SHARE your data

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

Come. Eat lunch. Accelerate the pace of science.

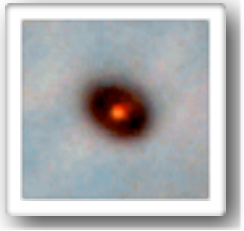
theastrodata.org

POWERED BY THE Dataverse Network™ PROJECT & SEAMLESS ASTRONOMY
Linking scientific data, publications, and communities

Questions? Contact Gus Muench at gmuench@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.

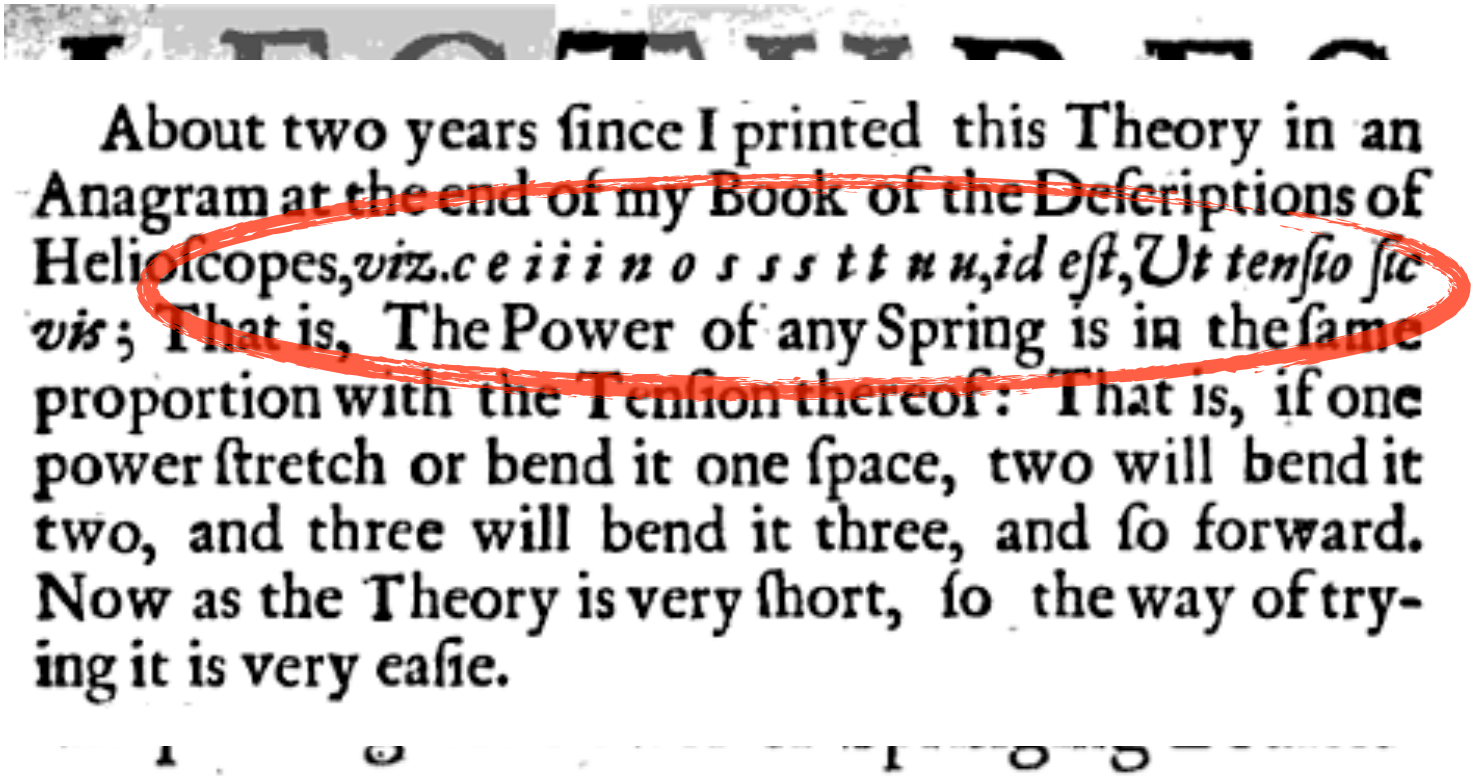
The Dataverse Network™ Project

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

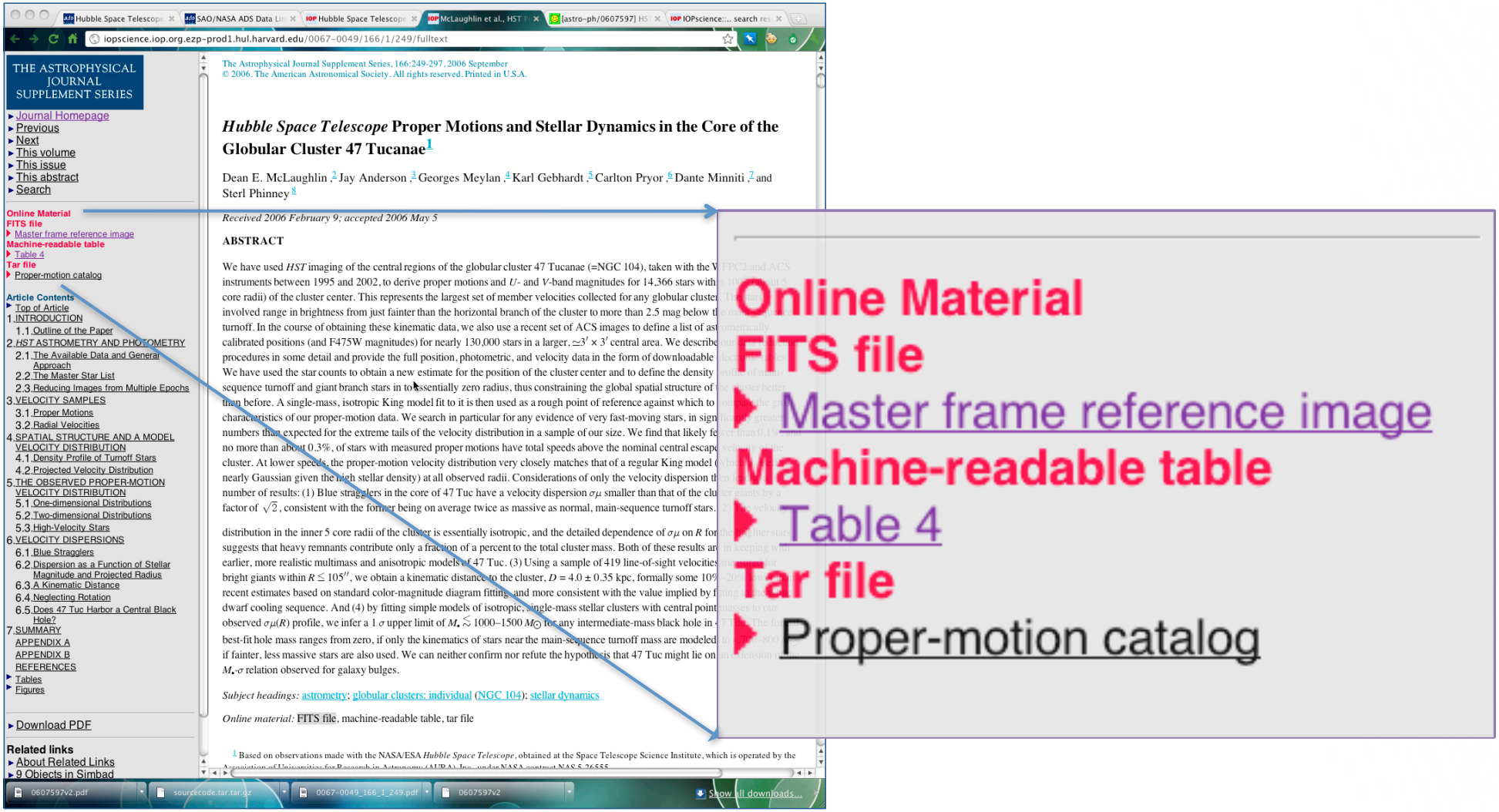


About two years ſince I printed this Theory in an Anagram at the end of my Book of the Deſcriptions of Helioſcopes, *viz. c e i i i n o s s s t t u u, id eſt, Ut tenſio ſic vis*; That is, The Power of any Spring is in the ſame proportion with the Tenſion thereof: That is, if one power ſtretch or bend it one ſpace, two will bend it two, and three will bend it three, and ſo forward. Now as the Theory is very ſhort, ſo the way of trying it is very eaſie.

scientists big data

References: Nielsen, M. "*The Future of Science*" http://michaelnielsen.org/blog/the-future-of-science-2/

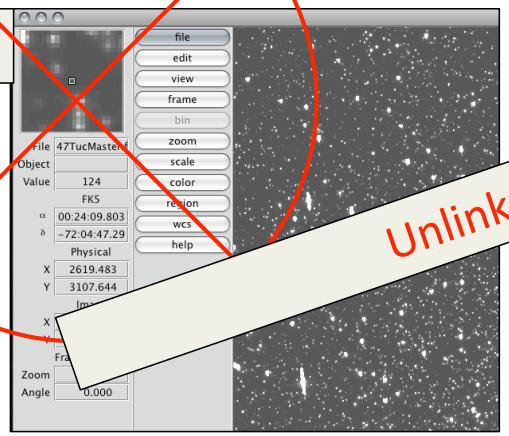
in the present data live...



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

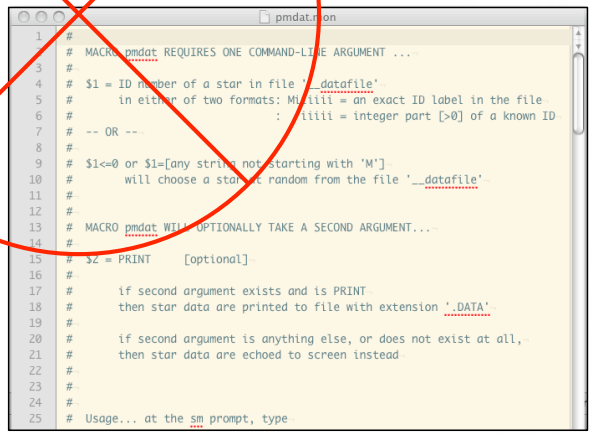
Some tables too

with a FITS file



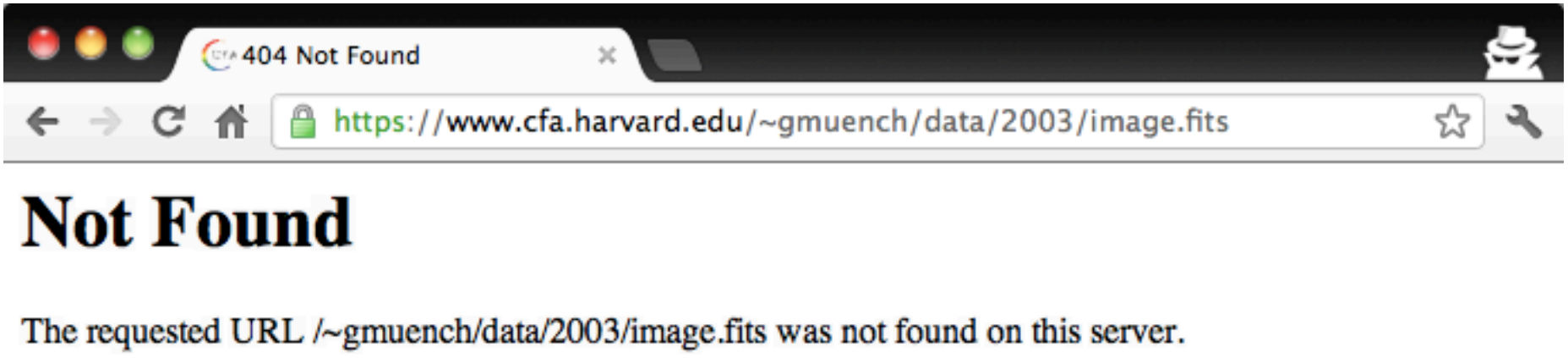
Unlinked Data is LOST data!

and some code,

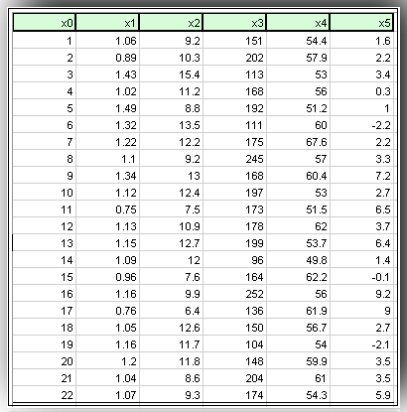


References: McLaughlin et al. 2006; http://adsabs.harvard.edu/abs/2006ApJS..166..249M

in the present data live...

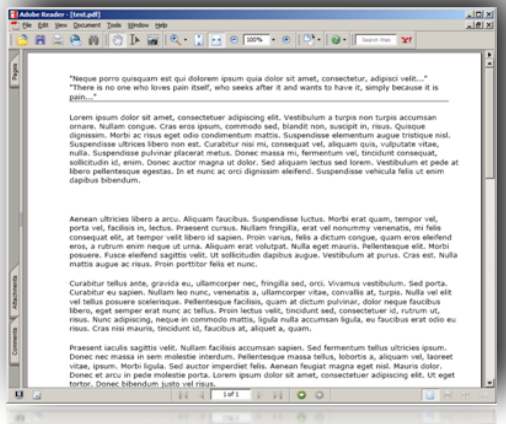


in the present data live...



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."

"Currently KB, MB (reduced)"

"Terabite-ish."

"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."

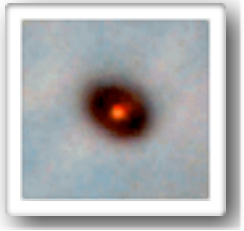
"if we were rich and organized, we would be like Sloan..."

"Visbility 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.

The Dataverse Network™ Project

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

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.

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

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

2.2.3. Astr

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.