Executive Summary on the 2016 Astrolabe Workshop

July 11-12, 2016
University of Arizona, Tucson, Arizona

Report prepared by:

P. Bryan Heidorn, PhD; heidorn@email.arizona.edu
Gretchen Stahlman; gstahlman@email.arizona.edu

Workshop sponsored by:

University of Arizona Office for Research & Discovery
American Astronomical Society

Workshop organizing committee:

Bryan Heidorn, University of Arizona School of Information
Chris Kollen, University of Arizona Libraries
Gretchen Stahlman, University of Arizona School of Information
Julie Steffen, American Astronomical Society

Astrolabe Co-PI's

Bryan Heidorn, University of Arizona School of Information
Dennis Zaritsky, University of Arizona Department of Astronomy and Steward Observatory

Workshop participants:

Scott Chernoff	Eiichi Egami	Angelo Fausti	Katie Frey
Jeremy Frumkin	John Glaspey	Richard Green	Robert Hanisch
Bryan Heidorn	Edwin Henneken	Barbara Kern	Chris Kollen
Robert McMillan	Susan Miller	Gus Muench	Joan Najita
Martha Narro	Jake Noel-Storr	Pat Norris	Kim Patten
Doug Roberts	Greg Schwarz	Rob Seaman	Edwin Skidmore
Gretchen Stahlman	Julie Steffen	Benjamin Weiner	Henry "Trae" Winter
Huanian Zhang			

Astrolabe Advisory Board Members:

Robert Hanisch, Bryan Heidorn, Edwin Henneken, Barb Kern, Chris Lintott, Doug Roberts, Julie Steffen, Frank Timmes, Benjamin Weiner, Henry "Trae" Winter, Dennis Zaritsky

Introduction and Purpose

The Astrolabe Project began in 2013 with discussions at the April AAS publishing strategy meeting between AAS meeting attendees and Dr. P. Bryan Heidorn, who researches the nature of uncurated “dark” data prevalent in the “Long Tail” of scientific research funding [Heidorn, 2008]. A workshop was held in July 2015 at the University of Arizona (UA) in Tucson to explore the idea of collaborating on creation of a hub for dark data in astronomy at UA. Outcomes of this workshop validated the researchers’ assumption that astronomers would benefit from an astronomy data hub that would provide both a resource for data storage and sharing as well as a computational environment for data analysis and reuse. Following this workshop, in January 2016 the University of Arizona’s Office for Research & Discovery granted the Astrolabe Project a one-year “Accelerate for Success” seed grant. Partnerships were established between the School of Information and the College of Social & Behavioral Sciences, the College of Science, the UA University Libraries, CyVerse, and the American Astronomical Society to provide computational and human resources to begin such a project. While the July 2015 workshop explored the concept and feasibility of creating a new data repository, the recent 2016 workshop focused on envisioning the Astrolabe system and establishing functional requirements, with a goal of creating a prototype data management environment populated with a variety of data products and driven by powerful science cases, over the next six months.

Workshop Structure

Workshop participants were identified through several channels. First, invitations were sent to most of last year’s attendees, considering this year’s focus on developing system requirements. Invitees who were unable to attend were asked to recommend a replacement. Second, authors of highly-cited papers referencing potential candidates for data or data-types that could be ingested into Astrolabe were invited. Word-of-mouth resulted in additional attendees. The final attendee list included twenty-two local and non-local participants - a group comprised of research astronomers, UA and CyVerse representatives, and other expertise related to astronomy data and publishing – along with five workshop organizers.

Just prior to the Astrolabe Workshop, the first meeting of the Astrolabe Advisory Board was convened. Board members broadly discussed the research challenges influencing the creation of Astrolabe, as well as the general function and operation of the board. This meeting guided the design of the workshop’s three breakout sessions and related group discussions. The workshop began at 1:00 pm on July 11 with a welcome and introduction from American Astronomical Society, including a recorded statement of support from AAS Executive Officer Dr. Kevin Marvel. Dr. Heidorn then updated the group on Astrolabe, the project’s activities and progress over the previous year, as well as the workshop structure and objectives. A series of invited talks provided an overview of the resources available through CyVerse, the Unified Astronomy Thesaurus (under the stewardship of AAS and a likely taxonomic tool for Astrolabe), and a general discussion of data management and visualization, and publishing in astronomy, along with a demonstration of the functionality and searching capabilities of the new ADS Bumblebee interface.

The first day's afternoon breakout session asked four breakout groups to discuss the most important and practical solutions, data types and astronomy subfields to direct focus for incorporating "dark" data that is both useful and accessible into Astrolabe. When the workshop re-convened on Tuesday morning, each breakout group briefly discussed the prior day's conclusions and reported out. The key areas of focus identified by the four breakout groups were listed, and participants voted on the four most critical and practical. Subsequent breakout sessions were organized around the following four topic areas recommended for the immediate attention of Astrolabe:

- Physical format of dark data
- Author websites archiving data
- LSST time domain and serendipitous data cases
- Searching the literature for dark data

Workshop participants explored these topics in depth during a breakout session directed towards the functional requirements that would be necessary to address each topic from both a user and developer perspective. The final breakout session was condensed to three breakout groups to discuss funding opportunities and overall strategies for the sustainable growth of Astrolabe. Several additional invited talks on Tuesday afternoon provided further context for Astrolabe through a discussion of the strengths and weaknesses of several case studies. The workshop concluded with a final whole-group discussion.

Outcomes and Recommendations

Physical format of dark data: Historical data could be made accessible through Astrolabe. Two datasets at NOAO and LPL would likely be of interest to the community and could function as prototypes for Astrolabe. Promising funding sources for these efforts were identified by UA Library participants, specifically for data preservation in the U.S. national library space. Similar efforts at the Harvard-Smithsonian Center for Astrophysics involving astronomical plates represent a possible partnership for archival expertise, and additional funding opportunities exist for such efforts. These data represent an ideal opportunity for citizen science projects as well.

Author websites archiving data: Authors have often linked to data on personal websites from the literature, and these websites are typically not long-lived. Astrolabe could provide authors with a website tool through CyVerse, ensuring the integrity of links in published literature. This would allow authors to freeze a site at publication to preserve a record of the research process. Several existing web pages should be ingested as a proof of concept. Search tools are necessary to locate files across collections. Templates could be created through CyVerse, and users could adopt templates for the appropriate data structure, including key presentation and descriptive information. Metrics are needed to demonstrate usage. DRYAD might be a business model for this project. Since the Moore Foundation funds research tackling grand challenge problems, this is a target funding opportunity for bringing together rich datasets in the Astrolabe environment through CyVerse.

LSST time domain and serendipitous data cases: A key opportunity for Astrolabe is follow up to LSST observations, and bringing together datasets and researchers from multiple telescopes, if possible. Another potential source of serendipitous discovery is NOAO historical data on nearly obsolete, institutional and individual removable media (see above), which could be converted to current formats and ingested into Astrolabe. Historical data represent an opportunity for citizen science and educational opportunities. The amateur astronomy community is also a potential target for data. It is recommended that Astrolabe start with a minimum of 100 TB of storage space, planning to scale up to 10 PB over 10 years. The system should include the following features: support for WWT interface; minting DOIs; checksum; licensing; capability to upload notes and other historical metadata related to data; public/private options and related policies; support for building tools; and capturing provenance of data. To build a sustainable framework, Astrolabe should consider federation with other models such as CADC and Data Lab, and highlight costs saved by utilizing CyVerse in proposals to NSF and private foundations.

Searching the literature for dark data: Text mining could locate references to dark or "gray" data in the literature. Authors could be incentivized to share such data with the availability of a resource for archival such as Astrolabe, and subsequent citation advantages. Highest impact papers and recent publications (within the past 10 years) could be targeted first. ADS has developed a good tool for this project, with relevant expertise and community support, and full text for text mining to locate papers without data links. Candidate data associated with an existing publication would advertise for Astrolabe and provide growth and overall sustainability for the project.

Final comments: Participants recommended incremental status reports through an established listserv. The envisioned timeline for Astrolabe aims to have a web ingestion tool based on CyVerse tools by October 2016, along with examples of websites that could be ingested. Before the initial seed grant period ends in January, proposals will be submitted to NSF and Sloan, and Astrolabe will have websites and data ingested, along with tiling tools, FITS metadata extraction, and UAT integration.

Objectives Moving Forward

Following the recommendations above, Astrolabe will be developed as a prototype data system in CyVerse over the next six months, with a limited variety of data to start. In all activities, Astrolabe must refine its focus and specify activities through clear goals and functional requirements. Proposal development will include the NOAO tape data rescue project, led by AAS and the UA Libraries. Astrolabe Accelerate Success staff member Huanian Zhang will investigate candidate dark data sets using existing lists and ADS search tools. Astrolabe Advisory Board meetings will be held quarterly and reports will be distributed to stakeholders through a listserv. A website will be created for Astrolabe, and project management tools and communication channels for the board and stakeholders will be developed. Partnerships will continue to be cultivated, including with LSST, NOAO and LPL.

Reference

Heidorn, P. Bryan (2008). Shedding Light on the Dark Data in the Long Tail of Science. Library Trends 57(2) Fall 2008 . Institutional Repositories: Institutional Repositories: Current State and Future. Edited by Sarah Sheeves and Melissa Cragin. (http://hdl.handle.net/2142/9127).