

# Report from the ICAC

*Ian Bird (chair), Peter Clarke, Suchandra Dutta (remote), Peter Elmer, Eric Lançon, Michel Jouvin*  
17th October 2019

## Introduction

The ICAC met at Fermilab on 15th and 16th October 2019 for a follow up to the March 2019 review. The agenda for the meeting presentations and discussions can be seen here: <https://indico.fnal.gov/event/22056/>. The committee saw presentations from the Fermilab CIO and from teams in the Core and Scientific Computing Divisions, and we thank the Fermilab management for their hospitality and very much appreciate their continued openness.

The main goal of this review was to evaluate the progress on the recommendations from the March 2019 review, and to advise on future directions.

The structure of this report thus follows the set of recommendations from the previous review and will comment on the status of each of those areas.

## General Comments

### Findings

Apart from the progress on the specific recommendations made in the March meeting, the Scientific Computing Division has undergone a significant reorganization, to enable a better level of communication, and to reduce the concerns over segmentation within the Division. This goes hand-in-hand with the efforts to reduce the “stovepiping” of hardware resources, commented upon later in this report, with the overall aim to improve effectiveness of the Division, both in terms of communication and in terms of usage of physical resources.

The re-organisation has removed the quadrants, has five distinct Departments, and created a virtual Department as a container for all of the cross-cutting projects. All of these Departments are overseen and report to Jim Amundsen, as Division Head. There are several associate heads for projects, science, CMS, DUNE, and facilities outside of the line management. The net result is a simpler organization, with less managers. The new organization has only been in place since mid-September, so it is too early to comment on effectiveness, but will be reviewed in a subsequent meeting.

### Comments

The committee regarded the re-organisation as a very positive action, with the potential to provide a simpler and more effective management structure, and significantly improving the overall level of communication between teams. We congratulate the management team for the speed with which they have designed and implemented this far-reaching change.

See further comments below on “Reducing Duplication”.

## Follow up of Previous Recommendations

### Resource Scrutiny Group

*Create a resources scrutiny group to review requests for computing resources and set priorities for allocations of resources between the experiments.*

The committee was pleased to see a concrete plan for setting up a resource scrutiny group. Fermilab will evolve the function of the existing internal scrutiny group (the SC-PMT) to have a wider remit to include fully internationalised projects such as DUNE. This requires introducing members from outside of Fermilab, particularly from the international community.

It was clarified and agreed that the purpose of this group is to:

- Receive “Resource Request Documents” from experiments. The RRDs should state the experiments’ usage over the last year, state the forward capacity requirements for the next year in detail, and the next n-years as preliminary requests. Resource requests should be based upon a sound computing model which should be described succinctly, but in enough detail to allow the panel to constructively scrutinise the requests.
- Scrutinise the requests to ensure the model is sound in terms of data access and replication policy, CPU campaigns, etc., and that the capacity provided is used appropriately.

It is anticipated that this should be an annual formal process. It should be clear that this group should be a standing body reviewing resource requests (and usage) for the global computing infrastructure (but not the computing models per se).

There are two particular needs to be addressed, not necessarily by the same scrutiny:

1. Scrutiny of DUNE requests on behalf of, and together with, the international DUNE funding agencies, for computing resources needed across the collaboration, and including Fermilab as the host lab (analogous to the LHC C-RRB);
2. A local scrutiny that would prioritise between Fermilab local resource needs between the various experiments. Assuming that CMS and DUNE have dedicated funding lines, this would be for all the other experiments that currently do not have specific funding.

### DUNE Funding

*A separate funding line for DUNE (as for CMS) would be useful in order to plan the resource profile appropriately. DUNE computing funding will need to be part of a long-term plan and not subject to squeezing by other competing demands.*

We noted the discussion in progress with DOE about pre-operation funding for DUNE, in particular the creation of a PEMP. This plan will include software and computing, and will

inform an eventual DUNE funding program. We regard this as a good step forward towards assuring a distinct funding line for DUNE computing resources.

## Funding lines for other non-LHC experiments

*Computing funding for non-CMS resources is not ring-fenced and is part of the detector and operations funding, thus gets low priority. This results in years where no resources can be acquired, or old systems replaced, despite demands for computing resources continuing to grow. Consider how a funding line for computing could be separated to ensure a manageable budget. The consequence of not doing so will be a gradual deterioration of services and equipment.*

We understand that it has not been possible to make progress on this (apart from the PEMP for DUNE).

We understand that ring-fencing non-CMS resource funding is not an easy task. Nevertheless, the aggregated computing capacity requirements of Fermilab experiments other than CMS are real and unavoidable if the data arising from them is to be exploited fully. In our experience it is unusual for a laboratory with the status of Fermilab to have no appropriate planning guidance for funding for computing. This lack makes it impossible to effectively or optimally plan for hardware and infrastructure requirements.

We suggest that Fermilab Scientific Computing Division should continue to petition the lab management for a more explicit funding line for non-CMS computing operations, in particular the pre-DUNE neutrino experiments.

This may be aided by a clearly advertised policy/expectation for the fraction of computing expected from outside Fermilab for international experiments, such that the Fermilab responsibilities are stated in a clear and transparent way.

We believe that having a documented resource plan for several years that addresses the anticipated needs of all the experiments' computing and storage needs, could help in convincing the funding body that a specific funding line is needed. We recommend that such a plan is established in the coming months. Such a plan can usefully be informed by the scrutiny group documents that will be required of (at least the larger) experiments as described earlier.

## Federated Identities

*Look at ways to speed up adoption of federated identity use as a building block of collaborative services, particularly needed for DUNE.*

The committee was presented a plan, a set of open questions and the current status. This raised two different issues that must be disentangled in order to make progress:

- Policy: federated identity is a cornerstone of the distributed computing infrastructure built in the last 15 years and is now essential for international scientific collaborations. To make the use of federated identities possible, a trust framework has been very successfully setup, through the IGTF (International Grid Trust Federation).
  - **The committee are very concerned that the DOE identity vetting process, as interpreted, at Fermilab, will break this trust model and will effectively terminate the ability of Fermilab to lead any international scientific computing collaboration.**
  - This would result in a loss of leadership and would prevent Fermilab from being a partner of choice for international collaborations.
- Technical implementation: a move is currently happening worldwide away from the current X509-based identity federation to the more pervasive token-based identity federation. There are several bodies involved in coordinating the various technical and operational aspects of this evolution, in particular to ensure the full interoperability of the various implementations. WLCG and OSG in particular have been two very active players in this area (e.g. via the WLCG Authorization working group).

Our recommendations in these two aspects are:

- Work with other DOE national laboratories at the CIO and lab Director level to address the policy aspects of the new vetting process with the DOE. A first step is probably to come to a common interpretation on the impact of the vetting policy for acceptance of federated identities, and the potential impact on scientific collaboration.
- At a technical and operational level, work closely with global efforts (WLCG, OSG, FIM4R in particular) to bring the Fermilab expertise and needs, and to benefit from the work already done or in progress related to the technical issues that you may have identified.

During this meeting, the committee did not see a report on progress towards use of federated identities for the collaborative services of Fermilab as a host lab.

- For the next ICAC meeting, the committee would like to see a plan on integrating the collaborative services, (e.g. Indico, video conferencing, etc.), with a federated identity, including the authorization part if relevant.

## HPC Strategy

*Draft a high-level plan for the strategy of use of HPC resources. What are the main goals of the work in this area? What are the highest priority developments to enable success? The close relationship with ANL could be useful in setting out this plan, and perhaps a more explicit common project with ANL could be envisaged.*

The work in progress around the use of HPC resources appears to be appropriate. The main activity is around the CCE proposal to exploit the resources available on the ASCR supported facilities, in particular by implementing the HEP complex workflows and specific identified software problems on those facilities. In addition, Fermilab work with the Chicago area group and in particular have been running use cases on the ALCF and OLCF facilities.

This is a topic to be followed at the next ICAC meeting.

## DUNE Computing Management

*DUNE needs a strong computing collaboration with visible management. This was found to be very important for the LHC experiments in managing a global infrastructure and having a long-term voice and plan. We recommend working with the DUNE computing management to encourage putting in place a clear management structure to interact with Fermilab and their other collaborating computing sites.*

At the time of the March ICAC meeting, the new computing consortium has just been put into. It is too early to evaluate the concrete impact on the relationship with Fermilab computing, but for the next ICAC meeting an update on the relationship between DUNE computing and the Fermilab Scientific Computing Division would be useful.

We note also that DUNE is now an associate member of the WLCG, and has the stated intention of leveraging as much of the work and experience of the LHC experiments as possible.

## DUNE Computing Model

*DUNE should be encouraged to draft a computing model, in order that Fermilab (and other sites) can plan their facilities. A draft plan will highlight the areas that need R&D or testing. Such a draft should be produced this year to enable Fermilab management to plan their services and organisation.*

We acknowledge the significant progress made by DUNE on its computing model definition in recent months. We think that it may be useful for Fermilab to get a more formal document from DUNE which clarifies the computing and storage requirements over the coming years for Fermilab and the other contributing sites. This document will be input for the new scrutiny group mentioned earlier.

We also recommend (to DUNE) that an independent review of its computing model be planned at a time seen as appropriate by the experiment (when the computing model has been finalised but enough in advance to be adjusted if necessary). It is not a Fermilab responsibility *per se* to decide on this review but we suggest that you advocate for it. We also want to clarify that the ICAC is not the appropriate body for this review.

## Fermilab as Host Lab

*Fermilab should have a plan for how it becomes an international laboratory for DUNE, what collaborative tools will be provided, etc. The plan should clarify the responsibilities of Fermilab as a host lab, and as part of the computing model.*

We have seen major progress in Fermilab thoughts about its responsibilities as a hosting laboratory for an international experiment, and establishing a list of the services where

Fermilab must have a leadership role and those where it could be either a contributor or not play a role. We suggest that the initial rather extensive list of services that Fermilab propose that they should be leading needs to be refined, in order to identify whether some of the responsibilities could be taken by other institutes.

We recommend that in parallel with refining this list of services (including anticipated new services to be provided), a resource loading plan should be established in order to identify potential lack of effort or expertise, or whether these responsibilities can be assumed by reallocating some of the existing effort as part of the operation optimisation currently in progress. To be presented if possible, at the next ICAC meeting.

## Storage Strategy

*The future storage strategy requires particular attention. In particular, a vision and a roadmap is needed to address the needs in the Public cluster and a plan should be elaborated to address concerns over the sustainability of Enstore, possibly by adopting a solution with greater support in the community.*

The recent evolution in storage infrastructure and services was presented but the committee felt that the actions taken so far, despite going in the right direction, tended to be opportunistic rather than driven by a long-term strategy. The concerns of aging hardware, in particular the hardware that is out of warranty, has started to be addressed but it doesn't really address long-term sustainability of the Public storage cluster.

Fermilab is looking at the future evolution of the tape system as they have assessed that Enstore probably has not a long-term future, but has no immediate shortcomings. Fermilab is ready to investigate the migration to something used in other parts of the community and to participate in development efforts, in particular those related to Fermilab specific needs. We note that one candidate considered is the new CERN CTA tape back-end as a replacement for Enstore, but Fermilab failed to receive a clear answer from CERN on the readiness to provide CTA to external institutes and to build a collaboration around it. It is clear that any migration will take a long time as it will require a media migration and will need careful planning.

We recommend clarifying the long-term strategy and suggest that it be presented at the next review. Ian Bird is ready to help clarify the possible collaboration with CERN around CTA.

We also encourage Fermilab to participate in the WLCG DOMA work on the future of data management and access tools, in a distributed environment. The team had already recognised this opportunity to collaborate.

## Reducing Duplication

*A plan to harmonise the three separate components of the facility should be created, to avoid unnecessary duplication of both staff effort and hardware solutions, recognizing the practical difficulties of achieving this quickly.*

We think that the progress in evolution of the Institutional Cluster (IC) has started to reduce stovepiping. We strongly encourage continuing the development of the IC that we see as the main approach to reduce this legacy inefficient use of resources and the duplication of effort that results from it.

“Stovepiping” in this context refers to the full stack implementation of clusters and services for each customer experiment, that has been a long-standing feature of the way Fermilab provided services to its users. While this had been very experiment-centric, it is no longer a very efficient way to provision resources as it does not easily enable sharing of spare capacity, and is costly in terms of support. The FermiGrid and FermiCloud models had changed that view, and the introduction of the Institutional Cluster moves further in this direction. The committee regards this as a very positive direction.

We consider that the principles behind the SCD reorganisation are good and that the proposed new structure has the potential to address some of the concerns we raised previously. As it is still being implemented, we will review the actual impact at a future meeting.

## Software R&D Strategy

*A big picture strategy for the software R&D should be made, in order to understand how the (many) various projects fit into the overall strategy of SCD in answering its challenges. In particular such a plan can be used to ensure that funding opportunities are actually focussed on priorities. The plan should benefit from leveraging work that is happening in the field outside of Fermilab, for example in the HSF, and projects such as IRIS-HEP.*

The committee appreciated the effort to provide a more detailed view on the R&D strategy but we think a real review of the global strategy would benefit if the projects were categorised into three areas:

- Operations (ops)
- Sustaining the capabilities (suscap), i.e. all the R&D activities required to evolve the current services
- Long term R&D (lt)

We perceive that there is a chronic structural funding problem in that the first of these (ops) is “sort of” funded in that the Laboratory channels funds for basic operations, the last of these (lt) necessarily responds to opportunities generated by funding agency national strategic priorities - which only align with Lab priorities in part and, which by their nature, tend to be “bleeding edge” and hence to not cover sustaining existing capability and quality through essential development and evolution (suscap). This results in a long-term funding gap in the middle-ground, leading to the serious risk of current capabilities becoming not-fit-for purpose.

Suscap R&D is a clear set of activities that are essential to keep operating as a facility, but there seems to be almost no funding for this, and it cannot easily be taken from the operations program without impact on ongoing operation. Some of this R&D is essential to set up for DUNE, facilities development is crucial for the future, and is vital in order to keep Fermilab as a world-leading facility for the future.

For the next ICAC, we recommend that Fermilab identifies how to resource suscap on a strategic basis rather than an opportunistic one: this may require in some cases a pulse of resources to fund time-limited software asset creation projects, which can/must then drop down to operations. It is also necessary to identify how operation optimisation may allow to find these additional resources for suscap R&D.

The ICAC was initially concerned that it could not easily see that the allocation of the FTE effort available was appropriately proportioned to the different areas and reflected priorities. This information was provided during the meeting and we thank the team for the quick response. From an initial look the allocation seems reasonable given the constraints; i.e. there is no disproportionate amount going to areas of only long-term or speculative relevance. The panel would like to see this information provided for subsequent meetings and could be a summary agenda item towards the end of the meeting.

## Facility Resources

*Within SCD we recommend that CMS and other projects should be less stovepiped. This is a source of duplication of effort and inefficiency. This must be avoided for DUNE. Facilities and services should be as far as possible common across supported experiments, focussing on function rather than specific requested solutions. We encourage the computing management to continue to re-evaluate the organisational structures in the light of constrained resources and with an eye to the evolving needs of the lab and the experiments.*

As stated in a previous comment, the ongoing reorganisation seems a good start that has much improved the potential for internal communication and integrated projects across the division. We think that monitoring and discussing the progress should be one topic for the future ICAC meetings.

The vision presented, is to develop the Institutional Cluster as a collection of resources – HTC, HPC, Storage, and networking, with users interacting via a scientific gateway (HEPCloud). There is a need to provide a medium-term funding and effort profile for what the facility would like to do to implement the vision, and to support the approved scientific programme. In the experience of the committee it is essential to have a medium-term funding and resource planning outlook (5 years or so), which although it may be updated as the scientific programme evolves, is nevertheless essential to provide a baseline for planning of the facility.

## Separation of Environments

*It is essential to have an open, collaborative scientific environment, based on federated identities and trust with other national and international partners. For this reason, ongoing separation of business and open scientific environments is important and must be actively continued.*

We have seen no major issue here. In particular, the network structure in place, and the recent evolution presented, seems to support this. This is important also in that being able to provide an open scientific collaborative environment to a global community, requires that key internal



services (e.g. Business, and HR related services) need to be well isolated. The plans we have seen are well aligned with this aim.

## Student Programs

*We suggest to investigate having a coherent programme of summer students (or graduate students?) as a potential source of new recruits. Potentially in partnerships with universities, particularly local ones such as University of Chicago where many links exist. Having students and R&D illustrates some leadership capabilities.*

The summer student program presented seems to have been successful and we encourage Fermilab to continue these activities, investigating possible other channels. We are wondering if there is a scope for more students as 2 per year for computing seems on the low end for a large laboratory like Fermilab.

## Summary of Recommendations

- Continue to discuss with laboratory management the need for an explicit funding line for non-CMS computing operations.
  - Support this with a policy for the fraction of computing expected from outside Fermilab, such that Fermilab responsibilities are clearly stated.
- Work with other DOE national laboratories at CIO and Director level to address the policy aspects of new vetting process for foreign national computer access.
- Provide a plan for integrating the collaborative services with a federated identity, including authorization mechanisms.
- Provide resource loading plan to identify the resources needed to establish Fermilab as a host lab for DUNE.
- Clarify the long-term strategy for storage at Fermilab; and participate in the WLCG DOMA work for the future of distributed data management.
- Identify mechanisms to resource the sustain R&D capabilities in a strategic rather than opportunistic manner.
- Provide a 5-year (or similar) facility resource requirement plan, as a baseline for planning the ongoing facility development, and highlighting the required funding profiles.