Facilitating cosmic probes and dark matter searches with improved data access and software tools for multi-wavelength and multi-messenger analyses.

E. Charles[1]

[1]KIPAC, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Abstract Almost every science topic covered in the Cosmic Frontier will at some point require combined analysis of data from multiple instruments. Any one instrument can only provide a limited picture of the sky; much of the most important information comes from correlating data between instruments. In particular, a number of existing and proposed instruments observe the high-energy, dynamic and transient phenomena at different wavelength or photon energies, or even different messangers, such as gravity waves and neutrinos. A key to maximizing the scientific return of both existing facilties and those built in coming years is to facilitate analyses that combine data from diverse types of instruments. Careful planning, flexibility and good communication between instrument teams are likely to substantially improve the utility of the data and software needed to analyzing that data that is released by the instrument teams. The Snowmass process is a good opportunity to take stock of data release plans and set up effective communication both between instrument teams and between the teams and member of the community that will analyze and interpret the data.

1 Introduction and Background An important factor contributing to the scientific successes of many instruments probing the cosmic frontier has been the combination of publicly available data and high-quality publicly available software to analyze that data.

This has enabled spectacular discoveries by scientists outside the instrument teams, e.g., the discover of the "Fermi bubbles" [1]. It has also facilitated powerful analyses that combine data from different facilities, e.g., the analysis of the cross-correlation between the DES weak-lensing maps and the extra-Galactic gamma-ray background [2].

Publicly releasing data and providing the tools to effectively use that data is a challenging task, typically requiring a dedicated group of scientists and software engineers to accomplish. Depending on the scope of the instrument and the amount of data, this might vary from a few people who also have a number of other responsibilities, (as in the case, e.g., for the *Fermi* Gamma-Ray Burst Monitor) to large dedicated groups working within a very large project (e.g., the Data Management group of the Vera Rubin Observatory).

As new facilities come online and scientists seek to exploit new synergies between those facilities, both the released data products and supporting tools need to evolve to best match the science drivers. Doing this well requires a significant amount of communication between the instrument team and the community of scientists using their data.

2 The Current Situation Many existing and upcoming cosmic frontier instruments are fully supported to release fairly low-level data publicly (i.e., cleaned images or photon lists and associated instrument response functions). This includes all NASA and ESA missions, LSST, DES, and many other facilities. Notably, however, many instruments that detect high-energy gamma-rays, cosmic rays and neutrinos, e.g., HESS, Veritas, AMS-02, HAWC, CTA and Ice Cube, plan to publicly release only high-level summary data. The implementation of the summary data releases vary from instrument to instrument, and is often sub-optimal for performing combined analyses with other instruments. It is important to note that this is despite extensive goodwill on the part of the instrument teams as they typically do not have either the manpower or computing and data storage resources to support large scale data release and software support.

3 Using the Snowmass Process to Improve the Situation A multi-pronged approach to finding relatively simple and low cost ways to obtain high-impact improvements in facilitating data analysis in the high-energy Cosmic Frontier can easily be integrated into the Snowmass process.

Specifically, as part of the earlier stages of the process process a small amount of extra effort could be applied to:

1. Comparing the data and software release plans of existing and upcoming facilities in the cosmic frontier and in particular in high-energy astronomy and astrophysics, with the set of scientific goals studied as part of the Snowmass project. Building on this study, identifying possible changes to the data release plans that could improve scientific return.

2. Engaging the stakeholders in discussions about the feasibility, costs and benefits of potentail changes to their public data and software releases models, and identifying possible high-impact changes and improvements.

3. Surveying the available and planned software that can be used to perform combined analyses of public data from different instruments, identifying any pressing unaddressed needs and identifying ways to meet those needs.

The findings from these surveys and discussions should be summarized in a contributed paper as part of the Snowmass process. Obviously, without having studied the needs of the community and the plans of existing and proposed facilities in detail, it is not known what conclusions and recommendations such a paper will make. However, it would be natural to expect that the paper will identify some actions that would greatly benefit the community, such as:

1. Forming a small working group whose specific charge is to work with the stakeholders (i.e., existing and upcoming instrument teams) to publicly release data that best serves the community give the available resources.

2. Identifying best practices for public data and sortware release; particularly best practices that are specific to instruments in the high-energy domain of the cosmic frontier.

3. Identifying specific software tools that fall outside the scope of any one instrument, but would provide important synergist benefits, so that they be included in the charges of upcoming calls for proposals.

It is likely that process described above will clarify the details of which of these actions should be undertaken and how best to do so.

References

[1] M. Su, T. R. Slatyer, and D. P. Finkbeiner, ApJ**724**, 1044 (2010), arXiv:1005.5480.
[2] S. Ammazzalorso *et al.*, PRL**124**, 101102 (2020), arXiv:1907.13484.