The WLCG collaboration

The Worldwide LHC Computing Grid (WLCG) is the global infrastructure developed and operated over the last two decades that provides the computing infrastructure for the processing and analysis of data from the LHC experiments. The notable achievement of WLCG has been to successfully meet the needs of the LHC experiments by the integration of globally distributed Exascale resources with services and software, within a trust framework that transcends site and national boundaries. WLCG is currently of the scale of over a million cores and more than an Exabyte of storage. It extends across 165 sites in 42 countries and is underpinned by 66 Memorandum of Understanding signed by Funding Agencies across the world. The success of WLCG has relied on innovation, leadership, collaboration, agility, and the confidence of the community and funders to commit to the endeavor. The future brings new challenges in terms of scale, technology, funding, sustainability, and the growth of other related communities. Within this context, WLCG must develop a strategy to enact over the next four years to prepare for HL-LHC Run-4 that will bring a step change in the amount and complexity of the data that will need to be processed.

Over the next four years, WLCG must continue to:

- Support the core use cases (HL-LHC) whilst evolving the infrastructure for HL-LHC.
- Evolve services to benefit from modern technologies with a careful evaluation of in-house vs commodity
- Optimize the person-power needed to run the infrastructure.
- Explore solutions that constrain the cost of our computing and/or attract new resources
- Increasingly embrace heterogeneous architectures and include diverse facilities
- Develop methods to reduce our energy footprint and/or the CO2 emissions
- Collaborate with other communities representing sciences, software, and digital infrastructure, to share R&D and operational aspects where mutually beneficial.

The WLCG strategy to achieve these objectives include:

- Minimising the cost and energy footprint of computing by being able to use a wide range of resources including Grid, public or private clouds, HPCs, and by expanding the architectures from the original x86 to include ARM, Power, GPUs, and other technologies that may become available. Our strategy is to build agility so that opportunities can be seized.
- Modernize tools and services, benefiting from standards that are less HEP-specific where applicable. Our strategy is to profit where we can from well supported software that is shared by other groups.
- Ensure continuity and evolution through a programme of data challenges and improved monitoring. Our strategy is to re-risk the project by avoiding technical debt without introducing disruptive changes.
- Increased collaboration with other communities, including formalizing the status of observers in WLCG and increasing synergies with the national and international initiatives developing large scale computing infrastructures. Our strategy is to benefit where we can from other relevant work.

- Review the structure and governance of WLCG; set up bodies, boards, and task forces to address problems and issues. Our strategy is to focus and optimise our existing strengths.

WLCG service operations

The WLCG service has reliably delivered the infrastructure to support the LHC experiments' computing needs since time before the start of Run-1. The WLCG sites provide services to the experiments with an availability and reliability generally exceeding the MoU[1] targets. The volume of resources in WLCG (CPU, storage, network) has increased by more than an order of magnitude since the start of LHC, with no impact on the service efficiency and performance. Operations are a shared activity between the WLCG sites and the experiments, with a large overlap between the two. The WLCG Operations Coordination team is in charge of overseeing activities and reporting to the WLCG Management Board. The activities include both the day-to-day operations (e.g. addressing of issues and reviewing incidents and target metrics) but also the execution of medium-term activities (e.g. upgrade campaigns, deployment of new features). The strong connection between experiments, sites and infrastructure providers allows WLCG to quickly react to rapidly changing conditions, as we have seen many times in the LHC physics program. The review committees (LHCC and RRB) regularly congratulate WLCG for the success in delivering computing services which fulfill the needs of the experiments.

The effort available for operations has decreased over recent years, both within the experiments and at sites. A non-conclusive discussion on this topic took place at a recent LHCC. The WLCG stakeholders should consider different ways to attract effort in this area and/or to optimize the use of existing effort. Possible options include finding more synergies between stakeholders,

[1] https://wlcg.web.cern.ch/mou

innovating the tools and services in use to run operations, reducing the complexity of some of the services. A dedicated initiative on this topic should be launched.

[OPS-1] WLCG to periodically examine (e.g. yearly) the effort to run operations. Investigate, if needed, ways to further optimize the operational effort needed to run the infrastructure, services and workflows.

WLCG technical evolution

The operational success of WLCG is based, in part, on a conservative approach to technical evolution. Since the facilities in WLCG are loosely coupled and the middleware development effort is distributed, new features are introduced to the infrastructure rather gradually to ensure continuity of operations. Examples are the multi-year plan for the transition from X509 to tokens and the decommissioning of SRM.

In terms of more disruptive innovation, there is sometimes resistance to changing a system that works. In the modern and rapidly evolving environment of distributed computing, WLCG risks accumulating additional technical debt if a more agile model is not put in place. There are examples of initiatives put in place in preparation for HL-LHC, particularly in the area of Data Management (DOMA), that led to a successful R&D program, with short and long term benefits. Broadening the scope of modernisation of tools and services will be an essential ingredient for the future technical sustainability of the WLCG infrastructure. The capability to use different computing architectures will also play a key role.

[TECH-1] WLCG to develop plans to broaden the scope of modernisation of tools and services as an essential ingredient for the future technical sustainability of the WLCG infrastructure.

WLCG should establish a structure that allows for innovation to take place without interrupting production operations. The Data Challenges in preparation for HL-LHC have been so far a successful initiative to commission the data management services at increasing scale, but also demonstrate the outcomes of R&D activities in a large-scale environment. The Data Challenge model should be retained. The program should be continued according to the established plan and expanded to more topics. Such a plan should be regularly reviewed to adapt to the potentially changing landscape.

[TECH-2] WLCG to develop/expand a structure for testing innovation, possibly based on the successful model of data challenges

WLCG would benefit from more coordination in the technical evolution area. In particular a technical coordination team should ensure a coherent process of modernization of services. This process should build a technical evolution roadmap spanning from short term goals to long term ambitions. A possible approach where a series of topical blueprints is produced and followed by a set of demonstrators proved to be effective in our community, and should be considered. While modernisation should be a continuous process, the challenges presented by HL-LHC deserve special attention. WLCG should consider producing a Technical Design Report based on the outcome of the demonstrators that will be more relevant for HL-LHC.

[TECH-3] WLCG to set up a technical coordination body (e.g. a Technical Coordination Board) to oversee evolution and innovation of services

[TECH-4] WLCG to produce a technical roadmap for the ramp-up to HL-LHC in time to be useful for the start of Run-4

Financial sustainability of WLCG services and infrastructure

WLCG continues to receive strong support from its Funding Agencies, with resources being provided according to a flat budget model. The improvements in hardware technology up to now, have ensured that more capacity - roughly 15% - could be committed each year to the experiments. In addition, more resources are available to the experiments beyond the pledged capacity, particularly in case of CPUs, either from additional resources at the same sites or via the utilization of external cloud and HPC centers, via opportunistic and managed patterns. The Funding Agencies have been also responsive in providing further resources when unforeseen or unprecedented circumstances have occurred (e.g. in 2016 with the unprecedented performance of the LHC; in 2023 with the uncertainties caused by the war in Ukraine). The processes established in the context of the C-RRB have ensured that pledged resources could be delivered in time for the start of the LHC run every year, even in years of very volatile markets. The newly established channel between WLCG, the LHCC and the CRSG offers now the opportunity to better understand how resource availability impacts the scientific program. There is generally a good dialog and a trust relationship between experiments, Funding Agencies, sites/federations and WLCG management.

Several financial challenges have been identified:

- HEP as a science is not expanding and we can expect the current yearly budget for hardware to remain available but not to increase significantly.
- In several federations, the same budget needs to cover some costs that are increasing over time (e.g. infrastructure and electricity).
- The hardware markets in the last five years have not been very favorable, which limits the increase in the amount of resources that can be purchased year-on-year.

To assure the financial sustainability of the WLCG infrastructure, several pillars have been identified:

- The evolution of hardware and market trends need to be properly monitored. Having a medium to long-term outlook of these trends (e.g. up to five years) will allow us to make decisions about the areas of development in which the WLCG community should invest.

WLCG should identify the proper structure to monitor these trends, globally and at national level, and ensure the effort is continued.

[FIN-1] WLCG to identify the proper structure to monitor hardware and market trends, globally and at national level

- Some of the Funding Agencies have the opportunity to profile the budget for computing over many years. WLCG should regularly produce a multi-year (e.g. 5 and 10 years) outlook of the resource needs in addition to the current yearly estimates through the C-RSG process. The multi-year predictions should be accurate enough for the Funding Agencies to profile the future spending within a few tens of percent of uncertainty.

[FIN-2] WLCG to develop and maintain multi-year resource planning

- WLCG collects the pledges of the Funding Agencies once per year, in time for the October C-RRB. The commitment of a Funding Agency to an experiment is normally proportional to the number of authors in the experiment from the institutes of that Funding Agency. This model is a guideline but WLCG is not entitled to enforce it by its MoU, while experiments can have more prescriptive MoUs with the Funding Agencies. WLCG however would benefit from a yearly review of the commitments per Funding Agency in proportion to the involvement into an experiment. This review would allow the WLCG management, the experiments and the CERN management to have a coherent discussion with the Funding Agencies about current commitments and future expectations. This kind of review was happening in the early days of WLCG but stopped during LHC Run-1.

[FIN-3] WLCG to monitor pledges against experiments authorship levels

- The commitments of the Funding Agencies are expected to be flat over the financial year (same level of resources from April to March of the following year). This prevents opportunities such as providing pledged capacity, and particularly CPUs at low cost for a limited number of months. Examples are opportunities from special HPC allocations or Cloud contracts. At the same time, the current model allows the experiments to predict the amount of resources that will be available, and to benefit from the required flexibility to use them. A model which allows bursty allocations to be provided, while not impacting the core business of the experiments, should be considered as complementary to the flat allocations.

[FIN-4] WLCG to enable finer-grain pledges with respect of current annually flat pledges, defining a model that can be validated by the experiments and efficiently serve their needs

- Considerable middleware development for the services in use in WLCG happens as a contribution of the institutes in the Collaboration. These contributions are not formally agreed nor acknowledged. In order to reduce the operational risk linked to these services a more formal process for collecting these contributions should be considered. It would also acknowledge and recognise the commitments of the institutes beyond providing hardware capacity, hopefully helping them to secure funding for development and support of the middleware. This process could start by simply collecting the information on existing efforts and reviewing it regularly. In a second phase, the commitments could be made more formal, e.g. through an MoU or a Collaboration Agreement.

[FIN-5] WLCG to more formally recognise the commitments (or list the contributions) of the Funding Agencies for middleware development and support

Heterogeneous Grid infrastructure

The role of the WLCG Tier-0, Tier-1s and Tier-2s

The organization of WLCG facilities as Tier-0, Tier-1s and Tier-2s offers a simple but effective way to define the expected quality-of-service and target metrics, specified in the WLCG MoU. The experiments can base the organization of the data and the execution of the workflows on these quality of services. The resources are pledged according to the tier hierarchy, providing a simple model for experiment and Funding Agencies to agree on the commitments.

The Tier-0 and Tier-1s constitute the core data archive and delivery infrastructure. While the network topology is today more flexible than originally anticipated, the connectivity across Tier-0 and Tier-1s is provided through dedicated links (the Optical Private Network - OPN). This guarantees enough capacity for the time-critical use cases. The vast majority of remaining links between Tier-1s and Tier-2s is provided by the LHCONE overlay network, ensuring enough separation of the traffic from other sciences in the Research and Education networks. The network infrastructure is one of the main assets in WLCG, enabling its core functions.

The simple organization of WLCG facilities in three tiers also has drawbacks and it is in fact partially obsolete. Sites exchange data for example as a full mesh rather than hierarchically. Within the same tier type, the capabilities that sites offer can be very different. The experiments would benefit from being exposed more directly to these capabilities. It would allow them to better shape the computing model and operations and take advantage of these capabilities. However, introducing a more complicated process to request resources with some capability, is not felt to be appropriate. WLCG could, however, regularly collect the information about site capabilities (e.g. memory per core, I/O, bandwidth from/to tape, size of buffers), for example as part of the C-RRB preparation, and expose them to the experiments. This would build the basis for a possible future model where resources are discussed not just in terms of volume but also capability.

[INFRA-1] WLCG to collect, organize and expose information about site capabilities (compute, storage, network)

Integration of Cloud resources in WLCG

The WLCG experiments started using resources provided directly through cloud interfaces more than 10 years ago. The focus has generally been on CPU rather than storage. There are also successful examples of integration of cloud storages into the experiment data management stack. Over the years, several models for the use of cloud-provided CPUs were demonstrated. The cloud resources could be integrated as an expansion of an existing facility (T1 or T2), or access could be granted through a gateway service (e.g. HEPCloud) or the cloud native interface was integrated directly with the experiment workload (and data) management system (e.g. the recent ATLAS project with a major commercial cloud provider). Some of the WLCG experiments also expose their HLT resources through a cloud interface to use the compute power of the online facility outside data-taking periods. The possibility to integrate commercial and academic/private resources with WLCG services allowed the experiments to benefit from considerable beyond-pledge capacity. The progress made in the WLCG benchmarking with the HEPSCORE deployment now allows comparison of commercial cloud resources with CPUs at WLCG sites.

The opportunities offered by cloud resources and the integration of cloud services into WLCG has never been pursued at an holistic level. More opportunities could be exploited in future by making progress in several areas:

- An increasing number of WLCG federations expressed the interest in committing resources through an academic cloud infrastructure. This would allow them to support a single multipurpose platform for many academic use cases including HEP. For this to happen, some technical progress should be made. At the least, it should be understood how these resources could be properly monitored and accounted for, as they would not expose Grid middleware and services. Some recipes for integrating cloud resources with different interfaces should also be defined and shared. This work would also be beneficial for integrating commercial cloud resources.

[INFRA-2] WLCG to identify and document the technical solutions to integrate cloud resources with WLCG

- The integration of commercial clouds in WLCG, or generally with the production and data management systems of the experiments, occurred in rather isolated cases. In various instances, an analysis of the costs was made and in recent activities, an attempt was made to compare them with the costs of running services at a WLCG site. The cost of cloud resources is very dependent on the boundary conditions, the contact with the cloud vendor and the workflows being executed. The connection with the Research and Education network is also a factor to consider. WLCG would benefit from a blueprint for the integration of cloud resources and an activity tracking the costs in different scenarios.

A recent blueprint was produced specifically for the US and it could be extended to cover other regions in WLCG. In addition, some scientific communities (e.g. LSST) with slightly different use cases than HEP moved to a model where cloud resources play a central role in providing storage and compute capabilities to the end users. While not completely applicable to HEP, WLCG should follow the progress and learnings of these communities and retain what is appropriate.

[INFRA-3] WLCG to produce a blueprint for the integration of cloud resources and launch activity tracking the costs in different scenarios

[INFRA-4] WLCG to establish channels allowing to follow the progress of other communities exploiting clouds

- There has been little use of cloud storage so far, while cloud storage interfaces have been integrated in the data management system of the experiments. While there is a very moderate risk in relying on cloud resources for data processing, the risks to consider when relying on cloud resources for storage - in particular when applied to Tier-1 sites (e.g. vendor lock-in) which are the storage backbone of WLCG - are larger. There is no such a thing as a WLCG cloud usage policy and the collaboration should consider defining one.

[INFRA-5] WLCG to define a cloud storage usage policy

- The benefits from a model where part of the pledged resources could be provided for a limited number of months in a year were described in the financial sustainability section above. They apply well to provisioning through commercial cloud resources.

Integration of HPC resources in WLCG

The HEP use cases have little need for High Performance Computing per se - there is little or no use of MPI applications. At the same time, the embarrassingly parallel nature of HEP workflows is such that these workflows can easily profit from cycles in a HPC machine. HPC resources have been integrated into WLCG for decades. In some cases (e.g. the NDGF Tier-1 and more recently the INFN Tier-1) part of the pledged capacity is provided through cycles from HPCs, completely integrated behind the Grid middleware. This allows the experiments to benefit from these resources transparently. In most cases however, the integration of the HPC system happens through a central gateway (e.g in the case of HEPCloud) or directly with the experiment workload management system (Alien, Condor GlideinWMS, Dirac, Panda). The HPC systems have provided a large amount of opportunistic resources to the experiments. In particular cases, and for limited periods of time, the scale of these resources has been comparable to the pledges in WLCG.

The HPC machines are very heterogeneous in many respects. The interfaces to the machines can vary and the policy to access the resources are very specific to the HPC center. Some are very permissive and similar to a WLCG site, some others are very restrictive and difficult to access in an efficient way. The CPU hardware deployed by the HPC centers is not always x86 (e.g. some centers deploy Power or ARM and in a close future Risc-V). In addition, a large processing capacity at HPC centers today is provided with GPUs. Using these architectures requires a considerable investment in software development for the experiments and the communities supporting the common libraries (see next section). In addition, the algorithms need to be regularly validated to ensure the physics results are not dependent on the architecture being used. At the same time, many of the funding agencies expect in future to commit part of the capacity in the form of cycles at HPC centers and are inviting WLCG to take the necessary steps to ensure these resources can be done efficiently. Today this is not the case and the HPC centers are generally used for a limited number of use cases and in an opportunistic fashion. In order to make progress, WLCG needs to work in several areas:

- At the technical level, there are solutions that are known to work, in order to integrate HPC systems with the experiments' workload management systems. These solutions are often experiment-specific, but several of the building blocks are common. WLCG should produce a blueprint for the integration of HPC systems and describe a reference implementation based on these building blocks, also to be used when handshaking with the management of these centers to explain our needs. It should also explain which policies at the HPC centers would simplify the integration work if implemented or relaxed, and which are a "no-go" for resource utilization. A recent blueprint was produced specifically for the US[2] and experiment specific documents on the use of HPCs also exist[3]. These documents can be extended to cover other experiments and/or regions in WLCG

[INFRA-6] WLCG to document existing solutions to integrate HPC centers and organize a knowledge base. When possible, propose one or more reference implementations via blueprint documents

- At the executive level, a stronger relationship needs to be established between WLCG and the HPC centers. Good progress was made at the national level through the relationship between some of the Tier-1 facilities and large HPC centers. Both in Europe and the US, WLCG needs to construct a dialogue with the Funding Agencies to drive the future allocation policies at HPC sites. Multi-year allocations for example would be highly beneficial and needed for proper planning in WLCG. Opportunistic access of HPC resources for beyond pledge use should also be better negotiated.

[INFRA-7] WLCG to construct a dialogue with the funding agencies and the relevant global bodies to drive the future allocation policies at HPC sites

[2] https://arxiv.org/abs/2304.07376

[3] https://cds.cern.ch/record/2707937/files/NOTE2020_003.pdf and https://cds.cern.ch/record/2707936/files/NOTE2020_002.pdf

- The strong relationship mentioned above should be leveraged at the time to discuss the architecture of future HPC centers. So far, such architectures would not favour data-intensive sciences, but this is about to change and HEP has the opportunity to contribute with its decades of experience in suggesting the most suitable future architectures. Initiatives such as ETP4HPC[4] and the EU funded project SPECTRUM[5] focus on this.

[INFRA-8] WLCG to leverage the relationship with the HPC centers and with the Funding Agencies to influence the architecture of future HPC centers also via interactions with the relevant international bodies

The three items mentioned above should be part of a HPC WLCG strategy that should be produced in the next year. This work should happen in synergy with other data-intensive sciences. In Europe, there is an opportunity to do so in the context of a recent Joint ECFA-NuPECC-APPEC (JENAA) initiative, in time for the next European Strategy for Particle Physics.

Several projects are being proposed in EU, USA and other contexts to sketch future directions in the collaboration between the HPC ecosystem and scientific domains. WLCG should encourage and foster the participation of its members to such projects, in order to guarantee the visibility and the coverage of its present and future needs.

[INFRA-9] WLCG to monitor the projects on national and global scale designing collaboration strategies for the collaboration with HPCs, and foster the participation of its members

Analysis Facilities

The discussion about specialized facilities for end-user analysis has been taking place under the umbrella of the HEP Software Foundation (HFS) for several years. Some large projects, for example IRIS-HEP, developed an ecosystem for interactive analysis and the functionalities of this ecosystem are being demonstrated through a set of challenges. The ROOT framework also is investing effort enabling the use of modern technologies for analysis. WLCG established links with these initiatives in the context of data management through the WLCG DOMA activity. This relationship needs however to be strengthened as the future development of the analysis facilities concept will likely have an impact on WLCG; we expect many of these facilities to be co-hosted with a WLCG site and share resources and services. The role of specialized facilities for Machine Learning, for example, have been only superficially discussed in WLCG and the resources for these activities have been provided opportunistically by some of the sites. In future, however, we expect the needs for ML/AI to increase both in terms of volume of resources but also variety of services. WLCG should therefore explore how such a facility could be hosted

[4] https://www.etp4hpc.eu/pujades/files/ETP4HPC%20SRA%203%20-%20Single%20Page%20-%201.pdf
[5] The project will start in Jan 2024 and will deliver a Strategic Research, Innovation and Deployment Agenda (SRIDA) which defines the innovation and deployment roadmap for data-intensive science and infrastructures.

synergically with WLCG services. At the same time, we are aware that such facilities are being designed and will be operated in many national environments, and WLCG should be able to profit from their processing capabilities.

[INFRA-9] WLCG to explore how facilities supporting future analysis models could be hosted synergically with WLCG services

General areas of work towards a more heterogeneous Grid infrastructure

- Progress should be made in several areas of the LHC offline software (common libraries and experiment specific software) to be able to leverage the heterogeneous architectures offered by the HPC systems. Tools and solutions for heterogeneous computing are still an evolving ecosystem, with projects being started and retired. A technology tracking and evaluation (at the level of WLCG and beyond) would benefit the experiments by sharing experience and findings. From the infrastructure side, WLCG needs to prepare for heterogeneous compute architectures (e.g. GPUs) being provided to complement the x86 and non-x86 CPU hardware. This has needed to be started in the context of HPC facilities but it applies to all kinds of facilities, and will soon apply very generally. While The main challenge is at the level of offline software a progressive strategy needs to be put in place also at the infrastructure level. Measuring the use of CPU and GPU resources in hybrid systems should be the first step. This will require progress in the benchmarking area but also in the accounting tools and services. This work should be prioritized as in future these systems will likely be part of the pledge.

[INFRA-10] WLCG to prepare for heterogeneous compute architectures: facilitate the development of the offline software and progress in the area of benchmarking and accounting

- WLCG needs to modernize services and protocols to benefit from modern standards that are less and less HEP specific, while retaining the advantages of the specific services developed by our community tailored to our use cases. WLCG already started moving successfully to less HEP specific protocols, such as HTTP/WebDAV for data management. The adoption of new standards implies decommissioning some of the existing protocols to ensure maintainability. WLCG should build a (better) process for this as legacy services and protocols tend to be difficult to eliminate in our infrastructure, and this increases the operational burden.

  More progress should be made evaluating and possibly integrating widely adopted open source products as building blocks of WLCG services. An example is Kubernetes, for which an initiative started in WLCG but did not progress to the level one would have hoped. The introduction of more external dependencies into the WLCG service stack obviously introduces potential risks. The risks should be properly identified, evaluated, cataloged and managed - see the following section about a risk catalog. Existing initiatives at WLCG sites were already established to evaluate the risks from external software dependencies. The finding of these initiatives should be reused rather than launching a WLCG-specific process.

[INFRA-11] WLCG to establish a process for adopting modern, non-HEP specific standards where appropriate and decommission legacy services and protocols. The process should include risk management for external dependencies

- WLCG should carefully follow the process of evolving facilities currently happening at national level. Ideally this process could reduce the cost in operating services and create opportunities for service managers to engage in new initiatives and/or contribute supporting the services while the hardware is located at other facilities. For example a datalake-based solution has been proposed as a possible blueprint for the data management of HL-LHC data and processing. WLCG should monitor closely its design, implementation and testing, via the DOMA initiative and the expected data challenges, in order to validate it as a viable solution well before the start of Run-4. In general, the process towards consolidation needs to be adiabatic as the professional profile of the service managers will change and we need to allow for enough time for this to happen. The process also needs a proper narrative to avoid demoting the current experts or passing a message to the Funding Agencies that less effort is required.

[INFRA-12] WLCG to follow and accommodate the national plans in terms of consolidating facilities, particularly when aimed at reducing complexity. Monitor closely the design, implementation, testing and validation of the datalake and its components as solutions for the HL-LHC era. Engage the service managers in the transition and retain expertise

Relationship with other HEP and non-HEP communities

WLCG today is the infrastructure to support the needs of the LHC experiments and this should remain its core mission. At the same time, it is strategically important for WLCG to ensure its scope is not too narrow around the core mission, in order to guarantee sustainability and continued support. WLCG should position itself as a science-driven organization that provides services to HEP, but creates an ecosystem of tools, services and capabilities useful to many other sciences. This will motivate the Funding Agencies to continue investing in WLCG as there will be a return of investment for other sciences. Depending on the sciences in question, and how close they are to the core mission of WLCG, the level of engagement will differ.

The WLCG Funding Agencies also support other experiments in HEP, NP and in some cases astronomy. Many of these large scale experiments are organized in global collaborations motivating a distributed computing model and this will be more the case for the future. There is an interest in HENP in not duplicating the infrastructure for scientific computing but rather using a common set of tools and services, with the flexibility to complement them with experiment-specific ones. WLCG is seen as a model of scientific computing infrastructure, and many HENP experiments express willingness to collaborate with WLCG and adopt its services or a large set of them. WLCG sees this as an opportunity to enhance its sustainability by building a larger pool of contributors and not diverting effort of the community into supporting many solutions. The vision to enlarge the WLCG infrastructure to serve the needs of a larger number of HENP projects and also other sciences was presented at the ESPP in 2019 and at Snowmass in 2022. Different possibilities were discussed in the Collaboration concerning the governance of such an infrastructure. The governance model that has been implemented is very lightweight: experiments can become WLCG observers by presenting their use cases to the WLCG Management Board and motivating the interest to join the Collaboration. Today DUNE, Belle-2, JUNO and VIRGO are WLCG observers and participate in the technical evolution of WLCG and its operations. They are invited to the Management Board and the Grid Deployment Board. The current partnership model works well and there is no perceived need to change it substantially or strengthen the governance. Possible improvements are:

- The process should be more structured and the endorsement should be done probably at a higher level than the WLCG MB (e.g. at the WLCG OB or even via CERN level MoUs).
- The criteria for becoming an observer - or partner - should also be better clarified: there has to be an interest of the WLCG institutes in supporting an experiment as WLCG partner and there has to be a large enough overlap between the WLCG services and what the partner would like to use.
- The resource needs of the partner communities are expected to be small compared to the LHC experiments. The impact on resources should also be evaluated and monitored as part of the process. This is particularly true for networks especially if the connectivity would be provided by the LHCONE.

[REL-1] WLCG to better formalize the process to become “partner” (a.k.a. “observer”): ratification moved to a higher body than the MB, clarification of the criteria, evaluation of the resource impact

Other sciences, particularly Astronomy and Astroparticle Physics, also are interested in using several of the WLCG technologies and services. Some sharing of the infrastructure can be expected but the engagement of these communities with WLCG is more at the level of technologies. It is important that WLCG continues collaborating with these communities and leveraging initiatives such as JENA and ESCAPE in Europe, the DOE programs in the US, and partnering with OSG and EGI (see next section).

Relationship between WLCG and OSG, EGI and other (inter)national grid initiatives

The Open Science Grid in the US and EGI in Europe have been WLCG partners since they were created. Both OSG and EGI provide federating services in use by WLCG and other sciences. The relationship between WLCG and EGI and the mutual benefits are described in a whitepaper[6]. The main areas of collaboration between the three entities are:

- The coordination of cybersecurity across the WLCG infrastructure (policies, threat intelligence and incident response). This collaboration is essential to ensure the proper level of security in a complex distributed ecosystem such as WLCG where local policies have precedence over the global ones.
- The provisioning of federating services such as Accounting, User Support, Information System. There is no WLCG common funding that would normally support the provisioning of these services. The mechanisms offered by EGI and OSG provide a solution. It is important therefore that this collaboration continues to ensure transparent access and use of the services for the communities.
- The support of middleware distributions: packaging, certification and deployment. Both OSG and EGI contribute to the deployment of services in the WLCG infrastructure. In the case of OSG, it coordinates the deployment activity across US sites. This is partially true for EGI as well.
- Facilitate the dialog with middleware providers. For example OSG is the main point of contact between WLCG and the HTCondor team.
- Support innovation of services across the infrastructure. EGI has been very successful in engaging with the European Commission and attracting funding for the benefit of its members. OSG is heavily involved in R&D initiatives such as IRIS-HEP in the US and receives funding to introduce modern and disruptive technologies in the infrastructure. WLCG has enjoyed considerable benefits from these funding streams.

There is a strategic interest for WLCG to continue its relationship with OSG and EGI. Both have been successful engaging the long tail of science and enabling them to access resources with a moderate investment of effort. WLCG should continue to leverage the collaboration with OSG and EGI to ensure that its service ecosystem is useful also for other sciences that are not its main partners.

At the same time, the interaction and interdependency between EGI, OSG and WLCG services should be reviewed regularly. More coordination between the three parties would be beneficial to build a common roadmap and ensure the sustainability of the whole ecosystem. The WLCG management should have a leading role organizing this process.

The dependency of WLCG on other partners also introduces risks that need to be evaluated regularly and mitigated. The sustainability model of OSG and EGI are rather different and WLCG should ensure operations are secured also in the case one of the partner initiatives faces funding issues. The section about financial sustainability suggests strengthening the commitments of the Funding Agencies and institutes for software development. The same should happen here for federating services provisioning, to mitigate the above mentioned risks.

The benefits highlighted above apply to most countries in WLCG. A limited number of countries however see less advantages in being part of a federated grid infrastructure such as EGI, but

[6] https://zenodo.org/records/4717751

still are willing to provide resources to WLCG experiments. These countries must be able to join the collaboration and provide resources, and it should be a priority of WLCG to negotiate particularly with EGI the conditions for this to happen.

[REL-2] WLCG to take a leading role reviewing regularly the interaction and interdependency with OSG and EGI and strengthening coordination across the three parties

[REL-3] WLCG to ensure that operations are secured in case one of the partner initiative faces funding issues

[REL-4] WLCG to promote the benefits in being part of EGI (or OSG) while ensuring that ultimately countries can join the collaboration and provide resources even if not part of EGI (or OSG)

Other actors are appearing in the landscape of (scientific) computing:

- EuroHPC Joint Undertaking has taken a driving role for the funding and operations of HPC centers in Europe, in great part substituting PRACE;
- The European Open Science Cloud (EOSC) wants to coordinate and organize national FAIR data infrastructures, and allow for interoperability and service management via a Core of federating services and a larger Marketplace of thematic and general-purpose services;
- GAIA-X wants to develop a digital governance that can be applied to any existing cloud/ edge technology stack to obtain transparency, controllability, portability and interoperability across data and services, for the benefits of the productive system.

[REL-5] WLCG to look for synergies with the national and international initiatives developing large scale computing infrastructures

(Offline) software and middleware development: relationship between WLCG and the development communities

The functioning of WLCG relies on middleware services connecting the high level applications with the core services - e.g. compute, storage - at the sites. The high level applications are normally developed by the experiments and implement experiment specific policies and functionalities. The middleware development in the past benefitted from funding from European and US sources, aimed at providing general services for a large number of sciences. These sources of funding partially drained away, particularly in Europe. Today a lot of the middleware development is supported by institutes and teams in HEP or with strong links to HEP. The great commitment and success of these teams supporting middleware development for the WLCG Collaboration and other sciences, is notable. The challenge is how to make this sustainable.

- Distributed computing is nowadays largely adopted by many communities. There are many large open source projects providing solutions that could and are being adopted by WLCG. The technical coordination team proposed above in this document should be in charge of following the technology trends, understand which solutions would be useful for WLCG and define an innovation plan to evaluate these solutions, measure the benefits and possibly introduce them in production. At the same time, the risks should be evaluated and properly taken into account. This modernisation process would make the WLCG service portfolio more appealing and usable for other sciences and it would therefore be a step forward also in the ambition to provide a musti-science infrastructure.

- WLCG should work with the partners providing today's middleware solutions and help them secure the funding to continue this work. This can happen in different ways and could be specific for each partner. The strategic directions proposed in the “Financial Sustainability” and “Impact of WLCG on science and society” sections of this document should however apply in general and would be a good starting point.

The offline software - e.g. event generation, simulation, reconstruction, analysis - and its evolution is a key ingredient for the success of the LHC program. A large amount of the offline software is experiment-specific. The development is a responsibility of the experiments and is supported by the institutes in the experiments. Another considerable amount of the offline software is more common and serves all experiments. This is the case for example for Geant4, ROOT, and the event generators by the theoretical community. The common software libraries evolve in line with the needs of the experiments and in collaboration with the experiments. The communities supporting the common software libraries are very heterogeneous. Some have solid grounding in the major WLCG institutes and labs, while others (e.g. the event generators) do not have a focus point. Progress in the area of offline software has a large impact on WLCG and is one of the main factors impacting the future resource needs of the experiments, both in volume and type. The experiments and the communities supporting the common software libraries have an excellent track record of software improvements that allowed a better use of the resources in the WLCG infrastructure. WLCG has a strategic interest in facilitating this process and ensuring the process is properly organized.

[SOFT-1] WLCG to continue encouraging the necessary software improvements to make the best use of the infrastructure, while not being in charge of such a development.

The HSF was established to support and organize this process. The HSF made very good progress in many areas, recognised by our review bodies. The loose structure of the HSF, with minimal governance, supports the community building and creates an excellent forum for discussion. It does not offer the tools for a rapid decision making process, which happen instead through natural selection. While exploring different avenues is important, a trade-off between fulfilling the needs of the community and not losing focus is needed. The natural selection needs to come primarily from the experiments, with input from WLCG as a whole.

[SOFT-2] The role and structure of the HSF needs to be re-discussed based on recent years of experience. WLCG to facilitate this process

Several national and international software projects today support the needs of the WLCG experiments. There is little coordination across those projects as the funding structure is different for each one of them, and HEP is not always the only focus. Forming a coordinating body on top of these projects is unrealistic. WLCG could create however a mechanism for which PIs of current large projects or influential visionaries for new projects can discuss with the experiments and generally WLCG on the main strategic directions. The focus should be on the LHC needs, while the outcome should also serve the other communities. WLCG should ensure that the core goals of its community are well-understood and can influence the future directions. The focus should be on the HL-LHC program but the vision on new projects should also consider how to bring early benefits.

[SOFT-3] WLCG to consider creating a mechanism for which PIs of current large projects or influential visionaries for new projects can discuss with the experiments and generally WLCG on the main strategic directions.

Engagement and Governance

The success of the WLCG Collaboration is based on building consensus between the parties: experiments, sites, middleware and services providers, national and international scientific computing initiatives. This consensus allows the implementation of an infrastructure of loosely-coupled facilities offering the capabilities that the experiments need. The consensus building in WLCG has succeeded very well so far. The WLCG governance structure needs to support this decision making process. The governance needs also to evolve together with the Collaboration and its level of maturity, while the goal of building consensus needs to remain a core objective. The WLCG governance should be reviewed with these principles in mind. Areas to consider:

- The WLCG Project Leader is today 1) the lead of the WLCG Collaboration and 2) the leader of the CERN LCG project. The two roles are not the same but 2) is there to ensure that CERN’s commitments to WLCG are fulfilled, particularly with respect to the support of the collaboration as a whole, in addition to the T0 capabilities. Today the two roles are coupled and it is useful to have them coupled.
- There is an overlap of functions between the Grid Deployment Board and the WLCG Operations Coordination activity. There are fewer discussions about deployment than in the past as the infrastructure reached a considerable level of maturity. At the same time there is no Technical Coordination Board to organise the discussion about the evolution of WLCG services. WLCG should consider re-scoping the GDB into a Technical Coordination and Deployment board, with a different mandate and an extended participation.
- The Management Board is in charge of the execution of the WLCG functions. It is de facto a very open meeting - open attendance, open material - and that helps the consensus building. It also limits the level of discussion on specific topics. WLCG should

define the proper structure to discuss some topics in a more restricted environment, while keeping the current benefits of an open Management Board meeting.

- WLCG has demonstrated its capability to tactically address the computing needs of the experiments. There is no structured process however to define and endorse the more strategic aspects. This role is today partially covered by the Management Board but it is not in its mandate and does not have the right representation from the Collaboration. The Collaboration Board would be the natural environment where items concerning the WLCG strategic directions are discussed, engaging experiments and sites. The strategic directions would be presented periodically to the Overview Board for endorsement. Several of the strategic items discussed above in this document should be a matter of discussion at the Collaboration Board. Examples are the cases of federations not committing to WLCG as expected, discussions about new policies for pledging resources, policies for integrating non-grid facilities into WLCG (eg.clouds and HPCs), aspects related to membership in the collaboration, relationship with partner infrastructures.

[GOV-1] WLCG to review its governance to better support this strategy, while preserving the "principle of consensus". Areas to consider include project leadership, OB, MB, GDB

Environmental sustainability: energy efficiency, green computing, carbon footprint

Environmental sustainability is a strategic area of priority for many of the WLCG Funding Agencies. WLCG started a process to estimate the energy consumption needed by the computing equipment for the LHC experiments. A study was produced in 2023 and presented at the CHEP2023 conference. Such a study analyzed the main factors impacting the energy needs of WLCG and predicted the trend for the remainder of the LHC programme. GridPP also studied in detail the energy efficiency of different CPU architectures and the benefits for the experiments and sites if the software could be ported to these architectures. Providing information, opportunity and incentive to help collaborators reduce the carbon footprint of computing needs should be a priority WLCG as it is a priority of its Funding Agencies and a demonstration of social responsibility. WLCG should elaborate a plan that covers the area of software, computing models, facilities and hardware lifecycle. The progress needs to be regularly measured and exposed and the proper metrics need to be defined. Particularly In the area of improving energy efficiency of the facilities, WLCG cannot mandate a schedule to the countries but should collect information and track the progress.

[ENV-1] WLCG to agree metrics and provide a framework to collect information related to energy efficiency

[ENV-2] WLCG to enable the use of more energy-efficient hardware where possible

[ENV-3] WLCG to develop and promote a sustainability plan to improve energy efficiency, covering software, computing models, facilities, and hardware technology and lifecycle.

Identify and review Risks and Mitigations

Several sections of this document highlight possible risks for WLCG and suggest defining mitigations. WLCG went through the process of identifying the main risks in 2020 and analysing them. The risks were assessed in terms of impact and likelihood, to define their severity. The risk assessment and the possible mitigations were cataloged in a risk register that has been periodically reviewed at the WLCG Overview Board. The structure of WLCG does not allow a formal risk management process due to the loose relationship between the partners. Somehow CERN accepts the risk that computing is organized through WLCG rather than being uniquely a host lab responsibility. WLCG should continue reviewing the risk register at least once per year and present it to the Overview Board.

[RISK-1] WLCG to continue reviewing the risk register at least once per year and present it to the Overview Board for endorsement.

Impact of WLCG on science and society, training and outreach

The role of WLCG in scientific computing allowed various countries to develop a research and education cyberinfrastructure for the benefit of other sciences. The most clear example is the network infrastructure that WLCG had to develop to support increasingly challenging needs. There are still regions in WLCG where the network is not as good as one would hope but the network consortia and NRENs supporting WLCG are expanding on them. WLCG has not always been effective at communicating its value and the benefits of a new region or site joining the collaboration. WLCG does not really have a communication plan or a communication officer. Because of that, many successes of the collaboration go unnoticed.

[IMPACT-1] WLCG to establish a proper structure for communication and develop a communication strategy that is reviewed periodically by the collaboration. Form a communication team.

It is relatively easy today for a site to get into WLCG from the technical perspective as long as the site is supported by a structure in the same country or region. In some countries however, the cyberinfrastructure for research and education is less developed. There is no central WLCG team supporting the installation and configuration of a new site. The experiments have an interest to support the academic environment in these countries but are not resources to support the installation of WLCG services. WLCG needs a strategy for supporting new countries willing to volunteer with limited resources. While there might be a limited return of investment from the pure hardware capacity perspective, broadening the scope to new countries should be seen as part of the social mission of WLCG and CERN to support science by enlarging their horizons. For this pursuit, WLCG should aim at reviving some of the support programs existing in the past, leveraging also national and international partnering initiatives. For example OSG and EGI conferences are a good opportunity for training service managers on established and new technologies. The same could be considered at HEPIX meetings. Evolving WLCG services and using less HEP-specific protocols, tools and services would facilitate the process and is therefore of major importance.

[IMPACT-2] WLCG to define a strategy for supporting new countries willing to volunteer with limited resources.

Properly supporting adequate training programs is also important in the area of offline software. Some initiatives in the scope of the HSF training program are very popular and received very good reviews from the WLCG scientific referees. They complement the experiment-specific programs well. What is missing today is more connection to the universities and laboratories to support proper careers as scientific software developers. The Software Institute for Data Intensive Sciences (SIDIS)[7] had the ambition to facilitate this process. ECFA is establishing an academic program for detector technologies and could consider a similar initiative for software development.

[IMPACT-3] WLCG to continue pursuing the objective to establish proper career opportunities for scientific software development, creating synergies with other sciences and leveraging existing organizations such as ECFA and national training programs

[IMPACT-4] WLCG to review the format of the WLCG Workshop, offering the opportunity for the community to present new ideas, for example through lightning talks and poster sessions.

[IMPACT-5] WLCG to maintain a list of job opportunities at institutes willing to advertise the openings. This will hopefully allow us to retain expertise in the community and attract new ones.

.

[7] https://sidis.web.cern.ch/