LHC tunnels, will be 20 meters in diameter when complete and weigh 7,000 tons. For reasons of technical complexity, scale, and facility scarcity, the human scale of HEP research is correspondingly large. The ATLAS experiment, a large experiment at the LHC, involves over 1,800 physicists at 140 institutes in 34 countries around the world.

While high-energy physicists have long worked in large collaborations and with shared instrumentation, the complexity of the LHC, the size of the collaboration, and the volume of the data that will result from the experiments have introduced the need for serious changes in the social and technical infrastructures that support HEP collaborations.

Social and Organizational Challenges

The increasing size and the complexity of the HEP community's scientific goals have introduced challenges to the ways that experiments are organized and managed due to underlying changes in the social and organizational structures of the field. New forms of authority, attribution of credit, and training have emerged in response to these challenges of scale and complexity.

Leading by Persuasion and Managing by Coffee

Consider first the challenge of leadership. CERN is governed and has historically been funded primarily by its twenty-five European member states. But its status as perhaps the leading HEP facility in the world puts it in a global spotlight, and its experiments have attracted participants from institutes around the world, thereby complicating the issue of governance. Participating institutes not in CERN member states must secure funding from their home country and volunteer to make some contribution to the experiment (e.g., constructing some component of the detector or writing software). But as the number of non-CERN-member-state participants increases, so too does the fraction of experimental resources not under the direct control of CERN and the elected leaders of the experiments (for a detailed description of the HEP collaboration structure, see Knorr Cetina 1999, 166–191).

As of 2004, only about 25 percent of the resources in the ATLAS experiment that come from CERN member states are routed through CERN's internal accounting systems. This means that only 25 percent is controlled directly by the formal experiment leadership. Researchers from other participating institutes control the remaining 75 percent of the resources. Some of these researchers may also be leaders of specific subprojects, but these resources are not formally controlled by CERN. Effectively, because any non-member-state institute is free to withdraw its voluntary contribution to ATLAS at any time (to be sure, there would be substantial intangible costs to the institute, but withdrawal is not unprecedented), this means that the elected leaders have little real power beyond persuasion and a technique that one project team leader

describes as "managing by coffee." In other words, leadership in this environment becomes an exercise in continuous consensus building through informal meetings (usually held over coffee at CERN), formal presentations, and peer review panels for making certain important decisions. The LHC experiments have evolved (and are continuing to evolve) a fascinating array of techniques for organizing themselves to take action and make decisions in an environment where leadership and power are highly ambiguous.

One example of this is the election of a "technical coordinator" for each of the large experiments. This individual has direct responsibility for the construction and installation of the detector, despite (as noted above) having minimal control over the resources. On the ATLAS experiment, this is accomplished through a constant process of technical review that is tied both to ATLAS's technical requirements and those of the funding agencies that support the work of participating institutes. These reviews, in other words, assess the suitability of work for installation in the detector, but can also be used by individual institutes in justifying their efforts to funding agencies in their home countries. While ATLAS members reported that it was difficult initially to get institutes to agree to submit to this extensive review process, these were later widely felt to be quite useful, and the technical coordinator indicated that institutes were eager to undergo these reviews.

Beyond this, the experiment has an elaborate hierarchy with coordinators of resources and efforts at multiple levels, but it is widely acknowledged that this hierarchy is not absolute. When there are disputes about how a particular component should be designed or constructed, multiple competing proposals may take shape and move forward until a crucial decision point is reached. At this point the leader of the work group in which the dispute has occurred may try to bring the group to a shared solution, or if there is insufficient common ground, may turn to a panel of peer reviewers to resolve the dispute. Some physicists also acknowledge that politics and economics play a role in this process. An institute that is contributing large amounts of money, materials, and/or effort toward the construction of a detector component will have more influence on how that component gets constructed, because in the end it is their responsibility to contribute to a working detector, and nobody wants to build something they think will not function.

There are similar disputes when it comes to the analysis of physics data. The experiment hierarchy assigns data to physics groups for analysis, and assigns shared computing resources for these analysis tasks. It is widely acknowledged, however, that some data sets will be more desirable than others, as they are more likely to result in a high-profile discovery. Thus, the assignment of data to physics groups is a highly contentious process that is watched closely by all collaborating institutes. Yet despite these assignments, the data are available (without accompanying computing resources for analysis) to anyone who wishes to analyze them. Some LHC physicists have therefore

described colleagues attempting to amass private clusters of personal computers for conducting their own analyses. Once again, there is the potential for conflict here in that multiple groups might be analyzing the same data and attempting the same discovery. To address this potential, the experiment has a strict policy that no results can be released without prior authorization and approval from the publication committee. In the event of multiple papers on the same topic, the publication committee will decide which result gets published or if the results must be combined. Additionally, all papers released by the collaboration must bear the name of all collaborators as authors.

Standing Out in a Crowd: Getting Credit When You Have 1,699 Coauthors

Another issue that becomes problematic as the size of HEP collaborations continues to increase is the attribution of credit for research discoveries. Historically, the scientific research enterprise has used reputation as its primary currency (Whitley 2000), and one's reputation is earned through first- and single-author publications, awards, and similar clear measures of individual contribution. Promotion and tenure, in turn, are awarded based on tangible evidence of scientific productivity provided by first-author publications as well as the peer assessment of a researcher's ability to carry out high-quality independent research. Where collaboration occurs in other fields, researchers usually demonstrate that they played an important role in the collaborative work by publishing some results as a first author. This mode of operation becomes quite difficult in HEP, where there is a long-standing tradition of listing all collaborators as authors on every paper, always in alphabetical order.

The idea behind this practice is to render individual reputation subservient to the collaboration, and recognize that everybody plays a crucial part in carrying out the work, even if they do not all participate in the final analysis that yields the highest-profile results. At the same time, individuals must nonetheless find ways to differentiate themselves from the crowd if they are to remain competitive in applying for jobs, promotions, and even desirable assignments within the experiment itself. As the author list on the current CERN experiments approaches two thousand names, one can easily imagine the difficulties in attempting to evaluate individuals via a list of publications alone. In fact, many in HEP argue that authorships have become meaningless; what really matters for individuals are letters of recommendation, informal reputation within the experiment, and the number of conference presentations given on behalf of the entire collaboration. This argument has given rise to a contentious debate within the HEP community—a debate that is described in greater detail in Birnholtz (2006).

On the one hand, some contend that the current system of authorship (long, alphabetical author lists) is the only way to ensure that credit is attributed to everybody involved in the project, such as those whose primary contribution was in the design of the apparatus and who will not be actively involved in the data analysis. On the other hand, others assert that this system, by virtue of rendering all individual contri-

butions highly ambiguous, does not really effectively recognize people like detector designers, and also deprives the truly clever contributors to specific papers and discoveries of the individual recognition they deserve. Many alternatives have been proposed, such as listing the major contributors to a paper before an alphabetical list of the rest of the authors or removing authors who do not feel they could effectively defend or explain the results presented in a particular paper. The system that ultimately takes shape will not likely be put to the test until the LHC becomes operational and collaborations of unprecedented size begin publishing results. It will likely have important implications, however, for credit attribution in many big science disciplines.

Designing for Our Progeny: The Impact of Long Time Horizons

Another critical issue that emerges in a discussion of the social effects of the large physical scale of HEP research is the concomitant increase in the time scale of experiments. One example of a long-lived project is the ATLAS experiment, which began its life around 1989 as a "proto-collaboration," a type of working group that is critical to the development of an experiment, called Eagle. The current experiment is not likely to have publishable results until at least 2008. This means that those most actively involved in this experiment will have a twenty-year gap in their publications based on data from "real" experiments (as contrasted with, say, Monte Carlo simulation data, which is often used as a supplement to instrument data). For junior faculty and graduate students in the United States, this is widely acknowledged to be tantamount to career suicide. (This is less of a factor in Europe, where publications and dissertations based on simulation data are more accepted.) The unsurprising effect of the long time horizon, therefore, is that there are few junior faculty or graduate students from U.S. institutes involved in the LHC experiments.

Before discussing the implications of this, it bears mentioning that time scale was a problem historically as well. In the past, though, it was common for graduate students to participate in the detector design for one experiment, while simultaneously taking and analyzing data from another. As such, students got exposure to both analysis and design tasks, had publications based on real experimental data, and had a logical career path to follow that frequently led them to work on the experiment for which they assisted with detector design. Today, experiments take so long and there are so few of them in progress that such arrangements are no longer possible.

It is thus primarily senior faculty from the United States that are involved in the LHC experiments—and many of them will be ready to retire (or will have already retired) when the experiment begins taking data and their junior colleagues become involved. This seems likely to put the U.S. institutes, which are located physically far from CERN, and whose junior faculty will have less experience and familiarity with the detectors than their European and Asian colleagues, at a significant disadvantage. The actual outcomes remain to be seen, however.

Building a Cyberinfrastructure to Support the LHC Experiments

Because of the significant amount of data anticipated from the LHC (over fifteen peta-bytes per year from the four experiments) as well as the large, globally distributed col-laborations (typically fifteen hundred to two thousand PhD physicists per experiment), the LHC physicists are striving to create a cyberinfrastructure that can harness more physical and intellectual resources to enable scientific discovery at the LHC. The com-putational needs for an LHC experiment are so large (typically about one hundred thousand of today's most powerful workstations) that LHC physicists need a system that allows their collaborations to utilize all the computational resources available, wherever they are physically located in the world. Given the extraordinary amount of data (about ten petabytes per year per experiment when simulated data are included in the total), the collaborations also must be able to access storage resources wher-ever they exist. The common component that ties storage, computers, and people to-gether is the network. The network is thus a critical component of the LHC global infrastructure.

Computational and Data Grids

LHC physicists have extensively studied how to build an infrastructure that will pro-vide the needed computational and storage resources, and that will ensure their ef-fective and efficient use. They have concluded that a grid-based cyberinfrastructure composed of hundreds of computing sites linked by high-speed networks offers the most cost-effective means of sharing resources and expertise, since large geographic clusters of users are likely to be close to the data sets and resources that they employ (chapter 1, this volume). Such a distributed configuration is also preferred from a socio-logical perspective as it enables distributed control, and therefore facilitates autonomy in pursuing research objectives. In the Compact Muon Solenoid detector (the other large detector in the LHC), for example, the computing resources will be arranged in a tiered "hierarchy" of regional computing centers, interconnected by regional, na-tional, and international networks. The levels include Tier-0, the central facility at CERN where the experimental data is taken, and where all the raw data are stored and initially reconstructed; Tier-1, a major national center (located typically at a major lab-oratory) supporting the full range of computing, data handling, and support services required by a large scientific community; Tier-2, a university-based system supporting analysis and reconstruction on demand by a community of typically thirty to fifty physicists; Tier-3, a work-group cluster specific to a university department or a single physics group; and Tier-4, an access device such as an individual user's desktop, laptop, or even mobile device. Each Tier-1 will have about 40 percent of the computing and storage capability of the Tier-0 CERN facility, and each Tier-2 site will have about 10 to 20 percent of the capability of a Tier-1.

The grid framework described above is expected to play a key role in realizing the LHC collaboration's scientific potential by integrating all resources, from desktops and clusters at universities to the high-performance computing centers and national labs, into a coherent environment that can be utilized by any collaboration member. Such a collaboration-wide computing fabric will permit enhanced participation in the LHC research programs by physicists at their home institutes—a point that is particularly relevant for participants in remote or distant regions. Since grids enable distributed resources to be fairly shared while taking into account experiment policies as well as local ownership and control, a highly distributed, hierarchical computing infrastructure exploiting grid technologies is a central element of the LHC worldwide computing model.

In the United States, several key initiatives were undertaken in support of this vision of building worldwide grid-based cyberinfrastructures (Avery 2003). The Grid Physics Network (GriPhyN) project, funded by the National Science Foundation in 2000 for $11.9 million, involved a collaboration of physicists, astronomers, and computer scientists from fifteen institutions, including universities and national laboratories. Its computer science research was aimed at developing grid "middleware" capable of supporting large grid-based cyberinfrastructures. The GriPhyN Virtual Data Toolkit—a comprehensive packaging of grid software from GriPhyN and other projects—has been adopted by the international grid community.

The International Virtual Data Grid Laboratory (iVDGL) was funded by the National Science Foundation in 2001 for $13.7 million and is composed of approximately twenty institutions, including universities and national laboratories. The iVDGL is deploying a grid laboratory where advanced grid and networking technologies can be tested on a large scale by multiple disciplines. In 2003, the iVDGL, in partnership with GriPhyN and the Particle Physics Data Grid (PPDG), deployed Grid3, a general-purpose grid of approximately thirty sites and thirty-five hundred processors that operated for two years, and supported applications in HEP, gravitational wave searches, digital astronomy, molecular genomics, and computer science. Grid3 sites are now part of the Open Science Grid, a distributed computing infrastructure for large-scale scientific research that integrates computing and storage resources for more than fifty sites in North America, Asia, and South America.

The PPDG is another example of a U.S.-based project to deploy, use, and extend grid technologies to serve the data management and computing needs of HEP (Bunn and Newman 2003). It began in 1999 as a joint project between several laboratories and universities funded by the U.S. Department of Energy under the Next Generation Internet program, and has continued through ongoing support from the Department of Energy's Mathematical, Information, and Computational Sciences Division base program along with funding from the Scientific Discovery through Advanced Computing program. The PPDG has played a critical role in hardening grid technologies,

promoting service specifications, deploying different service implementations, and developing the security policy and architecture that allows these different elements to be integrated into a common grid fabric.

The three grid projects described above each played a unique role in deployment, operation, and integration that contributed to the community's ability to build a production-ready cyberinfrastructure with the Open Science Grid. While there have been similar grid technologies efforts internationally as well, these projects illustrate the complexity of building a cyberinfrastructure to support a global-scale collaboration, and the effort required to direct proper attention and resources to the development, packaging, integration, and operation that are required to produce a production-ready cyberinfrastructure.

Networking

HEP has a long history of involvement in networking, and has been one of the primary proponents and developers of wide-area networks in support of science. This involvement started with analog 9.6 Kbit per second leased lines that composed HEPNet in 1985 and continues to multiple 10-gigabit transatlantic links supported by LHCNet today. Networks are critical to a discipline like HEP since they have such large distributed collaborations. Robust, ubiquitous networks are key enablers of large international collaborations.

Today's best-effort networks with their rapidly increasing bandwidth will be vital to the success of HEP in the future. Shared network infrastructures like Internet2 and ESnet have gone a long way in enabling HEP, but physicists are finding that even more capabilities are required to deliver an efficient, effective infrastructure to support LHC-scale collaborations. Current networks, though highly performing and reliable, are still only best-effort networks that are unable to adapt their behavior or modify their delivered service in response to demanding applications or high-priority tasks.

In preparation for the LHC turn-on in 2008, physicists, in collaboration with computer scientists and network engineers, are working on numerous projects to advance the network from a "black box" into a dynamic managed component of their infrastructure through projects in the United States (like UltraLight, LambdaStation, Terapaths, and UltraScienceNet) and internationally (like the Global Lambda Integrated Facility, national-scale user- and application-managed network projects such as the User-Controlled Lightpath Project (UCLP) with CA4Net in Canada, and many others). The goal is to create a network infrastructure that is dynamic, manageable, and integrated within the HEP infrastructure. Physicists need to have the network support numerous types of data flows: real-time, interactive uses like videoconferencing, remote-control rooms, shared interactive applications, and remote presence; high-bandwidth data transfers from storage elements to compute elements; and low-priority

bulk transfers between storage elements and varying priority user analysis applications accessing widely distributed copies of data sets.

Conclusion

The HEP community represents project-based collaborations of remarkable scale. Achieving the capacity to organize collaborations consisting of thousands of individual investigators took many decades to achieve, and introduces new social and technical challenges with each new experiment. Currently, the HEP community depends not only on a unique set of social and organizational processes but also an advanced cyber-infrastructure that the community continues to build and invest in that encompasses data, computing, and networking. Continued development of and reliance on this type of cyberinfrastructure is a critical enabling factor that allows HEP to conduct global-scale "big science".

References

Avery, P. 2004. Grid computing in high-energy physics. In *Proceedings of the 9th international conference on B physics at Hadron Machines—BEAUTY 2003: AIP conference proceedings, vol. 722*, ed. M. Paulini and S. Erhan, 131–140. Melville, NY: American Institute of Physics.

Birnholtz, J. 2006. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology* 57 (13): 1758–1770.

Bunn, J., and H. Newman. 2002. Data-intensive grids for high-energy physics. In *Grid computing: Making the global infrastructure a reality*, ed. F. Berman, G. E. Fox, and A. J. C. Hey, 859–906. New York: Wiley.

Galison, P. 1997. *Image and logic: A material culture of microphysics*. Chicago: University of Chicago Press.

Galison, P., and B. Hevly. 1992. *Big science: The growth of large-scale research*. Stanford, CA: Stanford University Press.

Heilbron, J. L., and R. W. Seidel. 1989. *Lawrence and his laboratory: A history of the Lawrence Berkeley Laboratory, volume 1*. Berkeley: University of California Press.

Knorr Cetina, K. 1999. *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Whitley, R. 2000. *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.