



Minutes of the 2nd Scientific Computing Forum (CERN, Geneva, 12 May 2017)

Present:

F. Antinori (representing ALICE)
M. Barroso Lopez (CERN, IT Department)
V. Beckmann (CNRS/IN2P3, France)
I. Bird (CERN, IT Department)
P. Buncic (representing ALICE)
S. Campana (representing ATLAS)
E. Elsen (CERN, Director for Research and Scientific Computing, Chair)
J. Flix Molina (PIC/Ciemat, Spain)
S. Foffano (CERN, IT Department, Scientific Secretary)
V. Guelzow (DESY, Germany)
F. Hemmer (CERN, IT Department)
B. Jones (CERN, IT Department)
M. Klute (representing CMS)
T. Kollegger (GSI, Germany)
E. Lancon (BNL, USA)
S. Roiser (representing LHCb)

Remote:

T. Boccali (INFN Pisa, Italy)
C. Brokema (ASTRON, Netherlands)
P. Clarke (University of Edinburgh, UK)
M. Delfino (Universitat Autònoma de Barcelona, Spain)
C. Grab (ETH Zurich, Switzerland)
A. Heiss (KIT, Germany)
J. P. Meyer (CEA Saclay, France)
B. Roser (FNAL, USA)
O. Smirnova (Lund University, Sweden)
A. Zoccoli (INFN Bologna, Italy)

Introduction and feedback since first meeting (E. Elsen)

E. Elsen welcomes delegates to the second Scientific Computing Forum (SCF). There are no comments on the minutes of the first meeting therefore considered approved.

E. Elsen summarises the input received since the initial meeting of the forum with expressions of support from several countries and delegates for the establishment of such a forum with CERN taking the lead. Specific opinions were shared from a university Tier-2 in Germany, India and STFC, United Kingdom. There is general interest, including from RRB delegates, to follow the discussions and evolution. All information about the forum is publically available from the SCF Indico page: <https://indico.cern.ch/category/9249/>.

On 11 May 2017, EIROForum members CERN, EMBL, ESA, ESO, ESRF, EUROfusion, European XFEL, and ILL approved the “The Federated Scientific Data Hub” strategy paper available from the SCF Indico page. F. Hemmer reported on a discussion in Brussels on 12 May 2017 between F. Gianotti on behalf of the EIROForm DGs, C. Moedas and R.J. Smit from the European Commission. They were very supportive of the initiative due to the clear needs, goals and deadlines such as HL-LHC, and will consider ear-marking funding for future related technical calls and applications.

Concerning SCF membership, E. Elsen points out the intention is not a fixed group, as discussions evolve new members will attend to contribute their technical expertise as appropriate.



CERN Scientific collaboration with SKA on computing (I. Bird)

I. Bird presents the relationship between SKA and CERN, active for over 10 years in computing for sharing experience and participation in reviews. In 2016 the Dataflow Advisory Panel proposed the concept of SKA Regional Centres, as a result more frequent discussions and workshops have taken place. In mid-May the SKA Project Office Computing team will visit CERN to discuss a number of topics including governance, resource management, networking, EOS and cloud computing. A SKA-CERN “Big Data” workshop is also planned for late 2017 or early 2018.

There are possible synergies between the ideas for future federated data centres, as discussed at the last SCF and exposed in the EIROForum strategy paper, and the SKA Regional Centres. A prototype with SKA involvement exploring areas of mutual interest for managing exabyte scale datasets is being discussed.

E. Elsen thanks I. Bird for his presentation mentioning the SKA initiatives in South Africa and Switzerland. Strengthening the collaboration is in line with the will of the Funding Agencies to support commonality. He invites comments and opinions from the participants.

Points arising from the round table discussion:

P. Clarke: For the UK working with SKA is important as there is limited funding for the expansion of computing. STFC management supports the sharing of infrastructures between the different fields of science they fund.

V. Beckmann: Within CNRS, SKA is driven by INSU therefore IN2P3 cannot currently commit resources to SKA.

A. Zoccoli: Discussions have started in Italy on computing collaboration for astrophysics with SKA and CTA to avoid competing for funding and duplicating the infrastructure. Coordination will take place at ministry level.

C. Grab: CTA and SKA are the big players in the Science Data Centre whose mandate is with ESA. Discussions are currently on-going to use the same infrastructure. E. Elsen: ESA is a member of the EIROforum, it is an appropriate time to see convergence.

C. Broekema: up to now discussions have focussed on technical instrument design and the Science Data Centre. The aim in the Netherlands is also to avoid duplication of resources and equipment, with discussions at an early stage. I. Bird points out at the upcoming meeting with SKA it should be possible to identify specific topics at a technical level. Perhaps future workshops to discuss these topics could go to the countries with the relevant connections such as the Netherlands and United Kingdom.

E. Elsen recognises that the joint use of computing resources requires appropriate agreements to be concluded. The partners and projects may differ by country but should make the support of scientific computing more affordable in the end.

P. Clarke points out that HEP transfers files whereas Astronomy puts data in databases; high performance databases could be an interesting topic for a workshop. The underlying software could be similar.

V. Guelzow underlines that for synergy, the operational aspect must be kept in focus. I. Bird emphasises the focus is rather on what can be done in common and what can be done to prepare for the 2025 timescale with exabyte-scale datasets.

Future computing models in Germany (V. Guelzow)

V. Guelzow summarises the current status in Germany comprising the Tier 1 site, 9 Tier 2 sites providing up to 25% of the computing and strong Tier 3 sites at some universities, with most sites connected by LHCone. Germany appreciates the efforts to consolidate scientific computing and data models however there are financial implications due to funding in Germany, and political implications.

Based on the suggested future LHC computing model connecting a limited number of powerful and efficient resource centres with smaller centres or commercial providers, a possible Helmholtz setup is being discussed with 3 large centres; DESY, GSI and KIT, to appear as a single resource centre similar to the Nordic data facility. All centres would support LHC, and each would support other projects such as Belle II, CTA, FAIR and



XFEL. Each of the three centres could upgrade to 5-10 MW. It is felt important to not only look at resources but also common software solutions, as certain Tier 2s complain the current infrastructure is too costly to run.

E. Elsen thanks V. Guelow for his presentation and invites comments on the separation of Tier 1 and Tier 2 centres in the coming years.

Points arising from the round table discussion

Care is needed when talking about future models as there are different potential models all pushing data to bigger centres; prototyping is necessary to test the feasibility. There is sensitivity around the Tier question and the naming – the emphasis should be on separating where the bulk data are stored, renaming should be based on the future functionality.

The future role of the Tier 2 is a difficult question. It is often important to distribute resources at universities and in different national regions. Various possibilities and scenarios need to be tested with the many available options with the focus on those feasible.

Storage and service availability are very important, both of which should be agreed on at an early stage. The current goal is to include non-WLCG resources, HPC centres and commercial clouds. Time is needed to learn how this works as the experience gained will have implications on the future model. Talking now about functionality in 2025 may be too early.

E. Elsen concludes the discussion adding the intention is evolution rather than revolution, however to knowing in which direction to push is important.

IN2P3 Computing outlook (V. Beckmann)

V. Beckmann presents the LCG landscape in France with CC-IN2P3 as the Tier 1, seven Tier 2s and a Tier 3. LCG-France is under the responsibility of IN2P3, with a significant contribution from CEA. The Tier 2s represent half the computing and disk storage, with all tape storage at the Tier 1. IN2P3 participates 100% to the Tier 1 and 60%-70% to the rest. In addition to the LHC experiments, other experiments such as ANTARES, AMS, LSST, Euclid, KM3NeT and CTA are using IN2P3's computing resources.

There is a strong interest from the ministry to see more harmonisation between HPC and HTC, and interest from IN2P3 to keep the Tier 2s in the loop on the long term. There is openness to new collaborations with other domains, (not yet SKA) working closely with the HPC centres and maintaining 10% LCG participation in future. The LCG computing budget might slightly increase.

E. Elsen thanks V. Beckmann noting the commitment to LCG asking if in future the Tier 2s would remain dedicated to science topics or become more commonly used. For the moment they are dedicated however this could evolve in future. Comments or questions are invited.

Points arising from the round table discussion

There is a discussion about the resources needed to run the software stack at the Tier 2s, whether to centralise or virtualise, and how to make the best use of the expertise available. A large fraction of CPU needs to be available for simulation, with good storage and networking capacity; the definition of a Tier 2 on a 10-year timescale becomes necessary.

Concerning the evolution of HPC and HTC there are extensive tests in Switzerland in 2017 to calculate the real cost of single jobs on a proprietary system or on HPC which should provide interesting results. E. Elsen closes the discussion expressing an interest in these performance measurements.



INFN Computing outlook & the Bologna Initiative (A. Zoccoli)

A. Zoccoli presents the research and education landscape in Italy underlining the difference in networking speed; 10 GB/s in the North, 100 GB/S in the South as a result of European funding. The current e-infrastructure comprises the Tier 1 and 10 Tier 2 sites, all of which provide resources for the LHC experiments and other disciplines. Much of the financing comes from external sources.

Regarding HPC, INFN has an agreement with CINECA and is participating in the H2020 HPC project to study network and storage for exa-scale facilities and to build a prototype of an exa-scale machine with a new network.

In 2016 European funds were used to enlarge the Tier 1 by allocating CPU for Monte Carlo production from the Tier2 in Bari.

Bologna has been selected to host the Computer Centre for the European Centre for Medium-Range weather forecasts, to be ready by end 2019. The current plan is to delocalise the Tier 1 by offloading Monte Carlo production to the Bari Tier 2, and to integrate CINECA computing into the Tier 1 by deploying fast network and software to transparently access storage and computing power. The Tier 1 will evolve into the LHC Data Centre, the Tier 2s will be reorganised in terms of numbers and configuration as a result of regional funding, and the INFN e-infrastructure will be integrated with HPC and network into a single e-infrastructure.

As part of the integrated cross-disciplinary national e-infrastructure strategy, INFN is participating in various initiatives to prepare the future of scientific computing such as middleware, HPC and Big Data integration, and collaborating with the Italian Space Agency to develop and implement a collaborative data hub. 12 post-doctoral positions will be opened to collaborate with LHC on future development such as large-scale science data management solutions, high performance data analysis and algorithms, machine and deep learning techniques.

A. Zoccoli concludes by pointing out 2 future challenges needing discussion due to their high impact on Funding Agencies; coordination of EC projects, and the computing contribution for HL-LHC with respect to detector operation.

E. Elsen thanks A. Zoccoli, remarks on the clear developments in Italy with several points to pick up on, and invites comments or questions.

Points arising from the round table discussion

The optimistic timescale for the move of the Tier 1 is discussed. A. Zoccoli replies once the infrastructure is in place and functioning the move can be gradual, and is timed to coincide with the LHC shutdown. The timing may stretch to 2020, ultimately it is up to INFN management to decide. New resources bought in 2019 will be installed in the new centre with a gradual transfer of resources. The use of Bari for the Tier 1 worked well for ALICE and CMS Monte Carlo; the plan is to continue to work like this in collaboration with the experiments to optimise resource use.

The importance of software is noted. The Community Whitepaper from the HEP Software Foundation is needed, with time to absorb the content before a dedicated Scientific Computing Forum on software. There is a question about where HL-LHC computing should be discussed; the WLCG Overview Board or Scientific Computing Forum? The WLCG Overview Board is restricted to HEP computing whereas the Scientific Computing Forum is intended to also discuss computing for other fields.

Considering the discussions in certain countries about reducing the Tier 2's, a clear vision, strategy and naming convention is needed to set the future direction. From the WLCG perspective the process is clear with the 2020 Technical Design Report defining the computing model with the HEP Software Foundation Community Whitepaper feeding into the Technical Design Report – much work is needed to achieve this. HL-LHC needs will have to be defined and cross-checked with national initiatives. The idea behind the Community Whitepaper is to ensure it engages all aspects of the community.

There is clarification that the role of the Scientific Computing Forum is for exchanging information as input to the Funding Agencies. Any sort of decisive authority would upset this process; views are exchanged and discussed at the Scientific Computing Forum to raise awareness and convince, however the Scientific Computing Forum cannot oblige Funding Agencies. Meetings should be often enough for profitable exchange, but should not interfere with the ground-work which is vital for input to the discussions.



AoB and date of next meeting

It is agreed to hold the next meeting of the SCF on 27th October in the afternoon. The next WLCG Overview Board was already scheduled for that date. The relevant information and agenda will be sent nearer the date.

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