

Implications of the Three Scientific Knowledge Discovery Case Studies – The User Perspective

International Online Astronomy Research, c. 2011

Alyssa Goodman
-Harvard University-

Dr. Conti provided a great overview of the challenges that face our field of astronomy. Instead of focusing on only one example, I will draw a map of the way that things work in our field, and then give a few brief examples.

Figure 2-7 shows water droplets on a spiderweb.

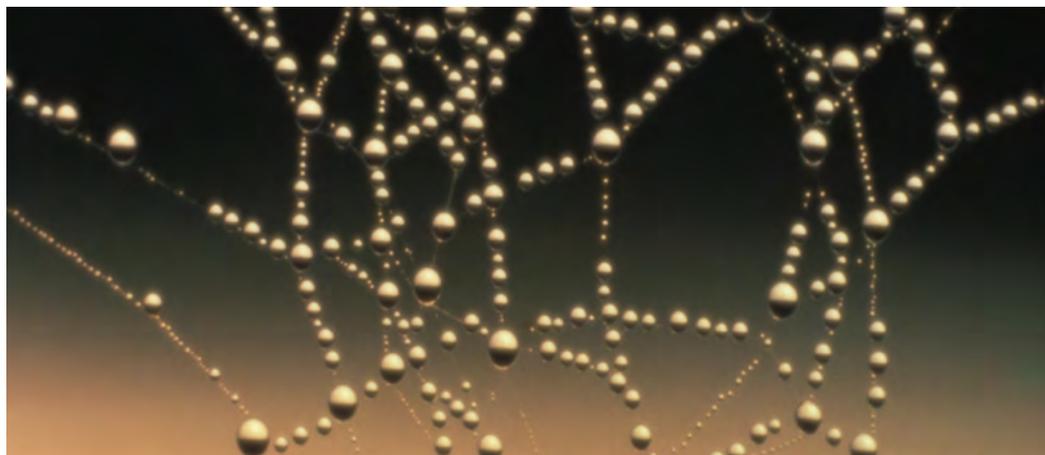


FIGURE 2-7 Image of water droplets on a spider web.
SOURCE: Alyssa Goodman

This is the way that I like to think of “the cloud” with the International Virtual Observatory Alliance. Basically each one of those little water droplets is some data repository, service, or person that provides something in this ecosystem of online astronomy research. The way that those kinds of services are connected on the World Wide Web is how I think that online astronomy research should be done—in a slightly less orchestrated fashion that people might have thought of a decade or more ago. In my work, we call this “Seamless Astronomy.”

There is a researcher in the middle of the image in Figure 2-8. Imagine that this researcher is drawing on two kinds of primary sources of information: data on the left, and literature on the right.



FIGURE 2-8 Realm of Seamless Astronomy.
SOURCE: Alyssa Goodman

In astronomy, our literature is organized very well by a service called the Astrophysics Data System (ADS). Much of what I am going to discuss is how the ADS team is working on connecting the data to the literature through the ADS Labs at the Center for Astrophysics in Cambridge.

I am going to take this Web-like approach and try to explain a few of the interconnections between the two sides of the picture: literature on the right and data on the left. I have reduced this model to just a few services on which I will focus, but as the disclaimer at the bottom says, this slide shows key excerpts from within the astronomy community and excludes more general software that is used, such as papers, graphing and statistic packages, data-handling software, search engines, and so forth. As I mentioned earlier, the literature is very organized, so I have put it all in one accessible repository, and it is accessed through the ADS system. As for the data, there is a set of archives that I will cover shortly.

Figure 2-9 is a view of the international online astronomy ecosystem that I will use to frame the rest of this talk.

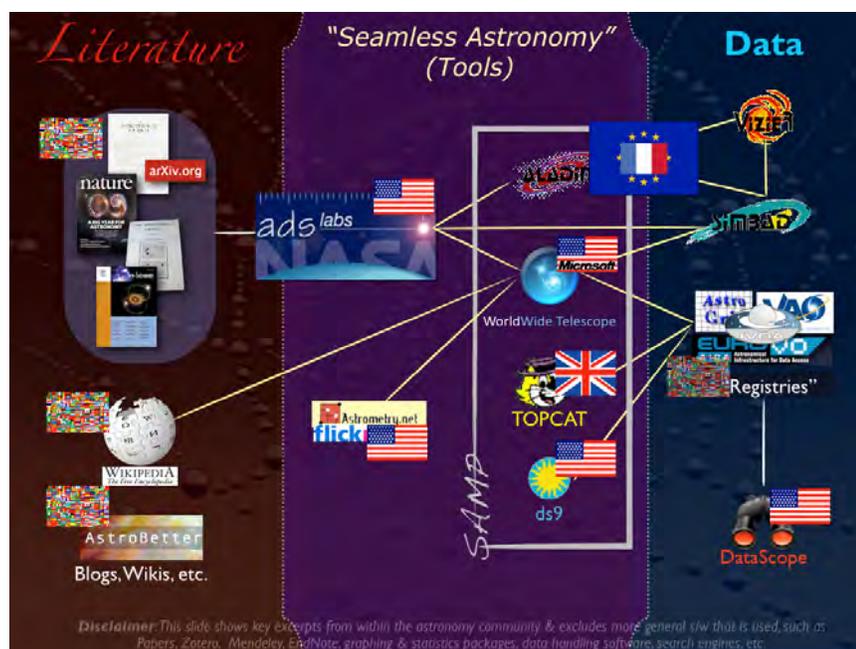


FIGURE 2-9 International online astronomy ecosystem.
SOURCE: Alyssa Goodman

In the middle, I have depicted some tools in a regime called Seamless Astronomy, which offers links between data and literature that are still partly missing, but are now forming and expected to lead to the increase in scientific productivity. The “ADS Labs” is a project in which the ADS team is taking what is already a successful system for accessing the research literature in astronomy and trying to draw that system closer to the data by using the articles themselves as a retrieval mechanism for data. Since ADS already stands for “Astrophysics *Data System*,” why not develop a way to retrieve articles using the literature as a filter for the data?

There are also linkages between many services that do different work. What you should notice, however, is that there are many yellow lines in the figure, which connect all of these services in a variety of ways. It is important to note that this is not a linear process. There are many different ways that the services can access each other without the user even knowing about it.

Four of these applications—the four highlighted in the “SAMP” box in the figure (SAMP is a message-passing protocol)—are actually connected to each other in a way that allows them to interoperate as a set of services and as a research environment.

I would like to point out that the flags in the figure represent the various countries involved. There is also an international flag that indicates this activity is not associated with any one particular country. Some of the services I am going to talk about are from the United States, but others are not, such as TOPCAT from the U.K. *e-Science Program*.

Figure 2-10 shows a screenshot of what happens when we run many of these services together as well as the SAMP hub that allows the services to talk to each other.

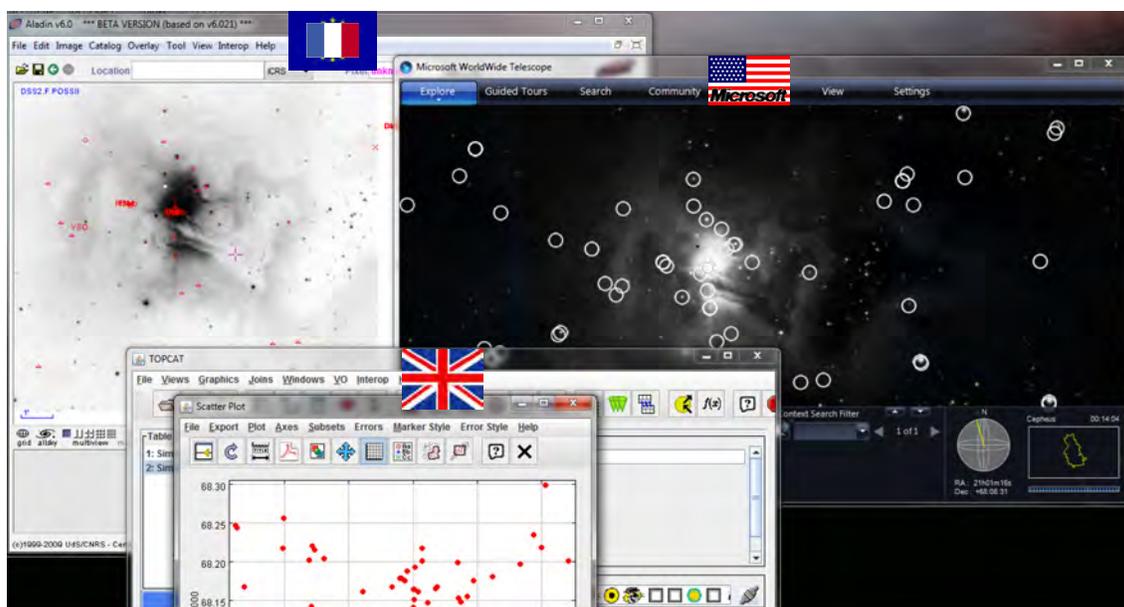


FIGURE 2-10 SAMP hub.
SOURCE: Alyssa Goodman

Suppose that I was looking at some particular region in the sky in one application. I can open another application that looks at the exact same region. Worldwide Telescope can superimpose catalogs in ways that are different from how the first application superimposes catalogs. And TOPCAT, which allows us to statistically manipulate those catalogs, will show the results through live links. It is important to note that these are from three different sets of countries, and the applications are working together through this open-source message-passing protocol. At the moment, my view is that probably only about 5 percent of astronomers know how to do this correctly.

Now let me show you the ADS Labs itself. I should point out that ADS is funded through the U.S. Virtual Observatory funding, provided by the National Aeronautics and Space Administration (NASA). It also has a mirror site in France run by *Centre de Données Astronomiques de Strasbourg* (CDS). ADS Labs was released in January 2011, at the American Astronomical Society meeting in Seattle. The idea was to take ADS and, instead of just doing a keyword search, do more intelligent searching. There are some interesting semantic aspects of this kind of search. Let me give an example of how the ADS can be used.

I was working on an observing schedule recently to observe an obscure quantum mechanical effect called the Zeeman effect in the molecule CH from a particular telescope. I wanted to make sure that no one has ever done this before, so I checked for related papers. If I use the most recent work function of ADS and type “Zeeman effect,” I get something that looks like the image in Figure 2-11. There is a filter list down the side, and I can choose particular authors, key words, missions, objects, and so on. The list of papers that I get will change depending on what is chosen.

The screenshot shows the ADS Labs Beta Search interface. At the top, there are logos for 'ads labs' and 'NASA', and navigation links for 'Home', 'Labs Home', 'ADS Classic', 'Help', and 'Sign on'. The search results are titled 'zeeman effect ch - Most recent'. On the left, there is a 'FILTER BY:' section with various categories like Authors, Keywords, Archives, Missions, SIMBAD Objects, and VizieR tables. The 'SIMBAD Objects' section is expanded, showing a dropdown menu for 'NGC 7027 (1)' with options: 'ADS Publications', 'SIMBAD Info', 'World Wide Telescope', and 'Aladin applet'. A green box highlights this dropdown menu, and a yellow arrow points to the 'World Wide Telescope' option. The search results list includes three entries, with the first one being '2010ApJ...716L...1A The J = 1-0 Transitions of 12CH+, 13CH+, and 12CD+ Amano, T.' and the second one being '2009ApJ...705L.176S Detection of the Zeeman Effect in the 36 GHz Class I CH3OH Maser Line with the EVLA Sama, A. P.; Momjian, E.'

FIGURE 2-11 ADS Labs Beta Search.
SOURCE: NASA

The green box in the figure highlights some linkages between the data and the literature. If I click on one of the particular objects that seems to be mentioned very frequently in these papers, I get four different choices: (1) I can look up all the publications about that object; (2) I can look up data on that object using the SIMBAD data service from CDS; (3) I can see that object in context using Worldwide Telescope; or (4) I can do a similar thing, although it is a narrower context, in another service called Aladin.

If I clicked on Worldwide Telescope, I would get a view like the one in Figure 2-12. The object turned out to be a planetary nebula that I did not know. Since this is not what I was interested in, I concluded that no one has observed what I was looking at and that this object was irrelevant. But perhaps I became curious about this object, and so I could click on the button at the bottom that says “Research.” By clicking there, I can get another set of choices, linking to yet other services.

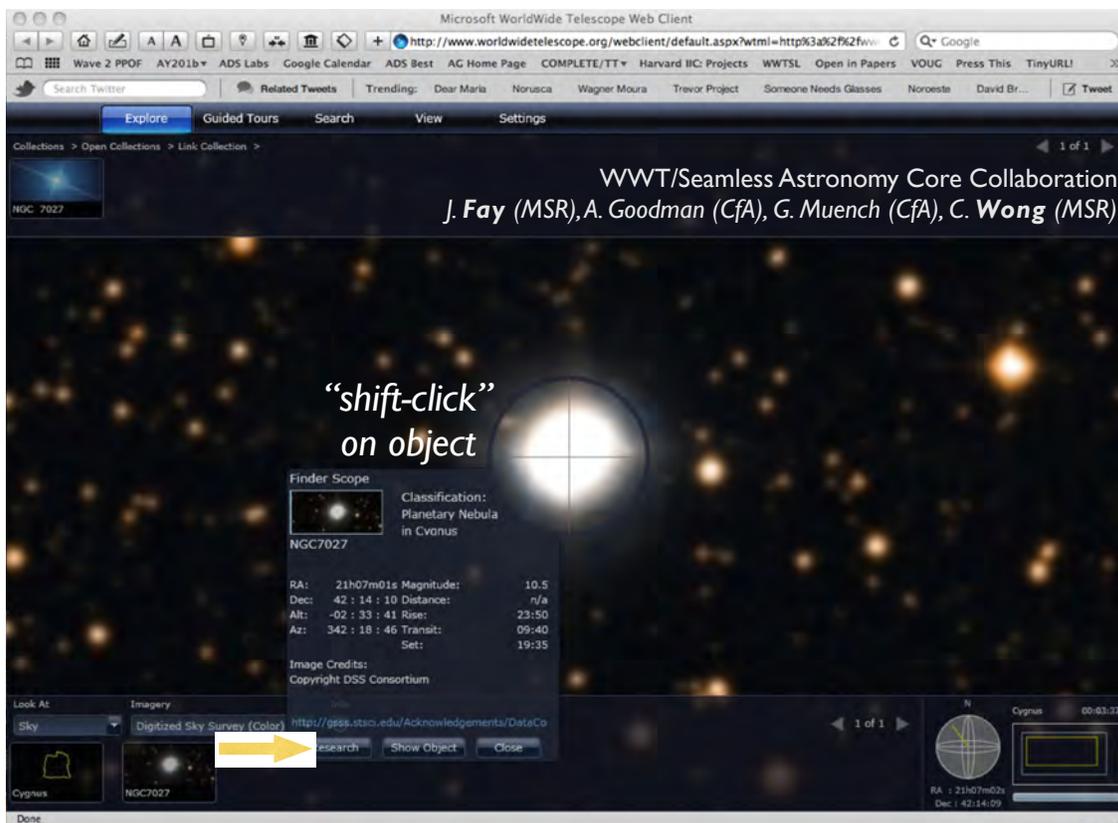


FIGURE 2-12 Worldwide Telescope.
SOURCE: Alyssa Goodman

Basically what we are being offered when clicking on the different options under “Research” is either more data or more literature. The links are such things as “look up all the publications about that particular object,” “look up that object on Wikipedia,” or “look up that object in astronomy databases.” This system offers a lot of flexibility. However, even though the system is supposed to list the most relevant papers, it is employing a very low level of semantics to do the search. For example, because I know this field, I can tell if the list of papers is incomplete. It is partially right, but it is not giving the full picture to a scientist who has never worked in this field, and that can be very misleading.

At this point, I should recognize the people who are involved in this project. Worldwide Telescope was envisioned by Curtis Wong and implemented by Jonathan Fay, who is a software engineer at Microsoft Research. At the Harvard-Smithsonian Center for Astrophysics, Gus Muench and I have been working closely with them since the beginning, but Curtis and Jonathan get all the credit.

Let me give another example. Assume that I find an astronomical image somewhere on the Web or that I go outside on a clear night with a camera and I take a picture of the sky. With this service called *astrometry.net*, I can submit a picture, and the service will take the image and miraculously calculate exactly where it goes in the sky and, if the picture is good enough, even figure out when it was taken.

There was a service that some researchers were hosting to do this work, but they did not have enough money or servers. So they made a group called “astrometry” on Flickr. As illustrated in figure 2-13, when you upload a picture to Flickr, not only does it host the image but it also offers some services that, for example, tell you what is in that picture. As soon as a picture finishes loading, if you add it to the astrometry group, you will soon see all sorts of little boxes pop up on the picture, and it will tell you famous objects in the picture that it recognizes. The reason Flickr can do this is because the *astrometry.net* program went in and blindly solved for the coordinates of the image. On Flickr, after an image is solved by *astrometry.net*, you will see a button at the bottom of the page that says “View in Worldwide Telescope.” If I click this button, then I will see that object overlaid on the sky where it goes, and I can compare it to 40 different layers of multiwavelength data. I can also use all the catalog-searching tools that I showed before and interact with all the statistical software, plotting software, et cetera, live.

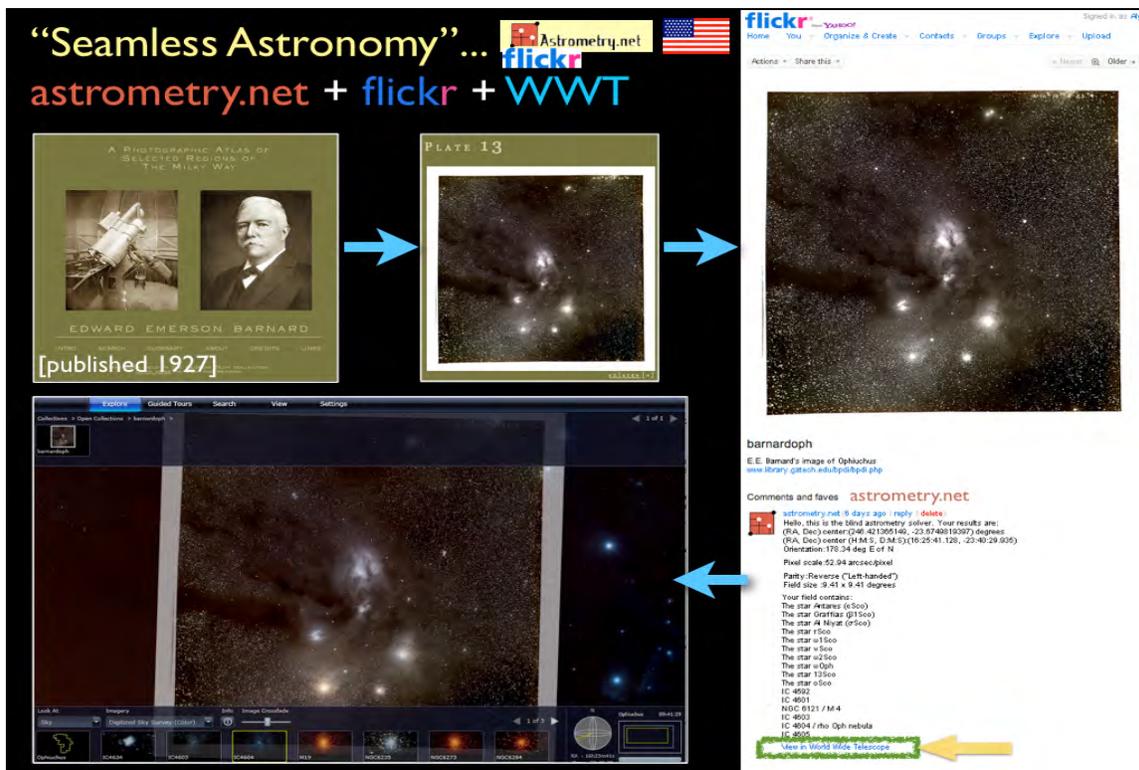


FIGURE 2-13 Seamless Astronomy.
SOURCE: Alyssa Goodman

There are two more projects in which Alberto Pepe is involved. In one of them, we are taking all of the historical images that are online, including all of the literature, and making a historical image map of the sky. With the second project, we can tell where particular papers on the sky are and why those papers were written, so you can get heat maps of articles on the sky.

I will conclude by highlighting a couple of examples where one field of study has borrowed from another. We published a paper (Goodman et al. 2009, *Nature*) 2 years ago⁹ that has embedded in it an interactive three-dimensional (3D) diagram; the idea is that you go there, click it, and you can put the data in the literature (Figure 2-14).

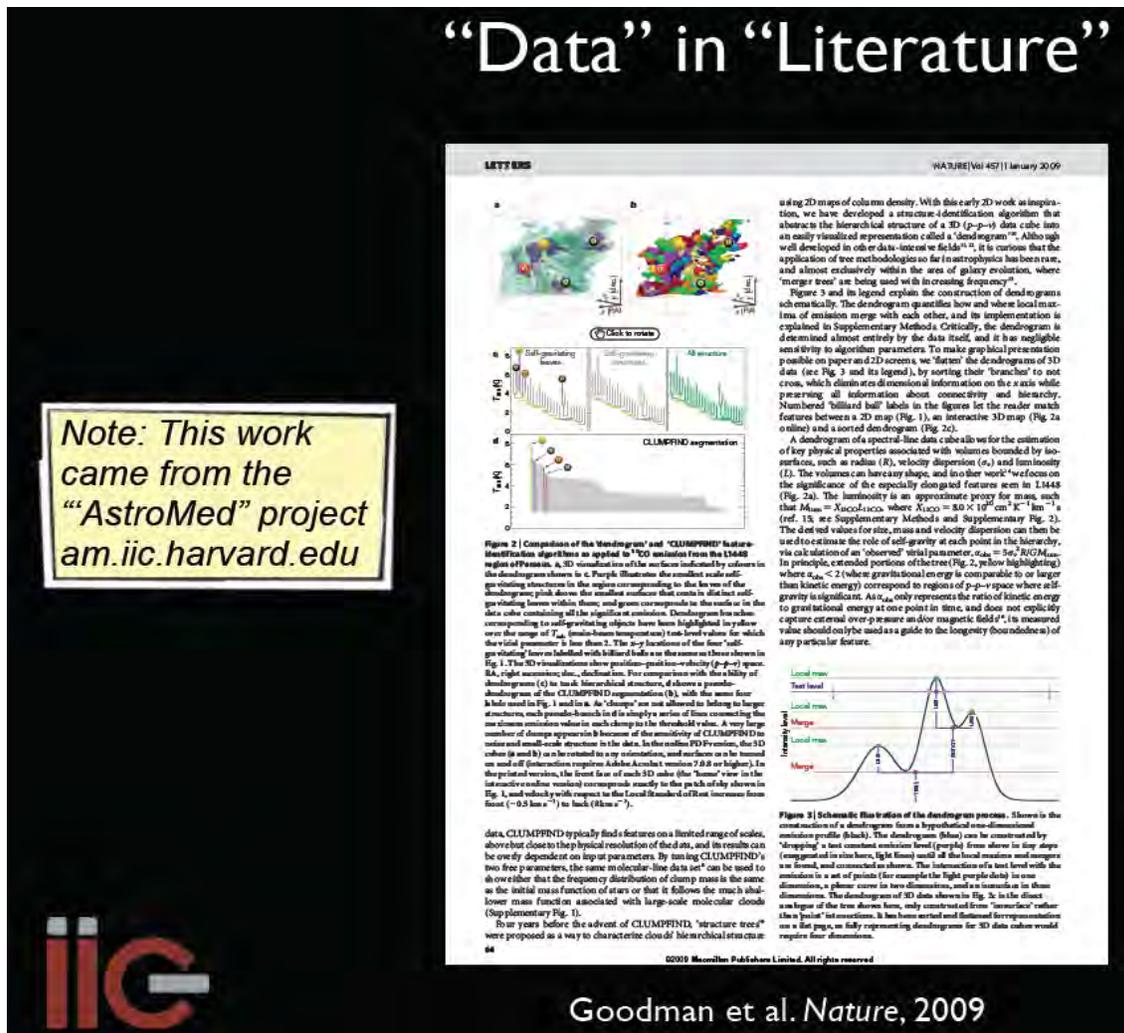


FIGURE 2-14 Data in literature.
SOURCE: Alyssa Goodman

The techniques used to create and study the 3D data were borrowed from medical imaging and used in astronomy. In the new Astronomy Dataverse (theastrodata.org), we have borrowed tools from the social science world as well. The Astronomy Dataverse tools allow individuals or groups to publish their data online in ways that can be retrieved in the future and that are permanent and linked to literature. At present, we are experimenting with the Astronomy Dataverse at the Center for Astrophysics, to let researchers who have small datasets and want to share them with everybody have a way to publish them that is linked to the literature and is persistent.

⁹ *Nature*, Volume 457, January 1, 2009.