**Astro2020 Science White Paper**

**Increasing the Discovery Space in Astrophysics**

**The Exploration Question for Cosmology**

**Thematic Area:** Cosmology and Fundamental Physics

**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), A. Avelino (CfA, Harvard), 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), E. Lusso (Univ. of Florence, Italy; CfA collaborator), W. P. Maksym (CfA), R. Martinez Galarza (CfA), J. Mazzarella (Caltech/IPAC), M. Ntampaka (CfA), G. Risaliti (Univ. of Florence, Italy; CfA collaborator), 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 cosmology.

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 cosmology. 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 Cosmology**

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 constraining the cosmological parameters. 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*

Cosmology is perhaps the best example of a field that was stimulated by observational discoveries, interplaying with theoretical models, beginning with the Hubble diagram and the Hubble-Lemaitre law. The current standard Lambda CDM (Cold Dark Matter) model of the observed Universe comes from the discovery of **dark matter** having several times the mass density of normal baryonic matter. The subsequent discovery of **dark energy** relegates baryonic matter to just below 5 % of the content of the Universe.

*Dark Matter -* The discovery of Dark Matter is based on serendipitous results of observational studies. The first – controversial - direct evidence was given by Fritz Zwicky in 1933 on the basis of the virial mass estimate in the Coma cluster while trying to refine the Hubble-Lemaitre relation. The idea did not get a broad acceptance at the time: many astronomers considered an alternative explanation to the mass problem by assuming that galaxy clusters are transient, non-virialized complexes. Convincing arguments for the need of the dark matter came from the studies of the rotation curves in spiral galaxies, but this study was originally undertaken to understand the dynamical properties and masses of galaxies throughout the Hubble sequence (Rubin et al. 1978). The gravitational attraction seen is clearly stronger than could be accounted for by the visible baryonic matter. There are expectations that this non-baryonic matter is in the form of some weakly interacting particle, but no direct detection of such particles has been achieved, and the arguments for detection of gamma-rays from their decay from astronomical sources have been questioned.

*Dark energy* was introduced by Albert Einstein in the form of a cosmological constant but estimates of its role based on the local and high redshift observations only started to appear in 1980s (Peebles 1988), to explain observational results: the growth rate of the large-scale structure seemed to require faster expansion of the Universe than predicted by the gravitational forces. *The convincing argument came from a totally independent line of study* – from Supernovae Ia (Riess et al. 1998, Perlmutter et al. 1999). SNIa are standardizable, and thus relatively easy to use for cosmology.

**2.2** *New Understanding from survey / archival studies: New cosmology probes*

In the last decade the values of the cosmological parameters in the standard Lambda CDM model have been determined precisely by combining the independent measurements of the Cosmic Microwave Background (CMB), SNIa, and Baryon Acoustic Oscillations (BAO). Combining the methods was essential, since there is considerable parameter degeneracy in any single model. This degeneracy is due to the limited range of redshifts in each method, the relatively large errors in individual measurements, and the limited number of measurements. Recently, tension is appearing between the newest measurements and the standard model. *These studies typically use and combine new extensive data bases, often obtained for different purposes.* Examples include:

* Riess et al. (2018), combining HST photometry and Gaia DR2 parallaxes of Cepheids to simultaneously constrain the cosmic distance scale and to measure the Gaia DR2 parallax zero point offset appropriate for Cepheids. They report a tension between the newest local measurements of the Hubble constant and those based on the CMB at close to the 4 sigma level.
* Risaliti & Lusso (2019), making use of a sample of 1,600 quasars from public archives (*XMM-Newton* and SDSS), complemented with new *XMM-Newton* observations, determined a deviation from the Lambda CDM model at the ~4 sigma level using the observed non-linear relation between AGN broad band UV/X-ray flux ratio and UV luminosity.
* Emami et al. (2019), studying the clustering amplitude of a complete sample of 7143 clusters in the Sloan survey, report that the observed correlation length exceeds pure CDM simulation prediction by ~6% for the standard Plank-based parameters. This excess may be explained by free streaming of light neutrinos.

These discrepancies suggest the possibility of new physics beyond the standard ΛCDM cosmology. The combination of different, sometimes new, approaches, supported by an increased data quality and sample statistics, is the way forward to solve the dark matter and dark energy problems. In the future, we expect cosmological results from a range of astronomical measurements, making use of publicly released surveys and using tailored analysis workflows, including weak lensing (e.g., Mandelbaum 2018), strong gravitational lensing events (e.g., Cao et al. 2015), Supernovae II (de Jaeger 2015), Active Galactic Nuclei reverberation mapping (AGN; Cackett et al. 2007, Watson et al. 2011, Haas et al. 2011), galaxy clusters through combined X-ray and Sunyaev-Zeldovich effect (e.g., Bonamente et al. 2006), gamma-ray bursts (Schaefer 2007), and gravitational wave events (Abbott et al. 2017). While the methods to use these probes are still in development, some
promising results are coming, paving the way for further advances in the 2020s

**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)[[1]](#footnote-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]](#footnote-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 environment, cloud computing). But, resources must be made available for full development, which will demand remote Science Platforms[[3]](#footnote-3) and Server-side analytics[[4]](#footnote-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 forflexible 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.***

**References**

Abbott, B.P. et al., 2017, Physical Review Letters, 118, 22110.

Bonamente, M. et al., 2006, ApJ, 647, 25.

Bowler, R. A. A., et al. 2017, MNRAS, 466, 361.

Cackett, E.M. et al. 2007, MNRAS, 380, 669.

Cao, S. et al. 2015, ApJ, 806, 185.

De Jaeger, T. et al., 2015, ApJ, 815, 121.

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.

Emami, R. et al. 2019, arXic:1711.05210.

Haas, M. et al., 2011, A&A, 535, A73.

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

Madelbaum, R. 2018, ARA&A, 56, 393.

 Riess, A. G. et al. 2018, ApJ, 861, 126.

Risaliti, G., Lusso, E., 2019, Nat.Astr. Tmp. 195.

Rubin, V. C. et al. 1978, apJ, 225, L107.

Schaefer, B.E. 2007, ApJ, 660, 16.

Watson, D. et al., ApJ, 2011, 740, L49.

Zwicky, F. 1933 (English translation by H. Andernach in arXiv:1711.01693).

1. The forum for the development of the interoperability standards used by major astronomy datacenters (<http://www.ivoa.net>) [↑](#footnote-ref-1)
2. <http://www.astropy.org/acknowledging.html> [↑](#footnote-ref-2)
3. See LSST Science Platform Design document <https://ldm-542.lsst.io> [↑](#footnote-ref-3)
4. NASA Big Data Task Force (<https://science.nasa.gov/science-committee/subcommittees/big-data-task-force>) [↑](#footnote-ref-4)