Future Plans - Indexing of Research Artifacts

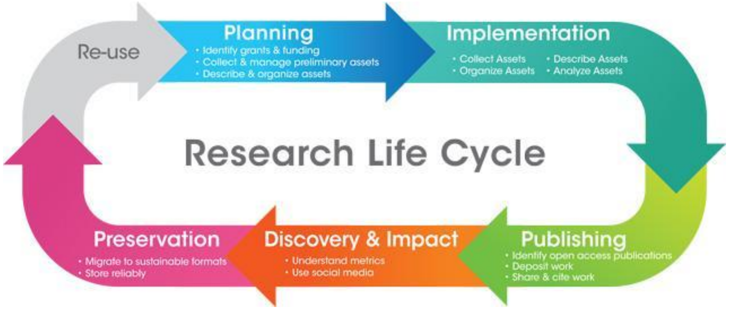
Edwin Henneken and the ADS Team

ADS Users Group Meeting, 19-20 Nov. 2020

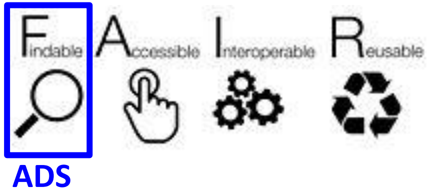
astrophysics data system

Research Objects Stage

Research in <u>data-intensive disciplines</u> is increasingly consuming and generating a variety of digital resources during the course of scientific investigations. This has steadily increased the need for means to systematically capture the lifecycle of scientific investigations, which at the same time provide a single-entry point to all the related resources, including data, publications, presentations, computational resources (software, Jupyter Notebooks, protocols), and the researchers involved in the investigation.



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



Goal: enhance the sharing, preservation and communication of data-intensive science, facilitating validation, citation and reuse by the community.

NASA: NASA SMD Strategy WP goal: "*Improve discovery and access for all SMD data to immediately benefit science data users and improve the overall user experience*"

ADS: data linking, ROR, ORCID, Asclepias, DOI → Research Artifacts

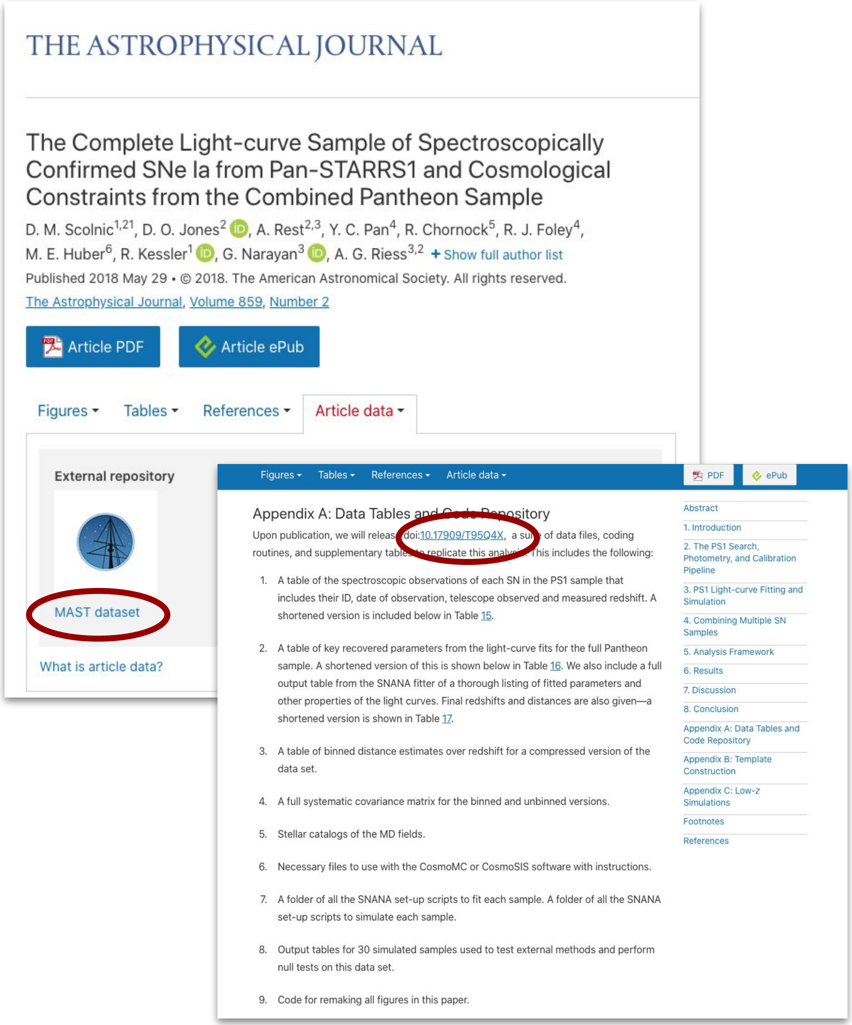
Scolnic et al, ApJ 859, 101 (29 May 2018)

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

Upon publication, we will release doi:10.17909/T95Q4X, a suite of data files, coding routines, and supplementary tables to replicate this analysis.

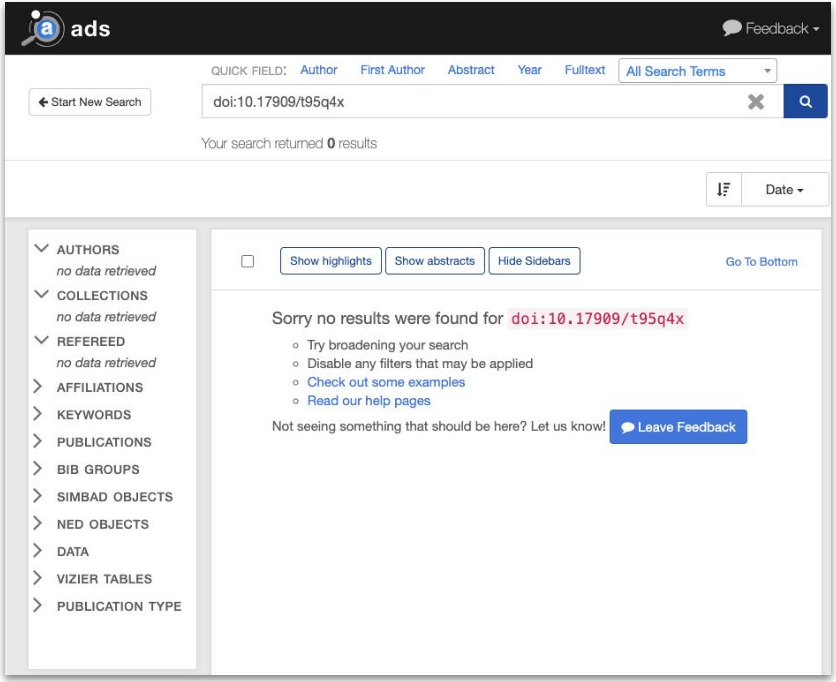
DOI 10.17909/t95q4x is mentioned 7 times in HTML and PDF document:

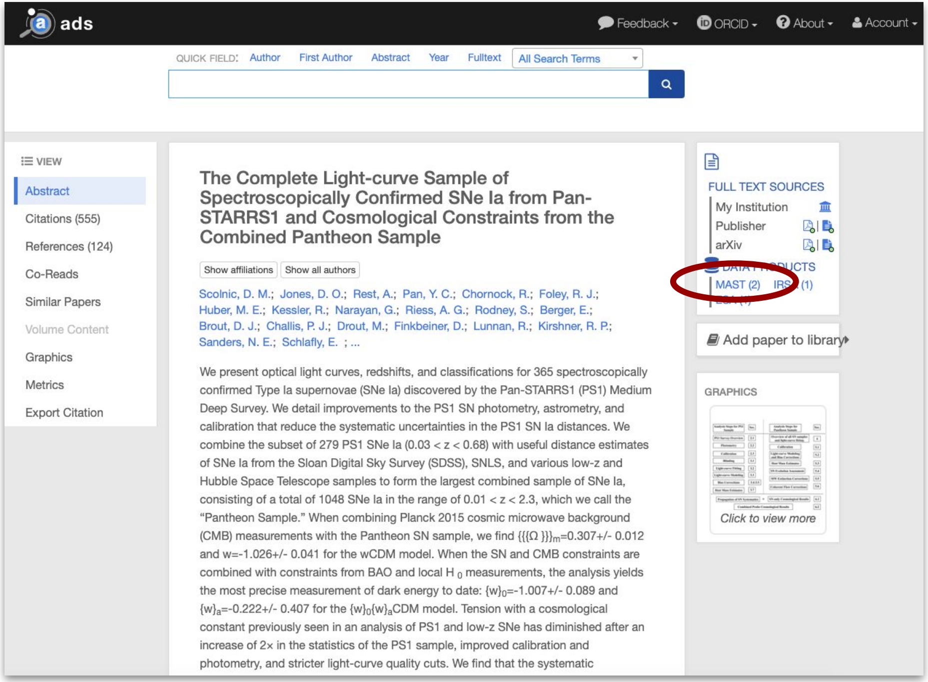
- Narrative (3 times)
- Table captions (3 times)
- Appendix A (data & code availability, 1 time)



How does this DOI show up in the ADS?

What does ADS know about 10.17909/t95q4x?



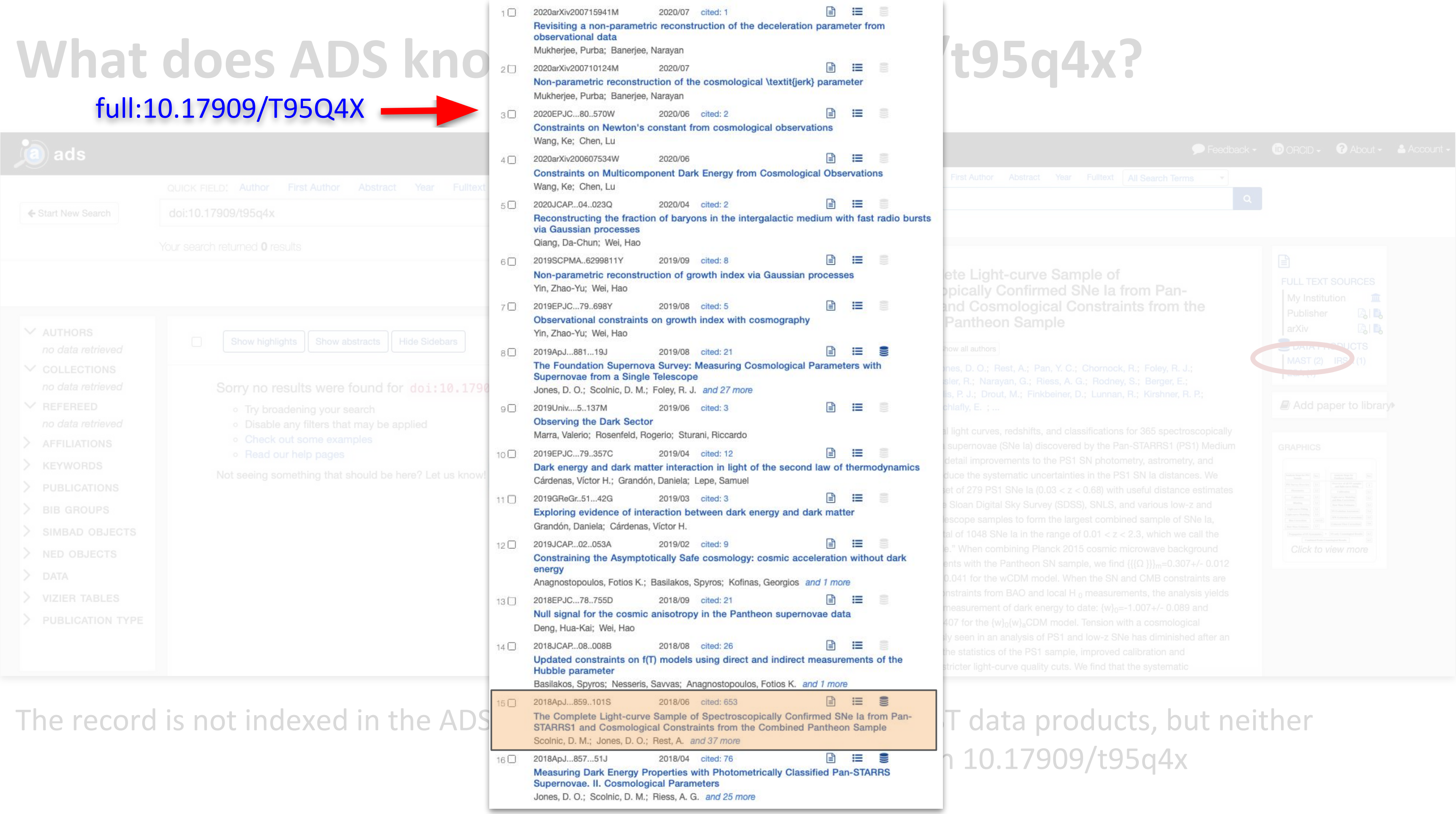


The record is not indexed in the ADS...

There are 2 MAST data products, but neither corresponds with 10.17909/t95q4x

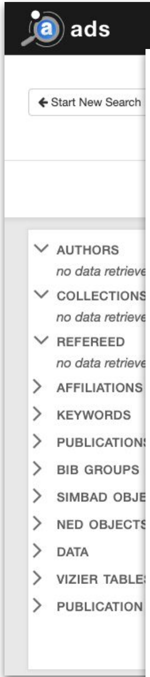
What does ADS kno … /t95q4x?

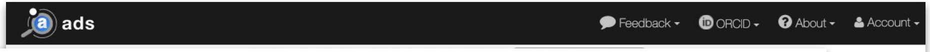
full:10.17909/T95Q4X



The record is not indexed in the ADS … T data products, but neither … n 10.17909/t95q4x

What does ADS know about 10.17909/t95q4x?





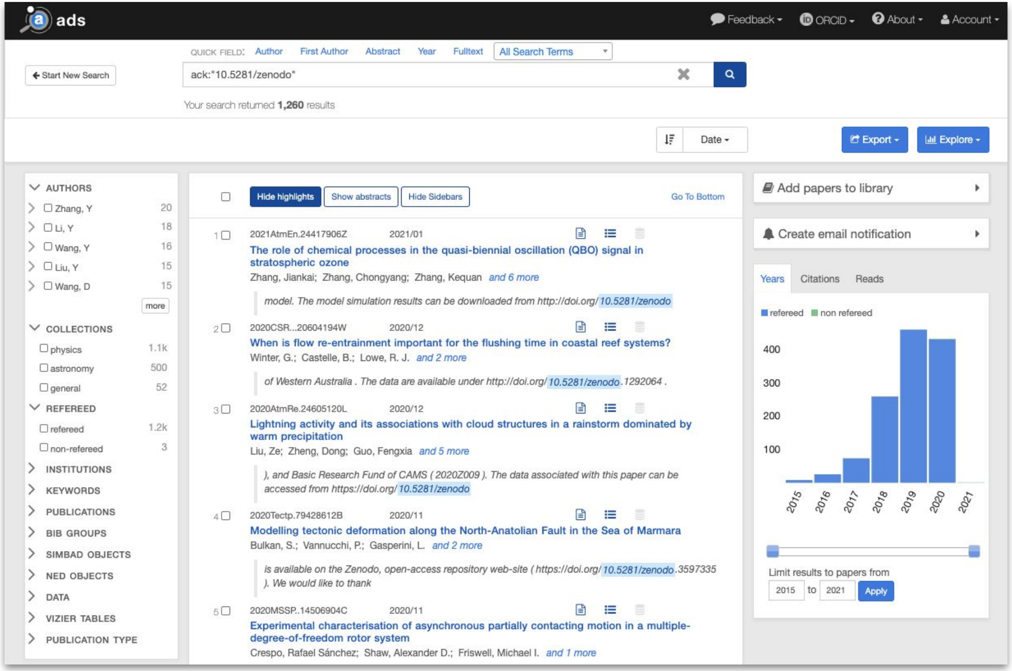
From full text:

```
<label>Appendix A</label>
<title>Data Tables and Code Repository</title>
<p>Upon publication, we will release doi:<named-content
content-type="dataset"
xlink:href="https://doi.org/10.17909/T95Q4X">10.17909/T95Q4X</named
-content>,  a suite of data files, coding routines, and
supplementary tables to replicate this analysis. This includes the
following:<list list-type="arabic" id="apjaab9bbl2">
<list-item id="apjaab9bbl2.1">
```

The rec

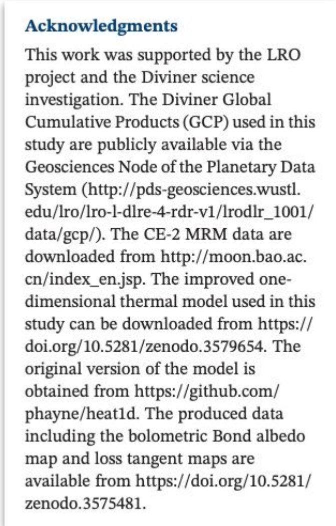
corresponds with 10.17909/t95q4x

More examples



ack:"10.5281/zenodo" finds 1,260 records
full:"10.5281/zenodo" finds 6,256 records

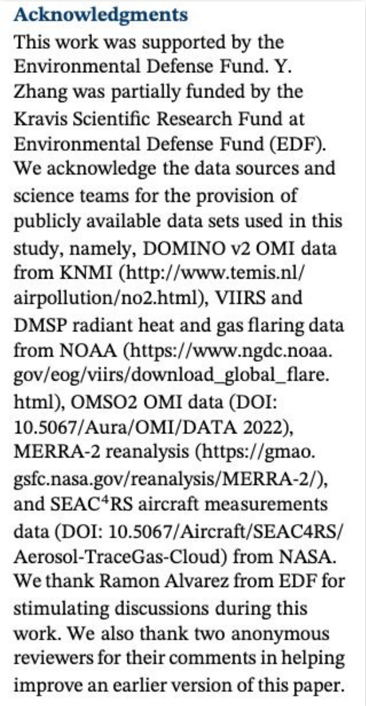
2020JGRE..12506130F



Acknowledgments

This work was supported by the LRO project and the Diviner science investigation. The Diviner Global Cumulative Products (GCP) used in this study are publicly available via the Geosciences Node of the Planetary Data System (http://pds-geosciences.wustl.edu/lro/lro-l-dlre-4-rdr-v1/lrodlr_1001/data/gcp/). The CE-2 MRM data are downloaded from http://moon.bao.ac.cn/index_en.jsp. The improved one-dimensional thermal model used in this study can be downloaded from https://doi.org/10.5281/zenodo.3579654. The original version of the model is obtained from https://github.com/phayne/heat1d. The produced data including the bolometric Bond albedo map and loss tangent maps are available from https://doi.org/10.5281/zenodo.3575481.

2019GeoRL..46.1879Z



Acknowledgments

This work was supported by the Environmental Defense Fund. Y. Zhang was partially funded by the Kravis Scientific Research Fund at Environmental Defense Fund (EDF). We acknowledge the data sources and science teams for the provision of publicly available data sets used in this study, namely, DOMINO v2 OMI data from KNMI (http://www.temis.nl/airpollution/no2.html), VIIRS and DMSP radiant heat and gas flaring data from NOAA (https://www.ngdc.noaa.gov/eog/viirs/download_global_flare.html), OMSO2 OMI data (DOI: 10.5067/Aura/OMI/DATA 2022), MERRA-2 reanalysis (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/), and SEAC^{4}RS aircraft measurements data (DOI: 10.5067/Aircraft/SEAC4RS/Aerosol-TraceGas-Cloud) from NASA. We thank Ramon Alvarez from EDF for stimulating discussions during this work. We also thank two anonymous reviewers for their comments in helping improve an earlier version of this paper.

Acknowledgements can contain a wealth of information, potentially linkable.

How about Jupyter Notebooks?

THE ASTRONOMICAL JOURNAL, 157:64 (29pp), 2019 February https://doi.org/10.3847/1538-3881/aae8e5
© 2019. The American Astronomical Society. All rights reserved.

CrossMark

starry: Analytic Occultation Light Curves

Rodrigo Luger[1,2,5], Eric Agol[2,3,6], Daniel Foreman-Mackey[1], David P. Fleming[2,3], Jacob Lustig-Yaeger[2,3], and Russell Deitrick[2,4]
[1] Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA; rluger@flatironinstitute.org
[2] Virtual Planetary Laboratory, University of Washington, Seattle, WA, USA
[3] Department of Astronomy, University of Washington, Seattle, WA, USA
[4] Center for Space and Habitability, University of Bern, Bern, Switzerland
Received 2018 July 18; revised 2018 September 28; accepted 2018 October 14; published 2019 January 23

Abstract

We derive analytic, closed form, numerically stable solutions for the total flux received from a spherical planet, moon, or star during an occultation if the specific intensity map of the body is expressed as a sum of spherical harmonics. Our expressions are valid to arbitrary degree and may be computed recursively for speed. The formalism we develop here applies to the computation of stellar transit light curves, planetary secondary eclipse light curves, and planet–planet/planet–moon occultation light curves, as well as thermal (rotational) phase curves. In this paper, we also introduce starry, an open-source package written in C++ and wrapped in Python that computes these light curves. The algorithm in starry is six orders of magnitude faster than direct numerical integration and several orders of magnitude more precise. starry also computes analytic derivatives of the light curves with respect to all input parameters for use in gradient-based optimization and inference, such as Hamiltonian Monte Carlo (HMC), allowing users to quickly and efficiently fit observed light curves to infer properties of a celestial body's surface map. (Please see https://github.com/rodluger/starry, https://rodluger.github.io/starry/, and https://doi.org/10.5281/zenodo.1312286).

Key words: eclipses – methods: analytical – occultations – techniques: photometric

1. Introduction

expected to dramatically push the boundaries of what can be inferred from these observations, potentially leading to the

This paper is organized as follows. In Section 2, we discuss the real spherical harmonics and introduce our mathematical formalism for dealing with spherical harmonic surface maps. In Section 3, we discuss how to compute analytic thermal phase curves and occultation light curves for these surface maps. In Section 4, we introduce our light curve code, starry, and discuss how to use it to compute full light curves for systems of exoplanets and other celestial bodies. We present important caveats in Section 5 and conclude in Section 6. Most of the math, including the derivations of the analytic expressions for the light curves, is folded into the appendices. For convenience, throughout the paper, we provide links to the Python[7] code to reproduce all of the figures, as well as links to Jupyter[8] notebooks containing proofs and derivations of the principal equations. Finally, Table 1 lists all of the symbols used in the paper, with references to the equations defining them.

2019AJ....157...64L

Software mentions

University Digital Conservancy Home / University of Minnesota Twin Cities / Data Curation Network / Data Curation Network Primers / View Item

Jupyter Notebooks: A Primer for Data Curators

Bouquin, Daina; Hou, Sophie; Benzing, Matthew; Wilson, Lee (Data Curation Network, 2019)

View/Download file
Jupyter_Primer_020619.pdf (800.6Kb application/pdf)

Persistent link to this item
http://hdl.handle.net/11299/202815

Services
Full metadata (XML)
View usage statistics

Title
Jupyter Notebooks: A Primer for Data Curators

Authors
Bouquin, Daina
Hou, Sophie
Benzing, Matthew
Wilson, Lee

Issue Date
2019

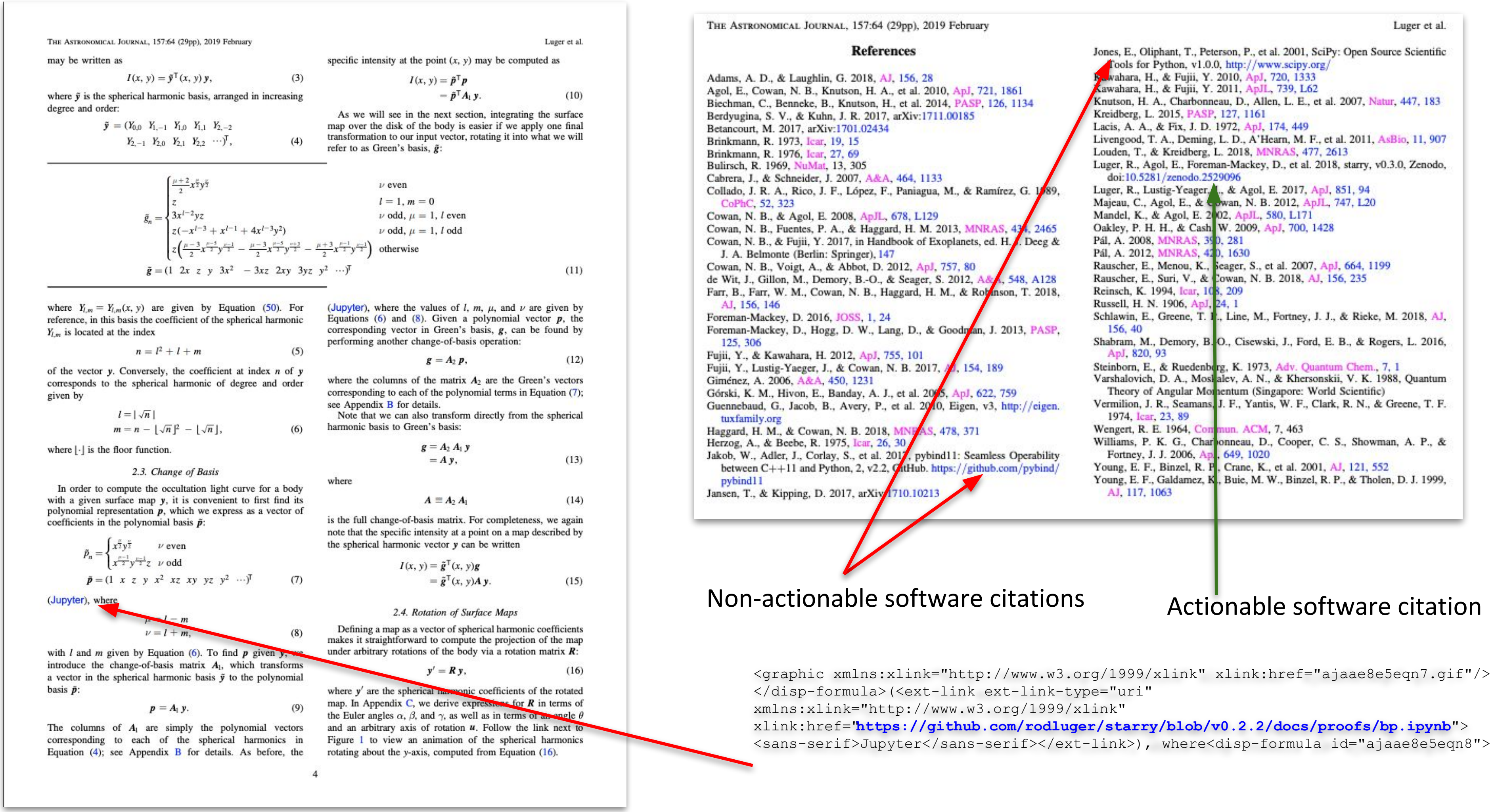
Publisher
Data Curation Network

Type
Manual or Documentation

Abstract
Jupyter Notebooks are composite digital objects used to develop, share, view, and execute interspersed, interlinked, and interactive documentation, equations, visualizations, and code. Researchers seeking to deposit software, in this case Jupyter Notebooks, in repositories do so with the expectation that repositories will provide documentation explaining "what you can deposit, the supported file formats for deposits, what metadata you may need to provide, how to provide this metadata and what happens after you make your deposit" (Jackson, 2018a). This expectation is not necessarily met by repositories that currently accept software deposits and complex objects like Jupyter Notebooks. This guide is meant to both inform curatorial practices around Jupyter Notebooks, and support the development of resources that meet researchers' expectations to ensure long-term availability of software in curated archival repositories. Guidance provided by Jisc and the Software Sustainability Institute outlines three different kinds of software deposits: a minimal deposit, a runnable deposit, and a comprehensive deposit (Jackson, 2018b). This primer follows this same conceptual framework in dealing with Jupyter Notebooks, which even in their static, non-executable form, can be used to document how scientific research was carried out or be used as teaching models among many other use cases.

Appears in collections
Data Curation Network Primers [23]

Description
This work was created as part of the Data Curation Network "Specialized Data Curation" Workshop #1 co-located with the Digital Library Federation (DLF) Forum 2018 in Las Vegas, Nevada on October 17-18, 2018.



may be written as

$$I(x, y) = \tilde{y}^{\mathsf{T}}(x, y)\, \boldsymbol{y}, \quad (3)$$

where $\tilde{y}$ is the spherical harmonic basis, arranged in increasing degree and order:

$$\tilde{y} = (Y_{0,0}\ Y_{1,-1}\ Y_{1,0}\ Y_{1,1}\ Y_{2,-2}\ Y_{2,-1}\ Y_{2,0}\ Y_{2,1}\ Y_{2,2}\ \cdots)^{\mathsf{T}}, \quad (4)$$

specific intensity at the point (x, y) may be computed as

$$I(x, y) = \tilde{p}^{\mathsf{T}} \boldsymbol{p} = \tilde{p}^{\mathsf{T}} A_1\, \boldsymbol{y}. \quad (10)$$

As we will see in the next section, integrating the surface map over the disk of the body is easier if we apply one final transformation to our input vector, rotating it into what we will refer to as Green's basis, $\tilde{g}$:

$$\tilde{g}_n = \begin{cases} \frac{\mu+2}{2} x^{\frac{\mu}{2}} y^{\frac{\nu}{2}} & \nu \text{ even} \\ z & l = 1, m = 0 \\ 3x^{l-2} yz & \nu \text{ odd}, \mu = 1, l \text{ even} \\ z(-x^{l-3} + x^{l-1} + 4x^{l-3}y^2) & \nu \text{ odd}, \mu = 1, l \text{ odd} \\ z\left(\frac{\mu-3}{2} x^{\frac{\mu-5}{2}} y^{\frac{\nu-1}{2}} - \frac{\mu-3}{2} x^{\frac{\mu-5}{2}} y^{\frac{\nu+3}{2}} - \frac{\mu+3}{2} x^{\frac{\mu-1}{2}} y^{\frac{\nu-1}{2}}\right) & \text{otherwise} \end{cases}$$

$$\tilde{g} = (1\ \ 2x\ \ z\ \ y\ \ 3x^2\ \ -3xz\ \ 2xy\ \ 3yz\ \ y^2\ \ \cdots)^{\mathsf{T}} \quad (11)$$

where $Y_{l,m} = Y_{l,m}(x, y)$ are given by Equation (50). For reference, in this basis the coefficient of the spherical harmonic $Y_{l,m}$ is located at the index

$$n = l^2 + l + m \quad (5)$$

of the vector $\boldsymbol{y}$. Conversely, the coefficient at index n of $\boldsymbol{y}$ corresponds to the spherical harmonic of degree and order given by

$$l = \lfloor \sqrt{n} \rfloor, \quad m = n - \lfloor \sqrt{n} \rfloor^2 - \lfloor \sqrt{n} \rfloor, \quad (6)$$

where $\lfloor \cdot \rfloor$ is the floor function.

2.3. Change of Basis

In order to compute the occultation light curve for a body with a given surface map $\boldsymbol{y}$, it is convenient to first find its polynomial representation $\boldsymbol{p}$, which we express as a vector of coefficients in the polynomial basis $\tilde{p}$:

$$\tilde{p}_n = \begin{cases} x^{\frac{\mu}{2}} y^{\frac{\nu}{2}} & \nu \text{ even} \\ x^{\frac{\mu-1}{2}} y^{\frac{\nu-1}{2}} z & \nu \text{ odd} \end{cases}$$

$$\tilde{p} = (1\ \ x\ \ z\ \ y\ \ x^2\ \ xz\ \ xy\ \ yz\ \ y^2\ \ \cdots)^{\mathsf{T}} \quad (7)$$

(Jupyter), where

$$\mu = l - m, \quad \nu = l + m, \quad (8)$$

with l and m given by Equation (6). To find $\boldsymbol{p}$ given $\boldsymbol{y}$, we introduce the change-of-basis matrix A_1, which transforms a vector in the spherical harmonic basis $\tilde{y}$ to the polynomial basis $\tilde{p}$:

$$\boldsymbol{p} = A_1\, \boldsymbol{y}. \quad (9)$$

The columns of A_1 are simply the polynomial vectors corresponding to each of the spherical harmonics in Equation (4); see Appendix B for details. As before, the

(Jupyter), where the values of l, m, μ, and ν are given by Equations (6) and (8). Given a polynomial vector $\boldsymbol{p}$, the corresponding vector in Green's basis, $\boldsymbol{g}$, can be found by performing another change-of-basis operation:

$$\boldsymbol{g} = A_2\, \boldsymbol{p}, \quad (12)$$

where the columns of the matrix A_2 are the Green's vectors corresponding to each of the polynomial terms in Equation (7); see Appendix B for details.

Note that we can also transform directly from the spherical harmonic basis to Green's basis:

$$\boldsymbol{g} = A_2\, A_1\, \boldsymbol{y} = A\, \boldsymbol{y}, \quad (13)$$

where

$$A \equiv A_2\, A_1 \quad (14)$$

is the full change-of-basis matrix. For completeness, we again note that the specific intensity at a point on a map described by the spherical harmonic vector $\boldsymbol{y}$ can be written

$$I(x, y) = \tilde{g}^{\mathsf{T}}(x, y)\boldsymbol{g} = \tilde{g}^{\mathsf{T}}(x, y) A\, \boldsymbol{y}. \quad (15)$$

2.4. Rotation of Surface Maps

Defining a map as a vector of spherical harmonic coefficients makes it straightforward to compute the projection of the map under arbitrary rotations of the body via a rotation matrix $\boldsymbol{R}$:

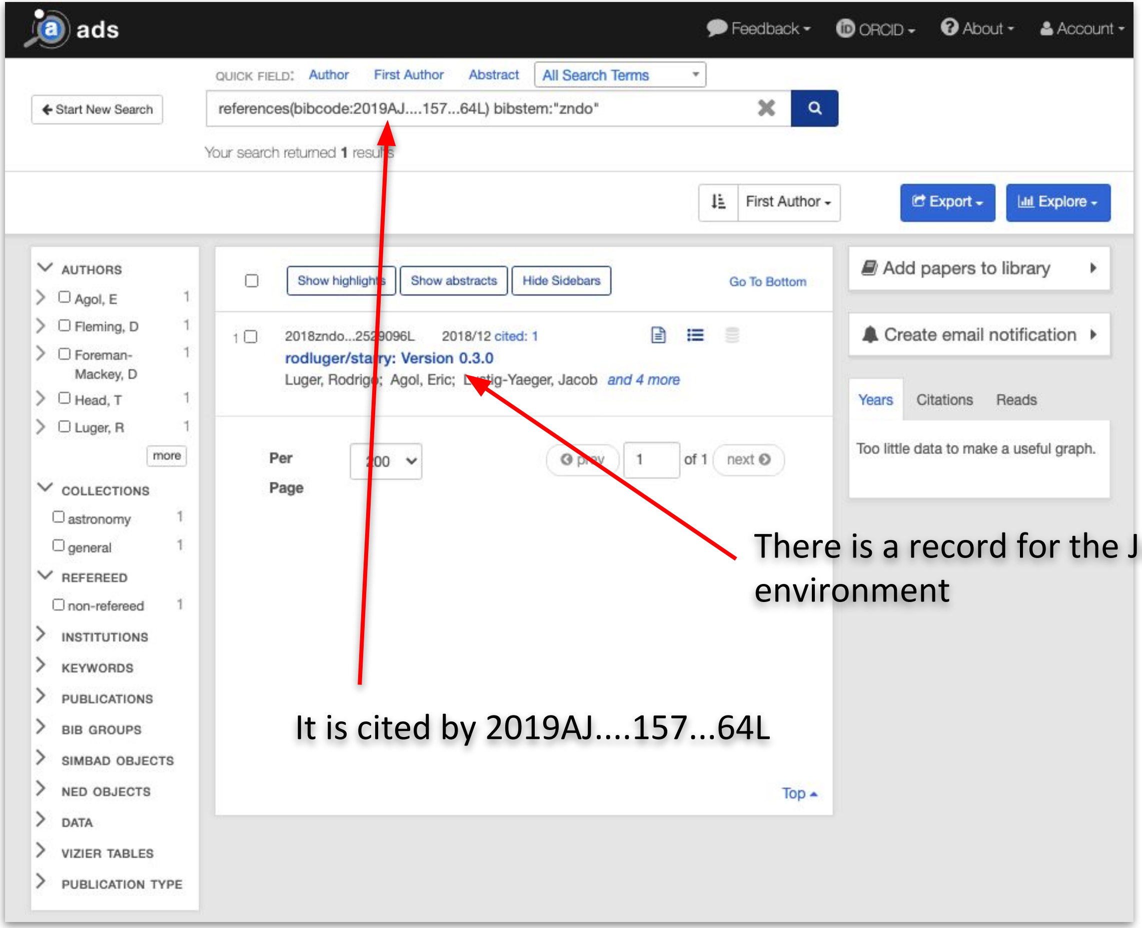
$$\boldsymbol{y}' = \boldsymbol{R}\, \boldsymbol{y}, \quad (16)$$

where $\boldsymbol{y}'$ are the spherical harmonic coefficients of the rotated map. In Appendix C, we derive expressions for $\boldsymbol{R}$ in terms of the Euler angles α, β, and γ, as well as in terms of an angle θ and an arbitrary axis of rotation $\boldsymbol{u}$. Follow the link next to Figure 1 to view an animation of the spherical harmonics rotating about the y-axis, computed from Equation (16).

References

Adams, A. D., & Laughlin, G. 2018, AJ, 156, 28
Agol, E., Cowan, N. B., Knutson, H. A., et al. 2010, ApJ, 721, 1861
Biechman, C., Benneke, B., Knutson, H., et al. 2014, PASP, 126, 1134
Berdyugina, S. V., & Kuhn, J. R. 2017, arXiv:1711.00185
Betancourt, M. 2017, arXiv:1701.02434
Brinkmann, R. 1973, Icar, 19, 15
Brinkmann, R. 1976, Icar, 27, 69
Bulirsch, R. 1969, NuMat, 13, 305
Cabrera, J., & Schneider, J. 2007, A&A, 464, 1133
Collado, J. R. A., Rico, J. F., López, R., Paniagua, M., & Ramírez, G. 1989, CoPhC, 52, 323
Cowan, N. B., & Agol, E. 2008, ApJL, 678, L129
Cowan, N. B., Fuentes, P. A., & Haggard, H. M. 2013, MNRAS, 434, 2465
Cowan, N. B., & Fujii, Y. 2017, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte (Berlin: Springer), 147
Cowan, N. B., Voigt, A., & Abbot, D. 2012, ApJ, 757, 80
de Wit, J., Gillon, M., Demory, B.-O., & Seager, S. 2012, A&A, 548, A128
Farr, B., Farr, W. M., Cowan, N. B., Haggard, H. M., & Robinson, T. 2018, AJ, 156, 146
Foreman-Mackey, D. 2016, JOSS, 1, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fujii, Y., & Kawahara, H. 2012, ApJ, 755, 101
Fujii, Y., Lustig-Yaeger, J., & Cowan, N. B. 2017, AJ, 154, 189
Giménez, A. 2006, A&A, 450, 1231
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Guennebaud, G., Jacob, B., Avery, P., et al. 2010, Eigen, v3, http://eigen.tuxfamily.org
Haggard, H. M., & Cowan, N. B. 2018, MNRAS, 478, 371
Herzog, A., & Beebe, R. 1975, Icar, 26, 30
Jakob, W., Adler, J., Corlay, S., et al. 2017, pybind11: Seamless Operability between C++11 and Python, 2, v2.2, GitHub. https://github.com/pybind/pybind11
Jansen, T., & Kipping, D. 2017, arXiv:1710.10213
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python, v1.0.0, http://www.scipy.org/
Kawahara, H., & Fujii, Y. 2010, ApJ, 720, 1333
Kawahara, H., & Fujii, Y. 2011, ApJL, 739, L62
Knutson, H. A., Charbonneau, D., Allen, L. E., et al. 2007, Natur, 447, 183
Kreidberg, L. 2015, PASP, 127, 1161
Lacis, A. A., & Fix, J. D. 1972, ApJ, 174, 449
Livengood, T. A., Deming, L. D., A'Hearn, M. F., et al. 2011, AsBio, 11, 907
Louden, T., & Kreidberg, L. 2018, MNRAS, 477, 2613
Luger, R., Agol, E., Foreman-Mackey, D., et al. 2018, starry, v0.3.0, Zenodo, doi:10.5281/zenodo.2529096
Luger, R., Lustig-Yeager, J., & Agol, E. 2017, ApJ, 851, 94
Majeau, C., Agol, E., & Cowan, N. B. 2012, ApJL, 747, L20
Mandel, K., & Agol, E. 2002, ApJL, 580, L171
Oakley, P. H. H., & Cash, W. 2009, ApJ, 700, 1428
Pál, A. 2008, MNRAS, 390, 281
Pál, A. 2012, MNRAS, 420, 1630
Rauscher, E., Menou, K., Seager, S., et al. 2007, ApJ, 664, 1199
Rauscher, E., Suri, V., & Cowan, N. B. 2018, AJ, 156, 235
Reinsch, K. 1994, Icar, 108, 209
Russell, H. N. 1906, ApJ, 24, 1
Schlawin, E., Greene, T. P., Line, M., Fortney, J. J., & Rieke, M. 2018, AJ, 156, 40
Shabram, M., Demory, B.-O., Cisewski, J., Ford, E. B., & Rogers, L. 2016, ApJ, 820, 93
Steinborn, E., & Ruedenberg, K. 1973, Adv. Quantum Chem., 7, 1
Varshalovich, D. A., Moskalev, A. N., & Khersonskii, V. K. 1988, Quantum Theory of Angular Momentum (Singapore: World Scientific)
Vermilion, J. R., Seamans, J. F., Yantis, W. F., Clark, R. N., & Greene, T. F. 1974, Icar, 23, 89
Wengert, R. E. 1964, Commun. ACM, 7, 463
Williams, P. K. G., Charbonneau, D., Cooper, C. S., Showman, A. P., & Fortney, J. J. 2006, ApJ, 649, 1020
Young, E. F., Binzel, R. P., Crane, K., et al. 2001, AJ, 121, 552
Young, E. F., Galdamez, K., Buie, M. W., Binzel, R. P., & Tholen, D. J. 1999, AJ, 117, 1063

references(bibcode:2019AJ....157...64L) bibstem:"zndo"



There is a record for the Jupyter Notebook environment

It is cited by 2019AJ....157...64L

Summary: what is **indexed** versus **linked** in the 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
- Upcoming: cited data products, other research artifacts such as Jupyter notebooks

Indexed records are scholarly research artifacts.
They are discoverable and can be used to create bibliographies

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, Now vs Upcoming

	Mention		Citation	
	URL	DOI	URL	DOI
Use cases	in **acknowledgments**: 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 acknowledgments: 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 the ADS	Searchable in ADS fulltext / ack fields	Searchable in ADS fulltext / ack fields; Link to data product created in ADS	URL ignored, citation not counted	New record for DOI created (if necessary) and citation counted against it

Note: some publications have "availability" sections for data and software to specify the locations of data and/or software used in the publication