An Iterative Approach to Library Data Services in Astronomy

NEASIST | C. Erdmann | @libcce
History background

→ Learned how to program
  → Joined a dotcom
    → Data mining, data linking, faceting
    → Became data savvy
Being able to program has opened the door to new possibilities for me in the library world.
European Southern Observatory (ESO)

- Multinational organization
- Headquarters in Munich
- Ground-based telescopes

http://i.livescience.com/images/i/000/027/364/i02/vlt-brunier-nuit-1600.jpg?1337613157
Telescope Bibliography (telbib): workflow

Scientific literature → Semi-automated search tool FUSE → Telbib database (back-end)

Telbib database (front-end) → Data archive → Tags + proIDs for telbib records

Image courtesy U. Grothkopf, ESO Library
1. 2013arXiv1312.7417W
   Measuring the X-ray luminosities of SDSS DR7 clusters from RASS
   Wang, Lei; Yang, Xiaohu; Shen, Shiyin; Mo, H. J. and 7 coauthors
   Published in Dec 2013
   ... noticeably the CFA redshift survey (e.g.Geller & Huchra 1983), the Las Campanas Redshift Survey (e.g.Tucker ...

2. 2013arXiv1312.4294S
   AGN surveys to study galaxy evolution along cosmic times
   Spinoglio, Luigi
   Published in Dec 2013
   ... selected sample of Seyfertgalaxies (25 Seyfert 1 and 23 Seyfert 2) has been extracted from the CFA Redshift ...
   ... Surveyof (2400 objects (1/50 or 2% efficiency) (Huchra & Burg 1992), Veilleux & Osterbrock1987 proposed ...

3. 2013PASJ...65...113T
   The 9 and 18 Micrometer Luminosity Functions of Various Types of Galaxies with AKARI: Implication for the Dust Torus Structure of AGN
   Toba, Yoshiki; Oyabu, Shinku; Matsuhara, Hideo; Malikov, Matthew A. and 5 coauthors
   Published in Oct 2013
   ... and the Centerfor Astrophysics (CFA) redshift survey (ZCAT: Huchra et al.1995). The flow chart used for sample ...

4. 2013MNRAS.434.2556W
   The impact of systematic uncertainties in N-body simulations on the precision cosmology from galaxy clustering: a halo model approach
   Wu, Hao-Yi; Hutener, Dragan
   Published in Sep 2013
   ... Peebles 1977) and the CFA Redshift Survey (Huchra et al. 1983; Geller Huchra 1989) revealing the cosmic ...

5. 2013arXiv1309.7786S
   Galaxy structures - Groups, clusters and superclusters
   Sepp, Tilt; Gramann, Mirt
   Published in Sep 2013
Harvard-Smithsonian Center for Astrophysics (CfA)

- Collaboration between institutions
- Based in Cambridge, MA
- Multinational projects
- Global & space-based facilities

http://commons.wikimedia.org/wiki/File:Center_for_Astrophysics_at_Harvard.jpg
Administrative Services

Policies & Copyright Advice  DataCite
DMPs & DMPTool  E-Science Institute
DM Training Programs  WH OSTP Response
DM @ Harvard Site  Survey: Story of Your Data?
Research Data Collaborative
Publishing Research Material

- Zenodo
- Dataverse
- Figshare
- Datahub
Lessons Learned

Low Barrier
User Experience
Author Workflow Integration
Social & Sharing Aspect
Citation Ease
Permissions & Copyright
Paper <-> Data Linking
Selling the Service

Objective Advice
Visual Customization (APIs)
Metadata Customization
Traditional vs Data Savvy Approach
Repository-centric
Tracking Work
Curation Advice
Learning Curve
Versioning
What If We Could Be More?

If We Assume

Could we be of more use to scientists?

Sympathize with their data needs and offer assistance?

Become data savvy?

The Pace of NSF Funded Research

Topics: academia, Astronomy, costs, statistics

Recently on Facebook I came across a note by Chris Erdmann that some handy folks at Harvard put together statistics on (nearly) every astronomy paper from 1995 to present that was funded through an NSF AST grant. This seemed like a really interesting dataset, especially for a young (read: financially uncertain) research such as myself.

So parsing through all 29,042 papers listed, here are two interesting things I’ve learned...
"By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions."

"Online help-wanted ads for data analysis mavens have shot up 46% since April 2011, and 246% since April 2009, to over 31,000 openings now."

Data Scientist Training for Librarians (DST4L)

Major goals:

- Hands on experience w/ research data lifecycle
- Inform new forms of library (data) services
- Upgrade librarian knowledge & skills
- Create a community of continuous learning by doing
- Experience new style of working
- Change library mindset
- Explore other learning moments
Course Outline

Extract
Wrangle
Analyze
Visualize
Technologies

Command Line
Git (GitHub)
Excel
OpenRefine
Python
Software Carpentry

R (RStudio)
Tableau
SQL (SQLShare)
MongoDB
Data Repositories
D3
More about the Course

1 class/week 3-4 months
* 1 hands on session (projects)
Lead organizer (context) w/ expert instructors
Open notes, participants blog (experience)
In-person, not virtual or recorded
Some homework, final presentations
Highlighted student comment:
http://www.youtube.com/watch?v=U5ZYM085bNo&t=1m21s

...very helpful, preparing for long career, going to see it more and more, will keep using skills set, like doing it, fun problem solving...
It’s Hard Work!
But worth it!

- NASA ADS Visualization
- Unified Astronomy Thesaurus
- CfA Bibliography
- DOE Grants
How Do I Start?

Software Carpentry Workshop
Invite Local Experts
Tap Your Own Community
A Challenge

If you dislike change, you're going to dislike irrelevance even more.

-- Eric Shinseki
Questions?