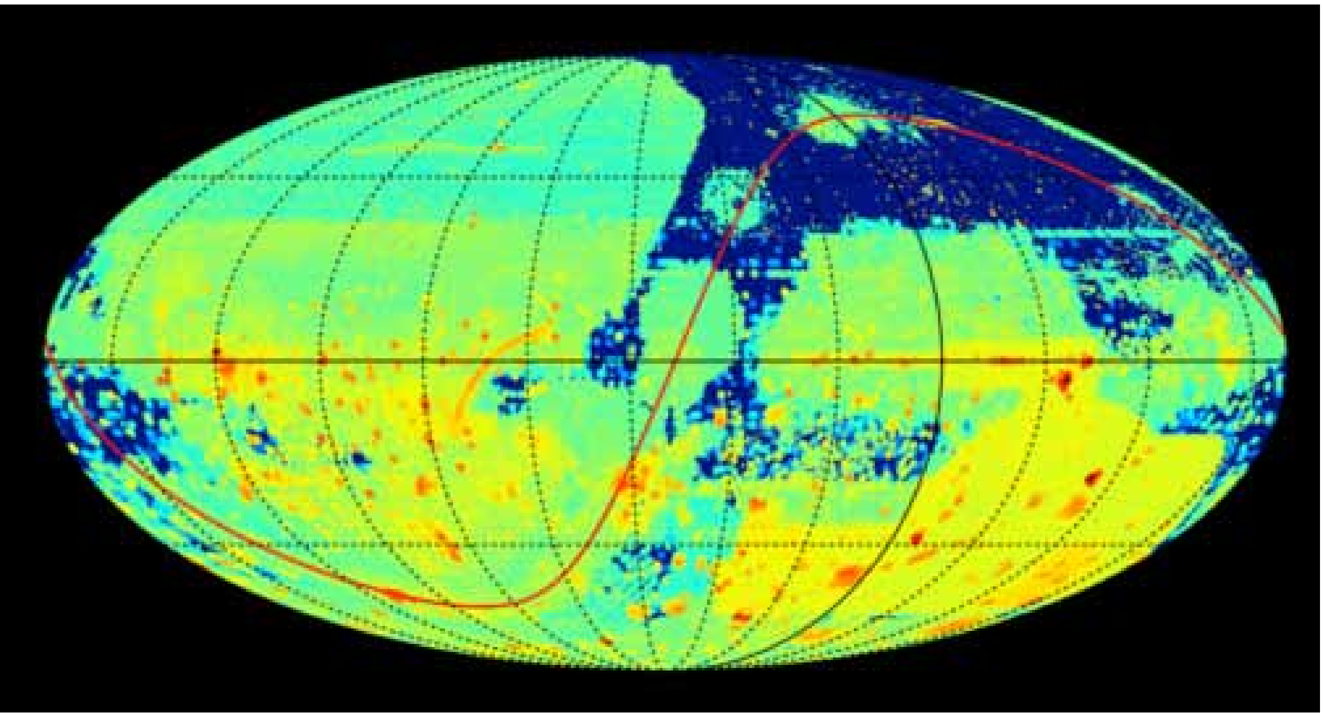
ASTRO BEAT

Can Big Data Lead an Inclusion Revolution?

Dara Norman (National Optical Astronomy Observatory)

July 2018



Crowd-sourced survey of the sky from the NOAO archives as seen in projection with the galactic plane shown as the red line. The image shows the total numbers of images taken with the DECam (south) and Mosaic (north) cameras from 2004-2017. Dark blue areas have no exposures while red areas have the maximal number (~1000 images). Image by K. Olsen.

Currently, there are two potentially paradigm-shifting trends taking place in astronomical research. The first is the move away from individuals or groups of observers obtaining data for a narrow scientific experiment, towards the use of grand surveys and large datasets and catalogs that enable a wide range of experimentation. The second trend is the recognition that the astronomical and astrophysical (ASTRO) community of researchers must become more inclusive in order to realize the best scientific innovation and productivity. Leveraging both of these trends now provides the field with a unique opportunity for both to be mutually supportive in the quest to advance scientific discovery! However, this can only happen if the necessary investments are made to provide the resources that support both of these ambitious movements.

Certainly, ever since humans began to look up at the night sky, we have wondered about the stars, planets and other celestial bodies. Stargazing included all types of people across the globe who have contributed star and constellation names, myths and folklore, navigation and harvesting rituals based on the patterns of the sky. However, within the US, professional ASTRO is not representative of the variety of people in the nation. For example, as of 2010, statistics from the American Institute of Physics show that women made up less than 20% of the faculty in astronomy and physics departments and the numbers for underrepresented minorities were, and remain, dismally low. More recently, there have been many efforts to increase the numbers of women and underrepresented minorities who opt to participate and continue on in ASTRO research. Conferences like Women in Astronomy I – IV (http://www.cvent.com/events/women-in-astronomy-iv-the-many-faces-of-women-astronmers/event-summary-589214b84ab94f26ac269ad9823ef977.aspx) have brought attention to the ways in which women have been marginalized, discouraged and shut out of ASTRO research and positions. These conferences engage the larger community in direct actions to remove the barriers faced by women in the field. Similarly, the 2015 Inclusive Astronomy Conference at Vanderbilt University (vu.edu/ia2015) grew as a grassroots movement to address these concerns (marginalization, discouragement, barriers) for intersectional [1] and other disenfranchised groups. The conference addressed four broad areas that the organizers hope will be instrumental in creating a better working environment and increased opportunities for all ASTRO researchers. In addition to these conferences, several programs have over the years been enabled to address some of the structural barriers and allow underrepresented groups access to ASTRO careers. Several institutions have followed the lead of the Fisk-Vanderbilt Bridge program and started other bridge programs that identify undergraduates potentially interested in a PhD program and link them, through a terminal Master's program to PhD granting institutions. Several of these bridge programs

The wide-field DECam camera on the Blanco telescope at Cerro Tololo Inter-American Observatory (CTIO) seen before being inserted into the instrument. Image by T. Abbott & CTIO/NOAO/AURA/NSF.

[1] Intersectional (coined by Kimberlé Crenshaw) refers to people who face discrimination and marginalization along multiple axes of their identity. As demonstrated in the landmark case, Degraffenreid vs. General Motors (1977), this discrimination comes not because of one identity or another, but because of the combination of both identities. That is, those who face increased discrimination because they are at the intersection of two or more marginalized identities.

Conference picture from Inclusive Astronomy 2015. On June 17-19, 2015 at Vanderbilt University, 160 astronomers, sociologists, policy makers and community leaders convened to discuss intersectional barriers and solutions to success in astronomy. The conference was centered around 4 themes: Removing Barriers to Access; Creating Inclusive Climates; Establishing a Community of Inclusive Practice; and Improving Access to Policy, Power, and Leadership.

have pioneered alternative criteria for the assessment of graduate school success of students (http://fisk-vanderbilt-bridge.org/toolkit/). In 2012 invited testimony for the National Academies' Seeking Solutions: Maximizing American Talent by Advancing Women of Color in Academia Conference, Norman et al (https://www.nap.edu/read/18556/chapter/18) identify nine barriers that women of color face as they pursue a PhD and career in ASTRO. At least two of these barriers (difficulty building networks and collaborations and difficulty achieving insider status), might be directly mitigated by tearing down barriers that limit access to ASTRO data and data products. Access to data products can fuel the ability of under-resourced scientists, like women of color, to have the opportunity to avoid nay-saying gatekeepers in pursuit of their research goals.

The Big Data movement in ground-based optical and near-infrared astronomy has been on the rise for several years. In the 1990s, the Sloan Digital Sky Survey (SDSS) embarked on an ambitious and groundbreaking exploration of more than 8,000 square degrees of sky to unprecedented depths. This survey ushered in a new way of approaching science with a large survey and public data products to service multiple experiments. The National Optical Astronomy Observatory (NOAO) has 'crowd sourced' a nearly complete survey of the sky through the capture of data from multiple individual surveys taken with modern wide field cameras like the Mosaic (at Kitt Peak

National Observatory) and Dark Energy Camera (DECam, at Cerro Tololo Inter-American Observatory), with a field-of-view 20 times the size of the moon as seen from earth. Coming in the future, the Large Synoptic Survey Telescope, expected to begin science verification in 2021, will be the flagship ground-based facility into the next decade, surveying the accessible sky every few nights and delivering 200 petabytes of data over ten years.

While telescope time for any individual project is difficult to obtain because of the limited number of nights available, weather, and scientific competition, an advantage of survey data is that all data taken is currently or will be archived and available for any researcher to use. Publications from archival data become a significant body of work over time. For example, statistics from the Hubble Space Telescope (https://archive.stsci.edu/hst/bibliography/pubstat.html) and the Spitzer Space Telescope, show that after about 8-10 years following the start of the mission, the number of publications by researchers using data obtained from the archives is about equal to the number of publications from those who received allocated telescope time directly. This is similar to what was seen with data from the ground-based SDSS survey (Chaomei Chen, Jian Zhang, Weizhong Zhu, Michael S. E. Vogeley, JCDL, 2007), where 43% of the bibliographical records between 1992 -2006 came from the 24 'most productive' institutions, which were generally the affiliations of original survey members. However, 57% of the papers came from other institutions, many of which were likely researchers at 'non-insider' institutions. Thus, if the necessary resources are made available, for archiving, software, user support and training, we will be able to make the pursuit of astronomical research more available, inclusive and open to researchers at a variety of institutions, including those of scientists with limited resources.

Archival databases are an important step in allowing researchers to access data, however, just putting data in a database does not make it useful to a broad section of the community. Older archives require users to know specifically what data they want and often they must be willing and able to retrieve more than what they want, in a piecemeal way to do their science. At NOAO, we are investing in making our archival datasets not just available, but discoverable and useable by the entire ASTRO community, through a data science platform called Data Lab (https://datalab.noao.edu/), that allows even those with limited computer and support resources access to data, data products and catalogs. The goal of this approach is to enable the researcher to keep their scientific question in the forefront of their work while they discover and identify the data needed for their study. With services that allow login space for registered users and thumbnail image cutouts of regions of interest, users need not have the network bandwidth to download and then sort through multiple gigabytes of data in order to narrow down the necessary experimental data. They can use existing software or design their own tools and use them close to the data.

Astronomical big data CAN lead an inclusion revolution, but only if we truly invest the resources—to enable the design, building and long-term curation of data archives; to sustain the work of programmers and scientists to develop, maintain and support software tools and; to provide opportunities for researchers to learn about and become proficient in the use of the tools and services developed. The goal of inclusion will not be reached without a sustained, committed backing of the full big data enterprise. It is crucial that the scientific community recognizes that investments made in pursuit of inclusion, by way of big data, are the way we advance the field of astronomy and astrophysics into the next century and beyond.

About the Author

Dr. Dara Norman is the Deputy Associate Director of the National Optical Astronomy Observatory's Community Science and Data Center in Tucson, AZ. She is also the AURA Diversity Advocate at NOAO. The duties of this position include creating and advancing opportunities at NOAO/AURA to bring more underrepresented minorities and women into the "astronomy enterprise," which includes research science, engineering, data science, and instrument building. Dr. Norman has been an active member of the AAS's Committee on the Status of Minorities in Astronomy and was chair of the astronomy and astrophysics section of the National Society of Black Physicists. She has participated in numerous public outreach programs including ASP's Project Astro at NOAO and University of Washington, and the NSF's Research Experiences for Undergraduates at NOAO and CTIO. Dr. Norman holds M.S. and Ph.D. degrees in Astronomy from the University of Washington and a B.S. in Earth, Atmospheric and Planetary Science from the Massachusetts Institute of Technology.

NOAO

AstroBeat

Number 162 • July 2018

Publisher: Astronomical Society of the Pacific
Editor: Greg Schultz
Designer: David Barker

The Astronomical Society of the Pacific increases the understanding and appreciation of astronomy by engaging scientists, educators, enthusiasts, and the public to advance science and science literacy.