Astro2020 Science White Paper

Increasing the Discovery Space in Astrophysics: The Exploration Question for Resolved Stellar Populations

Thematic Area: Resolved Stellar Populations and their Environment

Principal Author:
Name: G. Fabbiano
Institution: Center for Astrophysics | Harvard & Smithsonian (CfA)
Email: gfabbiano@cfa.harvard.edu
Phone: 617 495 7203 (but email is much better)

Co-authors:
M. Elvis (CfA), A. Accomazzi (CfA), G. B. Berriman (Caltech/IPAC-NExScI), N. Brickhouse (CfA), S. Bose (CfA), D. Carrera (Penn State Univ.), I. Chilingarian (CfA), F. Civano (CfA), B. Czerny (Center for Theoretical Physics, Warsaw; CfA collaborator), R. D'Abrusco (CfA), J. Drake (CfA), R. Emamia (CfA), J. R. Farah (CfA), G. Fazio (CfA) E. Feigelson (Penn State Univ.), F. Fornasini (CfA), Jay Gallagher (Univ. Wisconsin), J. Grindlay (CfA), L. Hernquist (CfA), D.-W. Kim (CfA), M. Karovska (CfA), G. M. Lacy (NRAO), J. Lazio (JPL, Caltech), W. P. Maksym (CfA), R. Martinez Galarza (CfA), J. Mazzarella (Caltech/IPAC), D. Sanders (Institute for Astronomy, Hawaii), N. Scoville (Caltech), I. Shapiro (CfA), A. Siemiginowska (CfA), A. Smith (STScI), H. Smith (CfA), A. Szentgyorgyi (CfA), S. Tacchella (CfA), A. Thakar (Johns Hopkins Univ.), V. Tolls (CfA), B. Wilkes (CfA), D. Willner (CfA), S. P. Willner (CfA), S. J. Wolk (CfA), Zhao, J.-H. (CfA)

Abstract:

We write in response to the call from the 2020 Decadal Survey to submit white papers illustrating the most pressing scientific questions in astrophysics for the coming decade. We propose exploration as the central question for the Decadal Committee's discussions. The history of astronomy shows that paradigm-changing discoveries were not driven by well-formulated scientific questions, based on the knowledge of the time. They were instead the result of the increase in discovery space fostered by new telescopes and instruments. An additional tool for increasing the discovery space is provided by the analysis and mining of the increasingly larger amount of archival data available to astronomers. Revolutionary observing facilities, and the state-of-the-art astronomy archives needed to support these facilities, will open up the universe to new discovery. Here we focus on exploration for resolved stellar populations.

1. The exploration question

There has been a long-standing tension in our discipline between the 'exploration' approach and the more physics-based 'question-driven' approach.

1.1 - *The* ***question-driven*** *approach*

This approach seeks to formulate the most important open questions in our discipline. It is based on our *present knowledge* of the field (both theoretical and observational) and is formulated usually as a way to constrain and/or advance a currently proposed cosmological or astrophysical scenario. 'Question/hypothesis-driven' has been the preferred approach in the last few decades and is used to justify both observing proposals and proposals for new instruments and telescopes. Most talks at conferences and papers are framed based on this question and answer approach. This is how we teach our students to approach research. This is the approach formulated in the Decadal Survey call for white papers. This approach addresses the '*known unknowns*': for example, the way to best constrain the cosmological parameters of our universe, and lately the search for dark matter and dark energy, and the definitive discovery of gravitational waves.

The question-driven approach continues to be fruitful, and it gives us a certain sense of control in our progress, but -by its own nature- is also a limited and limiting epistemology. For example, it can bias our knowledge. As expounded in a recent article with reference to extra solar planets "the key is to make sure that science policy permits discovery for the sake of discovery and not for finding Earth-like planets, which we have prejudiced to be of greatest interest (D. J. Stevenson, CalTech, Physics Today, Nov. 2018)". The same opinion can be easily shaped to apply to other fields of astrophysics.

The question-driven approach does not address the *'unknown unknowns'* that by their nature cannot be addressed as well-defined 'important questions'.

1.2 - *The* ***exploration*** *approach*

This approach, i.e. gaining the capability to find new questions, rather than solving known ones, is the only way we can address the *unknown unknowns*. Harwit (1984) calls this '*discovery space*'. The notion that most of science is undiscovered and that 'out of the book' thinking may be needed for real progress is making fast inroads (e.g., see the book 'Ignorance: How it Drives Science' by S. Firestein, 2012). How to best foster the discovery of *unknown unknowns* is particularly poignant for astronomy, which throughout its history has been first and foremost exploratory.

The real big paradigm shifts in astronomy and astrophysics have occurred when new approaches have significantly opened up the discovery space, revealing unforeseen views of the universe. These approaches may have been framed as a way to address important questions of the time, but the real advances were from serendipitous discoveries. The discovery space may have been increased by means of new telescopes and instruments (both hardware and software), and also by *unanticipated data repurposing*.

Famous examples of discoveries stemming from exploration include:

- The Galilean Moons of Jupiter, the metal composition of the Sun and stars, the HR diagram, the expansion of the Universe, large scale structure, hot Jupiters (driven by improvements in optical telescopes and spectrographs);
- Quasars, radio galaxies, the microwave background, pulsars, superluminal motion, fast radio bursts (following the invention of radio telescopes, VLBI, and search in the archives in the case of bursts);

- Black holes and their mass range, dark matter, dark energy, super-starburst galaxies (from the availability of new space-based observing windows, X-ray, IR, and high resolution optical imaging with HST, and availability of multi-wavelength archives).

These foundational discoveries for the present understanding of the Universe and its evolution were not in any way anticipated. Most of them were fostered by the use of increasingly larger telescopes and more sensitive instruments, able to explore different parts of the electromagnetic spectrum. Others were surprising results of the data analysis.

Given the increasing availability of large and survey data sets in our open archives, a new hybrid approach, **question-driven exploration**, has emerged, where astronomers have mined these data and researched the literature guided by relatively vague questions, finding answers, new questions, and surprises (a similar approach is making inroads in biology; Elliott et al 2016).

In this white paper we discuss the '*exploration question'*, providing examples relevant for the field of resolved stellar populations. We include both serendipitous discoveries and question-driven explorations, resulting from unanticipated analyses of multi-wavelength survey data (Section 2). In Section 3, we address our recommendations for increasing the discovery space.

2. Exploration in Resolved Stellar Populations

Given the nature of exploration, it is not possible to give definite questions that need to be addressed in the near future. Rather, we provide a few examples of (1) serendipitous unexpected discoveries (*unknown unknowns*) and their potential for changing established paradigms; and (2) new research avenues posed by asking very general questions (*known unknowns*). We do not mean to provide an exhaustive survey of such discoveries, but only to illustrate our case with a few representative studies.

2.1 *The Panchromatic Hubble Andromeda Treasury (PHAT) survey*

PHAT (Dalcanton et al. 2012) consists of an imaging survey of ~1/3 of M31's star-forming disk in six filters, UV to NIR, to study resolved stellar populations in a controlled environment, at a given distance and metallicity while also avoiding the difficulties of observing stars through the inclined disk of our own Milky Way. PHAT resolved M31 into millions of individual stars, providing excellent constrains on stellar temperatures, bolometric luminosities, and extinction. *Unanticipated discoveries from the subsequent exploration of the dataset, and from combining it with other multi-wavelength surveys*, include:

- The discovery of weak CN stars, a previously unknown type of carbon star that appears to be associated with the He-burning phase of relatively massive stars (Masegian et al. 2019).
- The apparent universality of the high-mass (M>1M_sun) Initial Mass Function (IMF) across a broad range of cluster masses, ages, and sizes (Weizs et al. 2013),
- The strong UV bump in the extinction curve in the central region of M31, which indicates that dust destruction by supernova explosions are common in bulges of spirals (Dong et al. 2014).

2.2 *Characterization of the structure of the Milky Way halo via data exploration.*

Unexpected discoveries about the morphology of the Galactic halo stem from the exploration and mining of several large data sets:

- Using astrometric data from *Gaia* (Lindegren et al. 2018), Prince-Whelan et al. (2018b) found a young, low-mass metal-poor stellar association while searching for groups of co-moving blue stars in the far end of the Galactic halo. Its age and location suggest that this association was formed when the leading arm of the Magellanic Cloud gas stream last encountered the M.W. disk.

- With a joint *Gaia* and Pan-STARRs study, Price-Whelan et al. (2018a) discovered an extension of GD-1, the longest known cold stream in the Galactic halo. They derived the position of the progenitor tidally disrupted globular cluster and detected over-densities in the stream that might be related to perturbation by the long-predicted dark-matter subhalos.
- Using Deimos spectra (Keck-II) and the Palomar Transient Facility Database, Cohen et al. (2017) traced over-densities in the outer Galactic halo via the identification of variable RR-lyrae stars, whose heliocentric distance was directly obtained from their light curve.
- While inspecting HST photometry of the cluster NGC 6752 for the presence of white dwarfs, Bedin et al. (2019) discovered a dwarf spheroidal galaxy near the galaxy NGC 6744.

2.3 *The bones of the Milky Way*

Determining the 3D structure of the Milky Way from our position very close to its mid-plane has been a perpetual challenge for astronomers. *Data exploration can provide a way around this*.

A good example is the discovery that the previously identified filamentary infrared dark cloud (IRDC) known as ***Nessie*** is in fact a very long structure that runs along the very center of the Scutum-Centaurus spiral arm of our galaxy, and acts as a "spine" that supports it. This realization came from re-analyzing archival *Spitzer* imaging data and combining it with kinematic properties derived from star-forming gas traced by CO and NH3 (Goodman et al. 2014).

The unexpected discovery of these dense structures that run along the spiral arms of the Milky Way and the realization that the Sun is offset by about 25 pc from the plane that contains them, provide a new advantageous perspective for the observer, and opens a new avenue for investigations of the 3D structure of our Galaxy.

2.4 *Multiple stellar populations in globular clusters*

Globular Clusters were traditionally thought as single, first generation stellar populations. Observational evidence over the last decade has completely changed this view. For a given globular cluster, distinct populations enriched in He, N and Na, and distinct populations depleted in O and C, create complex patterns in the color-magnitude diagram that cannot be explained by individual stellar mixing and stellar evolution (Bastian & Lardo, 2018).

This discovery was highly unexpected. It came about through the exploration of large photometric datasets from HST observations, originally intended to characterize stellar evolution of a single populations.

Although several ideas have been put forward in order to explain this ‘multiple populations problem’, e.g., several bursts of star formation within the cluster, none of these theories can explain all the available observational evidence. Exploration of existing and upcoming pan chromatic surveys (Pan-STARRS, *Gaia*, LSST) are likely to boost a new way of discoveries and possibly provide some needed answers. Research in this particular area may provide new astrophysical insight about stellar evolution and star formation in the early universe.

2.5 *Future data mining of the Hubble archive*

The great heritage of the *Hubble* Space Telescope includes multi-color imaging for hundreds of star clusters in the Milky Way and Magellanic Clouds, collected by different observing programs. The results from the many studies based on these images are hard or even impossible to compare, because they were processed using a heterogeneous set of tools to perform photometry and then interpreted using a variety of stellar population models (e.g. isochrones).

The *Hubble Legacy Archive* has produced a uniform set of data products from these data, so that now the photometry can be extracted homogeneously across the entire collection and then analyzed using one or several particular sets of stellar population models. Exploration of this uniform large archival data set may provide answers to questions such as: Why did the Large Magellanic Cloud not form any star clusters between 9 and 3Gyr ago? What is the binary star fraction in clusters, and is it affected by the clustered star formation? How did star clusters with multiple stellar population form?

3. Increasing the Discovery Space

3.1 *Observing facilities that expand boundaries*

Any new observing facilities/missions for the next decade should significantly improve performance in some key metric (e.g., energy range, sensitivity, exposure time, angular resolution, higher dimensional data, rapid response), and be well characterized and calibrated, so to provide flexibility for new observing avenues. *Hubble*, *Spitzer* and *Chandra* provide examples in the discovery of Dark Energy, the detection of z=11 galaxies, and the nature of Dark Matter (Bullet Cluster), respectively. Beyond hardware capabilities these discoveries require: mission longevity, community driven science, high-quality data products in readily accessible, interoperable, archives and a well-supported user/observer community.

3.2 *Multi-wavelength and multi-messenger capabilities*

Many historical examples also demonstrate a strong synergy between different wavebands and messengers. Having contemporaneous access to the entire electromagnetic spectrum was vital to finding the first counterpart to a gravitational wave source, for example. This multi-wavelength coverage of the sky that we are currently enjoying needs to be preserved.

3.3 *Curated Data Archives and Powerful Data Analysis tools*

These new facilities will generate increasingly larger and more complex multi-wavelength and multi-messenger data sets and catalogs. These data will need to be properly reduced and curated to fully enable their discovery potential. *Archives must provide both easy access to these data and (with the community) the means to exploit them*. These goals translate into:

(1) Ensure that any operational (old and new) facility/mission explicitly include in their scope the proper processing of software so to produce well documented and calibrated data products, as well as the capability for data recalibration and reprocessing.
(2) Organize these data products in well-maintained archives, following the International Virtual Observatory Alliance (IVOA)[1] standards, so to allow a basic level of access and *interoperability*, as well as *repurposing.* Much of this is already in place in the NASA archives, and they are collaborating in extending and evolving these capabilities to meet the demands of new data types and research methods through the 2020s. Data products should be replicable and reproducible, ranging from basic observation data to high-level aggregated data and catalogs.
(3) Ensure that data centers engage in the development and refinement of interoperability standards, via the well-established processes of the IVOA, and work with groups such as *Astropy*[2] to ensure support for these standards in present in community developed, open source software.
(4) *New facilities* (Sections 3.1, 3.2) *will demand a transformation in the way data are analyzed*. The early phases of this transformation are already underway (e.g., the use of *Python* as an

[1] The forum for the development of the interoperability standards used by major astronomy data centers (http://www.ivoa.net)
[2] http://www.astropy.org/acknowledging.html

environment, cloud computing). But, resources must be made available for full development, which will demand remote Science Platforms[3] and Server-side analytics[4], implementation of complex fault-tolerant workflows, data mining and machine learning, and advanced visualization.

(5) Foster the development of *next generation* interoperable, user-friendly visual interfaces, data mining tools, the ability to construct and implement analysis workflows easily, both via visualization and scripting, and the ability to work with data both locally and remotely (current-generation well-know examples include TOPCAT, DS9 and CSCView).

(6) Support interdisciplinary research in astrostatistics and astroinformatics and the transfer of methods from the statistics, computer science, and machine learning communities, for development and application of innovative data analysis methods and algorithms.

(7) Ensure that facilities and archives participate in curation efforts and initiatives to link together datasets, related ancillary data (e.g., atomic and molecular databases), objects, and the literature.

Data are an important legacy of major astronomical facilities, and proper data maintenance will insure that new science will be produced for the future, even after the first crop of scientific papers and discoveries have been published. Statistics of data usage from the NASA archives demonstrate that archival data is used for new published scientific work several times (Fig. 1).

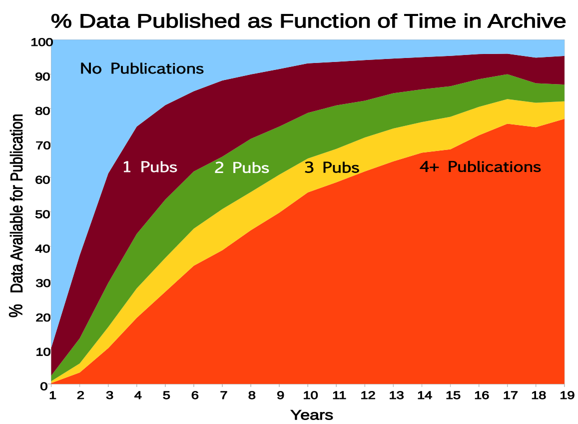

Figure 1. – Percentage of *Chandra* exposure time published versus years in the archive. The scientific use of archival *Chandra* data is increasing with time in the public archive. For example, 19 years from launch, ~75% of the observation have been published in more than 4 papers. A similar trend is observed for the HST data.

4. Conclusions

We propose *exploration* as the central question for the Decadal Committee's discussions. The history of astronomy shows that paradigm-changing discoveries were not driven by well-formulated scientific questions, based on the knowledge of the time. They were instead the result of the increase in discovery space fostered by new telescopes and instruments. An additional tool for increasing the discovery space is provided by the analysis and mining of the increasingly larger amount of archival data available to astronomers. We urge the Decadal Committee to **(1)** *keep multi-wavelength and multi-messenger exploration center stage* in their deliberations of new facilities, including consideration for flexible and well-calibrated modes of operation that could foster adaptation for use with new discovery space; and **(2)** *recognize the importance of data and their stewardship, and computational services*, as major elements of any new scientific development for the next decade. ***Revolutionary observing facilities, and the state-of-the-art astronomy archives needed to support these facilities, will open up the universe to new discovery.***

[3] See LSST Science Platform Design document https://ldm-542.lsst.io

[4] NASA Big Data Task Force (https://science.nasa.gov/science-committee/subcommittees/big-data-task-force)

References

Bastian & Lardo, 2018, ARA&A, 56, 83B

Bedin et al. 2019, MNRAS, 484L, 54B

Cohen et al. 2017, ApJ, 849, 150C

Dalcanton et al. 2012. ApJS, 200, 18D

Dong et al. 2014ApJ, 785, 136D

Elliot , K. C. et al. 2016, Bioscience, 66(10), 880. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5862324/

Firestein, S. 'Ignorance: How it Drives Science' 2012, Oxford Univ. Press

Goodman et al. 2014, ApJ, 797, 53G

Harwit, M. 1984, 'Cosmic Discovery: The Search, Scope and Heritage of Astronomy', The MIT Press

Lindegren et al. 2018, A&A, 616A, 2L

Price-Whelan et al. 2018a, ApJ, 867, 5P

Price-Weelan et al. 2018b, AJ, 156, 18P

Weisz et al. 2013, ApJ, 762, 123W