

# Astro2020 Science White Paper

## Cyberinfrastructure Requirements to Enhance Multi-messenger Astrophysics

- Thematic Areas:**
- Planetary Systems
  - Star and Planet Formation
  - Formation and Evolution of Compact Objects
  - Cosmology and Fundamental Physics
  - Stars and Stellar Evolution
  - Resolved Stellar Populations and their Environments
  - Galaxy Evolution
  - Multi-Messenger Astronomy and Astrophysics

**Principal Author:**

Name: Philip Chang

Institution: University of Wisconsin - Milwaukee

Email: chang65@uwm.edu

**Co-authors:** Gabrielle Allen<sup>1</sup>, Warren Anderson<sup>2</sup>, Federica B. Bianco<sup>3,4</sup>, Joshua S. Bloom<sup>5,6</sup>, Patrick R. Brady<sup>2</sup>, Adam Brazier<sup>7</sup>, S. Bradley Cenko<sup>8</sup>, Sean M. Couch<sup>9</sup>, Tyce DeYoung<sup>9</sup>, Ewa Deelman<sup>10</sup>, Zachariah B Etienne<sup>11,12</sup>, Ryan J. Foley<sup>13</sup>, Derek B Fox<sup>14</sup>, V. Zach Golkhou<sup>15</sup>, Darren R Grant<sup>9</sup>, Chad Hanna<sup>14,16,17</sup>, Kelly Holley-Bockelmann<sup>18,19</sup>, D. Andrew Howell<sup>20,21</sup>, E. A. Huerta<sup>22,1</sup>, Margaret W.G. Johnson<sup>22,1</sup>, Mario Juric<sup>15</sup>, David L. Kaplan<sup>2</sup>, Daniel S. Katz<sup>1</sup>, Azadeh Keivani<sup>23</sup>, Wolfgang Kerzendorf<sup>24</sup>, Claudio Kopper<sup>9</sup>, Michael T. Lam<sup>11</sup>, Luis Lehner<sup>25</sup>, Zsuzsa Marka<sup>23</sup>, Szabolcs Marka<sup>23</sup>, Jarek Nabrzyski<sup>26</sup>, Gautham Narayan<sup>27</sup>, Brian W. O’Shea<sup>9</sup>, Donald Petravick<sup>22,1</sup>, Rob Quick<sup>28</sup>, Rachel A. Street<sup>20</sup>, Ignacio Taboada<sup>29</sup>, Frank Timmes<sup>30</sup>, Matthew J. Turk<sup>1</sup>, Amanda Weltman<sup>31</sup>, Zhao Zhang<sup>32</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign, <sup>2</sup>University of Wisconsin - Milwaukee, <sup>3</sup>University of Delaware, <sup>4</sup>Center for Urban Science and Progress, NYU, <sup>5</sup>University of California, Berkeley, <sup>6</sup>Lawrence Berkeley National Laboratory, <sup>7</sup>Cornell University, <sup>8</sup>NASA GSFC, <sup>9</sup>Michigan State University, <sup>10</sup>University of Southern California, <sup>11</sup>West Virginia University, <sup>12</sup>Center for Gravitational Waves and Cosmology, WVU, <sup>13</sup>University of California, Santa Cruz, <sup>14</sup>The Pennsylvania State University, <sup>15</sup>University of Washington, <sup>16</sup>Institute for Gravitation and the Cosmos, PSU, <sup>17</sup>Institute for CyberScience, PSU, <sup>18</sup>Vanderbilt University, <sup>19</sup>Fisk University, <sup>20</sup>Las Cumbres Observatory, <sup>21</sup>University of California, Santa Barbara, <sup>22</sup>NCSA, <sup>23</sup>Columbia University in the City of New York, <sup>24</sup>New York University, <sup>25</sup>Perimeter Institute for Theoretical Physics, <sup>26</sup>University of Notre Dame, <sup>27</sup>Space Telescope Science Institute, <sup>28</sup>Indiana University, <sup>29</sup>Georgia Institute of Technology, <sup>30</sup>Arizona State University, <sup>31</sup>University of Cape Town, <sup>32</sup>Texas Advanced Computing Center

## Abstract

The identification of the electromagnetic counterpart of the gravitational wave event, GW170817, and discovery of neutrinos and gamma-rays from TXS 0506+056 heralded the new era of multi-messenger astrophysics. As the number of multi-messenger events rapidly grow over the next decade, the cyberinfrastructure requirements to handle the increase in data rates, data volume, need for event follow up, and analysis across the different messengers will also explosively grow. The cyberinfrastructure requirements to enhance multi-messenger astrophysics will both be a major challenge and opportunity for astronomers, physicists, computer scientists and cyberinfrastructure specialists. Here we outline some of these requirements and argue for a distributed cyberinfrastructure institute for multi-messenger astrophysics to meet these challenges.

## Scientific Motivation

With the identification of the electromagnetic counterparts of both gravitational wave and high-energy neutrino events over the past two years, the new era of multi-messenger astrophysics (MMA) has begun. The observations of a binary neutron star merger in gravitational waves (GW/GRB 170817A; Abbott et al. 2017a), along with subsequent worldwide electromagnetic discoveries (Abbott et al. 2017b), and of neutrinos and gamma-rays from TXS 0506+056 (IceCube Collaboration et al. 2018) demonstrate both the promise and the challenges of MMA. These events will allow us to answer questions such as the origin of the heaviest elements, the equation of state of extremely dense matter in mergers of compact systems, the demographics of the time-domain sky, and the rate of the expansion of the Universe.

The next decade will see dramatic improvements in MMA capabilities. Upgrades of existing and deployment of new GW detectors such as LIGO (Aasi et al. 2015), VIRGO (Acernese et al. 2015), and KAGRA (Kagra Collaboration et al. 2019) will enable the identification of dozens of neutron-star merger events per year. IceCube-Gen2 (IceCube-Gen2 Collaboration et al. 2014; Aartsen et al. 2017) and other enhanced neutrino and cosmic ray observatories will continue to survey the sky for high-energy cosmic particles. NANOGrav (McLaughlin 2013) and other Pulsar Timing Arrays (PTAs) will probe the gravitational wave spectrum at nanohertz frequencies. The Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008) and the Zwicky Transient Facility (Graham et al. 2019) will survey the optical sky with unprecedented speed and depth. Upcoming radio facilities will collect terabytes of data with unprecedented resolution. This range of astronomical observatories will gather data on candidate multi-messenger sources, source populations, and host galaxies throughout the visible Universe. A comprehensive cyberinfrastructure for MMA will coordinate the efforts of these facilities, while providing valuable benefits for all areas of astrophysics that are part of their core scientific missions. The explosive growth in the scale and rates of data acquisition, the importance of event follow-up in MMA, and the challenge of analysis over different messengers drive the need for cyberinfrastructure upgrades that will require the sustained collective efforts of astrophysics facilities and a broad-ranging group of physicists, astronomers, computer scientists, data science specialists, and software/infrastructure engineers. **Cyberinfrastructure will need to be designed, built, and operated in order to collect, process, analyze, synthesize, and enable interpretation time-critical and archival data from observing facilities across the multi-messenger spectra.** Robust tools for data sharing, collaboration, joint analysis, as well as training and education will also need to be developed. Investments will enable sustainable and

efficient operations leading to more compelling discoveries, scientific outcomes, public releases of data and supporting analysis codes, cross-disciplinary research, cross-collaboration work, training and education of a diverse STEM workforce, and building a community-driven integrated software infrastructure.

## Cyberinfrastructure For Multi-Messenger Astrophysics

By its nature, MMA requires diverse scientific teams to bring together their observational resources, data, analysis and modeling tools, and expertise to ensure the maximum scientific return. With a diversity of different instruments, systematic effects, data formats, analysis protocols, collaboration styles, proprietary and other restrictions, and scientific cultures come a myriad of challenges. Working together effectively in this heterogeneous and international scientific federation requires a common, scalable and extensible framework for cooperation, involving communication and engagement among instruments, collaborations, projects, individual scientists, and citizen scientists. These larger groups require the ability to form flexible teams on short time scales, to share data, codes and other digital objects in real time, and to self-organize into spontaneous collaborations of varying scope. Additionally, the differing hardware infrastructures and computational resources may also need to be shared and mutually understood. On the organizational level, teams may need to share proprietary data and information governed by memoranda of understanding (MoUs). This may be facilitated by providing a means to hold data in “escrow” and give policy-driven access permissions to that data. Although important, these are merely the “known unknowns” of cyberinfrastructure challenges for MMA. To map out the entire problem space, the Scalable Cyberinfrastructure for Multi-messenger Astrophysics (SCiMMA)<sup>1</sup> team has been meeting since August of 2018 to define the requirement and challenges in the areas of cyberinfrastructure that can provide common, value-added services for scientists across the multi-messenger spectrum, to discuss with major MMA stakeholders—such as the Laser Interferometer Gravitational-Wave Observatory, IceCube, LSST, Laser Interferometer Space Antenna, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), National Optical Astronomy Observatory (NOAO), Cherenkov Telescope Array (CTA; Actis et al. 2011), NASA and others—their major cyberinfrastructure requirements, and to formulate a plan to meet these challenges. In January of 2019, a number of astrophysicists, computer scientists, and cyberinfrastructure specialists met at Columbia University to discuss the cyberinfrastructure requirements and challenges of MMA. The two day workshop revealed a number of important open questions and areas of cross disciplinary research and interest. The workshop concluded that:

- **Software:** There is a need for community-oriented and -driven software that can 1) access MMA data, 2) analyze multiple types of MMA data, 3) model MMA sources, 4) minimize and facilitate new development by building on and contributing to common libraries and frameworks. In addition, some means to collectively (as a community) evaluate competing or new disruptive technologies, encourage commonalities, and guarantee maintenance of critical MMA codes would be another valuable service. Finally, there is an opportunity to improve practices in sustainable software development across the different communities that use different codes through an umbrella collaboration. In particular, there is a need to foster integration across (software) ecosystems, and reducing the pigeonholing of software

---

<sup>1</sup><https://scimma.org>

developers. This can be partially accomplished by fostering workshops, conferences, or hackathons across these software ecosystems, but coordination in making this happen is sorely needed.

- **EM Followup Challenges:** Overall science output is increased, and redundancy reduced, when information is shared promptly and efficiently. Tools for sharing information need to be developed or improved in several regimes. These include searching localization regions, candidate follow-up, scheduling telescopes or instruments, observing plans, and data sharing. Standardized methods of information and data sharing also need development, together with incentives for opening access to data. The scheduling of telescope resources could be improved if they are done in an algorithmically-driven, adaptive manner that coordinates observations on multiple facilities and provides feedback to users. Responsive systems could make and distribute new plans in light of ongoing data acquisition and analyses of events. Such scheduling of resources would both benefit time-critical transient events and long-term continual followup campaigns and can facilitate resource and data sharing. These scheduling algorithms must also be swappable and extendable as the network of instruments grow, and will need to be implemented in a robust manner to ensure the necessary trust from facilities and the community.
- **Storage, Archiving, Access, and Search of Data:** The major challenge of the data revolution is the volume of data that needs to be stored, archived, accessed, analyzed and interpreted. A community-driven body can lead the way in developing the cyberinfrastructure that allows for improved and simplified access to MMA data, respecting proprietary policies that applies to both data and descriptive metadata that vary greatly between facilities and experiments. For instance, community broker services can facilitate searching for data in an area of a sky filtered by time and metadata, together with the creation of custom alerts for end users. This service must be modularized with defined communications interfaces accessed via Application Programming Interfaces (APIs) to allow the underlying software to be swapped out, enhanced, or replaced, whilst ensuring that scientists can write their own clients if they wish. How this should be done and what descriptive metadata should be collected remain open questions and would interface well with work in the library/information sciences and computer science. Archiving issues include supporting the life-cycle of data such as providing archiving resources for data useful to the community.
- **Permissioning and MOUs:** A service that exposes subsets of data based on permissions, including derived data products, would be an important value added service for MMA. Centralization of MOUs and permissioning would help enable coordination of follow-up, enable observatory facilities to optimize scheduling, as well as relative prioritization of events. It would also provide large facilities a centralized and standardized way of publishing and managing their data. Such a service would also promote the implementation of a cross collaboration authentication/authorization framework, which creates opportunities for collaboration with researchers in computer science.
- **Efficient and effective collaboration across disciplines:** Ensuring a mutually rewarding collaboration between astrophysicists, astronomers, computer scientists and

cyberinfrastructure specialists requires an understanding of the common areas of interest, the nature of the relevant expertise, and that scientists express their requirements so that designs can be *targeted* at meeting those requirements. There is a clear need to facilitate collaboration between these groups together, establish a community-driven set of requirements, record and disseminate decisions and designs, create funding streams for the production of cyberinfrastructure and tools, and verify that the work satisfies the requirements.

- **Education and Training:** Hosting summer schools, developing a high quality set of interactive tutorials and notebooks, and enabling collaborative teamwork are major value-added activities to the MMA community, and are priorities for the larger astrophysics community. There is also a major need across astronomy and astrophysics to provide hooks into the deeper scientific and engineering knowledge, language, and cultural practices of the different major collaborations/facilities so that the subtleties of data analysis for different messengers are accounted for by external users. Establishing an unified approach to education on statistical techniques that will enable MMA analysis across the various messengers, and uphold a strong level of statistical analysis and results. Thoughtful design of these programs to be inclusive from the outset will also allow the MMA community to better reach under-represented groups as part of a larger effort.
- **Collaboration Tools:** Services that enable collaborations, including wikis, mailing lists, scheduling and calendaring, video conferencing, centralized code repositories, data stores and management platforms all require a managing body. Such an agency could also provide the means to hold proprietary data in “escrow”, or enable pre-registration of analyses to be executed on data as it becomes available, and can ensure that the APIs that enable communication between the diverse array of facilities are functioning as intended.
- **Enabling Data Analysis Services:** Effective data analysis requires orchestration, which in turn necessitates the co-location of data and computational resources in a suitable analysis environment. Providing for such scalable frameworks for analysis of MMA data would be a major value added service to the community. One of the challenges is to quantify the requirements of data storage persistence, incoming data rate, processing latency, model robustness and update frequency, and prediction (service) latency. Another challenge is to corral the scalable compute systems (servers, compute clusters) for analyzing and serving data (centralized or on the cloud, on publicly-funded or commercial services) to provide services such as low-latency localization calculations based on incoming counterpart search data or large scale correlation or data analysis computations across the MMA spectrum.

## **An Institute for Multi-messenger Astrophysics**

There are no simple solutions to these problems; however, they provide significant opportunities for collaborative research between MMA and a broad range of computational disciplines. The success of MMA cyberinfrastructure will depend on both theoretical advances in how we represent data and the information they contain as well as the development of applications and systems that can scale to the complexity and size of data from a heterogeneous network of astrophysical facilities. Close interaction between MMA and programs within the NSF and other

national agencies (DOE) such as the Office of Advanced Cyberinfrastructure (OAC), Cyber-Human Systems (CHS), and Information & Intelligent Systems (IIS) can provide the necessary collaboration among data science, statistics, and astrophysics.

The emerging issues of MMA cyberinfrastructure have been studied by a number of groups including an NSF-supported workshop at the University of Maryland in May 2018 (Allen et al. 2018) and SCiMMA. A key conclusion of these studies is the best way to meet the challenges is to establish a community-driven, decentralized institute for cyberinfrastructure for MMA.

**We advocate support for a cyberinfrastructure institute for Multi-Messenger Astrophysics, to meet these community-driven challenges.** Such an institute must both lead and pragmatically respond to community opinion and needs. The institute will have the technical capability to evaluate the success of existing solutions, identifying new requirements from community needs, and integrating existing or new cyberinfrastructure from other efforts. The institute will ensure the stable delivery and maintenance of services, prioritize new development according to its importance to scientific goals, and keep abreast of technical developments in the academic and commercial sectors. The institute will also integrate with other international infrastructures supporting multi-messenger astronomy and work to support the required level of inter-operation. The core mission of such an institute would be to provide value-added services and support to the emerging area of MMA without impeding the organic growth of the science. As such, this institute would focus on cyberinfrastructure services that add to current capability without reproducing already existing solutions. It will provide the glue between the different areas of MMA in a rational, flexible, and standardizable way. Such an institute must be distributed in order to access the broad cross-section of researchers and expertise required to develop and support enhanced cyberinfrastructure critical to the success of MMA.

## References

- Aartsen, M. G., Ackermann, M., Adams, J., et al. 2017, *Journal of Instrumentation*, 12, P03012, doi: 10.1088/1748-0221/12/03/P03012
- Aasi, J., Abadie, J., Abbott, B. P., et al. 2015, *Classical and Quantum Gravity*, 32, 115012, doi: 10.1088/0264-9381/32/11/115012
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *PhysRevLett*, 119, 161101, doi: 10.1103/PhysRevLett.119.161101
- . 2017b, *ApJ*, 848, L12, doi: 10.3847/2041-8213/aa91c9
- Acernese, F., Adams, T., Agathos, M., et al. 2015, in *Journal of Physics Conference Series*, Vol. 610, 012014
- Actis, M., Agnetta, G., Aharonian, F., et al. 2011, *Experimental Astronomy*, 32, 193, doi: 10.1007/s10686-011-9247-0
- Allen, G., Anderson, W., Blaufuss, E., et al. 2018, arXiv e-prints, arXiv:1807.04780. <https://arxiv.org/abs/1807.04780>
- Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, arXiv e-prints, arXiv:1902.01945. <https://arxiv.org/abs/1902.01945>
- IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018, *Science*, 361, eaat1378, doi: 10.1126/science.aat1378
- IceCube-Gen2 Collaboration, :, Aartsen, M. G., et al. 2014, arXiv e-prints, arXiv:1412.5106. <https://arxiv.org/abs/1412.5106>
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2008, arXiv e-prints, arXiv:0805.2366. <https://arxiv.org/abs/0805.2366>
- Kagra Collaboration, Akutsu, T., Ando, M., et al. 2019, *Nature Astronomy*, 3, 35, doi: 10.1038/s41550-018-0658-y
- McLaughlin, M. A. 2013, *Classical and Quantum Gravity*, 30, 224008, doi: 10.1088/0264-9381/30/22/224008