Astronomy Dataverse: enabling astronomer data publishing
http://theastrodata.org
in the past data were hidden...

1660 Robert Hooke “pre” published anagram:

- “ceiiinossstttuv”
- “ut tensio, sic vis”
- as the tension, so the force

in the present data live in papers


Monday, July 9, 2012
Tables, Tables in tar file

FITS Files

Tables, Tables in tar file

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

Monday, July 9, 2012
Losses (and some Gains) from Literature to Archives:

- **Data still leaks:**
  - data products that are not machined tables;
  - data in tar files;
  - data from external websites (linked as footnote URLs).

- **Recovery:** 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 (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.
☑️ 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
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 ?
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.
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
Why DVN?

- Friends
  - Work with DVN developers to evolve software:
    - Metadata/Data format support.
  - Link Dataverse “Studies”
    - NASA-ADS
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