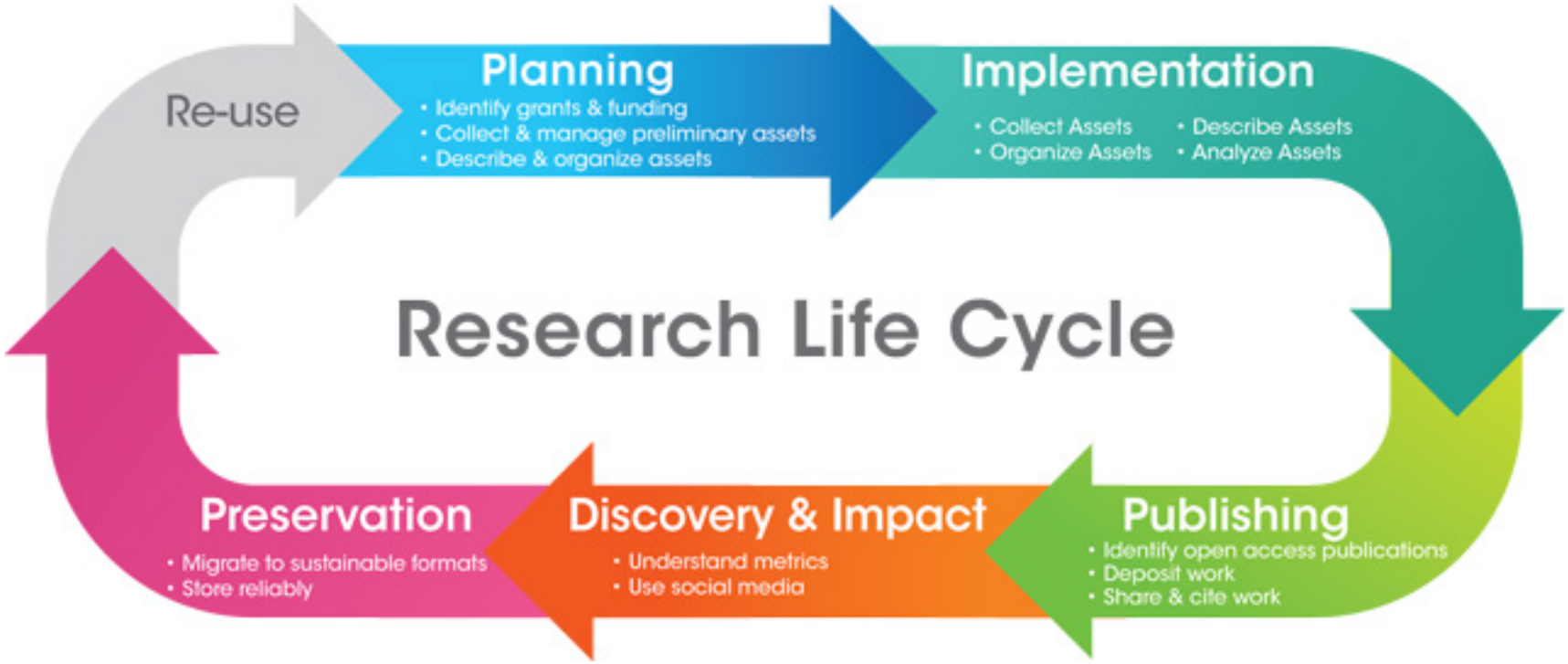
Policy RECommendations for Open Access to Research Data in Europe

Data Stewardship: The Ideal



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, http://dx.doi.org/10.1038/sdata.2016.18

Data creation and reuse: The Ideal



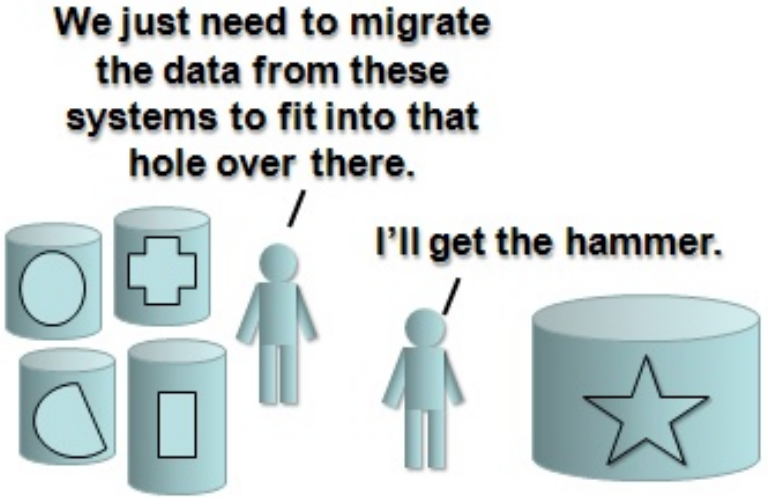
http://www.lib.uci.edu/dss/images/lifecycle.jpg

Data Stewardship: the Reality


http://www.information-age.com/cloud-computing-pharmaceutical-industry-123462676/


Getty Research Institute


Mount Wilson Solar Observatory, 2017



http://www.datamartist.com/data-migration-part-1-introduction-to-the-data-migration-delema


http://gsa.rice.edu/

Graduate students


https://med.nyu.edu/our-community/life-nyu-school-medicine/life-postdoc

Post-doctoral fellows

Astronomy Knowledge Infrastructures

- Observations: continuity over millennia
- Astronomy became digital: 1970s-
- Array of stakeholders: international
- Private and public funding: renewal
- Consensus mechanisms: Decadal survey

http://www.bbc.com/news/magazine-22105898

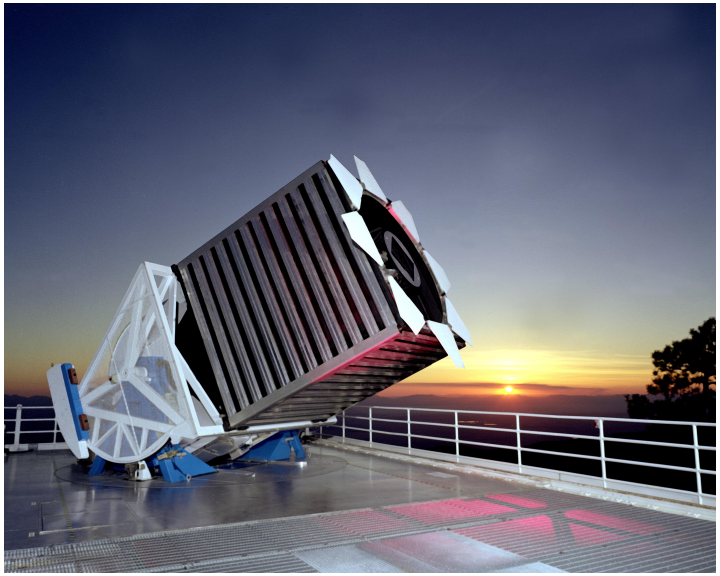
Astronomy Infrastructure Chronology

Astronomy Infra-structure	**1960-1985**	**1986-1990**	**1991-1995**	**1996-2000**	**2001-2005**	**2006-2010**	**2011-**
NASA funded infrastructure observatories missions centers archives & data tools	GSFC IPAC IRAS STSci	Hubble (HST) NED	ADS IRSA SATC	Chandra HEASARC 2MASS MAST NExScI Spitzer WIRE	SOFIA Spitzer Suzaku Swift	Fermi (FGST) Herschel Kepler Palomar (PTF) Planck WISE	NuSTAR
Other infrastructure observatories centers surveys data tools standards	SAO CfA CDS/ SIMBAD FITS		arXiv VizieR Keck	Aladin NVO IVOA SDSS Gemini Subaru		Dataverse GitHub LSST Pan-STARRS VAO WWT	ADSASS DDT Jupyter Note-book Zenodo DES TMT

Questions for CfA Residency

- What is astronomy knowledge infrastructure?
  - What features are most durable?
  - What features are most fragile?
- How do CfA astronomers keep their data alive?
- Where is the invisible work of data stewardship?
- Where are the path dependencies?
- How is astronomy KI evolving with open science, open access, and open data?

Sloan Digital Sky Survey (SDSS-I/II)

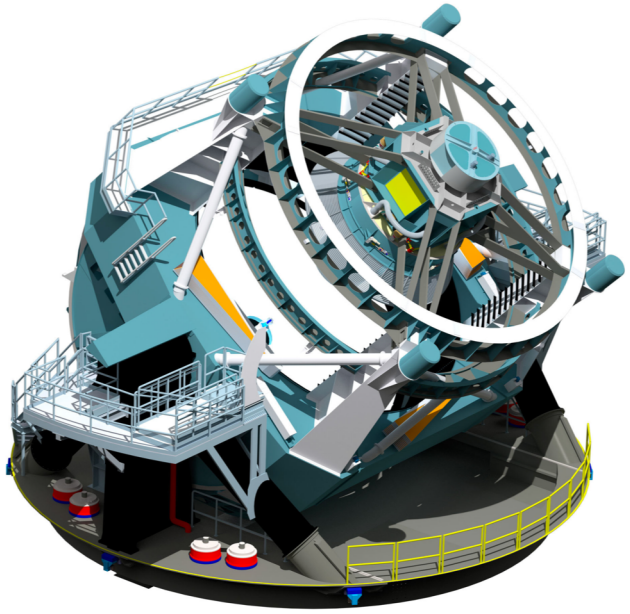


- Survey from 2000-2008
- 160+ TB data total
- Tens of millions of dollars
- Data open to all
- Proprietary software

Telescope for the Sloan Digital Sky Survey, Apache Point, New Mexico

Large Synoptic Survey Telescope (LSST)

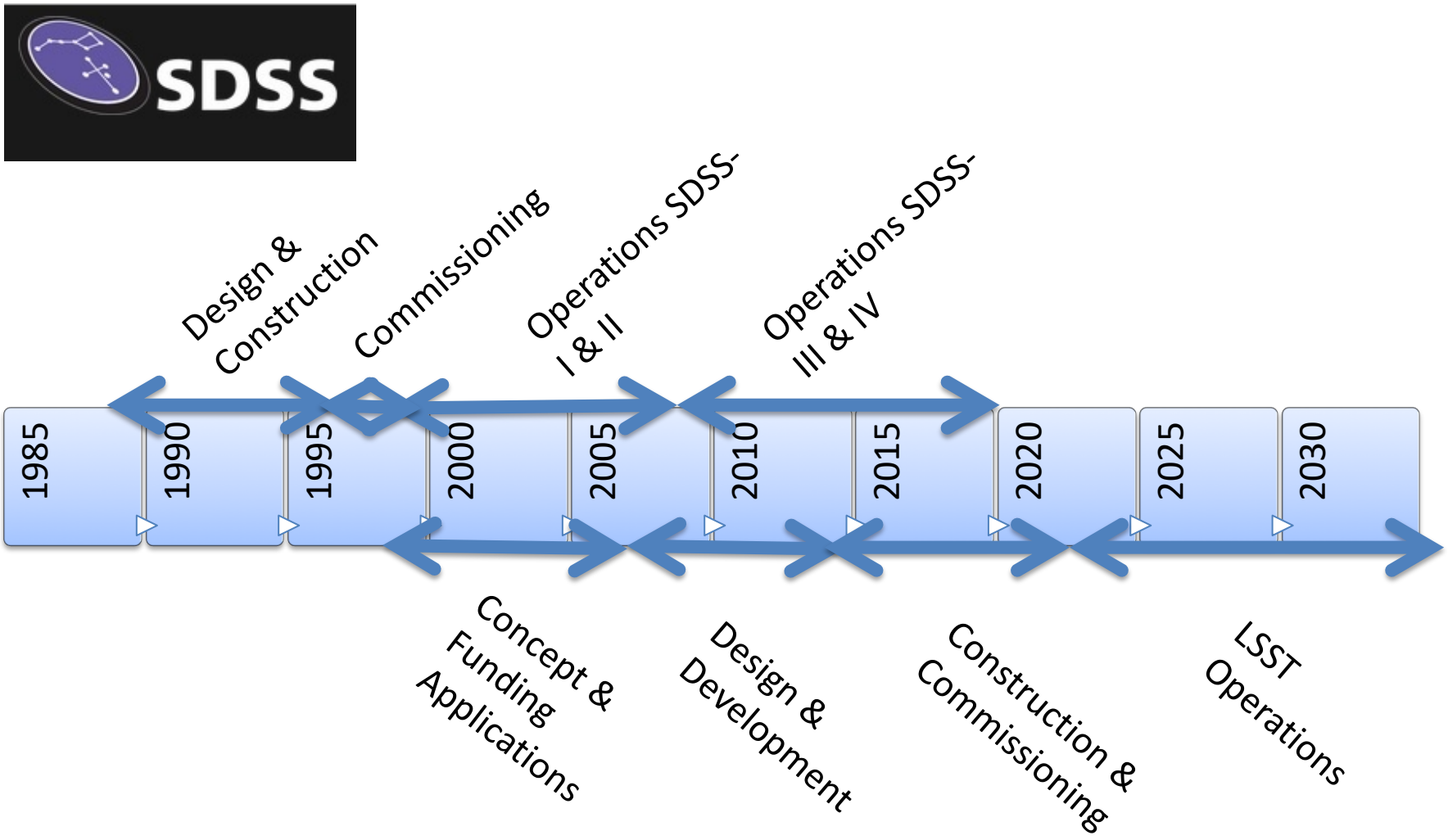
- Survey from 2022-2032
- 15 TB data per night
- 1+ Billion dollars
- Data open to partners
- Open source software



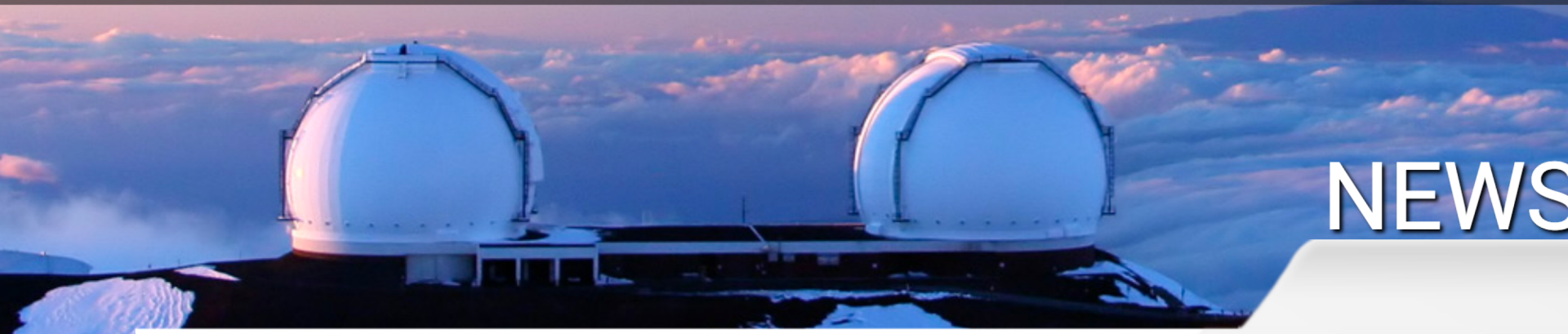
https://news.slac.stanford.edu/sites/default/files/images/image/lsst_h_0.jpg

LSST telescope, Chile

Project Timelines



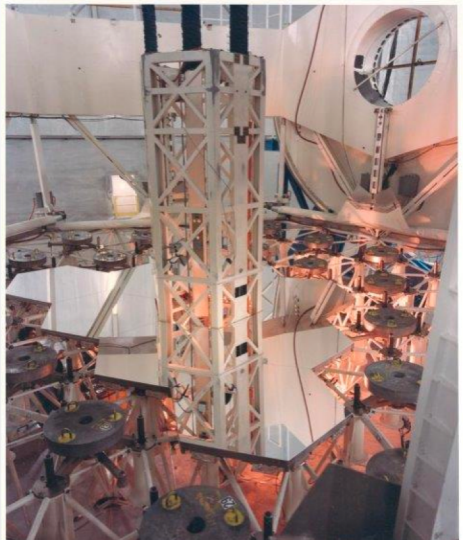
LSST Timeline: https://www.lsst.org/about/timeline



NEWS

"First Light" Marks 25 Years Of Hawaii's Leadership In Astronomy

Posted 12:36 pm October 2, 2018



By Sean Adkins

Every night, all over the world, people look up at the sky and wonder about the distant stars. Here in Hawaii we have the privilege of looking up at a very dark sky, but even here with the naked eye we can only see a few thousand stars. This is mainly because of the small size of the lens in our eye, which limits the amount of light it can gather, and also limits the detail we can see for those incredibly distant objects.

This week we will celebrate the 25th anniversary of

NEWS

NASA Honors Keck Observatory for Opening its Archive to the Public

DECEMBER 8, 2015

CREDIT: ANDREW RICHARD HARA / ENA MEDIA HAWAII

Keck I and Keck II silently opening the Universe to the world.

MAUNAKEA, Hawaii – The W. M. Keck Observatory in Hawaii has just been awarded the 2015 NASA Group Achievement Award for pioneering the Keck Observatory Archive (KOA) ten years ago, which has significantly increased the impact of Keck Observatory data. The award was received by Keck Observatory Chief Scientist, Dr. Anne Kinney at NASA headquarters on December 8, 2015.

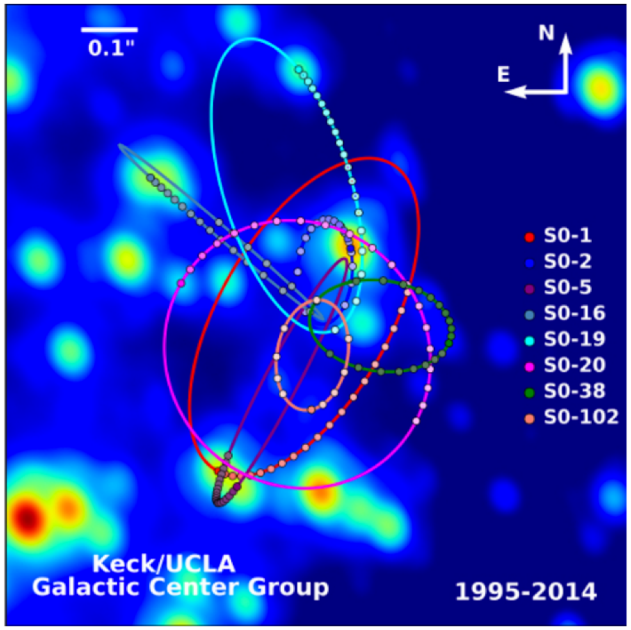
"For the past 10 years, the NASA KOA team has boosted the science value of data acquired at Keck Observatory by providing the scientific community with open access to WMKO data," said Mario Perez, Keck Observatory Program Executive at NASA Headquarters in Washington. "They helped set a standard that all new ground

Black Hole Properties

The Galactic Center Group members have been measuring the positions of thousands of stars in the vicinity of the Galactic Center for more than 20 years. This unique data set allowed us to measure directly short-period orbits of stars. In particular, a full phase coverage has been measured for two stars: S0-2 with an orbital period of 15.56 years, and S0-102 with 11.5 years. At the closest approach, S0-2 is only 17 light hours away from the center of the Galaxy, about four times the distance of Neptune from the Sun. From these orbital data, we can determine the mass of the central black hole in our own Galaxy.

The Milky Way Galaxy black hole is the closest example of the supermassive black holes, located only ~25,000 light years away from us. Its mass is estimated to be 4 million times the mass of the sun, which implies that the schwarzschild radius is about 17 times that of Sun's radius. As a comparison, Mercury's orbit is located at a distance of ~ 83 solar radii. Because the Galactic Center is the site of the closest supermassive black hole by a factor of 100, it is a unique laboratory for solving some of the greatest mysteries associated with the fundamental physics of supermassive black holes and the role that they play in the fomration and evolution of galaxies. Furthermore, it is the only galacitc nucleus in which direct measurements of stellar orbits is possible, with either the current or the next-generation instruments.

Observations of stellar orbits around the Galactic black hole also yields precision measurement of the distance to the Galactic Center, which is important as it affects almost all questions not only of Galacitc structure, dynamics and mass, but those of extragalactic distance scales and the value



The orbits of stars within the central 1.0 X 1.0 arcseconds of our Galaxy. These orbits provide the best evidence yet for a supermassive black hole. While every star in this image has been observed to move since 1998, estimates of orbital parameters are best constrained for stars that have been observed through at least one turning point of their orbits.

Once FITS, Always FITS? Astronomical Infrastructure in Transition

Michael Scroggins, Bernadette Boscoe

(Submitted on 24 Sep 2018)

The FITS file format has become the de facto standard for sharing, analyzing, and archiving astronomy data over the last four decades. FITS was adopted by astronomers in the early 1980s to overcome incompatibilities between operating systems. On the back of FITS' success, astronomical data became both backwards compatible and easily shareable. However, new advances in astronomical instrumentation, computational technologies, and analytic techniques have resulted in new data that do not work well within the traditional FITS format. Tensions have arisen between the desire to update the format to meet new analytic challenges and adherence to the original edict for FITS files to be backwards compatible. We examine three inflection points in the governance of FITS: a) initial development and success, b) widespread acceptance and governance by the working group, and c) the challenges to FITS in a new era of increasing data and computational complexity within astronomy.

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; History and Philosophy of Physics (physics.hist-ph)
Cite as: **arXiv:1809.09224 [astro-ph.IM]**
(or **arXiv:1809.09224v1 [astro-ph.IM]** for this version)

Submission history

From: Bernadette Boscoe [view email]
[v1] Mon, 24 Sep 2018 21:08:21 GMT (419kb)

Astrophysics Data System: 20 years

- 1993: Abstract Service launched implementing federated search to SIMBAD object database
- 1994: First web-based version of the ADS Abstract Service released
- 1995: Fulltext of ApJ Letters digitized and online
- 1996: Citation data incorporated in ADS; links between bib records and datasets created; first mirror site online
- 1997: Indexed astronomy preprints from arXiv
- 1998: The online readership via ADS surpasses the worldwide print readership for the main Astronomy journals
- 1999: Extracted and incorporated 1.2M citations from digitized literature via text mining
- 2000: Incorporated usage data and Citation ranking in search engine
- 2001: ADS's 10th mirror site comes online
- 2002: Digitized and placed online 300K historical scans from microfilms online
- 2003: myADS notification service launched
- 2004: Introduced fulltext search, private libraries
- 2005: Introduced daily database updates of arXiv content, RSS feeds
- 2006: Launched the ADS basic search, implemented openURL linking
- 2007: Implemented user login system
- 2008: Upgraded content in ADS with color and grayscale scans
- 2009: Introduced the ADS topic search
- 2010: Developed the ADS recommender
- 2011: The ADS Labs website is launched
- 2012: Incorporated metrics, visualizations in ADS Labs
- 2013: The ADS corpus reaches 10M bibliographic records, 3M fulltext documents, and over 50M citations.

http://conf.adsabs.harvard.edu/ADSXX/

Home

About CDS

People

Support
Help and Tutorials
Developer's corner
Publication support

myCDS

Virtual Observatory projects
IVOA
Euro VO
ASTERICS
CoSADIE
Past projects:
Euro-VO ICE - AIDA - DCA - VOTech
OV France
Europlanet

Other projects
RDA Europe
ARCHES
ASTRODEEP
Gaia

Authorities
Strasbourg astronomical Observatory
CNRS - INSU
University of Strasbourg

Links
ADS

Centre de Données astronomiques de Strasbourg

Strasbourg astronomical Data Center

Entry point to all services

Object database

Catalogue database

Interactive sky atlas

Object/position

Obj/position/bibcode

Keywords, target, ...

Object/position

Other services

X-match Dictionary Sesame SimPlay

Hosted services

ADS mirror A&A

TIPTOPbase
INES

Latest news

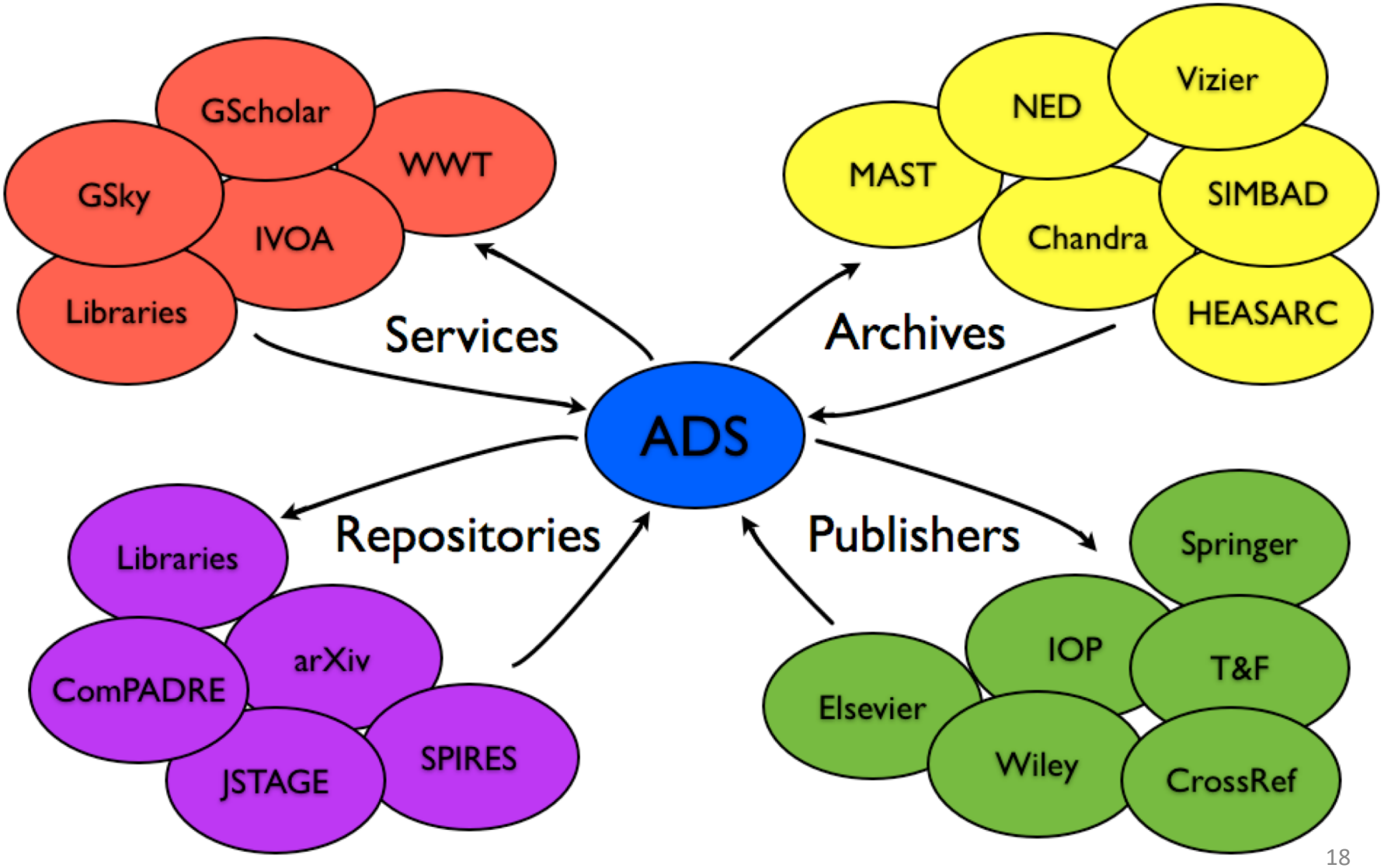
- Catalogs added between 22-Sep-2018 and 29-Sep-2018
- Catalogs added between 15-Sep-2018 and 22-Sep-2018
- Catalogs added between 08-Sep-2018 and 15-Sep-2018
- Facebook page, YouTube channel and Twitter
- Cross-identification of SIMBAD with Gaia DR2
- Simbad Tap upgrade
- CDS publishes the ESA Gaia mission Data Release 2 on April 25, 2018.
- PanSTARRS color HiPS available

More news

Featured news

Cross-identification of SIMBAD with Gaia DR2

ADS Collaborators



Slide: Alberto Accomazzi, ADS

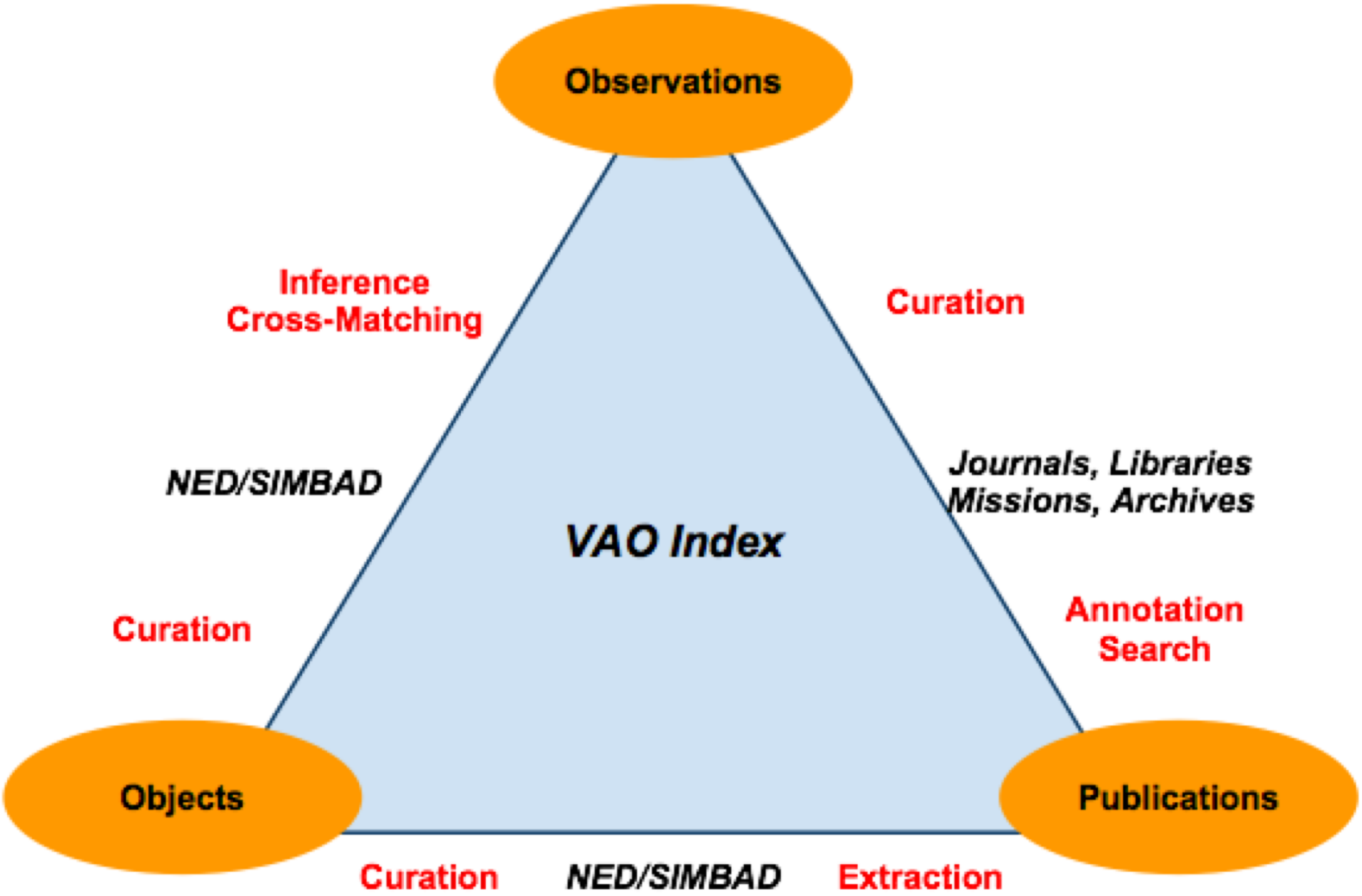


Figure 1. Relationships between Publications, Objects, Observations and the corresponding major actors in the curating process and their activities (in red).

Accomazzi, A., & Dave, R. (2011). Semantic Interlinking of Resources in the Virtual Observatory Era. *arXiv:1103.5958*

Data Standards and Infrastructure Fabric

VO Standards

- Standard Format: VOTable
- Access Protocols: ADQL, TAP, SIAP, SSAP

and others

Astronomy Infrastructure: Durability

- Collaboration and openness
- International coordination
- Long-term value of data
- Agreed standards
  - Units of measurement
  - Coordinate systems
  - Data structures
- Shared resources
  - Missions, instruments
  - Data archives
  - Tools and technologies

The Library of Congress
Operating a hand drill at Vultee-Nashville, woman is working on a "Vengeance" dive bomber, Tennessee (LOC) Palmer, Alfred T., photographer. 1943 Feb. Title from FSA or OWI agency caption. Transfer from U.S. Office of War Information, 1944.

Astronomy Infrastructure: Fragility

- Investment in data stewardship
  - Varies by mission
  - Varies by type of research
    - Space-based vs ground based
    - Large missions vs observing proposals
    - Shared vs custom instruments
- Access to data varies
  - Public archives
  - Local websites
  - Derived data

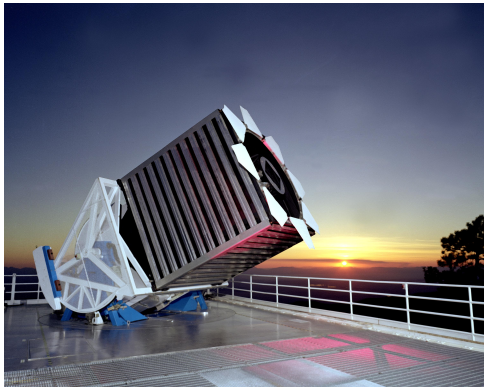
National Library of Scotland View from north end of gap Photograph of wrecked piers after collapse of part of the first Tay Bridge digital.nls.uk/74585062

Astronomy Infrastructure: Fragility

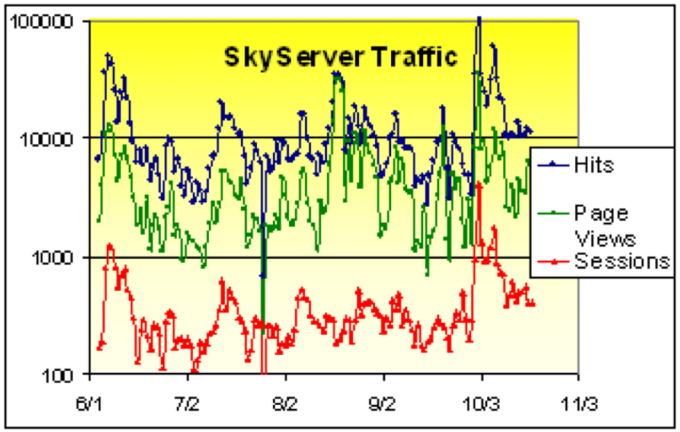
- Curation of tools and technologies
  - Shared investments
  - Open source
  - Proprietary tools
  - Local pipelines, tools, scripts
- International coordination
  - Investment in new instruments, missions
  - Investment in archives, infrastructure

National Library of Scotland View from north end of gap Photograph of wrecked piers after collapse of part of the first Tay Bridge digital.nls.uk/74585062

Conclusions so far…



- Infrastructures are fragile
- Durability is an accomplishment
- Visible infrastructure
  - Instruments
  - Institutions
- Invisible infrastructure
  - Data, metadata, provenance…
  - Information work



Telescope for the Sloan Digital Sky Survey, Apache Point, New Mexico

LSST All Hands Meeting, August 2014, Arizona State University. Arrow to Peter Darch

Research Methods

- Document analysis
  - Public and private documents and artifacts
  - Official and unofficial versions of scientific practice
- Ethnography
  - Observing activities on site
  - Embedded for days or months at a time
- Interviews
  - Questions based on our research themes
  - Compare multiple sites over time

UCLA OHRPP Office of the Human Research Protection Program

Center for Knowledge Infrastructures UCLA

Questions for CfA Residency

- What is astronomy knowledge infrastructure?
  - What features are most durable?
  - What features are most fragile?
- How do CfA astronomers keep their data alive?
- Where is the invisible work of data stewardship?
- Where are the path dependencies?
- How is astronomy KI evolving with open science, open access, and open data?

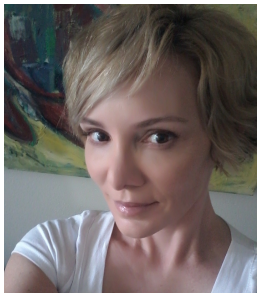
Acknowledgements

Harvard-Smithsonian Center for Astrophysics

UCLA Center for Knowledge Infrastructures

Alfred P. Sloan Foundation


Christine Borgman


Bernie Boscoe


Peter Darch


Milena Golshan


Irene Pasquetto

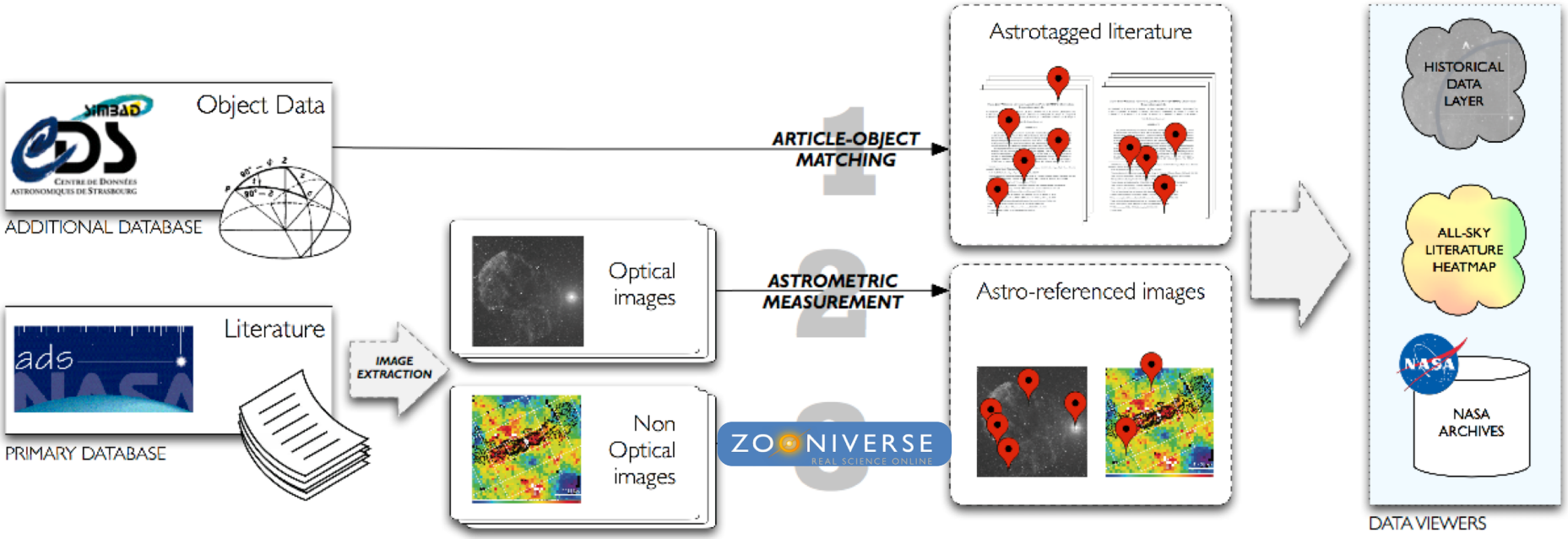

Michael Scroggins


Cheryl Thompson


Morgan Wofford

Borgman, C. L., Darch, P. T., Sands, A. E., & Golshan, M. S. (2016). The durability and fragility of knowledge infrastructures: Lessons learned from astronomy. In *Proceedings of the Association for Information Science and Technology* (Vol. 53, pp. 1–10). ASIS&T. https://doi.org/10.1002/pra2.2016.14505301057

Seamless Astronomy: ADS All Sky Survey



Slide: Alberto Pepe

W. M. KECK OBSERVATORY

UCO acts a managing partner
of the Keck Observatory on Mauna Kea, Hawaii.

Learn More

The University of California Observatories (UCO) is a multi-campus research unit. We operate the Lick Observatory, the technical labs at UC Santa Cruz and UCLA, and we are a managing partner of the Keck Observatory in Hawaii. We are the center for the UC participation in the Thirty-Meter Telescope (TMT) project.

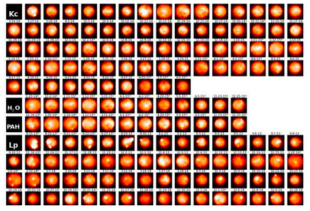
Learn More

Information for Observers:

Call for Proposals

Featured News

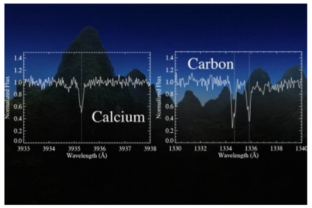
Gillian Wilson appointed Interim Deputy Director



Tracking volcanos on Jupiter's moon Io

OCTOBER 24, 2016

Follow-Up Of Kepler Data Yields More Than 100 Confirmed



Planet-Devouring Star Reveals Possible Limestone Crumbs

UCLA Center for Knowledge Infrastructures

The Center for Knowledge Infrastructures conducts research on scientific data practices and policy, scholarly communication, and socio-technical systems. Their latest project, funded by the Alfred P. Sloan Foundation, is studying data practices, policy, and infrastructure of multiple distributed scientific collaborations, exploring methods of data collection and management, data sharing and reuse, innovations in scaling and workflows, and multidisciplinary approaches to complex problems.

Peter Darch: astronomy, earth science
Bernie Boscoe: astronomy
Irene Pasquetto: biomedicine
Christine Borgman: astronomy, earth science, sensor networks
Milena Golshan: data science, archives
Ashley Sands: astronomy, archaeology
Not in photo
Michael Scroggins: astronomy, anthropology
Cheryl Thompson: earth sciences, astronomy
Morgan Wofford: earth science, data science