SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

Astronomy Dataverse:

enabling astronomer data publishing

http://theastrodata.org

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

CfA

Harvard-Smithsonian Center for Astrophysics

ads NASA

HARVARD LIBRARY

CDS Centre de Données astronomiques de Strasbourg

VAO

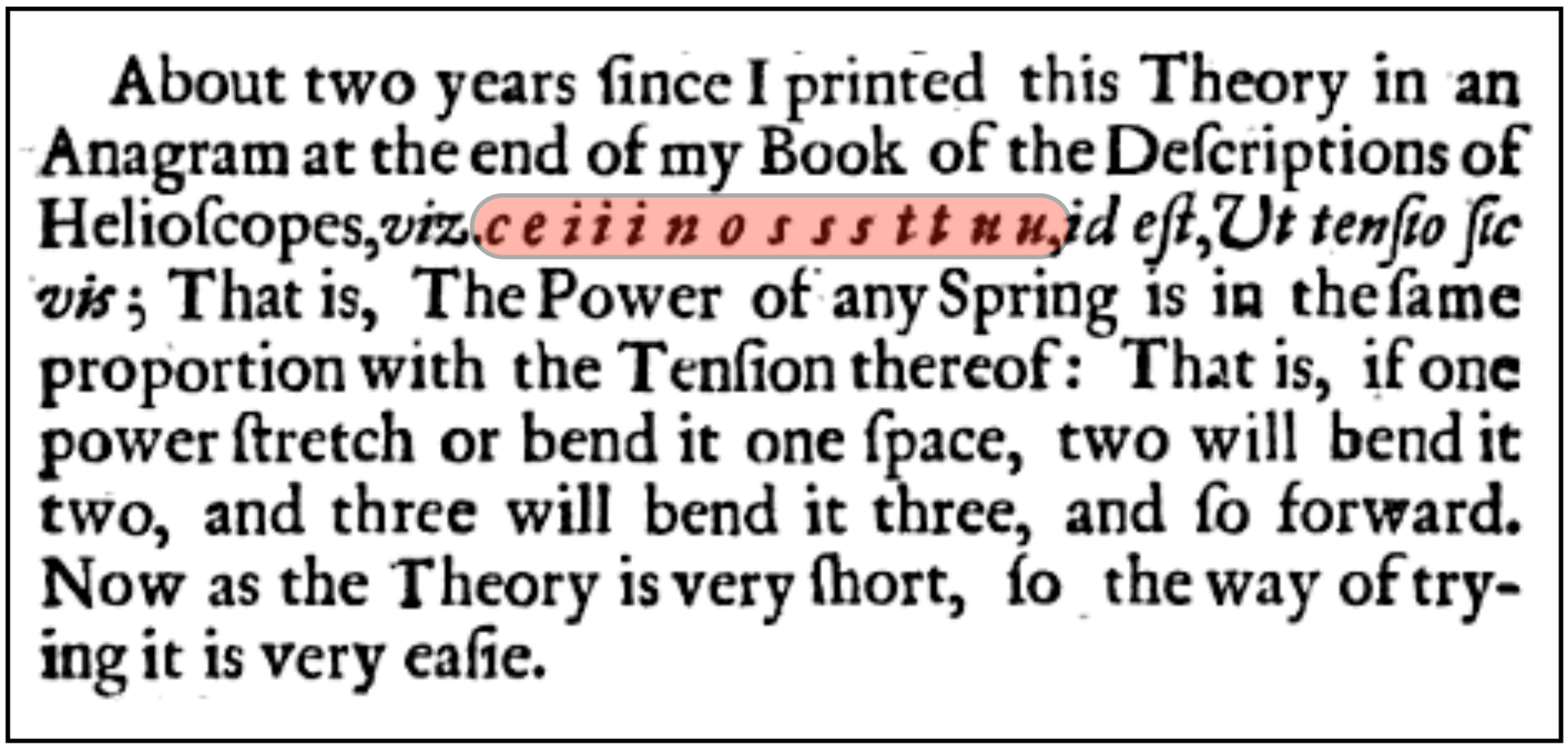
The Dataverse Network™ Project

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


LECTURES
De Potentia Restitutiva,
OR OF
SPRING
Explaining the Power of Springing Bodies.

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

in the past data were hidden...



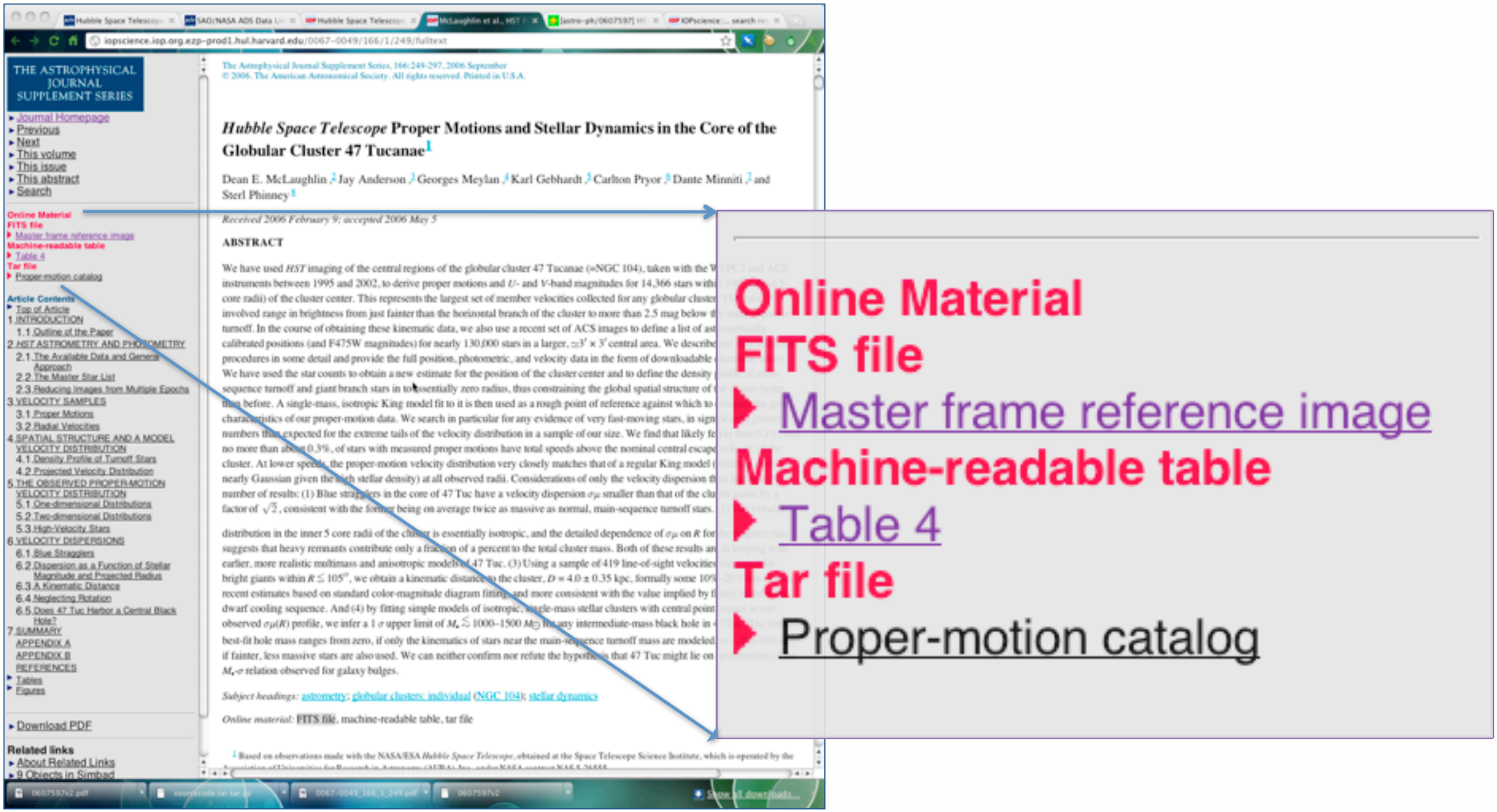
About two years ſince I printed this Theory in an Anagram at the end of my Book of the Deſcriptions of Helioſcopes, *viz.* *c e i i i n o s s s t t u u*, *id eſt*, *Ut tenſio ſic vis*; That is, The Power of any Spring is in the ſame proportion with the Tenſion thereof: That is, if one power ſtretch or 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.

1660 Robert Hooke "pre" published anagram:

- *"ceiiinosssttuv"*
- *"ut tensio, sic vis"*
- *as the tension, so the force*

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

in the present data live in papers

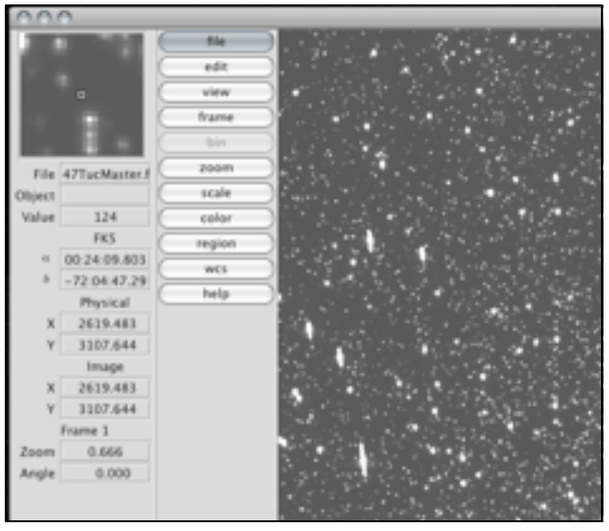


References: McLaughlin et al. 2006; http://adsabs.harvard.edu/abs/2006ApJS..166..249M

Tables, Tables in tar file

Code in tar file

FITS Files

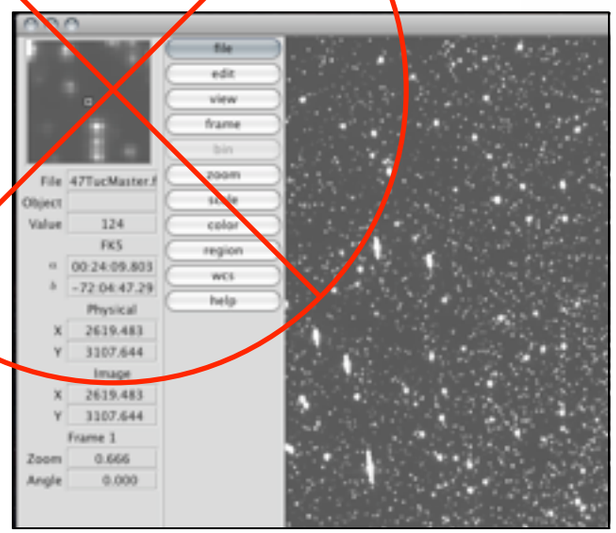


References: McLaughlin et al. 2006; http://adsabs.harvard.edu/abs/2006ApJS..166..249M

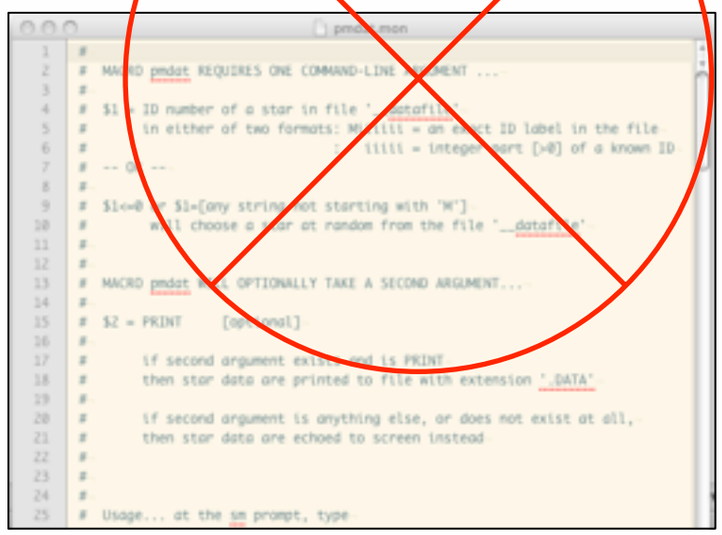
VizieR

Tables, Tables in tar file

FITS Files



Code in tar file

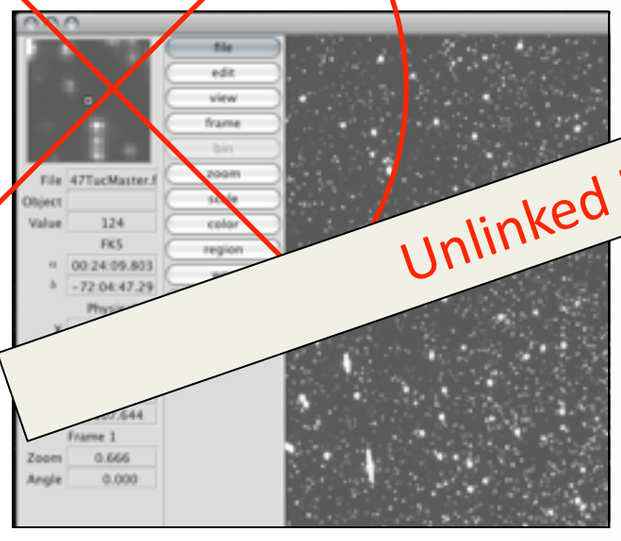


References: McLaughlin et al. 2006; http://adsabs.harvard.edu/abs/2006ApJS..166..249M

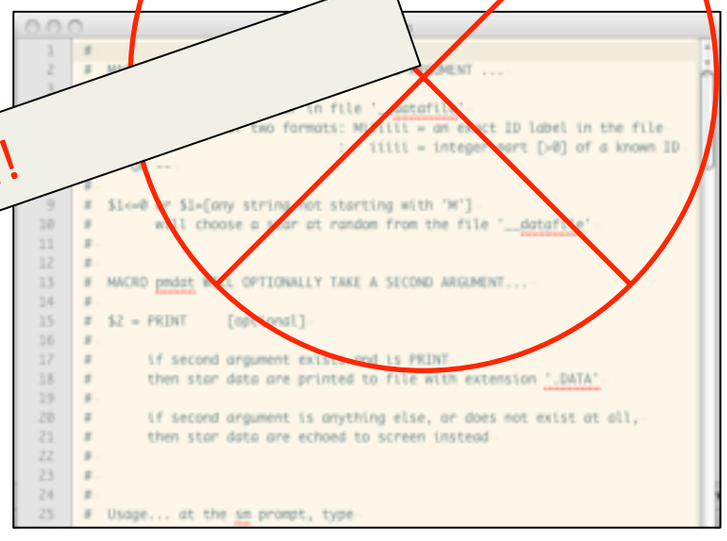
VizieR

Tables, Tables in tar file

FITS Files

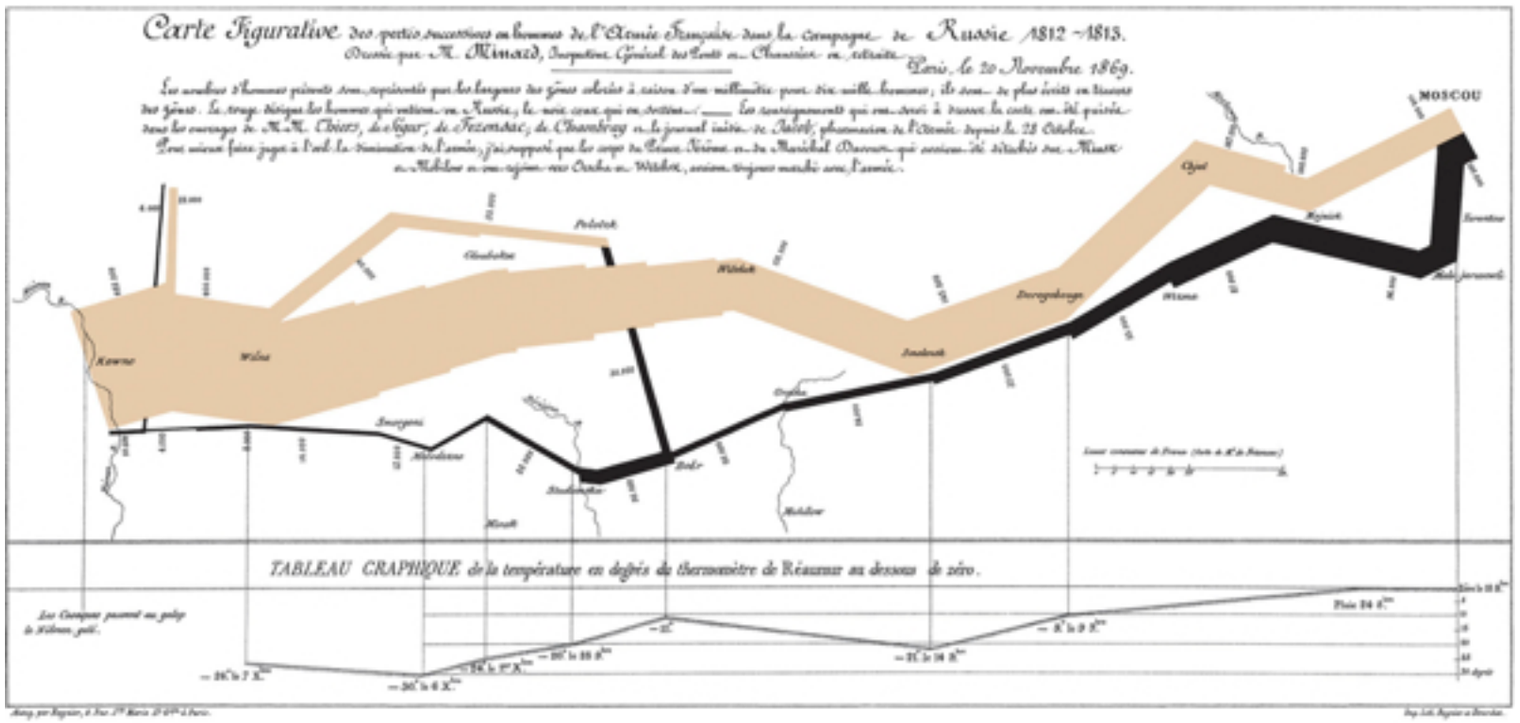


Code in tar file



Unlinked Data is LOST data!

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

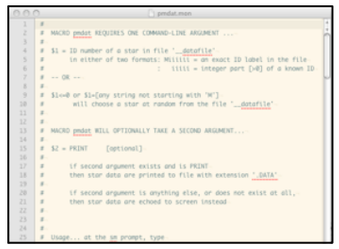


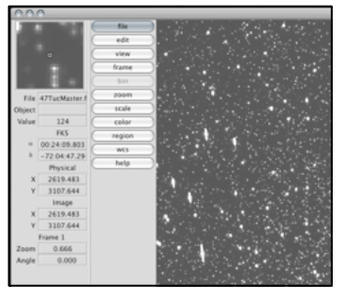
And now for a remix...

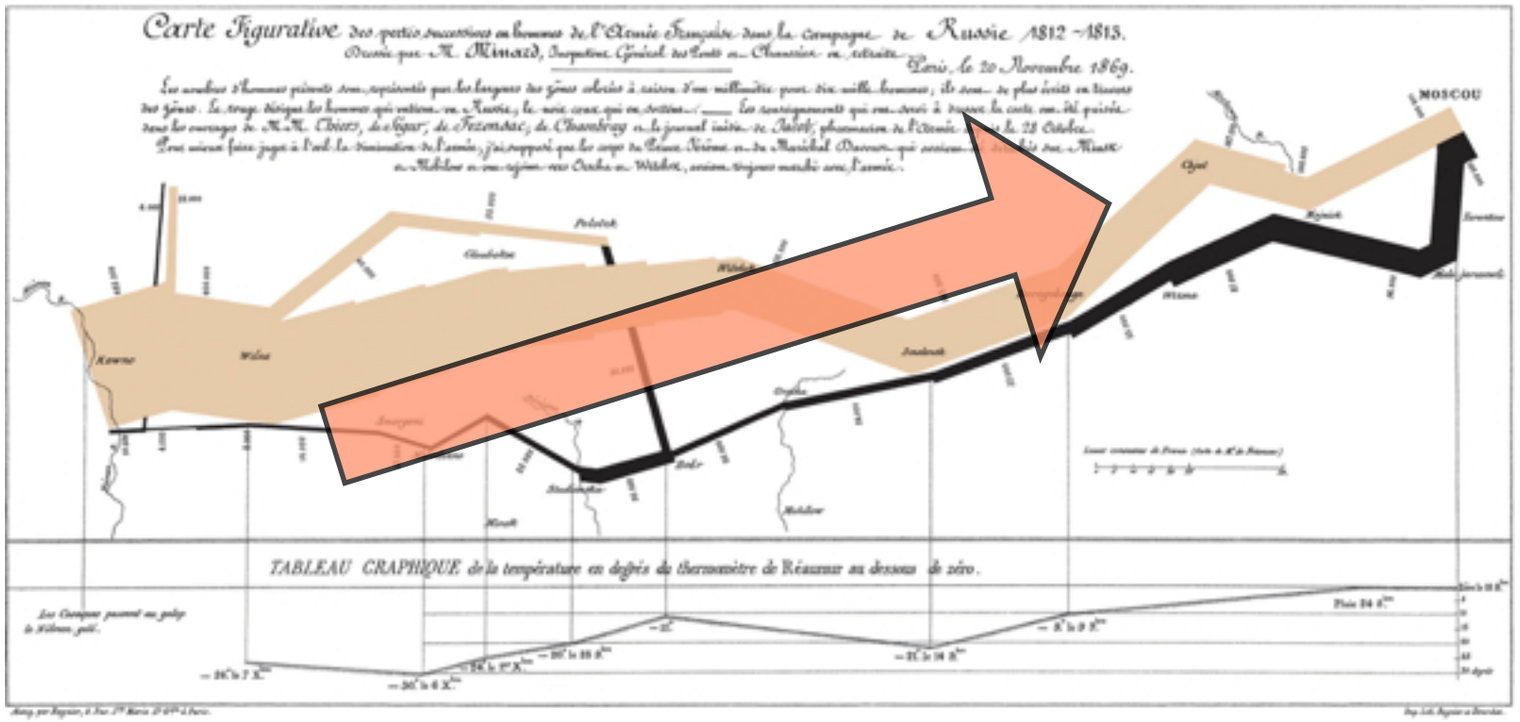
Consider Minard's charting of the demise of Napoleon's army on its roundtrip to Moscow...

except instead of losing soldiers, we ask about ***losing data behind or in a paper...***

References: Charles Minard (1781-1870) (see upload log) [Public domain], via Wikimedia Commons



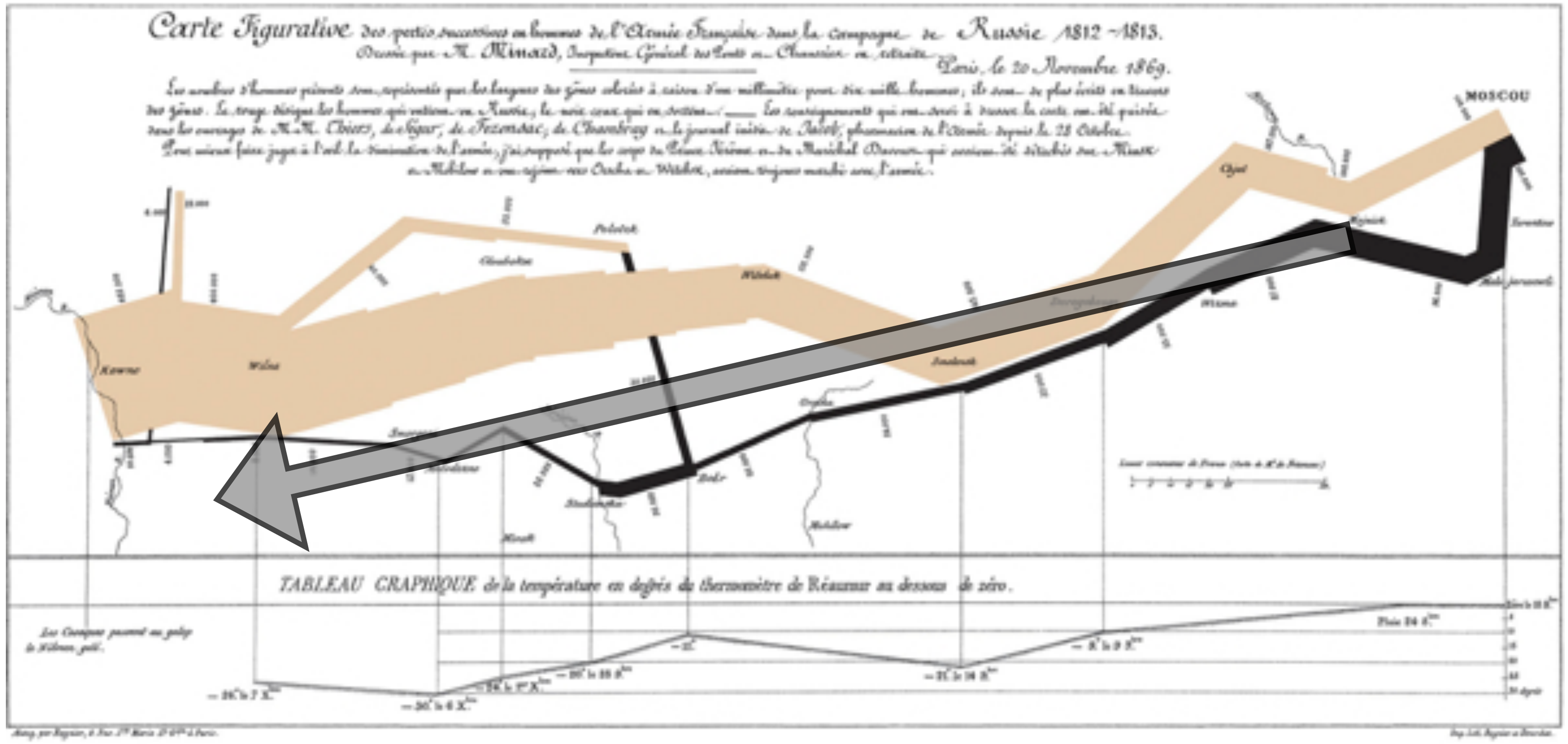


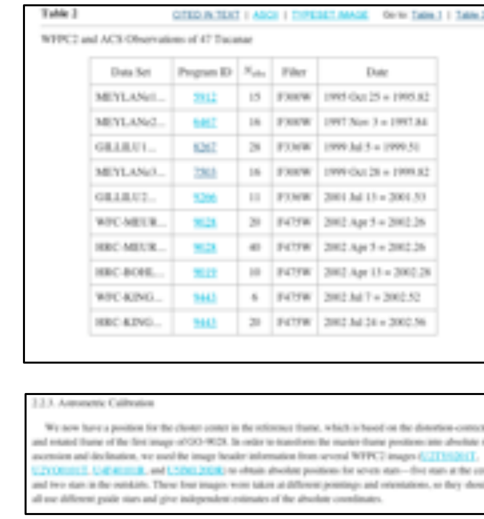


Losses from Data to Literature

- Raw data:
  - ➡ *might already be in a telescope archive*
  - ➡ *linkage partially fixed by post-pub curation*
- Theoretical data;
- Analysis codes and logs;
- Processed data:
  - ➡ *Reduced data; mosaics;*

References: Charles Minard (1781-1870) (see upload log) [Public domain], via Wikimedia Commons





Losses (and some Gains) from Literature to Archives:

- <u>Data still leaks:</u>
  - ➡ *data products that are not machined tables;*
  - ➡ *data in tar files;*
  - ➡ *data from external websites (linked as footnote URLs).*
- <u>*Recovery:*</u> Post-publication curation creates or captures:
  - ➡ *SIMBAD objects; big archive data references;*
  - ➡ *large machined tables captured by CDS.*

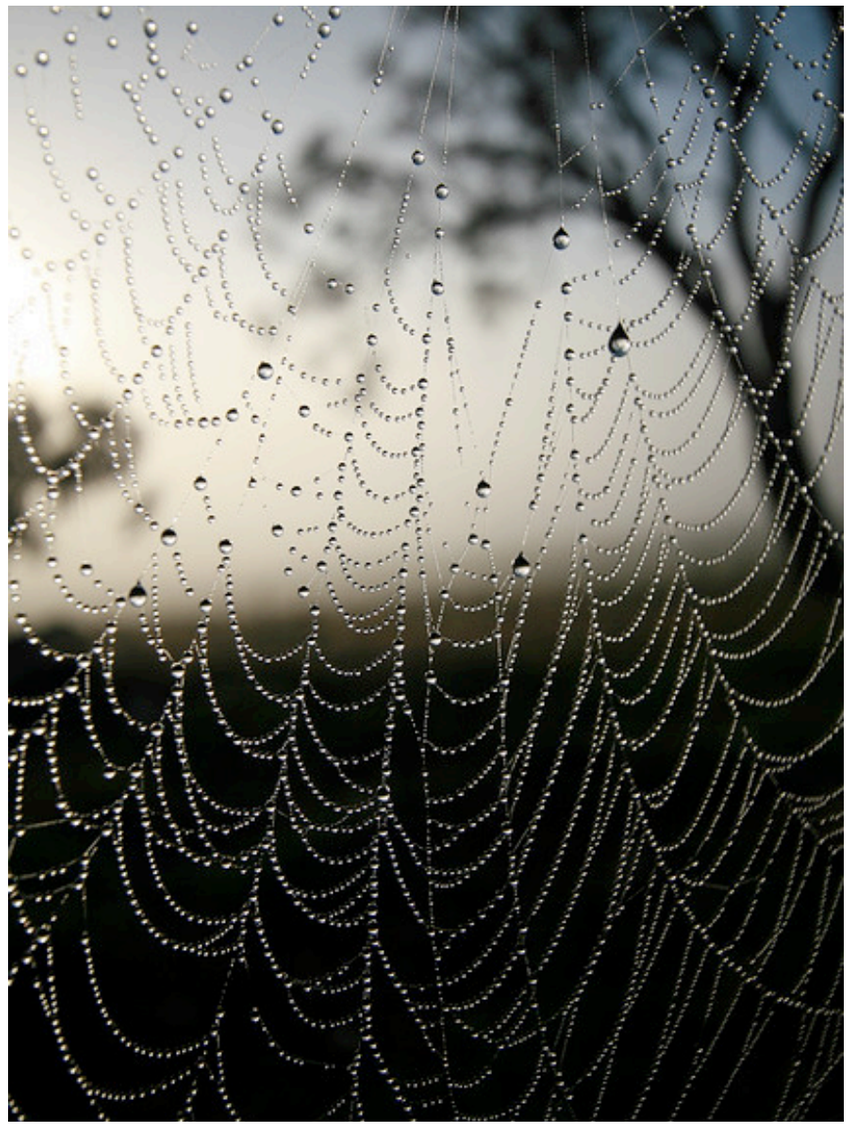
References: Charles Minard (1781-1870) (see upload log) [Public domain], via Wikimedia Commons

in the future data live...

- Refined data sets are published by scientists in long lived repositories;

- Scientist's data linked in ADS & are "searchable"

- Scientist's data is **reused** & **cited**, giving credit for that work.



http://theastrodata.org

The Dataverse Network™ Project

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

Created at IQSS at Harvard University

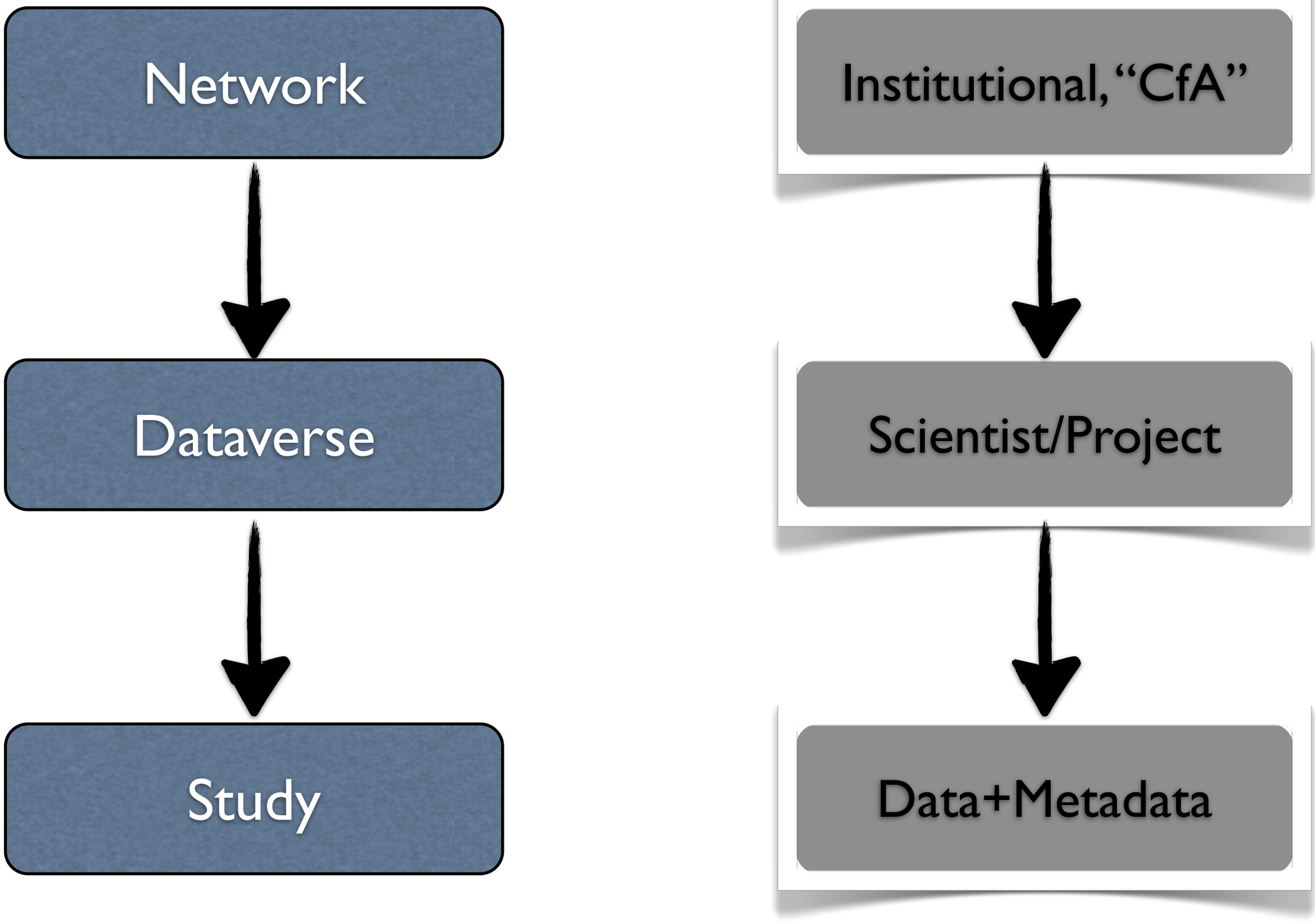
- ☑ The Dataverse Network (DVN) Project was built originally for managing Social Science Data;
- ☑ Collaboration between the Harvard/CfA "*Seamless Astronomy*" team and the DVN team to reuse this framework for Astronomy Data.
- ☑ Institutional support from Harvard Library for DVN infrastructure and training for Astronomy.

The Dataverse Network™ Project

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 your web site.
- ☑ Inter-operates with other systems using standards


Network
Dataverse
Study
Institutional, "CfA"
Scientist/Project
Data+Metadata

The Dataverse Network™ Project

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

Created at IQSS at Harvard University

We are:

- ☑ Metadata mapping between the Data Documentation Initiative (DDI) standard used by DVN and Astronomy's VO standards;
- ☑ Conducting Data "Interviews" with Astronomers to deduce their needs;
- ☑ Working with NASA-SAO ADS to expose data publications;
- ☑ Professional Outreach Training for CfA astronomers to use platform;
- ☑ Working on the DVN API for search & up/downloading of data products;
- ☑ Working with VAO to expose internal data products to VO indexing and search.

http://figshare.com ?

The Dataverse Network™ Project

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

Created at IQSS at Harvard University

Why DVN?

- ☑ Open Source (Java) Software Stack
  - ☑ Instantiate new Dataverse Networks:
    - ☑ Societal, Publishing, Institutional needs.
    - ☑ Copy our CfA work to new Astronomy DVN.
- ☑ Built in DVN "Universe" search and linking.

The Dataverse Network™ Project

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

Created at IQSS at Harvard University

Why DVN?

- ☑ Domain Specific
  - ☑ Metadata/Data Formats;
  - ☑ Use Astronomy Controlled Vocabularies for Curation;
  - ☑ Hook up DVN to VO and other Software tools.
  - ☑ Reuse DVN API for Astronomy specific software tools

The Dataverse Network™ Project

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

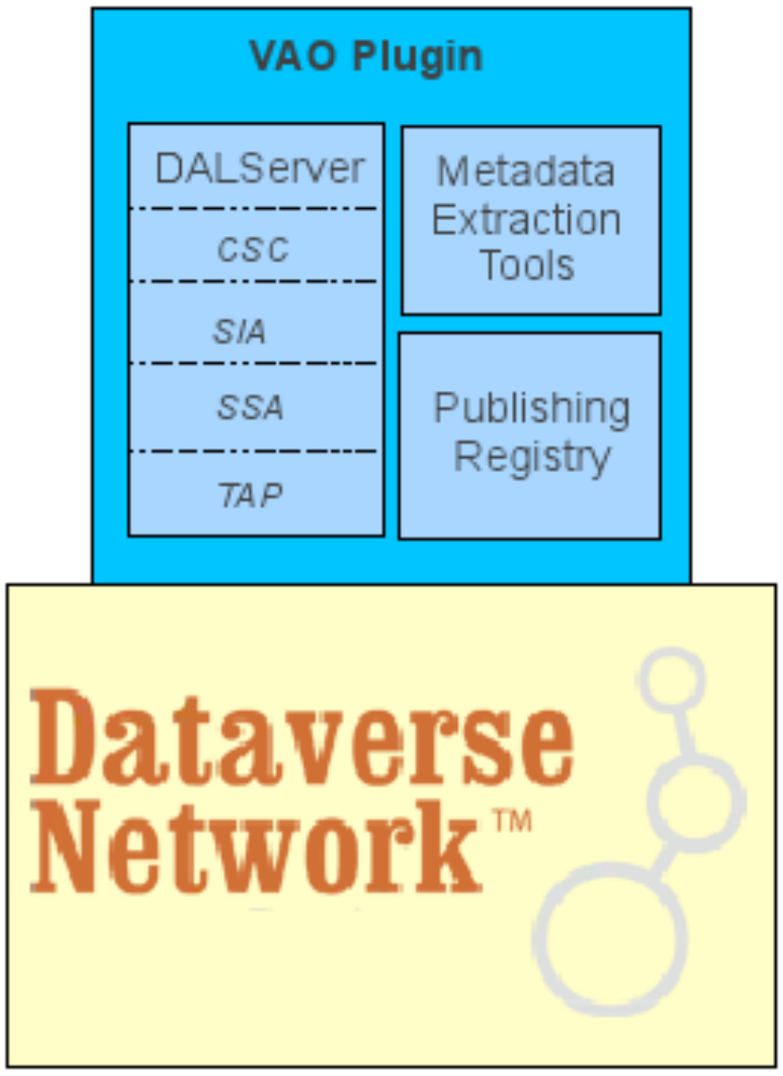
Created at IQSS at Harvard University

Why DVN?

- ☑ Friends
  - ☑ Work with DVN developers to evolve software:
    - ☑ Metadata/Data format support.
  - ☑ Link Dataverse "Studies"
    - ☑ NASA-ADS
    - ☑ American Astronomical Society Publications (ApJ, AJ...)

Virtual Observatory "Plugin" to DVN

- Index individual "datatypes" in a published data study;
- Expose services for datatypes;
- Manage publication registration to VO.



this problem

References: Ton Zijlstra; http://www.flickr.com/photos/tonz/2463875144/

http://theastrodata.org