

Seamless Astronomy

www.cfa.harvard.edu/~agoodman and worldwidetelescope.org

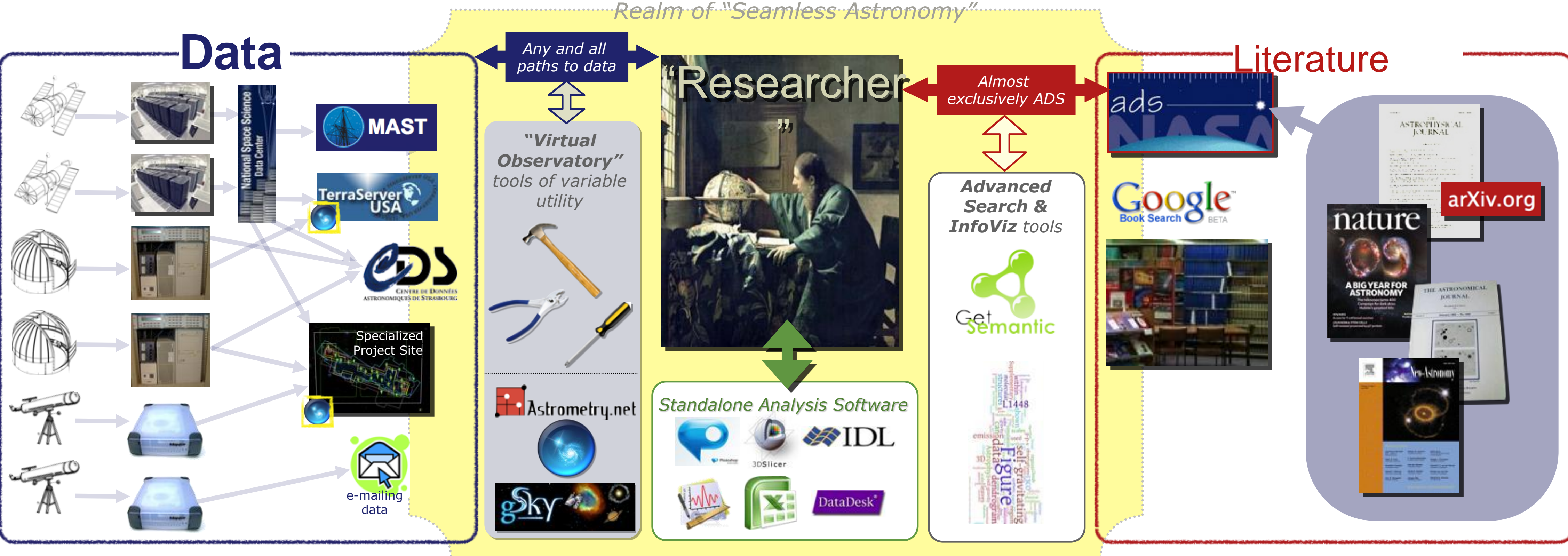
Harvard University

Alyssa Goodman



What's the Problem?

The Astronomical community is now in the situation of having a large body of information—both data and literature—available via the World Wide Web. New facilities are about to *drastically* increase the size of this data corpus, but existing search-and-retrieval systems were not designed for even the present quantity and diversity online data. A new approach is needed to guide astronomers to the information needed to make new discoveries, whether the sought-after bits represent a single number, or a million independent observations. The **current, rather overwhelming, situation facing research astronomers looks a bit like this:**



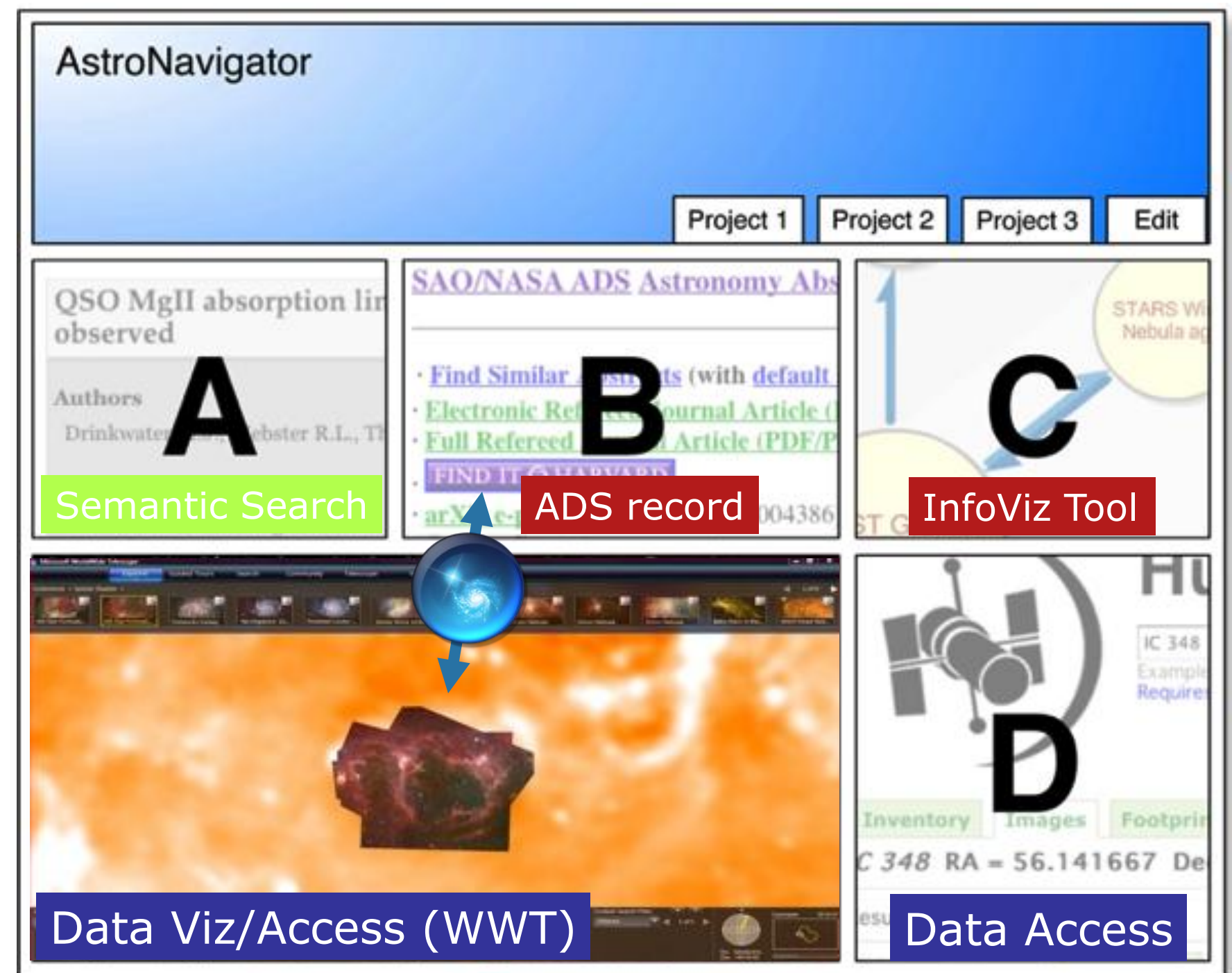
Where are the "seams"?

Some attempts (e.g. NASA's multi-mission archive (MAST), Microsoft WorldWide Telescope's data on the TerraServer, GoogleSky's data) have been made to **consolidate access to data**, but many data sets are still only discoverable available via custom searches and requests (even by email!). On the whole, the "data corpus" is laudably tending toward becoming a "cloud", but even that soon-to-be cloud is still essentially **sewn off** from **analysis-** and **literature-** focused research tools. The literature-access tool of choice in astronomical research is unquestionably "ADS" the **NASA Astrophysics Data Service**, which is based at the **Harvard-Smithsonian Center for Astrophysics**. Thanks to the combination of ADS and Google Books, physical library use is very infrequent, except in the case of access to rare holdings like image collections or other historical artifacts. ADS, at present, is essentially isolated from non-literature data sources. In spite of recent experiments with infoviz tools, and limited links (in <10% of papers) to "data" related to individual papers, astronomers presently think of ADS as a way to search for and retrieve Journal articles. The currently deployed ADS search tools are currently purely keyword-, and text-search based, and *do not use any semantic or infoviz strategies*. The **ADS interface** has been static for about a decade, and its use is **sewn off** from the rest of astronomers research tools. The goal of the national-scale "virtual observatory" projects in the US, Europe, and elsewhere, is to someday create a data/literature ecosystem where any researcher can access any information and any tool easily, online. Our goal is more limited, aimed at creating a tool that allows access to services more directly at the researcher's fingertips--in the **yellow box**.

Forays into Seamlessness

Thanks to plug-ins, mashups, and "cloud-think," plus a **generally more modular approach to computing**, it is possible to "seamlessly" attach existing and soon-to-be-available astronomy software and databases so that any astronomer can **create their own custom data retrieval/display/analysis tool**. The **WorldWide Telescope** from Microsoft is a "seamless" image browser offering much of the function those working on professional "virtual observatory" interfaces have long sought. **Google Sky** provides much of the same functionality, albeit with a less elegant interface. A full "seamless" astronomical research environment, though, needs more than image browsing--it needs non-image data, and access to literature. The mockup at right shows a mockup "AstroNavigator" mashup which can stitch together a custom suite of applications on the fly, via plug-in and/or message-passing approaches. An early attempt at this approach, developed in Europe, uses message-passing protocols and is called "SAMP." (Note that Microsoft's WorldWide Telescope is now SAMP-enabled.)

Our "Seamless Astronomy" project is focusing first on designing a system for "**Concept Oriented Retrieval of Astronomical Data**," or "**CORAD**." The goal is to create and implement a sample suite of tools to plug in to an "AstroNavigator"-like interface that allows a researcher to **use the astronomical literature as a fantastic filter for data search**. WWT is likely to be the data browser of choice for most users, and we can use it given the new "web version" (see below). A "right-click" in the present version of WWT already can bring up an ADS window listing publications about any particular object.



Accomplishments to Date (many thanks to Jonathan Fay of MSR...demos on request!)

- "Silverlight" (web) version of WWT to be launched Spring 09 (alpha version now available).
- Web version of WWT is now also used as a **plug-in** in the COMPLETE Survey's data coverage tool (as is Google Sky API).
- further WWT integration with (government-sponsored) "**Virtual Observatory**" project tools, including "**SAMP**" message-passing hub & **NVO cone-search**
- **First "3D PDF"** published (Goodman et al.) in *Nature* on 1/1/09, integrates dataviz tool (3D Slicer) and a paper in the refereed literature.
- Prototype **faceted search browser** for astronomical literature developed (by D. Burke, CfA), using lightweight **semantic ontology**
- proposal submitted to NASA AISRP for "CORAD," in **collaboration** with researchers at MSR (Wong), RPI (Hendler, McGuinness), STScI (Conti, Christian), ADS (Accomazzi, Kurtz) and in consultation with Chandra (Rots, Burke) and UCLA (Borgman) researchers

Coming Soon

- **astrometry.net** to be used as automated tool for extracting images from Journal PDFs (via ADS), with posting of resulting image to WWT
- **automated data collections** for WWT based on data utilized in specific papers in ADS

"Seamless Astronomy" is based at Harvard University, as a collaboration between researchers at the Harvard-Smithsonian Center for Astrophysics (CfA), and the Initiative in Innovative Computing (IIC). Dr. **Gus Muench** is the Project Manager at Harvard. The key collaborators at Microsoft are the creators of WorldWide Telescope: **Curtis Wong** and **Jonathan Fay**. WWT and this project were also inspired by the pioneering work of Jim Gray in his quest to see a "WorldWide Telescope." Additional participants in our overall "CORAD" effort are located at RPI, UCLA, and the Space Telescope Science Institute. The project is supported by the



External
Research

Microsoft
Research