

RCS-ICT Steering Committee Meeting #4 (14/06/2024)

Present: Simone Campana (WLCG, secretary), Xavier Espinal (Technical Committee), Joao Fernandes (CERN IT - Invited Speaker), Richard Hawkings (EP), Andreas Hoecker (ATLAS), Alex Kohls (CERN RCS-SIS - invited speaker), Pere Mato (Technical Committee), Patricia McBride (CMS), Andreas Peters (Technical Committee), Enrica Porcari (IT), Zhechka Toteva (Technical Committee), Vincenzo Vagnoni (LHCb), Marco van Leeuwen (ALICE), Pippa Wells (RCS directorate, chair)

Introduction

The Steering Committee plans to have the next meeting in Q4 2024. The Committee could also consider approval of urgent proposals via email exchanges in between these dates.

Action: Simone Campana to identify a suitable date

Scientific Information Landscape

The update of the Scientific Information Landscape project was presented. The Scientific Information Landscape is built on two layers: the analysis and publication lifecycle management layer, and the repositories layer. The project implementation proposal foresees the following steps: 1) Adopt both Glance (LHC experiments) and CAP (SMEs) for the lifecycle management 2) Enhance CDS to support the needs of the communities for versioning drafts and commenting 3) Establish a central common publication approval workflow 4) ensure information consistency between Glance/CAP and CDS through better integration and 5) Allow Glance/CAP to transmit the same set of metadata to downstream publications services (e.g. arXiv)

Resources: one GRAD was granted to RCS-SIS in the last MTP to support the migration from the old to the new CDS. More resources would be needed to complete the project. The priorities expressed as an input to the Steering Committee meeting are:

High priority:

- 2 GRADs in IT-CA for 2 years for CDS review workflows & approval + PDF commenting
- 1 LD in RCS-SIS to support CAP long-term

Medium priority:

- 1 GRAD in RCS-SIS for 1 year to develop new features in CAP for SMEs (author management)
- 1 LD in EP to support central Glance coordination long-term

Discussion:

It was clarified that the LD in RCS-SIS to support CAP long term would be both for CAP as the current Analysis Preservation Portal (about 60%) and the new function as lightweight publication lifecycle management system (about 40%). The Steering Committee believes that while the technology would be the same, the role of CAP for the two functions should be referred to in two different ways, to avoid confusion.

The possibility of using Glance as publication lifecycle management system also for SMEs was considered by the Scientific Information Landscape task force with the conclusion that the customisation of Glance for SMEs would require considerable effort. The conclusion was discussed and partially challenged at the RCS-ICT Technical Committee.

At the moment there is a person in EP who is doing the coordination of Glance ad-interim and on best effort basis, but this model is not sustainable given the role of Glance for the four large LHC experiments. It was discussed how a full LD for Glance coordination seems to be excessive. That LD if granted might need to take over other responsibilities as well. The experiments would really like more central and long-term support for Glance. The spokespersons of the large LHC experiments also question the relative priorities of coherent support for Glance compared to using CAP in addition as a publication tool for SMEs.

The CDS migration is a very important point. The experiments have not been exposed to the new CDS and its functionalities, nor to a detailed migration plan. A first discussion happened at the RCS-ICT Technical Committee in April. The experiments would like to be early engaged in the process and follow step by step the migration. This will allow them to identify possible issues with the new implementation and the migration process of large collections.

Decision: the migration to the new CDS and implications for the large LHC experiments should already be evaluated by the existing Scientific Information Landscape working group, which contains the right experts from the experiments. The membership can be adjusted as needed, if people change roles.

Action: Alex should write a new request to the working group to define this new task and circulate it to the RCS-ICT Steering Committee by July 15th. Approval will be done via email.

Action: Alex should report on the progress of the working group with its new focus at the next Steering Committee in Q4 2024.

Decision: there are still diverging opinions on extending CAP vs making a Glance-lite for the SMEs. Glance experts should make a study of what it would imply to develop the needed functionalities for Glance-lite and run it for the SMEs.

Action: Following the discussion after the meeting, Pippa will try to identify a set of Glance experts to evaluate the feasibility and effort needed for SMEs to use Glance.

Decision: The RCS-ICT Steering Committee considers that it is premature to define resource requests for the 2025 MTP. This should be discussed again at the next Steering Committee.

Document Management Systems update

An update on DMS was presented. IT, in collaboration with other departments produced a document listing the capabilities currently available at CERN with respect to Document Management Systems. The document is currently being reviewed by the STEPS Technical Committee and should be endorsed soon by the STEPS Steering Committee. At that point, the list of documented capabilities will be authoritative. This list of capabilities should cover the current needs of the CERN community. If some additional capability will be needed in future, the request should go through STEPS as it might expose CERN to extra costs. The goal of the exercise was not to merge systems but to explain what these systems can be used for. The document being reviewed by STEPS also explains the lifetime of the systems providing the aforementioned capabilities. The document will be publicly available once endorsed by the STEPS process.

Public Cloud Strategy

The progress in defining a Public Cloud Strategy was presented. The public cloud package will include 1) Cloud Policy 2) Architecture Usage Framework and 3) establishment of cloud and quantum contracts. The important milestones: signature of the cloud and quantum contracts (Q4 2024) and finalization of the Cloud Operating Model (after summer 2024)

Action: the RCS-ICT Steering Committee expects an update once the milestones mentioned above are met. Simone should schedule it.

Technical Committee: status of ongoing initiatives and new proposals

The list of activities of the Technical Committees in the last one year was presented.

- Commissioning of RNTuple: there are very positive outcomes: the adoption of RNTuple should lead to up to x3 speed improvements in reading and writing and 40% less storage space needed. This activity is now fully on track as acknowledged in the May 2024 LHCC.
- Support for Heterogeneous Architectures: IT has means to provide non-standard compute architectures but the full lifecycle model expected as an outcome of the PSO is still not in place. IT needs early warning if new and different resources will be needed

with respect to what is available today. Offering these resources will be simpler once the cloud operating model will be available.

- Scientific Data Management: the initiative is progressing very well and there is very good collaboration between IT and the experiments. It was noted that this initiative is at the level of a pilot. The current level of effort needs to be maintained both in IT and the experiments if RCS the ambition is to go beyond the pilot phase. ATLAS and CMS strongly rely on the success of this pilot.

Action: a more detailed presentation should be foreseen at the next RCS-ICT Steering Committee, in order to assess the progress. Simone to organize it.

- Micro-Electronics Cluster (MIC): the initiative investigated the possibility to consolidate some of the MIC ICT services in IT. The MIC management decided for a no-go as the level of effort required for the consolidation is too high with respect to the expected benefits.

Decision: the MIC cluster initiative should be considered closed. The Technical Committee will re-open it if the conditions that led to the decision change in the next years

- Collaborative tools: the initiative is progressing with several successes (e.g. enabling the CERN Google workspace and the Atlassian consolidation). One item to be understood is the access for external contributors to CERN software repositories and development ecosystems. IT is looking at that from the technical and security perspective

Action on the RCS-ICT Technical Committee to present a (set of) proposal(s) at the next RCS-ICT Steering Committee

- Centralized logbook for the SMEs: the exploratory phase has been completed. There is a strong request from EP for IT to run this service.

Decision: IT agrees to support this service as part of the IT service portfolio and the provisioning could be as quick as the next six months. Once the service is ready, it should also be advertised more widely, for example to the PS/SPS coordinator and the SPSC. If the scope of the service in future should go beyond the SMEs, the conditions will need to be re-discussed.

- Traffic Marking: the initiative is the continuation of an existing R&D. It is important in light of a future possible sharing of the network with different experiments and possibly sciences. The initiative can be carried over with the existing effort in IT. It was noted that in the future some effort from the experiments might be needed.
- Review of Critical Services: IT and the experiments are reviewing the criticality of IT-provided services. The objective is to align the IT offering with the experiment expectations and vice versa. The services offered by one of the IT groups were reviewed

and there is a plan for the remaining groups. A document should be produced by the end of the year with the results of the analysis.

Action: The RCS-ICT Technical Committee should share the Critical Services review document with the Steering Committee by the end of 2024.

- Hosting hardware in the CERN IT datacenter infrastructure. IT received some requests to investigate the possibility to host HL-LHC online hardware in the IT premises. In some cases it was an initiative of individuals in the experiments and not a centrally coordinated initiative of the experiment. These explorations require non negligible effort from IT. While IT is willing to consider requests, it should be ensured that the management of the experiment is supporting these requests.

Decision: IT will consider only requests communicated by the spokespersons of the experiments.

- Text Transcripts as a Service. The pilot service supported by IT has a dual function: “online” and “offline” use. In the “offline” use case, TTaaS indexes videos in CDS and supports writing minutes after the meeting (noticeably for Council). CERN IT is getting good feedback about these capabilities and is willing to continue the service. In the “online” use case, the pilot service is used to produce live transcripts for video streaming events. The engine of TTaaS is an application from University of Valencia for which CERN has a perpetual license. TTaaS was used for the “online” use case in several meetings including ATLAS and CMS weeks, the LHCP conference and the FCC workshop. IT is receiving mixed feedback. Some feedback reports that TTaaS produces better transcripts than ZOOM. Most recently some very negative feedback was provided. There are doubts about the quality of TTaaS for the “online” use case and the applicability of the solution beyond very VIP events, for which there are some (admittedly more expensive) solutions as well. The recent very negative feedback causes reputational problems for CERN. CERN IT is not willing to continue TTaaS in its current form for the “online” use case

Decision: CERN needs to support the capability to produce text transcripts online (live) for meetings being attended by the CERN community. The current TTaaS pilot is not considered adequate for this need and will be discontinued in its current implementation and the concerned communities informed. ZOOM text transcripts can be used as ad interim solution. CERN IT will investigate different solutions on the market. The evaluation of these solutions should be supported by the CERN community, at the least the large LHC experiments and the Diversity and Inclusion office at CERN.

Action: In collaboration with relevant communities, IT to start the process of a technical and financial evaluation of different solutions on the market for the online text transcript use case mentioned above. The report will be prepared by the end of the year.

Decision: TTaaS will continue being supported for the “offline” use case. The engine will however be reevaluated considering more modern technologies.