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

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.

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 Heliocopes, viz. ceiino sssttuu, 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.

with a FITS file

Some tables too


Unlinked Data is LOST data!

and some code,
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...

Data sets

Keep it in my laptop

Send it to an archive

Post it in my web site

Hard to share, not safe

Author doesn’t get enough credit

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?
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)"

“No. No Licensing; No obligations.”

“General public”
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.”

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

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

“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

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>
<thead>
<tr>
<th>Data Set</th>
<th>Program ID</th>
<th>N_H</th>
<th>Filter</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEYLANe1...</td>
<td>5912</td>
<td>15</td>
<td>F380W</td>
<td>1995 Oct 25 = 1995.82</td>
</tr>
<tr>
<td>MEYLANe2...</td>
<td>6467</td>
<td>16</td>
<td>F380W</td>
<td>1997 Nov 3 = 1997.84</td>
</tr>
<tr>
<td>GILLELU1...</td>
<td>8267</td>
<td>28</td>
<td>F336W</td>
<td>1999 Jul 5 = 1999.51</td>
</tr>
<tr>
<td>MEYLANe3...</td>
<td>2503</td>
<td>16</td>
<td>F380W</td>
<td>1999 Oct 28 = 1999.82</td>
</tr>
<tr>
<td>GILLELU2...</td>
<td>9266</td>
<td>11</td>
<td>F336W</td>
<td>2001 Jul 13 = 2001.53</td>
</tr>
<tr>
<td>WFC-MEER...</td>
<td>9024</td>
<td>20</td>
<td>F475W</td>
<td>2002 Apr 5 = 2002.26</td>
</tr>
<tr>
<td>HRC-MEER...</td>
<td>9028</td>
<td>40</td>
<td>F475W</td>
<td>2002 Apr 5 = 2002.26</td>
</tr>
<tr>
<td>HRC-BH...</td>
<td>9019</td>
<td>10</td>
<td>F475W</td>
<td>2002 Apr 13 = 2002.28</td>
</tr>
<tr>
<td>WFC-KING...</td>
<td>9143</td>
<td>6</td>
<td>F475W</td>
<td>2002 Jul 7 = 2002.52</td>
</tr>
<tr>
<td>HRC-KING...</td>
<td>9143</td>
<td>20</td>
<td>F475W</td>
<td>2002 Jul 24 = 2002.56</td>
</tr>
</tbody>
</table>

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

http://archive.stsci.edu/cgi-bin/mastpreview?mission=hst&dataid=U2V00101T

2.2.3. Ast

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, U2VC0101T, U4E0101R, and U5N1201R) 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.