Evolution of Scientific Computing in the next decade: HEP and beyond

Ian Bird, Simone Campana – CERN

Introduction

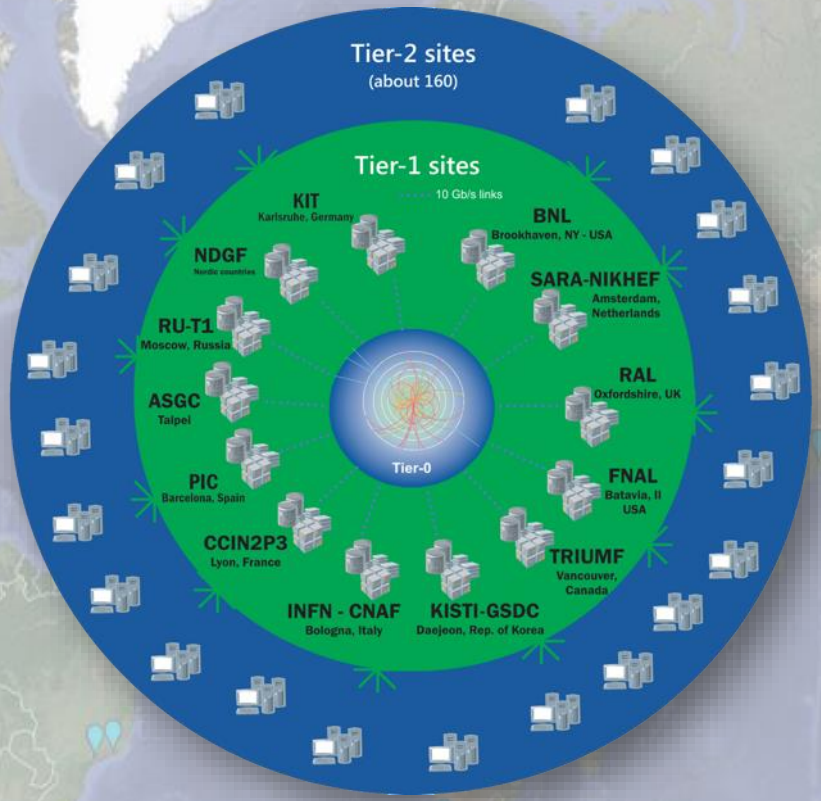
- Those ideas were presented and refined in several stages
  - The 1st scientific computing forum: https://indico.cern.ch/event/581096/
  - The November 2010 WLCG Overview Board: https://indico.cern.ch/event/734889/
- The outcome was submitted as whitepaper to the European Strategy for Particle Physics Open Symposium, on behalf of the WLCG OB
  - http://wlcg-docs.web.cern.ch/wlcg-docs/technical_documents/HEP-Computing-Evolution.pdf
- The proposal has been informally discussed with HEP projects other than LHC (read DUNE)

The Worldwide LHC Computing Grid

Tier-0 (CERN and Hungary): data recording, reconstruction and distribution

Tier-1: permanent storage, re-processing, analysis

Tier-2: Simulation, end-user analysis



~170 sites, 42 countries

~1M CPU cores

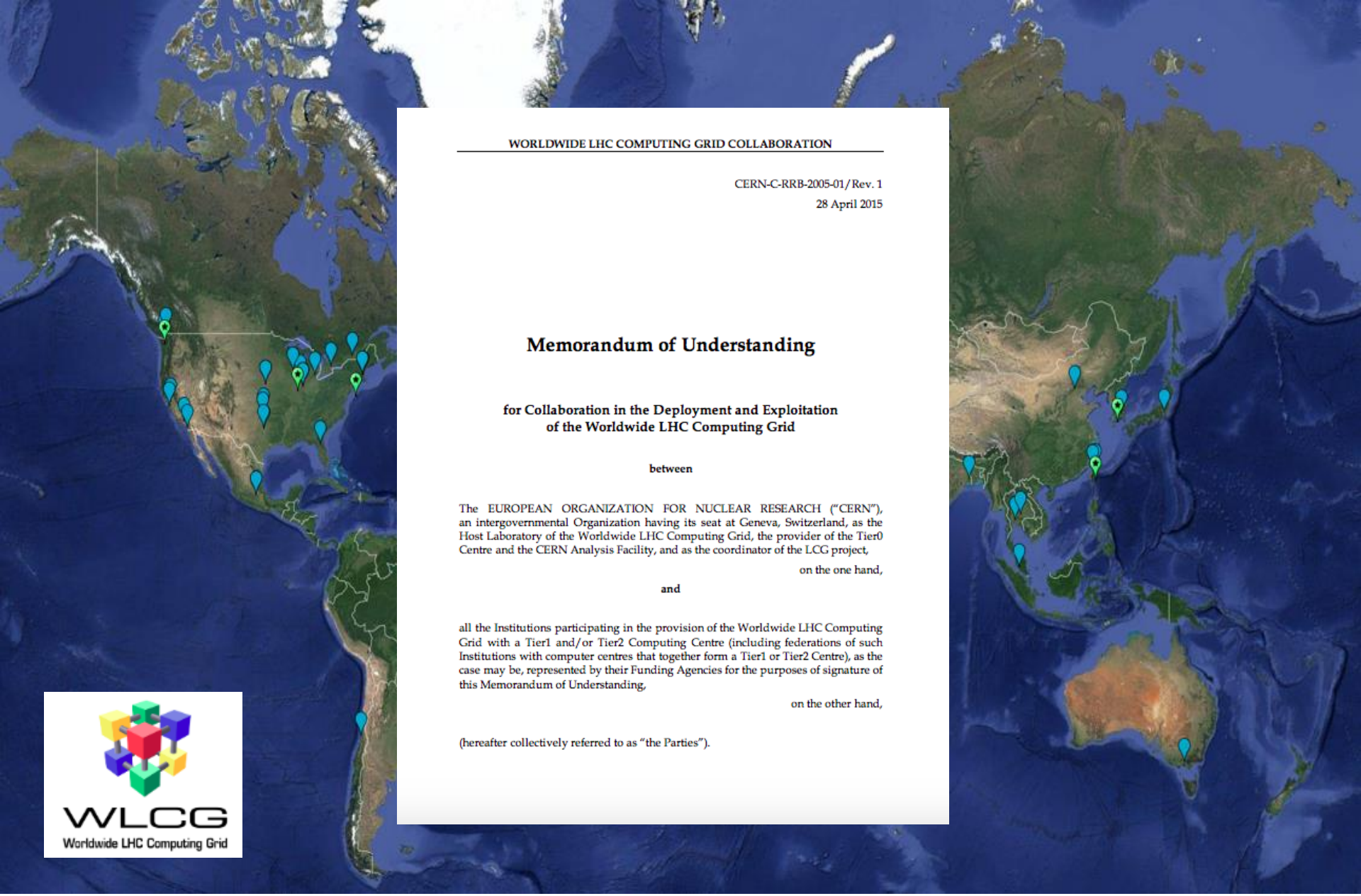
~1 EB of storage

> 2 million jobs/day

10-100 Gb links

WLCG: an International collaboration to distribute and analyse LHC data

Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists



WORLDWIDE LHC COMPUTING GRID COLLABORATION

CERN-C-RRB-2005-01/Rev. 1
28 April 2015

Memorandum of Understanding

for Collaboration in the Deployment and Exploitation of the Worldwide LHC Computing Grid

between

The EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH ("CERN"), an intergovernmental Organization having its seat at Geneva, Switzerland, as the Host Laboratory of the Worldwide LHC Computing Grid, the provider of the Tier0 Centre and the CERN Analysis Facility, and as the coordinator of the LCG project,

on the one hand,

and

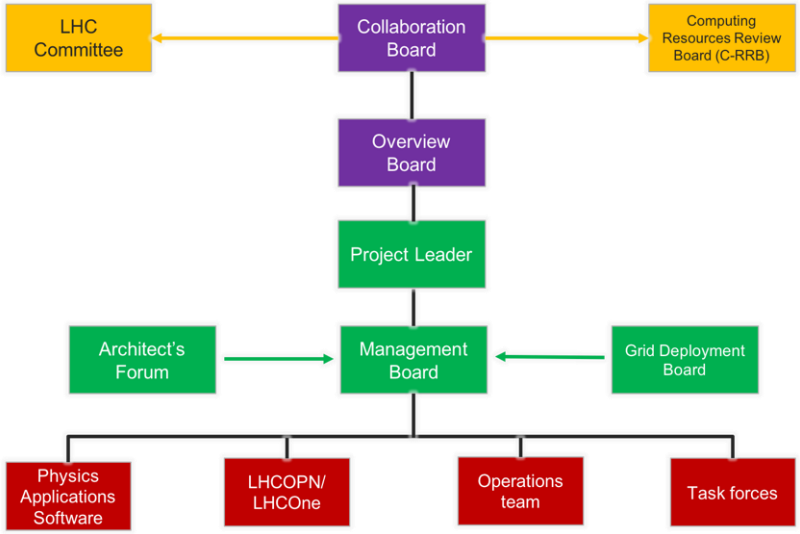
all the Institutions participating in the provision of the Worldwide LHC Computing Grid with a Tier1 and/or Tier2 Computing Centre (including federations of such Institutions with computer centres that together form a Tier1 or Tier2 Centre), as the case may be, represented by their Funding Agencies for the purposes of signature of this Memorandum of Understanding,

on the other hand,

(hereafter collectively referred to as "the Parties").

WLCG
Worldwide LHC Computing Grid

Governance of the WLCG collaboration (as in Annex 5 of the MoU)



- Collaboration Board: general assembly of WLCG members
- Overview Board: standing body of the CB. Strategic directions.
- Project Leader: executive director
- Management Board: organizes the program of work, ensure the execution of the CB/OB directions
- Grid Deployment Board: infrastructure planning and deployment
- Architects Forum: software planning and evolution

S.W.O.T analysis

Strengths

- WLCG established, trust network among stakeholders. Reputation of a success story.
- Lightweight decision making processes
- Consolidated infrastructure – common building blocks, policies, federation, resource sharing, etc.

Weaknesses

- Lack of investment in software for the future; initial investment in WLCG applications has levelled off; future need more weight on software
- WLCG does not have development effort – reliance on external providers – some developments and efforts were of marginal use

Threats

- Flat funding but expectation of more and more computing and data
- Other sciences now compete with WLCG for resources at similar scales; WLCG may lose influence
- Need to use facilities not particularly suited to our problems – HPC
- No financial support for software nor career recognition

Opportunities

- Leverage experience and be central to evolution of scientific computing and become the reference, for the benefit other sciences, resource providers, funding agencies
- Leverage extra resources (HPCs) by investing in software and building relationships
- Retain significant leadership role in setting direction

Scientific Computing Evolution: the strategy

1) Leverage the existing HEP computing infrastructure and evolve it to serve as a common computing system for HEP and beyond HEP

2) Evolve the facilities and services to build a HEP Data Cloud

3) Invest in common software and software techniques including training, dissemination and recognition

For this to happen, the current governance of the WLCG organization should be changed to factor out the infrastructure related aspects

Leverage the existing HEP computing infrastructure and evolve it to serve as a common computing system for HEP and beyond HEP

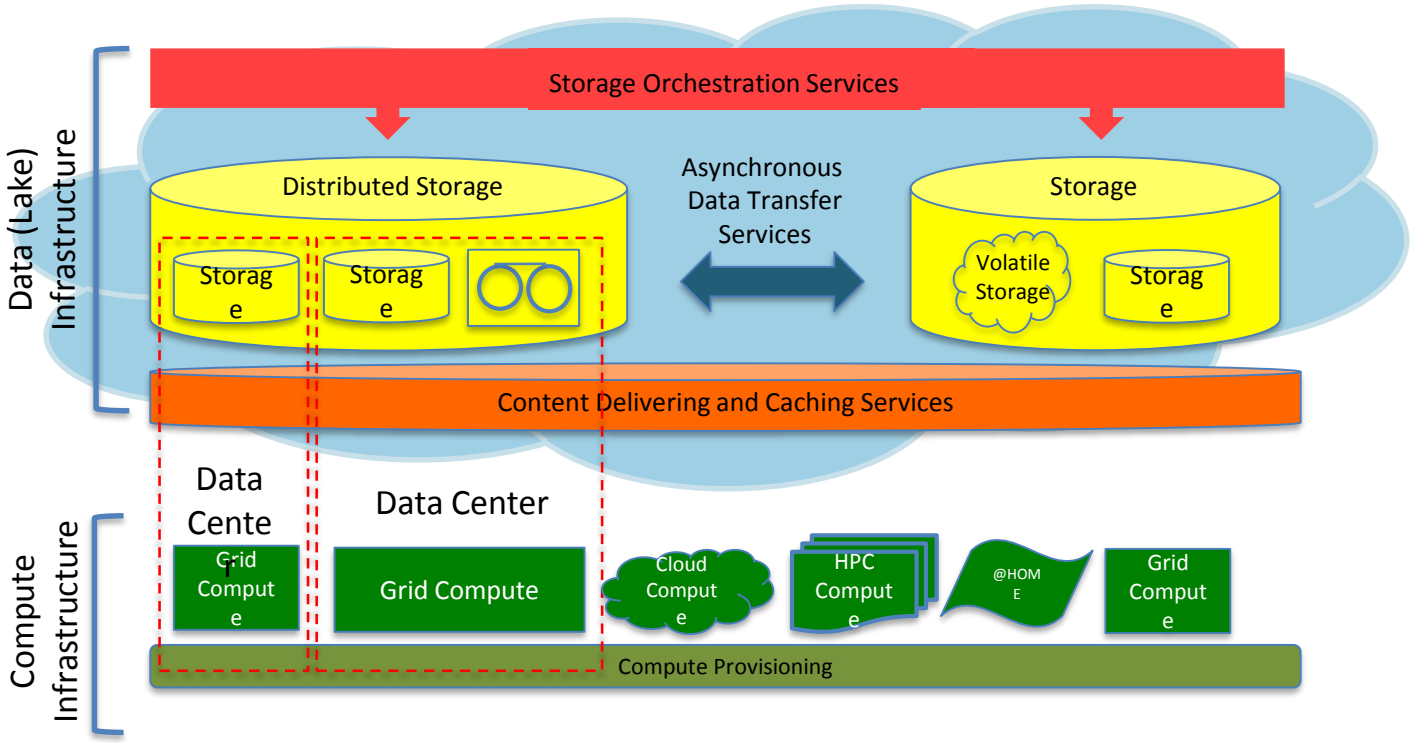
- Expose the well established WLCG services to other HEP and non HEP sciences into a marketplace. Different projects will flexibly adopt different service sets

- Complement such marketplace with services developed and supported by non WLCG sciences, for the benefit of all

- Consolidate the infrastructure around the current **core services** at WLCG facilities. Strongly encourage commonality across sciences.
  - Authentication, Authorization and Identity management. Security
  - Networks, Storage and Data Management
  - Operations

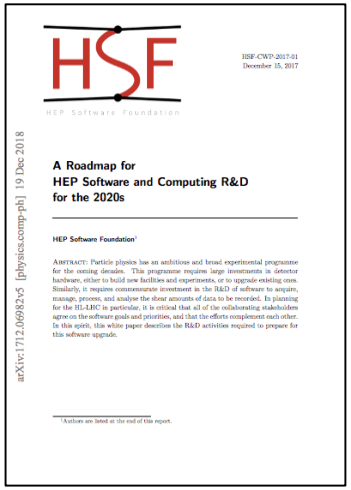
Favor, without imposing, sharing of infrastructure, tools, service and support as far as possible.

Evolve the facilities and services to build a HEP Data Cloud



Invest in common software and software techniques including training, dissemination and recognition

- Complement the project specific effort with common effort: tools and procedures, training, actual libraries
- The HSF: a bottom up approach collecting a set of tools contributed, developed, maintained and evolved by the community.


HSF

HEP Software Foundation

HSF-CWP-2017-01
December 15, 2017

A Roadmap for
HEP Software and Computing R&D
for the 2020s

HEP Software Foundation[1]

Abstract: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

[1]Authors are listed at the end of this report.

arXiv:1712.06982v5 [physics.comp-ph] 19 Dec 2018

Develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the HL-LHC and other planned HEP experiments of the 2020's.

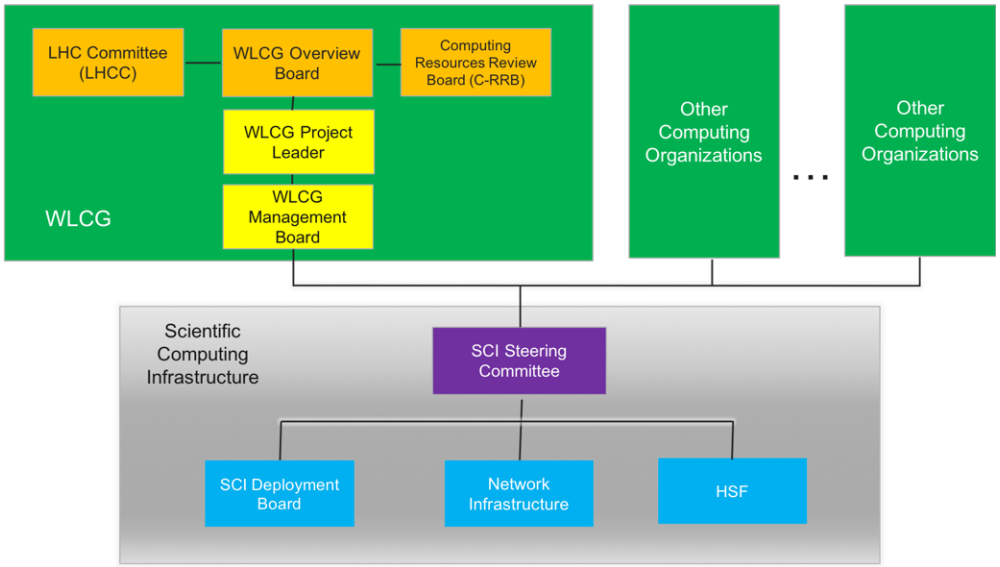
25M USD

e.g.

Governance, Organization and Steering

- Governance must remain **science-led** and **lightweight**

- The Scientific Computing Infrastructure gets established, factoring out the aspects common to all organizations, as infrastructure and software
- WLCG collaboration retains the LHC specific aspects
- Other organizations implement their own equivalent



The Scientific Computing Infrastructure

Driven by the sciences with a stake on the common infrastructure, as part of the **Scientific Collaboration Steering Committee**

- The steering committee defines/steers direction, encourages funding, brokers needs on licensing, policies, joint procurements
- Members of the committee:
  - The heads of Information and Communication Technology of the major HEP laboratories being part of the Infrastructure
  - The computing project leaders from the major projects and experiments
- The committee reaches decisions by consensus, while a voting procedure would formally exist. It reports to the involved sciences through their representatives

The Scientific Computing Infrastructure

The Scientific Steering Committee overseas the activities of and provide recommendations to different bodies:

- The SCI deployment board, replacing the WLCG Grid deployment board
  - Broadened to all parties of the SCI
  - Covering all integration and deployment aspect of services
  - Evolving in a more forward looking body in terms of computing technologies
- The HSF would be the vehicle by which common software tools, libraries and techniques is fostered across the SCI community

Risks and Mitigation

Two main risks were identified in the proposed model and possible mitigations were considered

1) Risk of loss of control

- Today, the tight coupling of the WLCG organization with the infrastructure ensures control
- In the new model the infrastructure and WLCG are loosely coupled and more projects compete with WLCG
- Minimized building up the SCI steering committee on the backbone of the current WLCG OB, bringing on board the new key stakeholders (sciences)

2) Risk of dominant role

- Other organizations than WLCG would refuse a model imposed upon them
- Minimized by presenting the SCI as a marketplace of tools/services rather than a top-down monolithic system

Initial steps

- Use DUNE as a first example
- Invite DUNE to GDB (~open anyway), MB as observer/associate (tbd)
  - Will need agreement of MB members
- Expect DUNE to set up their own computing organization and resource pledge/management process
- Evolve GDB towards the “steering committee” for the infrastructure
  - Will revitalise its role
- MB should eventually remain as a WLCG body but the steering committee should eventually take the decisions of the infrastructure/operation coordination
- Can initially do this informally – need more formality once DUNE moves towards production
- Assess/review how it works