Astro2020 Science White Paper

Increasing the Discovery Space in Astrophysics
The Exploration Question for Galaxy Evolution

Thematic Area: Galaxy Evolution

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), B. Diemer (Cfa), J. Drake (CfA), R. Emami-Meibody (CfA), J. R. Farah (CfA), G. 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. Wilner (CfA), S. P. Willner (CfA), S. J. Wolk (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 galaxy evolution.

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 galaxy evolution. 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 Galaxy Evolution

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 recent 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*), such as exploring the evolution of galaxies, which was the main driver of the COSMOS survey. 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 *Unanticipated discoveries from improved observational capabilities*

Below we give four examples of **unanticipated** important discoveries, stemming from the opening of new observational windows due to the availability of new capabilities. The first three are well-known discoveries, which have led to the accepted scenario of joint galaxy-supermassive black hole evolution. The fourth example is a recent serendipitous discovery that may change the way we think about AGNs and their interaction with the cold ISM.

Black Holes became an observational reality with mass measurements of accreting BH X-ray binaries (first, Cyg X-1, Bolton 1972). The discovery of quasars and AGNs led to the successful model of accretion onto massive BHs in galaxy nuclei (Pringle et al 1973). With HST dynamical measurements, the widespread existence of supermassive BH became a generally accepted fact (Magorrian et al 1998), leading to our present understanding of galaxy-BH co-evolution.

The hot gas of clusters and galaxies is now a key 'observational' property of simulations of galaxy formation and evolution. It traces the universe's dark matter concentrations and is responsive to stellar and AGN feedback. However, until ~50 yrs ago nobody had thought of its existence. Observations with the first X-ray astronomy satellite, *Uhuru* (Giacconi et al. 1971) led to the discovery of the hot intra-cluster medium (Kellogg et al. 1971; Gursky et al. 1971). An imaging X-ray telescope, the *Einstein Observatory* (Giacconi et al 1979), was needed to discover hot halos in elliptical galaxies (previously believed to be devoid of ISM), and hot outflows from active star-forming galaxies and mergers (see review, Fabbiano 1989). With the spectral imaging capability of *Chandra* + ACIS, the interaction of AGNs with these hot halos (radio feedback; e.g. Paggi et al 2014) and the interaction of these hot halos with the intra-cluster medium are being mapped, providing a detailed picture of the dynamical universe.

Extremely intense star-formation is a key stage of galaxy evolution. This field of studies followed the discovery of extreme objects (ultraluminous infrared galaxies, ULIRGS) with the infrared mission IRAS (Soifer et al 1984). Their preponderance at high redshift was demonstrated by deep HST observations (Madau et al 1998). ALMA has uncovered a population of high-z, high-mass, dusty star-forming galaxies, with star-formation rates > 1000 Msol/yr (Blain et al 2002). These results are the foundation of our present understanding of galaxy merging evolution.

The extended hard continuum (> 3 keV) and Fe Kα emission of Compton Thick AGNs. This is a very recent, surprising discovery of X-ray sub-arcsecond spectral imaging, only possible with *Chandra* (Fabbiano et al 2017, Fig. 1). These extended components question the accepted model of simply torus-shrouded active nuclei and open a new avenue for exploring the observational intricacies of the AGN-galaxy interaction.

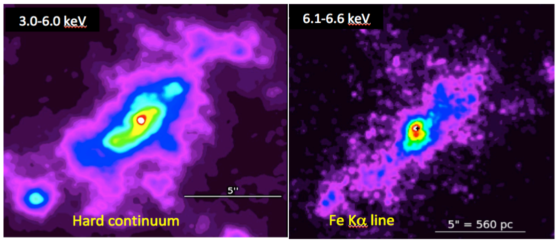

Figure 1. – ESO 428-G014 - > 2 kpc-scale hard continuum and ~1 kpc Fe K line emission (Fabbiano et al 2017)

2.2 *New understanding from archival studies*

Several astronomical databases and data sets are freely available from astronomy archives and their community use is vigorous. Frequently these data sets are mined in combination with other databases or complemented with new observations. For example:

- The discovery of a new class of superluminous spiral galaxies, as optically luminous as first-ranked ellipticals in galaxy clusters, was made by mining multiwavelength data synthesized within the NASA/IPAC Extragalactic Database (NED) (Ogle et al. 2016).
- Workflows based on IVOA[1] standards and tools led Chilingarian et al. (2009; 2015) to expand to over 200 the sample of compact elliptical (cE) galaxies, of which only 6 examples were known, using archival HST images of nearby galaxy clusters and additional information from archives and databases, and mining SDSS and GALEX survey data. These works suggest formation by tidal stripping and cD ejection from host clusters and groups by three-body encounters.

A few illustrative recent examples based on the COSMOS multi-wavelength survey have probed the evolution of galaxies and nuclear activity out to large redshift:

- A connection between AGN activity and merging out to z~3.5: the fraction of Compton Thick (CT, $N_H > 10^{24}$ erg s^{-1}) AGNs in mergers/interacting systems increases with luminosity and redshift (Lanzuisi et al. 2018).
- Mature quenched bulges, discovered in star-forming galaxies at z~2, by mapping COSMOS galaxies with HST and VLT/SINFONI (Tacchella et al. 2015).
- A massive, dusty starburst in a galaxy protocluster at z = 5.7, serendipitously discovered in the COSMOS Field (complemented by ALMA and VLA), forming stars at a rate of at least 1500 $M_\odot$ yr^{-1} in a ~3 kpc compact region (Pavesi et al. 2018).
- Massive proto-clusters of galaxies at z~5.7 and z~4.6, discovered in the COSMOS field, using spectroscopic observations taken from Keck and the Visible Multi-Object Spectrograph (VIMOS) Ultra-Deep Survey (Capak et al. 2011; Lemaux et al. 2018).
- Over-massive BH (~10^9Msol) discovered in a 10^{10}Msol star forming galaxy at z~3.3 (COSMOS + Keck), providing an example where BH growth may not be symbiotic with galaxy growth (Trakhtenbrot et al. 2015).

---

[1] International Virtual Observatory Alliance; the forum for the development of the interoperability standards used by major astronomy datacenters (http://www.ivoa.net)

- The nature and luminosity function of galaxies with z~7-9 were explored using the COSMOS/UltraVISTA database complemented with HST imaging and Spitzer (Stefanon et al. 2017; Bowler et al. 2017).

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 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)[2] 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*[3] 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 environment, cloud computing). But, resources must be made available for full development,

---

[2] The forum for the development of the interoperability standards used by major astronomy datacenters (http://www.ivoa.net)

[3] http://www.astropy.org/acknowledging.html

which will demand remote Science Platforms[4] and Server-side analytics[5], 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. 2).

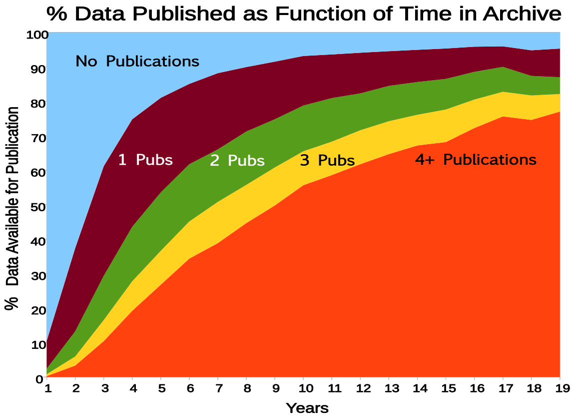

Figure 2. – Percentage of *Chandra* exposure time published versus years in the archive. The scientific use of archival *Chandra* data is increasing with time in the public archive. For example, 19 years from launch, ~75% of the observation have been published in more than 4 papers. A similar trend is observed for the HST data.

4. Conclusions

We propose <u>*exploration*</u> 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.***

---

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

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

References

Bolton, C. T. 1972, NPhS, 240, 124.

Bowler, R. A. A., et al. 2017, MNRAS, 466, 3612.

Capak, P. L., et al. 2011, Nature, 470, 233.

Chilingarian, I. et al. 2009 Science 326 1379.

Chilingarian, I., Zolotukhin, I. 2015, Science, 348, 418.

Cooke, K. C., et al. 2018, ApJ, 857, 122.

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

Fabbiano, G. et al 2017, ApJ, 842L, 4.

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

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

Lanzuisi, G. et al 2018, MNRAS, 480, 2578.

Lemaux, B. C., et al. 2018, A&A, 615A, 77.

Madau, P. et al. 1998, ApJ, 498, 106.

Magorrian, J., et al. 1998, AJ, 115, 2285.

Paggi, A., et al. 2014, ApJ, 787, 134.

Pavesi, R., et al. 2018, ApJ, 861, 43.

Pringle, J. E., et al. 1973 A&A, 29, 179.

Stefanon, M., et al. 2017, ApJ, 851, 43.

Soifer, B. T., et al. 1984, ApJ, 278L, 71.

Tacchella, S. et al. 2015, Science, 348, 314.

Trakhtenbrot, B. et al. 2015, Science, 349, 168.