# Dear Colleague Letter: Request for Information on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030)

Call: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031
Deadline: April 5

## Title

Cyberinfrastructure needs for the Compact Muon Solenoid experiment at the Large Hadron Collider

## Authors

Kenneth Bloom University of Nebraska-Lincoln
Lothar Bauerdick Fermilab
Brian Bockelman University of Nebraska-Lincoln
Barry Blumenfeld The Johns Hopkins University
John Cumulat University of Colorado
Sridhara Dasu University of Wisconsin-Madison
Peter Elmer Princeton University
Oliver Gutsche Fermilab
Burt Holzman Fermilab
James Letts University of California San Diego
Kevin Lannon University of Notre Dame
Daniel Marlow Princeton University
David Mason Fermilab
Norbert Neumeister Purdue University
Harvey Newman California Institute of Technology
Elizabeth Sexton-Kennedy Fermilab
Paul Sheldon Vanderbilt University
Eric Vaandering Fermilab
Peter Wittich Cornell University
Frank Wuerthwein University of California San Diego
Avi Yagil University of California San Diego

## Abstract

The Large Hadron Collider (LHC) produces multi-petabyte datasets in every year that it operates. With a twenty-year program planned, and an upgrade to the collider that will allow for higher beam intensities, extracting physics discoveries from the data will require significant and sustained advances in cyberinfrastructure. This contribution to the request for information describes the physics motivations for the LHC, and the computing functions that require excellent cyberinfrastructure. Research and development work is needed in areas of distributed high-throughput computing, data management and access, software and hardware performance, wide-area networking, and cyberinfrastructure. The authors are all members of the Compact Muon Solenoid (CMS) Collaboration.

## Research Challenge

The exploration of particle physics at the energy frontier is one of the most promising routes to understanding the structure of matter and its interactions at the most fundamental level. Open questions include the mechanism of electroweak symmetry breaking, the nature of dark matter, and whether there are new particles and interactions to be discovered. The Large Hadron Collider (LHC) at CERN, the European particle physics laboratory, gives us access to the highest-energy particle collisions ever produced in a controlled environment, and offers the opportunity to discover and study new particles. The discovery of a relatively light Higgs boson in 2012 indicates that there are yet more new particles to be observed. Such observations could change our most basic understanding of the universe.

Particle physics is a data-intensive science that relies on the technologies of distributed high-throughput computing (dHTC). The Compact Muon Solenoid (CMS), one of the major experiments at the LHC, records multiple petabytes of raw data each year. These datasets must be processed, stored and transferred to multiple computing sites around the world so that they can be made available to the thousands of scientists who analyze the data for physics measurements. There is also an extensive program of simulation of different physics processes and how they manifest themselves in the experiment, which leads to additional datasets that are several times larger than those recorded by the experiment. CMS now manages 100 PB of real and simulated data, and expects this number to increase by 60 PB in 2017 alone. These datasets are now significantly larger than those available when the LHC began regular operations in 2010, and while the doubling rate has slowed, scientists will still need to regularly access a significant fraction of the total recorded data for each individual measurement.

In the longer term, the computing challenges will become even greater. The upgraded LHC that is anticipated to run at higher luminosity (HL-LHC) starting around 2026. Each beam crossing will have about 200 proton interactions, compared to about 30 right now. CMS will record more events at a 7.5 times higher rate, and each event will be more complex due to having more proton interactions superimposed on the event. The first year of HL-LHC operation will produce

130 PB of raw data alone for CMS, and the simulation needs will scale up correspondingly. Many innovations in software and computing will be needed to retain the same functionalities with this much data within finite computing budgets.

Advanced cyberinfrastructure will be required to support the three basic computing functions within CMS:

A. Data management: Multi-petabyte datasets must be stored, catalogued, transferred among computing sites and ultimately made accessible to a world-wide collaboration of scientists.
B. Central processing: The reconstruction of detector data into information about the particles produced in collisions is executed through centralized mechanisms, in which a small number of operators submit processing jobs to the computing system, which means the usage can be controlled in an orderly fashion. This is also true for the creation of the simulation samples. The outputs are “official” data samples of the experiment that are made available for any collaborator to analyze further. On average these activities, which are typically CPU-bound, currently occupy about 60,000 computing cores distributed around the world.
C. Multi-user distributed analysis: Further processing of the reconstructed detector and simulation data is performed by individual scientists, each of whom creates his or her own code workflow that is then submitted to the computing systems at a time of his or her choosing. This processing is “chaotic” in comparison to the central processing. On average these activities, which are typically I/O-bound, occupy an additional approximately 60,000 computing cores distributed around the world.

Underlying all of these functions is the need for access to the data. This includes matters such as efficient data input and output for the applications (particle physics workflows are more I/O intensive than many other scientific workflows) and the creation and operation of global data federations that allow the access to any experiment data from any computer, anytime, anywhere.

## Cyberinfrastructure needs

Since the startup of the LHC, the CMS experiment has succeeded through the use of dHTC, which has turned out to be a key piece of cyberinfrastructure. With each LHC collision event being statistically independent from all the others, the computational problem is pleasantly parallel, leading naturally to a scheme in which work can be distributed across many computing systems at once. dHTC has been a perfect match to the LHC, through an implementation that carries concepts of establishing trust and cybersecurity through a virtual organization (VO). dHTC also provides a natural way of resource sharing through opportunistic usage. When the owner of a computing resource leaves it partially idle, other VOs can straightforwardly have access to it, allowing for improved overall resource usage. This has proven critical for providing sufficient resources to the LHC experiments, and also has provided side benefits to non-LHC scientists who have gotten access to LHC-owned resources. It has allowed small institutions to participate and connect to a worldwide infrastructure using only reasonable amounts of local effort. In the U.S., the national cyberinfrastructure is operated by the Open Science Grid (OSG), and it is important that its services be maintained in the future, both for the benefit of the LHC experiments and a broad array of other sciences.

However, there are several categories of cyberinfrastructure that must continue to be developed and improved to meet the scale demands of the HL-LHC. These efforts would be best carried out by teams of computer scientists and “domain scientists” (i.e. the users of the cyberinfrastructure) to deliver products that bring the best computer science to bear on the problems faced by domain scientists.

For instance, CMS relies on a job-submission infrastructure that is used to provision processing resources for functions B and C above. The current infrastructure used in CMS, a combination of glideinWMS and HTCondor, has been demonstrated to reach the scale of 500K simultaneous CPU cores in use. We anticipate that a factor of four greater scale will be needed to handle the large datasets at the HL-LHC. Achieving that will require either continual improvements to the existing product or the development of a replacement product with the right scaling qualities.

Data access is a clear challenge for the HL-LHC era, given the anticipated size of the datasets. The current CMS technology for data management is subscription-based, i.e. datasets are explicitly placed at certain sites and thus the locations of files must be carefully tracked. Workflows are then typically executed at the sites that host the needed input dataset. Alternative models should be explored, including those based around data caches that are located at each computing site. Files would only be transferred to a site when needed as inputs to a workflow running at the site; afterwards it would remain cached at the site and available to other workflows, until it is deemed more useful to have a different dataset in the cache. Such caching mechanisms are already under study, and continued efforts are needed to make sure they achieve the needed scale.

The analysis of reconstructed event data for the preparation of physics measurement, function C above, is a task performed by many CMS scientists. As only a limited number of calculations are performed on each event during this stage, the speed of processing is limited by the I/O bandwidth available. Thus, a thorough program of improvements to I/O is necessary, ranging from the low-level performance of file systems. through the design of computing clusters that can maximize throughput, to the I/O performance of the processing executable itself.

The wide-area network (WAN) infrastructure that connects far-flung computing sites impacts all of the functions that must be provided. While the line speed of WANs has continued to grow steadily, the performance of the endpoint computing systems and of the software has not always kept up. Networking research must continue, including in the area of Software Defined Networks and Named Data Networks that would allow the construction of content delivery networks for the efficient movement of data. This would move many of our systems built on top of the network stack into the stack itself and allow for deeper optimization.

Cybersecurity is critical to the integrity of the worldwide computing system. Users are given access to the computing infrastructure through a system of identity management that validates the authenticity of each user. This will be simplified through a system of federated access in which users are validated by their home institutions. Efforts to simplify identity management must continue, especially considering the sometimes cumbersome authentication with personal certificates.

All of the challenges of the LHC and the HL-LHC require a broad array of cyberinfrastructure development: advanced computing, data infrastructure, software infrastructure, applications, networking, and cybersecurity. It is worth noting that while these elements of cyberinfrastructure are needed directly by particle-physics projects such as CMS, they can be used by data-intensive researchers in many fields of science. Indeed, the glideinWMS system is already used throughout the OSG, and OSG is also developing a system of data caches for all of its users.

## Other considerations

In this contribution, we have concentrated on needs of the CMS experiment related to the cyberinfrastructure itself and did not describe implications for the physics software, namely the simulation, reconstruction and analysis software needed for the successful analysis of LHC data. We are confident that the described needs for the dHTC infrastructure for CMS will enable the future software stack to perform well and be sufficiently flexible to accommodate future developments. A strategic plan for particle physics software is being developed through the S2I2-HEP effort, http://s2i2-hep.org, and will be delivered in Fall 2017.

Given the complexity of cyberinfrastructure, education and workforce development will always be a concern. First, physicists who will be using the cyberinfrastructure to analyze CMS data and make scientific discoveries are not trained as computing professionals. They will need some additional training to make best use of all the systems. It would be best if this sort of training could be incorporated into the undergraduate curriculum, as computing skills will be useful to physics students in all sorts of career paths. In addition, some fraction of physicists who go on to long-term careers will need to lead the development of this cyberinfrastructure. They will require more advanced training in this area, which could be facilitated through focused programs like summer schools. There also needs to be a proper career path for physicists who contribute to cyberinfrastructure, so that their achievements are recognized and respected as they advance through the ranks at universities and laboratories.