Evermore
Seamless Astronomy

Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics

with Alberto Accomazzi, Douglas Burke, Gus Muench & Michael Kurtz (Harvard-Smithsonian CfA); Eli Bressert (U. Exeter); Tim Clark (Massachusetts General Hospital/Harvard Medical School); Chris Borgman (UCLA); Jonathan Fay & Curtis Wong (Microsoft Research)
Realm of Seamless Astronomy
3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Galileo, 1600

Reber’s Radio Telescope, 1937

Observatory Tower, Lincolnshire, UK, c. 1300

The “Scientific Revolution”

Long-distance remote-control/“robotic” telescopes 1990s

NASA/Explorer 7 (Space-based Observing) 1959

“The Internet”

“Virtual Observatories” 21st century
What can today’s Astronomer’s “Research” look like?

Research

In my Astronomy research, I am primarily interested in how the gas in galaxies constantly re-arranges itself over huge time spans to constantly form new stars. I have also had a long-standing interest in data visualization, and in improving the use of computers in all aspects of scientific research. I teach a course at Harvard called "The Art of Numbers," and I am very involved in the WorldWide Telescope Project, which brings astronomical data to everyone through an interface that demonstrates data delivery for the 21st Century of "e-Science."

http://www.cfa.harvard.edu/~agoodman/
Publishing

Data

Simulation

e-Science Tools

Viz

WorldWide Telescope
Seamless Data/Literature Connections (e.g. ADS)

“Modular Craftsmanship” (e.g. flickr)

Collections, Communities & Guided Tours

Created by Curtis Wong and Jonathan Fay at MSR; AG is “Academic Partner” on the WWT Project
The (US) Backstory


Science News

$10 Million NVO
ScienceDaily (Oct 19, 2008) -- Its users the world over: research institutions starting an ambitious project to make the universe online.

See Also:

(NVO), headed by astronomer Alex

NVO senior personnel:
and meanwhile...
What/where are/is “Data”?
What/where is literature?
Seamless Astronomy

But, that was 2009...
Realm of "Seamless Astronomy"

Data

Advanced Search & InfoViz tools

Get Semantic

Literature

“Researcher”

Standalone Analysis Software

2010

Evermore Seamless Astronomy
This simple argument, first made at the 2009 WWT session at AAS, seems to be working:

“Astronomy research tools should work as seamlessly as travel research tools.”
“Astronomy research tools should work as seamlessly as travel research tools.”

When the concept of a "Virtual Observatory" (VO) was first discussed by future-looking astronomers in the mid-1990s, all thoughts were about distributed data and a common system to access it. But, information access on today's web primarily works in the reverse: distributed tools accessing common data centers. Capability and ease-of-use improvements to the web typically now come in the form of nesting, aggregating or connecting tools. Think kayak.com, iGoogle, or Bing Maps. In the "Seamless Astronomy" view to be discussed, today's "VO" should be thought of as the ever-improving set of data archives, tools, interconnections, and standards that strive to make astronomical research as "seamless" as travel research. The good news is that the cutting-edge of the astronomical research environment is moving rapidly in this seamless direction. The most savvy institutions are beginning to realize that the original VO model of data distributed on thousands of individual researchers' desktop hard drives is not a sustainable model, and that they need to offer data hosting, archiving, and stewardship services the way libraries offer such services for printed matter. Software tools are becoming much more interoperable thanks to protocols for message-passing such as "SAMP." And, the improved speed of web applications is to some extent removing platform-dependence as an obstacle to programmers and users alike. The bad news is that most astronomers are largely unaware of the tools that this new nirvana offers, and instead still conduct online research in the same way they did a decade ago. In this talk, I will focus in particular on how our recent work on connecting Microsoft's WorldWide Telescope program to other commonly-used astronomical research tools--most notably literature searching tools--has made the astronomical research environment more seamless. More generally, I will emphasize and demonstrate that an ever-increasing diversity of tools allow researchers to carry out a particular research task, so that the important research for the future lies in figuring out how to make the tools, their interconnections, and their connections to data and literature resources useful and well-known to the astronomical community.
Astronomers can see parallels...
Literature Handling: Diverse Apps, Common Data
What fraction of astronomy researchers know about these tools?
“writemypaper.org?”
ASA/NASA Astrophysics Data System (ADS)

Query Results from the ADS Database

Selected and retrieved 200 abstracts.

<table>
<thead>
<tr>
<th>#</th>
<th>Bibcode</th>
<th>Authors</th>
<th>Score</th>
<th>Date</th>
<th>Subject Area</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1995RvMP..67...661B</td>
<td>Biskovati-Kogan, G. S.; Silich, S. A.</td>
<td>19.000</td>
<td>Jul 1995</td>
<td>A E</td>
<td>Shock-wave propagation in the...</td>
</tr>
<tr>
<td>2</td>
<td>1999NewAR..43...31F</td>
<td>Frank, A.</td>
<td>18.000</td>
<td>May 1999</td>
<td>A E</td>
<td>Bipolar outflows and the evolution of stars</td>
</tr>
<tr>
<td>3</td>
<td>2007ARA&amp;A..45...177C</td>
<td>Crowther, Paul A.</td>
<td>13.000</td>
<td>Sep 2007</td>
<td>A E F X</td>
<td>Physical Properties of Wolf-Rayet Stars</td>
</tr>
<tr>
<td>4</td>
<td>2002ARA&amp;A..40...439B</td>
<td>Balick, Bruce; Frank, Adam</td>
<td>13.000</td>
<td>n/a 2002</td>
<td>A E</td>
<td>Shapes and Shaping of Planetary Nebulae</td>
</tr>
<tr>
<td>5</td>
<td>2008A&amp;ARv..16...209P</td>
<td>Puls, Joachim; Vink, Jorick S.; Najarro, Francisco</td>
<td>12.000</td>
<td>Dec 2008</td>
<td>A E X</td>
<td>Mass loss from hot massive stars</td>
</tr>
<tr>
<td>6</td>
<td>2005ApJ...631...435R</td>
<td>Ramirez-Ruiz, Enrico; Garcia-Segura, Guillermo; Salmonson, Jay D.; Perez-Rendón, Brenda</td>
<td>12.000</td>
<td>Sep 2005</td>
<td>A E F X</td>
<td>The State of the Circumstellar Medium Surrounding Gamma-Ray Burst Sources and Its Effect on the Afterglow Appearance</td>
</tr>
<tr>
<td>7</td>
<td>1992ARA&amp;A..30...235C</td>
<td>Chiosi, Cesare; Bertelli, Gianpaolo; Bressan, Alessandro</td>
<td>12.000</td>
<td>n/a 1992</td>
<td>A G T R C S U</td>
<td>New developments in understanding the HR diagram</td>
</tr>
</tbody>
</table>

This page generated using ADS faceted Search, as developed by Michael Kurtz, Alberto Accomazzi & Jonathan Fay
<table>
<thead>
<tr>
<th>#</th>
<th>Bibcode Authors</th>
<th>Score Date</th>
<th>List of Links</th>
<th>Title</th>
</tr>
</thead>
</table>
| 1 | 2009ApJ...700..1609M  
Myers, Philip C. | 1.000 08/2009 | A Z E F L X R C S U | Filamentary Structure of Star-forming Complexes |
| 2 | 2009ApJ...700..1190D  
Desai, Vandana; Soifer, B. T.; Dey, Arjun; LeFloc'h, Emeric; Armus, Lee; Brand, Kate; Brown, Michael J. I.; Brodin, Mark; Jannuzi, Buell T.; Houck, James R.; and 8 coauthors | 1.000 08/2009 | A Z E F L X R C S U | Strong Polycyclic Aromatic Hydrocarbon Emission from \( z \approx 2 \) ULIRGs |
| 3 | 2009MNRAS.396.1851N  
Nutter, D.; Stamatellos, D.; Ward-Thompson, D. | 1.000 07/2009 | A Z E F L X R S S U | The initial conditions of isolated star formation - IX. Akari mapping of an externally heated pre-stellar core |
| 4 | 2009A&A...502..175B  
| 5 | 2009MNRAS.395.1695H  
Hernán-Caballero, A.; Pérez-Fournon, I.; Hatziminaoglou, E.; Afonso-Luis, A.; Rowan-Robinson, M.; Rigopoulou, D.; Farrah, D.; Lonsdale, C. J.; Babbedge, T.; | 1.000 05/2009 | A Z E F L X R C S U | Mid-infrared spectroscopy of infrared-luminous galaxies at \( z \approx 0.5-3 \) |
**Available data**

Basic data:

**NGC 7023 -- Open (galactic) Cluster**

- **Other object types:** C1+ (C1+[BDB2003]), B2059+6755, B2059+6755, LBN, P20599+6755, B20599+6755, IRAS (IRAS)
- **ICRS coord. (epoch 2000):**
  - Declination:
    - 21 01 36.9 +68 09 48
    - 21 01 36.9 +68 09 48
  - Right Ascension:
    - 104.0616 +14.1926
  - 104.0616 +14.1926
- **Fluxes:**
  - B: 7.20
- **Identifiers (11):**
  - NGC 7023
  - C 2059+679
  - CI VDB 139
  - IRAS 20599+6755
  - IRAS P20599+6755
  - LBN 487
  - OCISN 50
  - AAVSO 2046+67
  - B20599+6755

**Plots and Images**

- **plot around** with radius
  - **radius:** 10 arcmin

**References (371 between 1983 and 2009)**

Simbad bibliographic survey began in 1950 for stars (at least bright stars) and in 1983 for all other objects (outside the solar system).
“alpha” Faceted Topic Search in ADS (courtesy of Michael Kurtz & Alberto Accomazzi)
List of objects with links to WWIT browser
(thanks to ADS team & Jonathan Fay)

<table>
<thead>
<tr>
<th>Bibcode</th>
<th>Score</th>
<th>Date</th>
<th>List of Links</th>
<th>Access Control Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007ApJ...657...810D</td>
<td>100.000</td>
<td>Mar 2007</td>
<td>A</td>
<td>E</td>
</tr>
<tr>
<td>Draine, B. T.; Li, Aigen</td>
<td>Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007ApJ...663...866D</td>
<td>96.842</td>
<td>Jul 2007</td>
<td>A</td>
<td>E</td>
</tr>
<tr>
<td>Draine, B. T.; Dale, D. A.; Bendo, G.; Gordon, K. D.; Smith, J. D. T.; Armus, L.; Engelbracht, C. W.; Helou, G.; Kennicutt, R. C., Jr.; Li, A.; and 10 coauthors</td>
<td>Dust Masses, PAH Abundances, and Starlight Intensities in the SINGS Galaxy Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engelbracht, C. W.; Gordon, K. D.; Rieke, G. H.; Werner, M. W.; Dale, D. A.; Latter, W. B.</td>
<td>Metallicity Effects on Mid-Infrared Colors and the 8 μm PAH Emission in Galaxies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
And now we got to NGC 7023 by using the literature as a filter.
Seamlessness through...
flickr
+ astrometry.net
+ WWT !?
HH4647

Embedded Outflow in HH 46/47

Spitzer Space Telescope • IRAC

NASA / JPL-Caltech / A. Nonaka-Crespo [SSC/Caltech]
Coming Soon from ADS (I hope!)

Faceted Heat Map of Articles on the Sky

Historical Image Layer Extracted from ALL ADS holdings (using astrometry.net)
The future is here... data IN articles

Note: This work came from the "AstroMed" project am.iic.harvard.edu

Fig. 1. The 3D visualizations show position–position–velocity (x–y–z) space within a 3D data cube containing all the significant emission. Each 3D visualization shows a pseudo-dendrogram that abstracts the hierarchical structure of a 3D data cube into an easily visualized representation called a 'dendrogram'. Although well developed in other data-intensive fields, it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency. Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 3D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the x–y–z plane while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius (r), velocity dispersion (σ_v), and luminosity (L). The volumes can have any shape, and in other work we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that M_{\text{obs}} \approx \frac{L_{\text{obs}} \text{c}^2}{8 \pi \text{G}} \text{Km s}^{-1} \text{K}^{-1} \text{cm}^{-2} \text{K}^{-3} \text{km s}^{-1} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, \text{S}_{\text{vir}} = \frac{M_{\text{obs}} \text{c}^2}{8 \pi \text{G} \text{L}_{\text{obs}}} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2).

Figure 2. Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to 12CO emission from the L1448 region of Perseus. A. A 3D visualization of the surfaces indicated by colours in the dendrogram shown in Fig. 1. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating boxes within them; and green corresponds to the surfaces in the data cube containing all the significant emission. Dendrogram branches corresponding to gravitating objects have been highlighted in yellow over the range of $T_{\text{mb}}$ (main-beam temperature) test-level values for which the virial parameter is less than 2. The dendrogram process is schematically shown in Supplementary Methods. The volumes can have any shape, and in other work we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{\text{obs}} \approx \frac{L_{\text{obs}} \text{c}^2}{8 \pi \text{G}} \text{Km s}^{-1} \text{K}^{-1} \text{cm}^{-2} \text{K}^{-3} \text{km s}^{-1} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2).

In principle, extended portions of the tree (Fig. 2, yellow highlighting) can be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\text{S}_{\text{vir}} = \frac{M_{\text{obs}} \text{c}^2}{8 \pi \text{G} \text{L}_{\text{obs}}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) can be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\text{S}_{\text{vir}} = \frac{M_{\text{obs}} \text{c}^2}{8 \pi \text{G} \text{L}_{\text{obs}}}$. The volumes can have any shape, and in other work we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{\text{obs}} \approx \frac{L_{\text{obs}} \text{c}^2}{8 \pi \text{G}} \text{Km s}^{-1} \text{K}^{-1} \text{cm}^{-2} \text{K}^{-3} \text{km s}^{-1} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2).

The volumes can have any shape, and in other work we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{\text{obs}} \approx \frac{L_{\text{obs}} \text{c}^2}{8 \pi \text{G}} \text{Km s}^{-1} \text{K}^{-1} \text{cm}^{-2} \text{K}^{-3} \text{km s}^{-1} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\text{S}_{\text{vir}} = \frac{M_{\text{obs}} \text{c}^2}{8 \pi \text{G} \text{L}_{\text{obs}}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) can be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\text{S}_{\text{vir}} = \frac{M_{\text{obs}} \text{c}^2}{8 \pi \text{G} \text{L}_{\text{obs}}}$. The volumes can have any shape, and in other work we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{\text{obs}} \approx \frac{L_{\text{obs}} \text{c}^2}{8 \pi \text{G}} \text{Km s}^{-1} \text{K}^{-1} \text{cm}^{-2} \text{K}^{-3} \text{km s}^{-1} (\text{ref. 15}; see Supplementary Methods and Supplementary Fig. 2).
"Old Data" ➔ astrometry.net/flickr/WWT ➔ "New Data"

"Your Data" ➔ WWT/ADS/SIMBAD/NAO ➔ WWT as API ➔ "My Data"

"Old Data" ➔ WWT as API ➔ "New Data"

"Old Data" ➔ 3D PDF ➔ "New Data"
Jim Gray (& Alex Szalay) had it right (in 2004)

All Scientific Data Online

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity
Jim Gray (& Alex Szalay) had it right (in 2004)

The World Wide Telescope
an Archetype for Online-Science

Jim Gray (Microsoft)
Alex Szalay (Johns Hopkins University)
Microsoft Academic Days in Silicon Valley

http://research.microsoft.com/~gray/talks

World Wide Telescope
Virtual Observatory
http://www.ivoa.net/

- Premise:
  - Most data is (or could be) online.

- The Internet is the world’s best telescope:
  - It has data on every part of the sky
  - In every measured spectral band: optical, x-ray, radio
  - As deep as the best instruments (2 years ago)
  - It is up when you are up.

  The “seeing” is always great
  (no working at night, no clouds no moons no...).

- It’s a smart telescope:
  links objects and data
to literature on them.

All Scientific Data Online

- Literature
- Derived and Recombined Data
- Raw Data

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity

The Big Picture

- Experiments & Instruments
- Other Archives
- Literature
- Simulations

- Query and Vis tools
- Support/training
- Performance
  - Execute queries in a minute
  - Batch query scheduling

The Big Problems

- Data ingest
- Managing a petabyte
- Common schema
- How to organize it?
- How to reorganize it
- How to coexist with others
How do we increase the fraction of astronomy researchers who know about these tools?

User Groups (CfA now has one)

+Suggestions?!
User Groups (CfA now has one)
How do we increase the number of people who create and interlink new tools?

Kiva model proposed here in 2009...

Now being implemented through VAO “Associates” and WWT Partners.
How do we organize such diverse tools, so as to make them interoperably useful?....

“SAMP” is a great technical start, but offers a very significant user interface challenge.
Think about the “modules” needed to make this work...but do the details matter, to your research, if the system works seamlessly?
Seamless Astronomy

Mockup based on work of Eli Bressert, excerpted from NASA AISRP proposal by Goodman, Muench, Christian, Conti, Kurtz, Burke, Accomazzi, McGuinness, Hendler & Wong, 2008
Fiction
(for now)

Fact
(right now)
George will show you more...

http://www.skynow.org/static/wwt.html
Evermore
Seamless Astronomy
Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics

with Alberto Accomazzi, Douglas Burke, Gus Muench & Michael Kurtz (Harvard-Smithsonian CfA); Eli Bressert (U. Exeter); Tim Clark (Massachusetts General Hospital/Harvard Medical School); Chris Borgman (UCLA); Jonathan Fay & Curtis Wong (Microsoft Research)