

RCS-ICT Steering Committee Meeting #3 (09/10/2023)

Present: Simone Campana (WLCG, secretary), Xavier Espinal (Technical Committee), Richard Hawkings (EP), Andreas Hoecker (ATLAS), Michelangelo Mangano (CERN-TH - invited speaker), Pere Mato (Technical Committee), Patricia McBride (CMS), Enrica Porcari (IT), Zhechka Toteva (Technical Committee), Vincenzo Vagnoni (LHCb), Bedanga Mohanty (ALICE), Pippa Wells (RCS directorate, chair)

Introduction

The Steering Committee plans to have the next meeting in early Q2 2024. The Committee could also consider approval of proposals via email exchanges in between these dates.

Action: Simone Campana to identify a suitable date

Action List Review

RCS-ICT-010. The proposal on all aspects of the information landscape, including the priorities for the required work, is in preparation. The RCS-ICT Steering Committee expects to hear on that at the next meeting, **noting** that implementation should not start before the plan is agreed by the Steering Committee.

Action on Alex Khols to present the plan at the Q2 2024 Steering Committee.

RCS-ICT-017. Enrica presented the preliminary results of the analysis of the document workflows ecosystem. The initiative started from a discussion at the Enlarged Directorate, related to possible savings by removing possible duplications.

Summary of the analysis and future plan: 1) there are many overlapping capabilities but there are different ways of using these capabilities; 2) IT expects to decommission some systems and particularly DFS and AISMedia (both to be replaced by CERNBox); 3) IT will provide a matrix of capabilities mapping of the needs from different personas/workflows with the appropriate system (aspects related to Data Privacy and Data Protection were looked into and taken into account in the mapping); 4) there was no identified need to increase the current offering 5) If new needs will be expressed, they will be evaluated by STEPS (Steering Committee for the Externally Provided Systems and Software); 6) IT will extend the analysis to other systems.

It was clarified that for the decommissioning of DFS, the RTS communities are not among the main interested parties. At the same time, RCS is more involved in the CERNbox/Google discussion and particularly on the use of the Google contract with CERN. It is in the IT workplan to issue an official communication by the end of the year about the use of Google tools in the context of the CERN contract. While Overleaf is not in the list of tools being examined so far, it is

a very useful LaTeX editor and heavily used by the experiments. It will be part of the next phase, extending the analysis to a larger set of tools.

Action on Enrica to follow up on the communication about Google tools and provide final report by next SC

RCS-ICT-001 and RCS-ICT-002. The NextGen Trigger project was approved by Council in October 2023. The project is expected to start in January 2024. The Steering Committee agrees to close those two actions and expects a progress report at the next Steering Committee meeting.

Action on Alberto Di Meglio to report on the progress of the NextGen Triggers project at the Steering Committee in Q2 2024.

Evolution of Event Generators and the role of CERN

Michelangelo explained the plan to form a Monte Carlo working group under the umbrella of the LPCC. The vision behind the initiative is to have a more structured approach to the evolution of Event Generators: this area is strategic in our field and must be approached with the same priority as accelerators and detectors; CERN has the responsibility to acknowledge this challenge and is asked to commit resources to drive and support its implementation. The authors of the Monte Carlo generators welcome the initiative of the LPCC and are on board with the proposal. The LPCC is organizing a MC generators workshop in November 2023 to discuss the technical evolution of the generators.

The presentation of this vision to the RCS-ICT Technical Committee concluded with acknowledgement that there is the need for this activity.

The RCS-ICT Steering Committee supports this effort in organizing the generator community. The generators workshop organized by the LPCC at CERN in November 2023 should identify the main areas of work and what level of resources would be needed. Each Monte Carlo generator group has its own source of support. At the same time, effort across the board is instrumental to evolve in a specific direction. For example, the fundamental role of CERN IT experts in porting generator code to GPUs was clearly acknowledged and should be continued. The NextGen Trigger project also allocated effort in the generators area. The existing (CERN IT) and new (NGT) efforts would likely be sufficient in terms of CERN contribution to the evolution of generators in the next phase. RCS however needs to ensure that the existing effort is continued and the new effort is not diverted to other tasks. For both efforts, milestones should be defined and followed.

Action on Michelangelo to report back on this activity at the RCS-ICT Steering Committee in Q2 2024.

Technical Committee: status of ongoing initiatives and new proposals

The RCS-ICT Technical Committee chairs reported on the discussion about the Preveessin Data Centre (PDC) at the last meeting. The PDC will be in use at the end of 2023, with 4 MW of initial power capacity available, expandable to 12 MW in future. The new infrastructure, together with the existing one, creates new opportunities to extend collaboration with experiments. Hosting the online farms and storage of the experiments on the IT premises was discussed with the LHC experiments, noting that there are very stringent time constraints in preparation for Run-4. The Technical Committee did not identify technical showstoppers in enlarging the use of the IT facilities beyond the Tier-0 needs, so the discussion is more at the strategic level. In particular IT would like to discuss with the communities on what are the service needs and how to provide them, rather than restricting the discussion to hosting of hardware on IT premises. The discussion should involve LHC and non-LHC experiments.

ATLAS and CMS clarified that the HLT farms in Run-4 will be hosted at P1 and P5 respectively, as the preparation work has been ongoing for a long time and commitments were made. ATLAS also noted that the current (Run-3) HLT farm cannot be kept online in P1 during LS3 because of infrastructural work, and it would be interesting for ATLAS if that hardware could be hosted on the IT premises. Other experiments might see similar opportunities. For the LHC experiments the Steering Committee concluded that rather than focusing on the online farms hosting, there are other processes and needs that should be considered, such as the point brought up by ATLAS. For the non-LHC experiments, the North Area consolidation should be discussed in this context (the timescale is Run-4) and particularly the aspects of where the online systems should be hosted. The Steering Committee also noted that hosting of experiment hardware and services of the experiments will require a discussion at the level of funding with the directorate.

Action on the Technical Committee to discuss with the LHC and non-LHC experiments the future service needs (beside the Tier-0) that could be fulfilled by the new IT facility. An update is expected at the Q2 2024 Steering Committee.

Technical Committee: status of ongoing initiatives and new proposals

- Initiative on Heterogeneous Architectures to define a model to provide access to non-X-86 architectures. Resources needed from IT: 2 FTEs to provide and maintain the service. Currently two graduates are working on this initiative, but the need for a more senior contributor (as part of the two) was identified. It was clarified that the quoted effort would be for providing the service and not for the development of the applications. Action RCS-

ICT-006 can be closed (DONE). Status: APPROVED but not yet resourced (PENDING discussion)

- Initiative on Collaborative Tools. The PSO is aligned to what Enrica reported with the document workflows ecosystem above. In particular the PSO foresees to identify if there are gaps in the current offering and build a matrix of capabilities, properly documented. The prioritization of the Google workspace deployment is covered by the action on Enrica above. Action RCS-ICT-011 can be closed (DONE). Status: APPROVED
- Initiative on the Large Scale validation and optimisation of RNTuple. The estimation of the personpower has now been done and is part of the PSO. The initiative could be covered with existing personpower in IT but a lead should be identified. Status: APPROVED
- Initiative on the Centralized Electronic Logbook. There is an urgent need for SMEs to have a centrally supported logbook system and this is the priority of the PSO. LHC experiments could also benefit from the same service if they so desire in the future. Status: PENDING (see action below)

Action on IT Technical Delivery to propose a properly sized shared service for the SMEs to be reviewed by the Technical Committee. The Steering Committee expects an update at the Q2 2024 meeting.

- Initiative on the Micro Electronics Cluster. On hold. A go-nogo decision needs to be taken by EP-ESE management as there are some technical challenges that might impact the way of working. The Steering Committee takes note and expects an update at the Q2 2024 Steering Committee meeting. Status: PENDING
- Initiative on Improving the Security of the Software Supply Chain. The initiative is about having a service that centralizes the control of software repositories and prevents security issues. There is a position paper that was presented to the Technical Committee and might mature to a PSO. Enrica highlighted that a CERN cybersecurity audit identified several areas where IT should strengthen the security of its service offering. This item should be done as one response to the audit, ensuring that the implementation is useful for the communities. Status: APPROVED but not yet resourced (PENDING discussion)
- Initiative on Critical Services. The Technical Committee started a discussion with the RCS communities to understand the criticality of the services and then have a dialog about that criticality and the possible mitigations. This refers mostly to non-WLCG services as the expectations on the WLCG services are generally covered by the WLCG MoU. There is an ongoing process collecting information from the experiments. A summary is expected at the Q2 2023 Steering Committee Meeting. Status: PENDING

Technical Committee: scientific data management

Scientific Data Management was identified as highly strategic by the Technical Committee. A proposal has been elaborated about forming an initiative across the experiments, IT and EP for the development and operations of a scientific data management ecosystem. Today the effort supporting Rucio beyond ATLAS is on a best-effort basis and is becoming a risk for success.

The ask is for CERN IT to contribute in DevOps activities for operating Rucio for ATLAS and CMS, and to deploy a scientific data management Reference Implementation service based on Rucio and the IT data management service ecosystem (e.g. FTS, EOS, CTS). The reference implementation would serve as a platform for all communities to evaluate functionalities of the system. The communities include CERN LHC and non-LHC experiments, and non-CERN communities (e.g. other HEP experiments and sciences in the ESCAPE cluster). The estimated person power is 3 to 4 FTEs with one long-term staff for sustainability.

At the Steering Committee there was strong support from ATLAS as the ecosystem is a real asset that CERN is providing to its experiments and partner communities. CMS also favors the initiative. It was clarified that the effort mentioned above (3 to 4 FTEs in IT) does not replace the existing effort from the experiments but rather complements it. Status: APPROVED but not yet resourced (PENDING discussion)

Prioritization of RCS-ICT initiatives

In preparation for the IT Program of Work, the Steering Committee went through the ongoing initiatives and prioritized them in view of defining a medium term (3 to 5 years) roadmap.

High Priority Initiatives:

- Support for Heterogeneous Architectures (RCS + ATS)
- Scientific Data Management (RCS)
- Physics Analysis Environments (including RNTuple) (RCS)

Several Initiatives are not a specific priority of RCS, but they are considered necessary as part of a broader scope:

- Collaborative tools consolidation (RCS+ATS) - Document Management Systems (RCS)
- Filesystems and Storage consolidation (RCS)
- Improving the Security of the Software Supply Chain (RCS)
- BC/DR, Enterprise Architecture and Service Level ("criticality") (RCS/ATS/FHR)
- Linux support (ATS/RCS)
- AAI (Grappa/Egroup) (RCS/FHR)

Some initiatives are currently not a priority for RCS, but they might be impactful in future for RCS needs and their development should be monitored

- Kubernetes and related ecosystem (ATS)
- Air-Gapped Environment (a.k.a Technical Network, Network Isolation) (ATS)

The Scientific Information Landscape project is currently seen as part of the work on Document Management Systems. A proper prioritization will be done once the analysis and a plan will be presented.

Items from the Steering Committee

The needs for a TTaaS (Text Transcript as a Service) was discussed. IT currently runs a pilot based on a software under a contract expiring in August 2024. The Steering Committee agrees that offering TTaaS is a need of the organization. The current implementation should be however re-evaluated in terms of architecture and building blocks also considering the technical improvements in many commercial solutions. The use cases should also be re-evaluated in order to understand the needs of the communities and the possible future adoption of the service. The software was successfully used in the LHCP conference in 2023 and it is unlikely that a new solution will be available for LHCP in 2024. The Steering Committee agrees therefore to support the current implementation up to LHCP 2024 in early June 2024, as this is compatible with the current contract. The experiments commit to contributing to evaluating other solutions on the same timescale and possibly in parallel during/after the conference. The D&I teams of the experiments will be the contact point. The IT engagement team will re-iterate the collection of use-cases with the communities using the TTaaS. The Steering Committee expects an update at the Q2 2024 meeting.

Action on the IT Engagement Team (Simone Campana) to re-iterate the collection of use-cases with the communities using TTaaS by Q2 2024

Action on the chair of the IT Architecture Review Board to review the architecture of the TTaaS service by Q2 2024