A Science Platform Network to Facilitate Astrophysics in the 2020s

Vandana Desai, Mark Allen, Christophe Arviset, Bruce Berriman, Ranga-Ram Chary, David Cook, Andreas Faisst, Gregory Dubois-Felsmann, Steve Groom, Leanne Guy, et al.

A Science Platform Network to Facilitate Astrophysics in the 2020s

Type of Activity: Technological Development

Vandana Desai	Caltech/IPAC	desai@ipac.caltech.edu
Mark Allen	CDS	mark.allen@astro.unistra.fr
Christophe Arviset	ESA-ESAC	Christophe.Arviset@esa.int
Bruce Berriman	Caltech/IPAC	gbb@ipac.caltech.edu
Ranga-Ram Chary	Caltech/IPAC	rchary@caltech.edu
David Cook	Caltech/IPAC	dcook@ipac.caltech.edu
Andreas Faisst	Caltech/IPAC	afaisst@ipac.caltech.edu
Gregory Dubois-Felsmann	Caltech/IPAC	gpdf@ipac.caltech.edu
Steve Groom	Caltech/IPAC	sgroom@ipac.caltech.edu
Leanne Guy	AURA/LSST	leanne.guy@lsst.org
George Helou	Caltech/IPAC	gxh@ipac.caltech.edu
David Imel	Caltech/IPAC	imel@ipac.caltech.edu
Stephanie Juneau	NOAO	juneau@noao.edu
Mark Lacy	NRAO	mlacy@nrao.edu
Gerard Lemson	Johns Hopkins University	glemson1@jhu.edu
Brian Major	National Research Council Canada	brian.major@nrc-cnrc.gc.ca
Joe Mazzarella	Caltech/IPAC	mazz@ipac.caltech.edu
Thomas Mcglynn	NASA/GSFC	thomas.a.mcglynn@nasa.gov
Ivelina Momcheva	STScI	imomcheva@stsci.edu
Eric Murphy	NRAO	emurphy@nrao.edu
Knut Olsen	NOAO	kolsen@noao
Josh Peek	STScI	jegpeek@stsci.edu
Alexandra Pope	UMass, Amherst	pope@astro.umass.edu
David Shupe	Caltech/IPAC	shupe@ipac.caltech.edu
Alan Smale	NASA/GSFC	alan.p.smale@nasa.gov
Arfon Smith	STScI	arfon@stsci.edu
Nathaniel Stickley	Caltech/IPAC	nrstickley@ipac.caltech.edu
Harry Teplitz	Caltech/IPAC	hit@ipac.caltech.edu
Ani Thakar	Johns Hopkins University	thakar@jhu.edu
Xiuqin Wu	Caltech/IPAC	xiuqin@ipac.caltech.edu

Endorsements

Chuanfei Dong	Princeton University	dcfy@princeton.edu
Susan Mullally	STScI	smullally@stsci.edu
Reed Riddle	Caltech	riddle@caltech.edu
David Rapetti	CU Boulder/NASA Ames	David.Rapetti@colorado.edu

Abstract

Astronomical facilities will produce petabytes of observational data in the 2020s. Simulated data sets created to plan and interpret the data from these missions will match or exceed these volumes. Mining such new petabyte-scale data sets to meet planned science goals and to explore discovery space will require astronomers to adopt new approaches and to develop new tools. Increasingly complex search criteria, necessary for identifying objects of interest within billion-row catalogs, will strain query response times. Modern statistical methods will result in data-reduction methods that actually increase data volumes. Visualization techniques that have worked well for decades will be inadequate in this regime. The current network infrastructure will be inadequate for downloading the vast quantities of multiwavelength observational and simulated data that should be jointly analyzed. Analysis tools will need to be augmented with scalable machine learning algorithms and data analytics. To meet these challenges, astronomers will require access to large volumes of high-performance storage and high-throughput computational resources, as well as the training to use them. **In this white paper, we advocate for the adequate funding of data centers to develop and operate "science platforms", which will provide storage and computing resources for the astronomical community to run analyses near the data. Furthermore, these platforms should be connected to enable cross-center analysis and processing.** Providing such resources will build on unrestricted data access to realize properly resourced data analysis, thus allowing scientists to explore and implement their research ideas regardless of their own institutional facilities.

1. Key Science Goals & Objectives

The 2020s will see large increases in data volumes from observational facilities and from simulations (**Figure 1**). In the optical and infrared, LSST, Euclid, and WFIRST will generate hundreds of petabytes of data. In the radio, surveys from the VLA, ASKAP and MeerKAT, and, near the end of the decade, SKA and ngVLA will also total hundreds of petabytes. Simulations supporting the planning and interpretation of these observational projects can meet or exceed these data volumes. These data sets have immense science potential, which will only increase when used together.

"Big data" resources are central to the astronomical science goals of the 2020s. This importance is reflected in the large fraction of submitted science white papers that reference these data sets (LSST 158; Euclid 68; WFIRST 144; VLA 71; ASKAP 13; MeerKAT 14; SKA 84; ngVLA 85; total ~600). While the science opportunities described in these white papers are diverse, the inherent (often unacknowledged) data challenges in realizing these opportunities are common "big data" problems. These include fundamental activities such as discovering, querying, visualizing, downloading, storing, reprocessing, analyzing, federating, and sharing large data sets across multiple archives. (Section 2.) Below are a few examples of science white papers that illustrate the opportunities and challenges of astronomical big data in the 2020s and beyond.

Paladini et al. (2019) describe how data from Gaia, WFIRST, LSST, Euclid, and CASTOR can be combined to accurately measure the initial mass function of stars and the related mass function of molecular cores across a range of environments and distances. These are fundamental measurements which impact our understanding of stellar evolution, galaxy evolution, and planet formation. Making these measurements will require a multi-wavelength, multi-center analysis of imaging data across large areas of the sky. High-level catalogs produced by the projects are unlikely to meet the science requirements in crowded fields with variable backgrounds, especially for the extended cores. Analysis on such a large volume of pixels will present significant data challenges for many users.

Kirkpatrick et al. (2019) describe the science potential of an all-sky infrared version of Gaia to determine the low-mass cutoff of star formation. Such a survey would produce an enormous data set with matching legacy value and data challenges. The associated potential and challenges are increased when this survey is used in combination with other large-area surveys, such as Gaia, 2MASS, WISE, ZTF, PanSTARRS, etc. Already, giant single catalogs like those from Gaia and WISE strain most users' abilities to fully visualize and analyze.

Wrobel et al. (2019) argue that observing Intermediate-mass black holes in globular clusters could shed light on the early formation of seed black holes and inform predictions of gravitational wave events. To achieve these science goals, they advocate for a facility like the ngVLA. Nyland et al. (191) also describe how the ngVLA will support studies of AGN feedback. The massive datasets from current and future radio interferometers will produce individual datasets with sizes ~0.1-1 PB. This volume makes it difficult for users to tune their image products to meet their science needs, given that traditional mission computing models cannot produce the entire range of possible processing choices. Analysis of the resulting image products will also be challenging. Already, image cubes from ALMA can reach ~1 TB in size, impossible for most users to view using existing tools and facilities.

Wang et al. (2019) describe how systematics in measurements of Dark Energy can be minimized with a new galaxy redshift survey (ATLAS) producing ~200 million galaxy spectra, two orders of magnitude larger than the SDSS Legacy Survey spectroscopic sample. Such a survey would have enormous legacy potential, which would best be realized by supporting the community in meeting the associated big data challenges.

Chary et al. (2019) argue that the loose tension between different measurements of some fundamental cosmological parameters could be resolved by processing data from Euclid, LSST, and WFIRST together *at the pixel level*. Pixel-level projects run into "big data" challenges very quickly, even when full data sets are not being analyzed.

Chang et al. (2019) describe how multi-messenger astronomy in the 2020s will require the sharing of data, code, modeling tools, and facility-specific expertise in near-real time

to localize the electromagnetic counterparts to gravitational wave (GW) events. Currently, several collaborations and data centers (e.g., the NED-GWF service at IPAC) combine data across the electromagnetic spectrum to facilitate prompt GW localization campaigns. However, in the 2020s, it will be essential to extend their capabilities in a collaborative analysis environment and prepare for the third generation GW and neutrino observatories, which will produce tens of events per hour (Reitze 2019). This will require developing the cyberinfrastructure needed to combine several large-area follow-up surveys (i.e., LSST and ZTF) with real-time alerts (LIGO/Virgo, IceCube, and LISA) and analysis software tools.

The white papers above provide concrete examples of how large data sets will be vital to make progress in specific science areas spanning astrophysics. Moreover, in an additional series of 6 science white papers, Fabbiano et al. (2019) emphasize that many paradigm-shifting discoveries in the 2020s will not be made through well-formulated hypotheses based on knowledge of the time, but rather by an exploratory discovery approach enabled by new telescopes and instrumentation, as well as by high-quality data products in easily accessible and interoperable science archives.

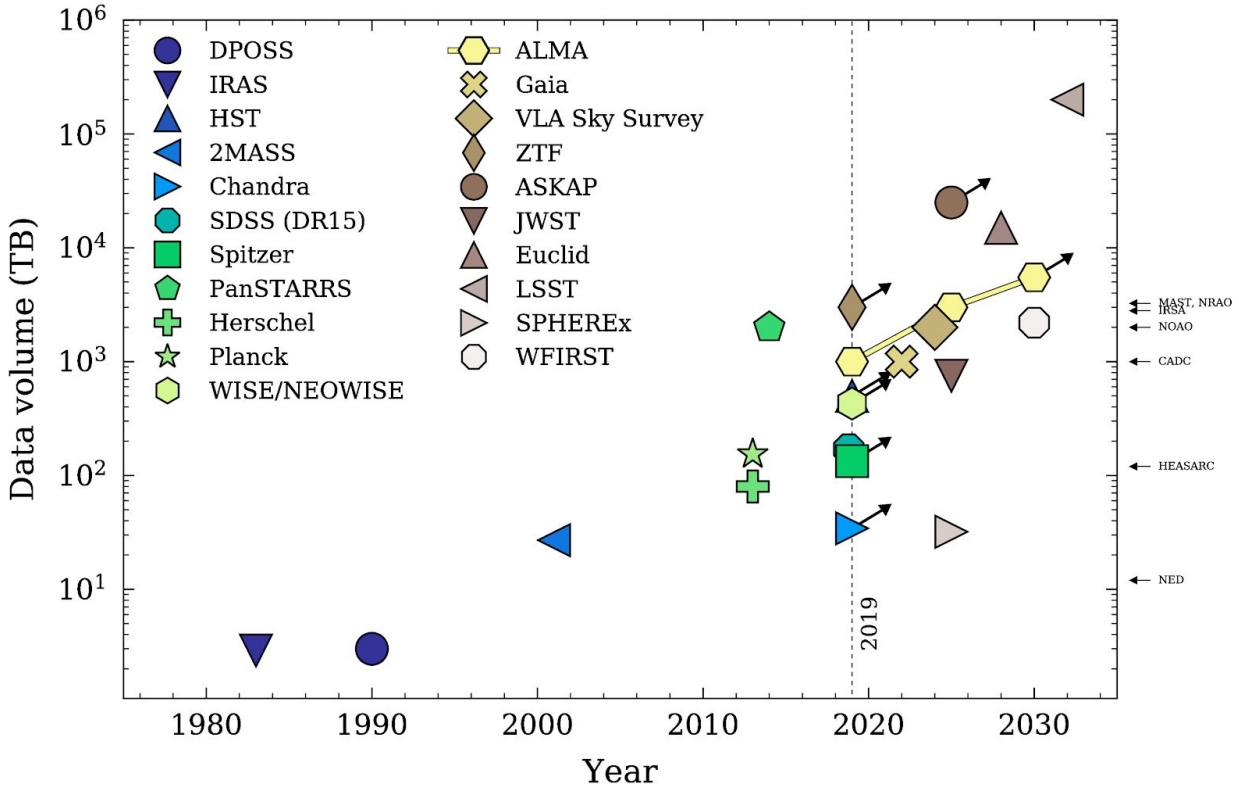

Figure 1. The 2020s and beyond will see large increases in data volumes. Approximate expected data volumes in terabytes of selected astronomical observational facilities and surveys are shown as a function of time. Symbols are plotted at the (expected) end of operations. Ongoing surveys as of this writing are plotted in 2019 with an arrow. The current size of major data centers are shown on the right axis.

2. Big Data Challenges in the 2020s

One of the ways that “big data” enable science is by providing statistically significant samples. Large samples underpin future advances in nearly all fields of astrophysics, including studies of exoplanet systems, stellar populations in the Milky Way, galaxy evolution, dark matter, and dark energy. In the current archive model, scientists can search for data satisfying a number of constraints and visualize the results using a web application, but must ultimately download the data for analysis. As samples grow, each of these steps becomes more challenging. The obstacles are even larger when science goals require a combination of large data sets housed at different data centers. Indeed, a survey conducted by the NASA/IPAC Infrared Science Archive in 2018 reveals that astronomers are already struggling with “big data” problems (see Table 1).

The following is a list of major challenges that will need to be met in order to realize the full science potential of astrophysics in the 2020s.

- **Searching through large data sets:** To identify populations within very large data sets, users need to set a number of search constraints. Traditionally, the most common constraints entered by users have been spatial coordinates or object names. In the era of large and all-sky photometric and spectroscopic surveys, these constraints are being swapped out or augmented by more complex constraints on color, flux, redshift, time, and other observables. During the 2020s, maintaining reasonable response times even for traditional queries -- as the volumes of archives continue to explode -- will be a major challenge.
- **Visualizing and interpreting large data sets:** Data visualization is an important step in conceiving ideas, exploring new data sets and parameter spaces and connections between parameters, checking analysis, understanding and summarizing results, and communicating conclusions. Because visualization is so integral to science, most astronomical archives provide some visualization facilities (e.g., FITS viewing, XY charts, catalog and footprint overlays). As data sets become larger, more complex, and highly multi-dimensional, new methods of visualization are needed in order to understand them. Multi-resolution projection techniques to visualize networks of image surveys that cover large sky area offer a good starting point for next generation visualisation (e.g. Fernique et al. 2015, McGlynn et al. 2019). Techniques such as dimensionality reduction using unsupervised machine learning (e.g., self organizing maps[1] or t-Distributed Stochastic Neighbor Embedding [2]) are becoming more important as they provide tools to visualize correlation in highly multi-dimensional data sets at a small cost of CPU/GPU time. Practical examples include the inspection of billions of images (e.g., of galaxies) or classification of light curves using machine learning, which would be an impractical task for a visual inspection by humans.

[1] Self organizing maps (SOM), https://en.wikipedia.org/wiki/Self-organizing_map
[2] t-Distributed Stochastic Neighbor Embedding (t-SNE), https://lvdmaaten.github.io/tsne/

- **Downloading and storing large volumes of data**: Currently, once a data set is identified, users typically download it to an individual computer or a departmental cluster to proceed with scientific analysis. However, as data sets continue to grow at an unprecedented rate, and scientists tackle problems that require larger and larger data sets to solve, this download step is becoming more challenging and in some cases completely impractical. First, the time it takes to download data through the network will become prohibitively long, even with improvements to the networking infrastructure. Second, downloading even a small percentage of a petabyte-scale data set requires a large amount of storage, both for the downloaded data and for intermediate and user-generated data products that may also be at petabyte scales. While raw storage is becoming increasingly affordable, it is not an out-of-the box solution for most astronomers because of the difficulty of making storage perform well enough to support efficient high-performance computation. For these reasons, *the bulk of the premier big data sets of the 2020s will be stuck at the data centers.*
- **Analysis:** The availability of "big data" in the 2020s will cause a change in the nature of astronomical analyses.
    - **Astroinformatics & Astrostatistics:** Siemiginowska et al. (2019) describe how advances in statistics, computer science, and machine learning are becoming standard in the workflow of astronomical research. The growing complexity of analyses coupled with the growing sizes of data sets will result in a growing need for computational power, at a level that will be out of reach for many astronomers.
    - **Multi-archive interoperability:** Joint analyses of data sets that are hosted at different locations can become challenging when large numbers of pixels need to be processed together, when huge catalogs need to be cross-matched, and when constraints must be applied across multiple large catalogs. Since these represent a large number of the desired use cases, data centers need to be interoperable in finite, carefully selected, and well-defined ways. In the future, some data sets may be distributed across data centers (e.g., the SKA regional centre model), which stresses the need for interoperability.
    - **Reproducibility & Replicability:** It is essential to science that results can be verified by independent groups. Verification in the 2020s will be a challenge because few groups will have the resources and data skills to reproduce (run the same software on the same data to get the same result) or replicate (write software based on analysis descriptions to get similar results) complex analyses on huge data sets spread over multiple data centers.
- **Accessing adequate computing facilities**: The computational aspects of big data analytics can be challenging to astronomers in several ways. First, basic computational techniques that have been in use throughout the 2010s will be too inefficient to deal with the larger data sets of the 2020s. Analysts of the 2020s will need to learn to optimize code for speed and parallelism. Second, astronomers will need access to the considerable computing resources that are required to support petabyte-scale analyses. Compute resources at adequate scale will be located at sites remote from the data host. For this reason, the use of services such as those at the Texas Advanced Computing Center (TACC) and those offered by commercial

cloud providers such as Amazon Web Services (AWS) will accelerate in the 2020s. In particular, fast provisioning and unlimited power to scale will make clouds especially appealing. Astronomers are already taking advantage of this (e..g Toomey et al 2017). Paying for cloud resources and moving petabyte-scale data from the data centers to a cloud resource will remain challenging.

- **Collaboration:** Research has long been a highly collaborative effort. This is now being reflected in the rise of collaborative software tools. From calendars to messaging systems to document writing, scientists expect their tools to support collaboration. New technologies like Jupyter notebooks are currently opening up powerful new opportunities for collaboration during data analysis. The importance of collaboration will increase as data analysis becomes more complex to support the science of the next decade. The challenge will be to support the need for collaboration within the "big data" computing environment of the 2020s.

Table 1: Survey Results: "Which of the following problems have you encountered when dealing with astronomical data sets?"

"Big Data" Challenge	**Respondents**
I struggle to find the data sets I need.	39%
The data set I want takes too long to download.	35%
I struggle to obtain enough storage space to manage the data I want.	30%
I struggle to obtain enough computing power to analyze my data.	22%
I struggle to scale my analysis routines to large volumes of data.	27%
I struggle to visualize data.	30%

3. Vision for the Future: Science Platform Network for Astronomy

We advocate for the development of a "science platform network" to address the data challenges of reaching the science goals of the 2020s and beyond. Science Platforms give the user control over all aspects of scientific work, as needed, including data discovery, data access and data exploration, data and model integration, deploying new analysis tools and integrating them with existing tools, running production analysis and modeling, and collaboration and sharing results. This evolution of community data services, in combination with workforce training (Norman et al. 2019, Besla et al. 2019) are necessary to meet the science goals enumerated in the submitted science white papers, as well as science goals that have not yet been imagined. Science platforms increase the accessibility of both data and analysis resources, supporting new types of scientific analysis demanded by petabyte-scale data sets, as well as the inclusion of astronomers at all institutions (Peek et al. 2019). A science platform has three essential aspects: an analysis environment close to the data, scripted data access, and web portals. It is also essential that users can share results between these aspects.

1. **Interactive and batch analysis environment co-located with data.** This aspect represents the largest change to the current data center model. Science platforms offer large-scale storage and computing resources at the data center itself, eliminating the need to download large volumes of data and to procure storage and

computing resources for petabyte-scale analyses. Computing resources are provided as a secure, interactive, and collaborative environment, allowing users to develop and share analysis tools and products. Analysis tools could include the command line, common community-created analysis packages (e.g. CASA, Tractor, ds9), Jupyter notebooks, and whatever tools the user chooses to upload. Once analysis is complete, users can download the summary results. New software tools developed by the community can be integrated into science platforms, but this does require training to ensure that new tools are designed and developed with modern software engineering processes (See the white paper "Elevating the Role of Software as a Product of the Research Enterprise" by Smith et al. 2019). We anticipate that providing these resources will likely rely heavily on adopting cloud technologies (see below).

2. **Application Program Interfaces (APIs).** APIs allow scripted access to data. Currently, most data centers provide web APIs that can be used to search for and download data from the data center to a user's local computer. APIs will continue to be necessary in the science platform context, and should follow (when practical) the standards established by the International Virtual Observatory Alliance (IVOA).
3. **Web Portals.** Web portals are the traditional access points for astronomical data. Most started with simple search and download capabilities, but have evolved to accommodate more sophisticated search criteria, to allow previews of data products, to provide visualizations of catalogs and survey footprints, and to offer other data exploration capabilities. In the 2020s, these web portals will be providing more sophisticated analysis tools designed to run at the data center, in addition to the above described interactive and batch analysis environments.

The development of science platforms will absorb many of the challenges described in Section 2. This transition is already underway; existing science platforms are listed in Table 2. However, joint analyses of data sets that are held across data centers will be far easier if these individual science platforms are *interoperable*. We therefore advocate for a science platform **network**, by which we mean that (1) astronomers will have access to a network of loosely coordinated science platforms that together provide access to a wide array of astronomical data; (2) these science platforms will share common elements (e.g. data models and APIs) so that astronomers do not encounter vastly different user experiences when accessing different data sets; and (3) moderate volumes of data can be transferred between science platforms to facilitate multi-wavelength science.

While we anticipate that a large fraction of science analysis will be performed within and across platforms, the most effective implementation of a platform network would support scalable access to standardized tools and APIs that would allow researchers to use the same approaches -- and often the same scripts and codes -- on all scales from personal laptops to global analysis. Scientists can test analyses on locally available subsets and then easily transition analyzing complete holdings on platforms. By ensuring that the analysis capabilities in and across platforms mirror those at smaller scales -- with

standardized software and interfaces adapting to the changing environment -- we can greatly ease the adoption of platform resources across the community.

To optimize the science benefits of a science platform network for astronomy, data centers will need to make progress in the following areas:

Collaborative Workspaces. To enable analysis environments co-located with the data, science teams will require workspaces that consist of secure, permanent storage volumes for results that need to persist, and much larger temporary storage volumes for scratch space. These workspaces will be most effective for science if they support collaboration. The IVOA VOSpace concept and WebDAV are two options for supporting collaboration that have been adopted by existing or developing science platforms. The challenge for data centers will be to manage the safety, security, and fair distribution of workspace resources in the face of increasing and variable demand. In the future, these workspaces will most likely be inside commercial cloud services.

Cloud Computing. Individual scientific investigations often require large amounts of computing for limited times. In addition, data centers currently see spikes in the number of queries and download volumes in the period immediately following a large data release. In the science platform context, these spikes will be accompanied by spikes in the demand for computational resources. Because it is not cost effective for data centers to provision for peak demand, they will need to consider augmenting in-house computing resources with dynamically-scalable commercial cloud computing resources, perhaps using a cloud-bursting model. While we anticipate that the cost of commercial cloud services will decrease over the years, these costs will need to be negotiated to fit within the financial model of the data centers. Regardless, there will be research teams that require computational resources that are beyond what can be provided through data centers. Science platforms will therefore be most effective if they can be containerized and deployed into the commercial cloud by individuals as needed.

Containers. Container technologies (such as Docker) encapsulate dependencies between components of a software system, allowing users to run an application reliably in multiple computing environments. Containers reduce the cost to deploy, operate, and maintain instances of science platforms. They also enable the exchange of user-written software between science platforms. In wide commercial use, they are an essential technology for the development of science platforms.

Networking and Data Transfer Tools. Although science platforms will absorb much of the burden of moving petabyte scale data, it will still be convenient for users to be able to move moderate amounts of data to their desktops or between science platforms. Therefore, data centers should ensure that they have state-of-the-art networking infrastructure and data transfer tools, which we anticipate will provide significant performance increases over the decade.

Standardized data models and web APIs. An astronomical data model is a standard way of describing a data set. This standardization allows data centers to easily provide search and retrieval options for data sets physically held at other data centers. One example is the Common Archive Observation Model (CAOM) that was developed by the Canadian Astronomy Data Centre (CADC) and is being taken up by a growing number of data centers. An astronomical web API provides access to a data center's holdings over the web using HTTP protocol. Many astronomical data centers have implemented web APIs that follow the IVOA's Table Access Protocol (TAP) for searching catalogs and tables of image metadata. CAOM, along with standard APIs, enables CADC's Multi-Archive Query option[3], which offers users access to data hosted at multiple data centers, in addition to the data held at CADC.

Table 2. Examples of Astronomical Science Platforms in Existence or Development

CANFAR: Canadian Advanced Network for Astronomical Research canfar.net CFHT, JCMT, HST, queryable observation metadata from multiple archives in the Common Archive Observation Model (CAOM)
ESA Science Exploitation and Preservation Platform (In Development) Initially Planck, Gaia, as well as GNSS Science Support Centre. Potentially all ESA space science (astronomy, planetary and heliophysics) datasets
HEASARC: High Energy Astrophysics Science Archive Research Center Science Platform (in Development) heasarc.gsfc.nasa.gov/ Fermi, Swift, XMM-Newton, INTEGRAL, Suzaku, NuSTAR, NICER, etc.
IPAC Science Platform (in Development) ipac.caltech.edu IRSA (Spitzer, WISE/NEOWISE, SOFIA, 2MASS, IRAS, etc.), NED (thematic, extragalactic object database), NASA Exoplanet Archive (thematic, exoplanet data)
LSST: Large Synoptic Survey Telescope Science Platform (in Development) https://ldm-542.lsst.io/
NOAO Data Lab datalab.noao.edu Survey datasets from NOAO 4-m telescopes (e.g. DES, Legacy Survey, NOAO Source Catalog, in future DESI); high-value datasets from other facilities (e.g. AllWISE, unWISE, Gaia, SDSS)
NRAO Archive Access Tool/Science Ready Data Products Initiative archive-new.nrao.edu VLA, ALMA
SciServer https://www.sciserver.org SDSS databases and DR7 images and spectra. Specific data sets for collaborative

[3] http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/search/

groups, e.g. mirror Millennium simulation database and raw data; Kepler light curves; Non-astronomical data sets.
STScI Science Platform (in Development) https://mast-labs.stsci.io/2019/02/zero-to-jupyterhub-with-ansible HST, TESS

4. Recommendations

Key Data Centers in the U.S. ground- and space-based astronomy community should be funded to deploy, maintain and extend science platforms by the end of the 2020s. The funding includes (a) development of software components, (b) investment in computing and storage resources for user computing, (c) operating expenses for the resulting systems and (d) resources for coordination.

Platform facilities should be specifically funded to work together to ensure that platforms are built to maximize collaborative capabilities. Since the required underpinning technologies are fast evolving and dependent on broader industry standards, we propose implementers design functional and interoperable systems, and then iterate within the IVOA structure to find workable long-term standards, as has worked with protocols like CAOM. While it is premature to propose a specific collaboration structure, the NASA Astronomical Virtual Observatories group (NAVO; https://heasarc.gsfc.nasa.gov/vo/summary) may be a useful model to build on.

Funding agencies should negotiate rates or credits with cloud infrastructure and computing service providers for U.S. researchers and for U.S. data centers building and supporting cloud-based archive and science platform services.

5. Schedule & Cost

Because of the diversity of data centers, in terms of current and future data holdings, adopted technologies, and funding streams, the cost and schedule of collaboratively implementing a science platform network is difficult to summarize without significant study. We suggest that this study be undertaken by the individual data centers as soon as possible. However, we can safely say that the schedule for development of the science platform network would ideally be driven by the schedule for the "big data" missions and projects of the 2020s, as described in Section 1. We suggest that a phased development is important to slowly acclimate the community to the change in community data infrastructure that will be needed. In the first phase, through FY23, development could focus on implementing individual science platforms, essentially continuing the trend seen in Table 2, while keeping in mind that the second phase, FY23 and beyond, will involve the implementation of the "network" capabilities described above.

References

- **Besla et al. (2019)**, “State of the Profession: Training the Future Generation of Computational Researchers” (APC White Paper)
- **Chang et al. (2019)**, Cyberinfrastructure Requirements to Enhance Multi-messenger Astrophysics: (Science White Paper)
- **Chary et al. (2019),** Cosmology in the 2020s Needs Precision and Accuracy: The Case for Euclid / LSST / WFIRST Joint Survey Processing (Science White Paper)
- **Fabbiano et al. (2019)**, Increasing the Discovery Space in Astrophysics - A Collation of Six Submitted White Papers (Science White Paper)
- **Fernique et al. (2015)**, A&A 578, 114. DOI:10.1051/0004-6361/201526075
- **Kirkpatrick et al (2019)**, The Need for Infrared Astrometry of Brown Dwarfs in the Post-Gaia Era (Science White Paper)
- **McGlynn et al. (2019)**, ApJS 240, 22. DOI:10.3847/1538-4365/aaf79e
- **Norman et al. (2019)**, “The Growing Importance of a Tech- Savvy Astronomy and Astrophysics Workforce” (APC White Paper)
- **Paladini et al. (2019)**, “On the Origin of the Initial Mass Function” (Science White Paper)
- **Peek et al. (2019)**, “Robust Archives Maximize Scientific Accessibility” (APC White Paper)
- **Reitze et al. (2019), “**The US Program in Ground-Based Gravitational Wave Science: Contribution from the LIGO Laboratory” (Science White Paper)
- Siemiginowska et al. (2019) “*The Next Decade of Astroinformatics and Astrostatistics*” (Science White Paper)
- **Toomey et al. (2017)**, CSIRO Astronomy and Space Science Report, EP172634
- **Wang et al. (2019)**, Illuminating the dark universe with a very high density galaxy survey (Science White Paper)
- **Wrobel et al. (2019)**, Intermediate-Mass Black Holes in Extragalactic Globular Clusters (Science White Paper)