

SHARE your data

it's good for you, and for the world.



Come. Eat lunch. Accelerate the pace of science.

CfA, PHILLIPS Auditorium, 11:45 MONDAY 4/2/12



theastrodata.org

POWERED BY THE **Dataverse
Network**™ PROJECT

&

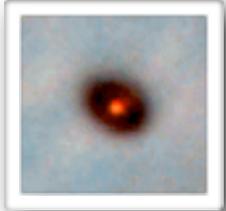
 **SEAMLESS
ASTRONOMY**
Linking scientific data, publications, and communities

Questions? Contact Gus Muench at gmunch@cfa.harvard.edu

Archetypes in a **Dataverse**



Asteroid You have small, data sets you'd like to see stay in reliable orbits.



Protostar You're young and eager to become a full-grown star, so you want to share all the data you can, and embed links to it in your publications.



Main-sequence Star You've been at this for a while, so you have long data history and a good future. You'd like to upload important data to go with "old" papers now, and more in the future.



Cluster You collect things in catalogs and lists, and you want to group the catalogs for the greater good.



Supernova Your disks are **EXPLODING** with data, and you don't know what to do with it. You want to permalink vast data sets directly to papers, and more...



Pulsar You really like it when things *change*. Time-domain astronomy is your thing, and you want online identifiers that understand time.



Galaxy You love everything, but you're organized. You make and collect Surveys you don't want to lose, and you want people to find them from far away.



Quasar Your energy is nearly unlimited, so you suck up (mine) and spit out as much data as you can find. And you like to share in showy ways.

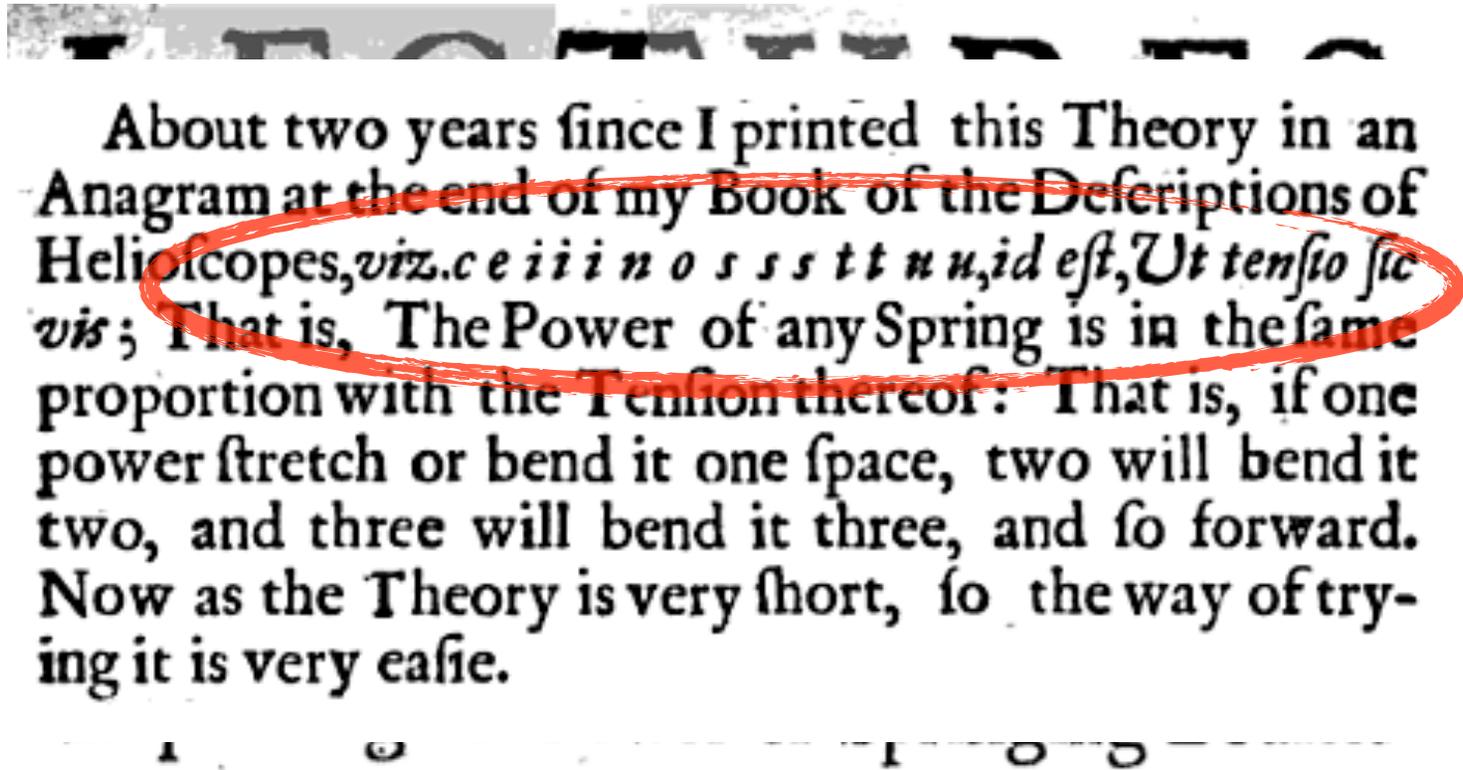


Black Hole You suck down any and all data, with unbridled appetite. Dataverse is *NOT* for you.

The Astronomy Dataverse

or how you can change the world & start publishing YOUR data

1660 Robert Hooke “pre” published as anagram



scientists ~~big~~ things .data

in the present data live...

The screenshot shows a web browser window with several tabs open. The active tab is titled "iopscience.iop.org.ezp-prod1.hul.harvard.edu/0067-0049/166/1/249/fulltext". The page content includes the journal title "THE ASTROPHYSICAL JOURNAL SUPPLEMENT SERIES", the article title "Hubble Space Telescope Proper Motions and Stellar Dynamics in the Core of the Globular Cluster 47 Tucanae", the authors "Dean E. McLaughlin, Jay Anderson, Georges Meylan, Karl Gebhardt, Carlton Pryor, Dante Minniti and Sterl Phinney", and the date "Received 2006 February 9; accepted 2006 May 5". The abstract text is visible, starting with "We have used HST imaging of the central regions of the globular cluster 47 Tucanae (=NGC 104), taken with the V...". The left sidebar contains navigation links such as "Journal Homepage", "Previous", "Next", "This volume", "This issue", "This abstract", "Search", "Online Material" (with links for FITS file, Master frame reference image, Machine-readable table, Table 4, Tar file, Proper-motion catalog), "Article Contents" (with a detailed table of contents), "Download PDF", and "Related links".

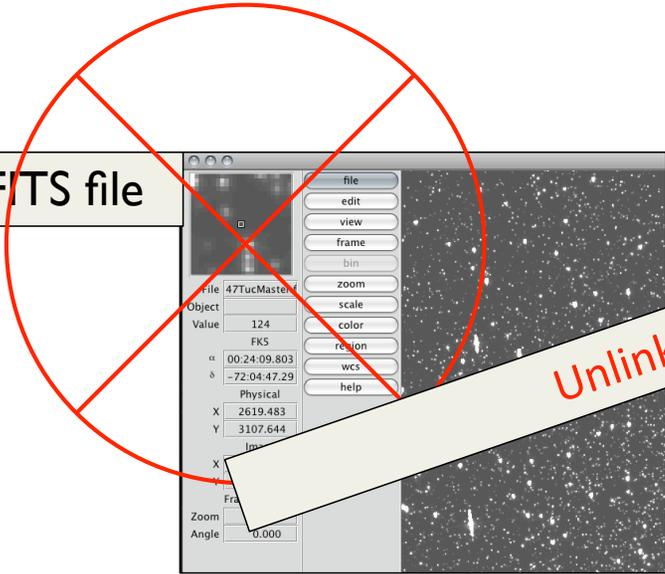
Online Material
FITS file
Master frame reference image
Machine-readable table
Table 4
Tar file
Proper-motion catalog



Some tables too

Table Name	Description
<input type="checkbox"/> J/ApJS/166/249/table3	(c) Stars used for astrometric calibration
<input type="checkbox"/> J/ApJS/166/249/table4	(c) 129733 stars in the Master Frame
<input type="checkbox"/> J/ApJS/166/249/table5	(c) Proper motions and displacements

with a FITS file



and some code,

```

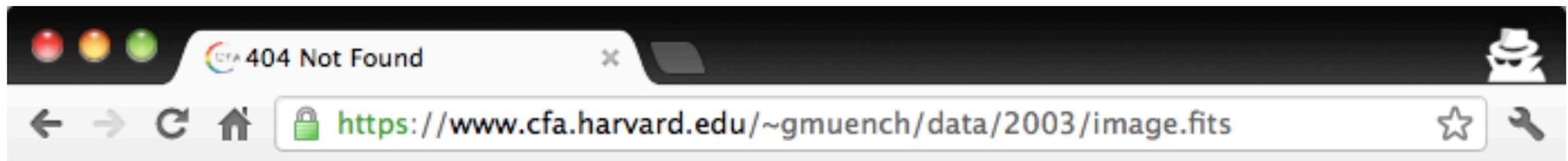
1 #
2 # MACRO pmdat REQUIRES ONE COMMAND-LINE ARGUMENT ...
3 #
4 # $1 = ID number of a star in file '___datafile'
5 # in either of two formats: M11111 = an exact ID label in the file
6 # : 11111 = integer part [>0] of a known ID
7 # -- OR --
8 #
9 # $1<=0 or $1=[any string not starting with 'M']
10 # will choose a star at random from the file '___datafile'
11 #
12 #
13 # MACRO pmdat WILL OPTIONALLY TAKE A SECOND ARGUMENT...
14 #
15 # $2 = PRINT [optional]
16 #
17 # if second argument exists and is PRINT
18 # then star data are printed to file with extension '.DATA'
19 #
20 # if second argument is anything else, or does not exist at all,
21 # then star data are echoed to screen instead.
22 #
23 #
24 #
25 # Usage... at the sm prompt, type

```

Unlinked Data is LOST data!

References: McLaughlin et al. 2006; <http://adsabs.harvard.edu/abs/2006ApJS..166..249M>

in the present data live...



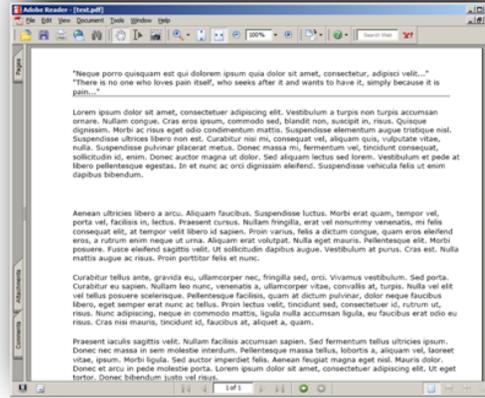
Not Found

The requested URL `/~gmuench/data/2003/image.fits` was not found on this server.

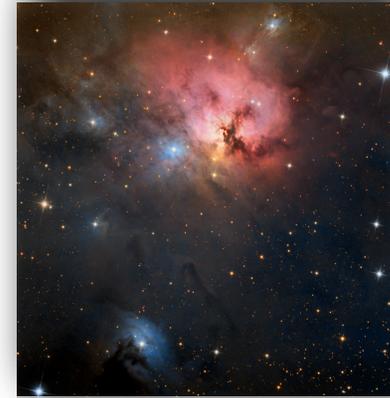
in the present data live...

x0	x1	x2	x3	x4	x5
1	1.06	9.2	151	54.4	1.6
2	0.89	10.3	202	57.9	2.2
3	1.43	15.4	113	53	3.4
4	1.02	11.2	168	56	0.3
5	1.49	8.8	192	51.2	1
6	1.32	13.5	111	60	-2.2
7	1.22	12.2	175	67.6	2.2
8	1.1	9.2	245	57	3.3
9	1.34	13	168	60.4	7.2
10	1.12	12.4	197	53	2.7
11	0.75	7.5	173	51.5	6.5
12	1.13	10.9	178	62	3.7
13	1.15	12.7	199	53.7	6.4
14	1.09	12	96	49.8	1.4
15	0.96	7.6	164	62.2	-0.1
16	1.16	9.9	252	56	9.2
17	0.76	6.4	136	61.9	9
18	1.05	12.6	150	56.7	2.7
19	1.16	11.7	104	54	-2.1
20	1.2	11.8	148	59.9	3.5
21	1.04	8.6	204	61	3.5
22	1.07	9.3	174	54.3	5.9

Data sets

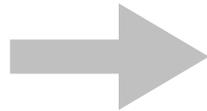


Text



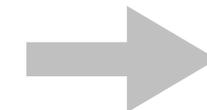
Images

Keep it in my laptop



Hard to share, not safe

Send it to an archive



Author doesn't get enough credit

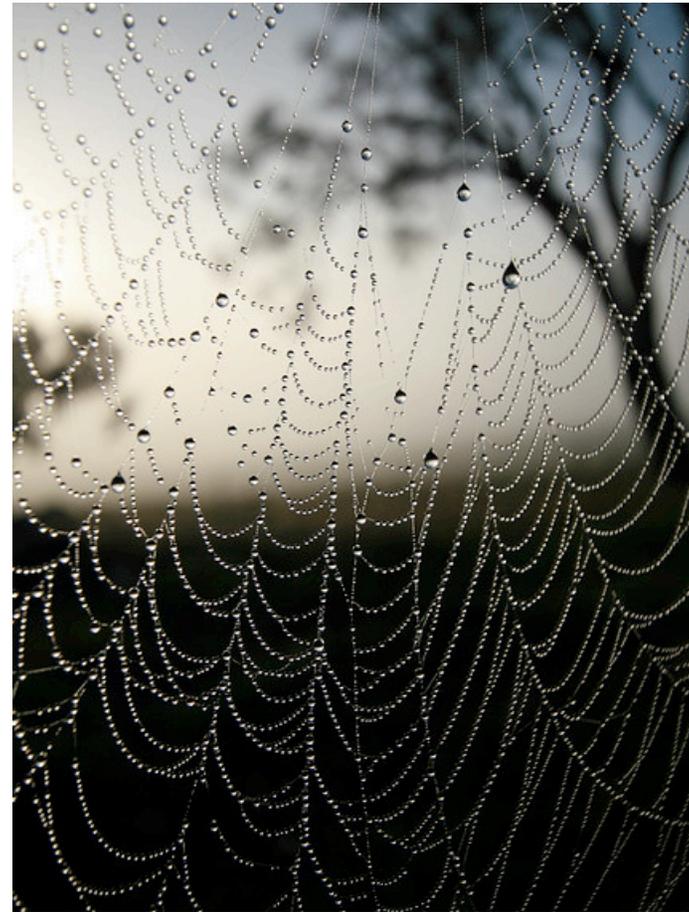
Post it in my web site



No professional archiving services
(backups, recovering, cataloging,
preservation, etc)

in the future data live...

- Refined data sets are published by YOU in long lived repositories;
- Your data appear in ADS & are “searchable”
- Your data are **reused** and **cited**, giving you credit for that work.



So how do we get
there?



An Open-Source Application for
Publishing, Citing and Discovering Research Data

Created at  **IQSS at Harvard University**

- ☑ Gives **ownership and recognition** to data owner
- ☑ Generates a **persistent data citation**
- ☑ Converts data sets to a **preservable** and verifiable format
- ☑ Distributes data to the **public**, but also supports **restricted** access
- ☑ Indexes all metadata for quick data **discovery**
- ☑ Supports **subsetting and analysis** for (some) data files
- ☑ Can be branded as yours
- ☑ **Inter-operates** with other systems using **standards**

How do Astronomers Use a Dataverse?

or what do you need to publish your data.

we asked a bunch of you some questions

“Mostly FITS”

“thousands of lines, hundreds of columns. hundreds of MBs at most.”

“Terabite-ish.”

“Currently KB, MB
(reduced)”

“General public”

“No. No Licensing; No obligations.”

we asked a bunch of you some questions

“I don't have a website where I store these data. Most of it is in various stages of mess.”

“if we were rich and organized, we would be like Sloan...”

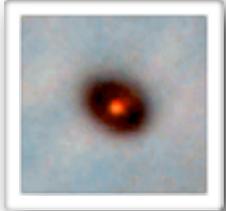
“Visibility from ADS, Vizier, arXiv... Interface: 1. ability to retrieve the data, 2. simple visualization, 3. VO-interoperability”

“We don't anticipate any fancy interactive data browsing capability. You just download the data and you do anything you like with it.”

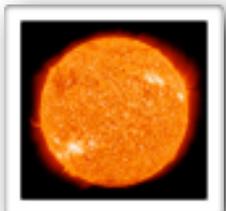
Archetypes in a **Dataverse**



Asteroid You have small, data sets you'd like to see stay in reliable orbits.



Protostar You're young and eager to become a full-grown star, so you want to share all the data you can, and embed links to it in your publications.



Main-sequence Star You've been at this for a while, so you have long data history and a good future. You'd like to upload important data to go with "old" papers now, and more in the future.



Cluster You collect things in catalogs and lists, and you want to group the catalogs for the greater good.



Supernova Your disks are **EXPLODING** with data, and you don't know what to do with it. You want to permalink vast data sets directly to papers, and more...



Pulsar You really like it when things *change*. Time-domain astronomy is your thing, and you want online identifiers that understand time.



Galaxy You love everything, but you're organized. You make and collect Surveys you don't want to lose, and you want people to find them from far away.



Quasar Your energy is nearly unlimited, so you suck up (mine) and spit out as much data as you can find. And you like to share in showy ways.



Black Hole You suck down any and all data, with unbridled appetite. Dataverse is *NOT* for you.

- Example run through of using the DVN
- Personal Experience from a CfA Astronomer
- Panel Q&A



Data Linking

Table 2 [CITED IN TEXT](#) | [ASCII](#) | [TYPESET IMAGE](#) Go to: [Table 1](#) | [Table 3](#)

WFPC2 and ACS Observations of 47 Tucanae

Data Set	Program ID	N_{obs}	Filter	Date
MEYLANe1...	5912	15	F300W	1995 Oct 25 = 1995.82
MEYLANe2...	6467	16	F300W	1997 Nov 3 = 1997.84
GILLILU1...	8267	28	F336W	1999 Jul 5 = 1999.51
MEYLANe3...	7503	16	F300W	1999 Oct 28 = 1999.82
GILLILU2...	9266	11	F336W	2001 Jul 13 = 2001.53
WFC-MEUR...	9028	20	F475W	2002 Apr 5 = 2002.26
HRC-MEUR...	9028	40	F475W	2002 Apr 5 = 2002.26
HRC-BOHL...	9019	10	F475W	2002 Apr 13 = 2002.28
WFC-KING...	9443	6	F475W	2002 Jul 7 = 2002.52
HRC-KING...	9443	20	F475W	2002 Jul 24 = 2002.56

http://archive.stsci.edu/cgi-bin/proposal_search?mission=hst&id=5912

2.2.3. Astr

<http://archive.stsci.edu/cgi-bin/mastpreview?mission=hst&dataid=U2VO0101T>

We now have a position for the cluster center in the reference frame, which is based on the distortion-corrected and rotated frame of the first image of GO-9028. In order to transform the master-frame positions into absolute right ascension and declination, we used the image header information from several WFPC2 images ([U2TY0201T](#), [U2VO0101T](#), [U4F40101R](#), and [U5JM120DR](#)) to obtain absolute positions for seven stars—five stars at the center and two stars in the outskirts. These four images were taken at different pointings and orientations, so they should all use different guide stars and give independent estimates of the absolute coordinates.

Someone added these article links to point to archive websites.

Links not present in *arXiv* preprint

Post publication curation? Why?

Author create links to their own data.