



Minutes of the 1st Scientific Computing Forum (CERN, Geneva, 15 February 2017)

Present:

U. Bassler (CNRS/IN2P3, France)
I. Bird (CERN, IT Department)
D. Britton (University of Glasgow, UK)
P. Buncic (representing ALICE)
J. Butler (representing CMS)
E. Elsen (CERN, Director for Research and Scientific Computing, Chair)
S. Foffano (CERN, IT Department, Scientific Secretary)
F. Hemmer (CERN, IT Department)
K. Jacobs (representing ATLAS)
B. Jones (CERN, IT Department)
P. Levai (Wigner RCP, Hungary)
A. McNab (representing LHCb)
J. Templon (NIKHEF, Netherlands)
P. Wells (CERN, International Relations)
A. Zoccoli (INFN, Italy)

Remote:

G. Chen (Chinese Academy of Sciences, China)
M. Delfino (Universitat Autònoma de Barcelona, Spain)
C. Grab (ETH Zurich, Switzerland)
V. Guelzow (DESY, Germany)
A. Heiss (KIT, Germany)
E. Lancon (BNL, USA)
J. P. Meyer (CEA Saclay, France)
B. Roser (FNAL, USA)
E. Ryabinkin (NRC Kurchatov Institute, Russia)
A. Streit (KIT, Germany)

Welcome (E. Elsen)

E. Elsen welcomes delegates to the first Scientific Computing Forum (SCF). Following brief round-table introductions, the background and aim of the forum is presented by E. Elsen:

- WLCG has been very successful and has served the LHC computing needs well so far. The Tier 1/Tier 2 hierarchy proved successful initially but the strict hierarchy is now less important for data movement. Operational costs are an increasing concern;
- Looking ahead to the high computing demands for HL-LHC, and profiting from compute, storage and network technology developments, a different model emerges:
 - HEP and CERN are invited to lead developments to evolve WLCG towards fewer and larger federated centres with heterogeneous architectures including High Performance Computing (HPC) and commercial clouds based on heterogeneous funding models, and to provide input to the European Commission;
 - Such developments could serve a wider scientific community and require a strategy to ensure future scientific computing costs are contained, technology developments are fully exploited, and various national boundary conditions are accommodated in an optimal way;
 - The SCF is proposed to develop the technical basis for such a strategy.



During the discussion that follows the initiative is welcomed by several participants, and considered timely. Questions are raised about funding, the role of the European Commission, and the involvement of the CERN experiments. The importance of the CERN Management working with the experiment management on a common strategy is emphasised. For the short-term needs a consultation process is set-up with CERN IT and experiments; technical and scientific assessment will be provided by LHCC. On a wider scale HEP will produce a series of white papers over the course of this years. Beyond HEP, involvement with neighbouring fields of science will be sought.

Scientific Computing Forum Strategy paper (I. Bird)

I. Bird begins by pointing out the very different context at the start of the LHC computing grid endeavour with no large internet companies available and therefore a need to build the infrastructure and tools from scratch. If one were to build the WLCG today it would look quite different and integrate infrastructure now available. Despite the recognized success of WLCG an extrapolation of the services into the future is difficult. Changes are needed with respect to volumes of data and operational costs. Today WLCG represents high data volumes and wide distribution with 2/3 of the global cost (of some 100 MCHF/year) in disk. The future infrastructure should become less specialised, be able to use commercial and opportunistic resources including HPC, cloud and special architectures, and should recognise and leverage commonalities between experiments, across disciplines, and with industry.

Building on the successful WLCG collaboration, technology, sociology and funding lessons were learned which can be used to propose a future computing infrastructure with the following key components:

General Infrastructure such as data stores, compute facilities, networking and associated services such as security, operational support and base monitoring, all capable of supporting different computing models and agile to changes in technology. The general idea is to have fewer, larger data centres (existing large Tier centres and eventually commercial centres) networked together with high speed networks for data storage and processing, accessed remotely by users perceiving this as a single cloud. Having all data logically in the same place opens new and different data analysis possibilities;

A **Software** stack ranging from workflow and data management tools to application level with common tools, libraries and operational tools developed and maintained by the community. The HEP Software Foundation (HSF) is already addressing a lot of this component;

Steering and governance with buy-in from heads of major laboratories, projects, facilities and experiments mandated, for example, by ICFA, to organise and evolve the infrastructure and software addressing issues such as joint procurement, general agreements and community licensing, ensuring direction and feedback, particularly on funding, to the relevant stakeholders.

As a practical way forward one or several prototypes could be built, possibly using the opportunity of the Data Centre in Prévessin, with investment in building and retaining software skills and provision of demonstrators (such as FTS, CERNBox, data archiving, Zenodo, Vidyo and Indico) as services to others sciences.

Points arising from the round table discussion:

Several participants expressed support for this future direction with strong support for reduction of operational costs, encouragement for common solutions, and a suggestion to ensure a way of monitoring the cost of the overall system in future. The importance of software development to get to such a vision is underlined as is the need to be self-sustaining; looking for synergies and learning from external companies and projects, without relying on them.

Shared, common solutions are key: currently the experiments have different solutions for workflow management. A single workflow common to all experiments should be a future goal as should be shared data management. A blueprint architecture for such a federated data cloud defining functionality and services could help to trigger discussions on implementation and realization. Such an architecture should be based on the needs of the HEP community, also considering the needs of other communities.



Future of the Scientific Computing Forum and Membership (E. Elsen)

In order to make progress, E. Elsen suggested to continue discussion in a forum like SCF with a limited number of attendees to allow for discussion. The focus should be both on LHC and technology drivers, including existing knowledge for networks and protocols, embracing other scientific fields.

Points arising from the round table discussion

Targeting SKA is considered useful as SKA has significant requirements on a similar timescale. LHC and SKA also overlap with other experiments such as LIGO, both of which already produce data. Targeting the climate community could be a way to tap into resources stream for neighbouring science although more common requirements may have to be developed.

There is a discussion on the approach needed to serve the HEP needs while taking those of other communities into account. It is agreed to remain restrictive initially to make progress on the large infrastructures, working on an architectural view based on the strategy paper to prototype something that can then be presented to other communities. The European Commission has asked for input for FP9 therefore need to be kept informed.

A question is raised about the mandate, membership and reporting line of the committee. It is intended to be an advisory committee whose work will help shape future investments; the SCF will hence advise funding agencies and initially foremost CERN. As an example, currently some national agencies are questioning the approach with respect to HPC investments. High-level technical expertise is needed amongst the committee members to guide the work and technical direction. There should not be too many members to ensure focussed discussion, reporting will be to Funding Agencies via the Resources Review Board.

AoB and date of next meeting

It is agreed to hold the next meeting in May 2017; the date **12.5.2017, 10:00-13:00** is proposed.

Participants are asked to provide feedback as soon as possible via email on the boundary conditions, composition, and any other input to E. Elsen.

Minutes summarising the discussion will be made available and posted to the [Indico page \(https://indico.cern.ch/e/Scientific-computing-forum-2017-02-15\)](https://indico.cern.ch/e/Scientific-computing-forum-2017-02-15) with the presentations; none of this information will be restricted.

E. Elsen thanks participants for their attendance and contributions, and closes the meeting.