US DUNE Computing and Software Review

Review Committee Findings

July 19, 2023 (final)

Overall Finding

Findings:

The DUNE Collaboration and US DUNE Software and Computing Consortium have done well in the first steps to identify the elements required for a successful Phase-1 experiment. They have also done well to identify international contributions for hardware resources. However, more work is needed for management to better identify, organize, and assign people to the large number of computing and SW tasks required.

Q1: *Does the Phase-1 DUNE Software and Computing system meet the needs of the collaboration?*

a. What are the milestones necessary to establish the reliability and readiness of the DUNE Software and Computing system? Are the requirements and timing of each milestone clearly defined?

b. What computing elements have international partners agreed to support?

Findings:

Milestones have been presented and are identified in the Phase-1 milestone document and in review talks. They are broken down into the 3 groups corresponding to the three major detector milestones. In the milestone document, the milestones and timelines are listed. However, the detailed timing, requisite activities and dependencies are not described.

Comments:

US DUNE Software and Computing should evaluate the strategy of contributing to most tasks and consider prioritizing support for US DUNE's high-priority tasks while seeking international support for others. US DUNE Computing and SW may want to consider identifying tools, services, or subprojects needed by DUNE that present the opportunity for international collaborators to support completely.

It is difficult to understand how the specific physics goal of the collaboration map onto the US DUNE Software and Computing systems technical requirements. One suggestion is to align software and computing activities and learn from physics needs by defining joint milestones related to the production of a "Physics Book" or other planned benchmarking of analyses before data taking.

Q1: *Does the Phase-1 DUNE Software and Computing system meet the needs of the collaboration?*

a. What are the milestones necessary to establish the reliability and readiness of the DUNE Software and Computing system? Are the requirements and timing of each milestone clearly defined?

b. What computing elements have international partners agreed to support?

Recommendations:

Addressed in other questions.

***Q2:** For US-supported elements of the Phase-1 DUNE Software and Computing system: a. What are the requirements of the Phase-1 elements? When must those requirements be met to achieve the Phase-1 needs? b. What is the current technical readiness? c. What future developments are necessary to meet the Phase-1 DUNE scientific needs? d. What opportunity is present in the development of the element? e. What risks are present in the development of the elements and the mitigation plan to reduce risks? f. What is the prioritization of the elements?*

Findings:

The US is expected to provide 50% of the total computing capacity in compute, disk storage, and tape archival. This expectation was essentially met during the last year.

The existing software and computing is expected to evolve to produce a system that meets the physics requirements for operations without beam by 2028, and with beam by 2031. The first high profile measurement with the former could be a supernova as early as day 1 of no-beam operations. First high profile measurement for the latter could be the neutrino hierarchy as early as 6 months after start of physics data taking with beam in 2031.

The project makes heavy use of common tools from across FNAL and HEP more generally, using DUNE effort to adapt these tools to DUNE needs.

The US DUNE Software and Computing Consortium is presently funded at the level of ~15 FTE. Most of the existing funding is set to expire within the next year. The project estimates that 25 FTEs are needed in 2024, and presented an effort envelope for the next decade or more that is essentially flat..

Q2: *For US-supported elements of the Phase-1 DUNE Software and Computing system: a. What are the requirements of the Phase-1 elements? When must those requirements be met to achieve the Phase-1 needs? b. What is the current technical readiness? c. What future developments are necessary to meet the Phase-1 DUNE scientific needs? d. What opportunity is present in the development of the element? e. What risks are present in the development of the elements and the mitigation plan to reduce risks? f. What is the prioritization of the elements?*

Comments:

The very fact that the US DUNE Software and Computing Consortium is capable of processing, simulating, and analyzing the existing prototype DUNE detector across a globally distributed computing system, including NERSC, indicates that they have significantly functional mature software, algorithms, services and computing infrastructure in operations today.

We consider their computing, disk storage, and tape archival storage planning as adequate for the stage the project is in. Their expectations to evolve this plan as their knowledge about input and the model itself evolves are as expected.

Using common tools from across FNAL/HEP and adding DUNE adaptations is a good strategy for sustainable software development. However, dependences from and requirements for external software deserves more attention. Work with external collaborations developing tools DUNE depends on to understand and negotiate plans and deadlines and, if necessary, make the in-kind contributions to ensure tool capabilities are available for DUNE on time.

There are essential tools, such as visualization end analysis tools, that do not currently appear in the DUNE Software and Computing planning. A complete survey of required software and a plan to identify responsible institutions within DUNE would be helpful.

Beyond the functionality of the existing systems and the integration validation via data challenges, it was not clear how physics goals map onto technical requirements for the algorithms, software, and system and its development. We are thus unable to answer the charge question regarding requirements for the Phase 1 minimal system, and what technical developments are necessary to meet the opportunities for early physics in either 2028 or 2031.

If funding for the increase in effort is delayed, US DUNE Software and Computing expects a large bulge in staffing in a later fiscal year needed to catch up, and deliver the current system in time for the start of Phase 1 data taking in 2028. This planned response to increase effort in later years may not be able to meet the milestones and timelines necessary to deliver the planned system. The US funding agencies should work to meet the project’s SW and computing staffing needs and US DUNE Software and Computing Management should consider how to respond based on the prioritized list of projects provided in a way that minimizes the impact to DUNE physics goals.

***Q2:** For US-supported elements of the Phase-1 DUNE Software and Computing system: a. What are the requirements of the Phase-1 elements? When must those requirements be met to achieve the Phase-1 needs? b. What is the current technical readiness? c. What future developments are necessary to meet the Phase-1 DUNE scientific needs? d. What opportunity is present in the development of the element? e. What risks are present in the development of the elements and the mitigation plan to reduce risks? f. What is the prioritization of the elements?*

Recommendations:

While we acknowledge the limitations of the current funding level, we recommend the project re-prioritize or add effort such that an appropriately detailed plan for how physics goals flow down to technical requirements on software & computing can be produced by February 2024. Such a plan should become the basis for future reviews, and be refined as progress is made.

US DUNE Software and Computing management should develop a plan, by February 2024, for how to respond to future software tasks and projects that arise but not currently identified.

US DUNE Computing and SW should clearly identify the minimal functioning system required to support US DUNE physics needs by February 2024.

US DUNE Computing and SW has done a good job identifying upcoming staffing needs and should develop a more formal plan for the staffing needs to be identified by February 2024. This plan should include contingencies if support for the desired staffing is not available.

Q3: *How are the US-supported computing elements managed? a. How is progress of the elements tracked and assessed? b. How does the US DUNE Software and Computing management respond to opportunity and risk? c. Are US DUNE Software and Computing responsibilities well-coordinated with international collaborators' responsibilities and other related computing efforts? d. Are management lines of authority and available resources clear?*

Findings:

US DUNE Computing and SW has worked hard to develop relationships with international partners, especially the close effort with the UK which can serve as a model for other relationships.

We were provided an impressive list of software that DUNE is leveraging from FNAL, the HEP community, the broader scientific research community and industry.

Data Management and Production are organized in teams with members from all DUNE regions (US and international), with large use of collaborative tools. US and non-US technical leads in the DUNE Computing Consortium both report to the consortium lead.

There is currently no formal agreement for computing institutional commitments. For hardware there is an informal pledging process in place and it is planned to have it more formal oversight through the Resources Review Board. For personpower, there is a table maintained by resource coordination recording the FTE effort in each area including software and computing.

Risks linked to dependencies from external (non-DUNE) components are partially understood and so is the mitigation strategy. A small number of components are under the sole responsibility of non-US DUNE institutes.

Resource needs estimates lack information to evaluate need-resource gaps, as well as the impact of R&D.

Q3: *How are the US-supported computing elements managed? a. How is progress of the elements tracked and assessed? b. How does the US DUNE Software and Computing management respond to opportunity and risk? c. Are US DUNE Software and Computing responsibilities well-coordinated with international collaborators' responsibilities and other related computing efforts? d. Are management lines of authority and available resources clear?*

Comments:
We emphasize the importance of an appropriate level of formal management for Software and Computing, including dependencies, decision-making and risk characterization, to enable the experiment to maintain an appropriate level of control and overview.

The US DUNE Computing and SW Consortium is encouraged to explore and identify the most efficient funding method and work with the funding agencies to develop the appropriate plan.

Suggestion: Add to resource projection plots the impact of R&D in reducing system needs, mitigating risks, as well as computing budget assumptions as a baseline to measure potential resource gaps.

Recommendations:
A more formal management structure appropriate to a collaboration of this size should be developed by February 2024. This structure should ensure progress tracking, assessment of opportunities and risks, coordination and communication with other US DUNE consortia and international collaborators, and clearly define management lines of authority. Additional responsibilities and needs are detailed in Recommendations to other questions.

Q4. *What is the strategy for training, mentoring, and retaining a diverse pool of early-career scientists with sufficient technical expertise to maintain the US-supported elements over the life of the DUNE collaboration?*

Findings:

An innovative and creative scheme for leveraging the needs of DUNE and the opportunities it represents in order to recruit and train a broad range of early-career scientists and students is presented. The strategy for mentorship, especially of a one-to-one nature, and retention in general, appears to be somewhat less clearly fleshed-out.

The inclusion of smaller institutions to leverage their expertise was noted as a distinct advantage to the US DUNE Computing and SW Consortium.

There is a core team of experts working on critical areas of SW and computing that are now hired through the DOE grant. There is currently no long-term plan to retain this expertise.

Spending that is explicitly targeted towards specific younger cohorts is proposed: summer internships; undergraduate research positions; graduate student secondments to labs; postdoc ‘pipeline’, and funded recruitment visits.

Q4. *What is the strategy for training, mentoring, and retaining a diverse pool of early-career scientists with sufficient technical expertise to maintain the US-supported elements over the life of the DUNE collaboration?*

Comments:
The relatively small amounts of funding that are sought to support engagement with students and early career scientists will have a very high positive impact, both in terms of PIER as well as for the DUNE project in general, and should be pursued vigorously.

A survey of needed expertise and skills would be valuable in long-term success of the project.

The PIER Plan includes many valuable descriptions of principles and implementation ideas, but as a document, contains much room for improvement in style and organization. The individual institutional contributions are somewhat varied in quality, and it would be helpful if the institutions, and DUNE as a whole, could improve their plans by learning from each other's contributions.

Recommendations:
Bring in systematic tracking mechanisms for the progression of individuals at all stages; from undergraduates who are present at recruiting visits and research positions, to graduate students and postdocs who receive training, to quantify the impact of the strategies being pursued and to allow them to be improved.

Similarly, the tracking of the number of active instructors, and the development of new instructors, should be instituted such that their number and quality grows in proportion to the needs of the collaboration, rather than burdening the same individuals more and more.

Means to track both of these should be implemented prior to February 2024.

Q5. *What is the process and timeline to develop the DUNE Collaboration's software and scientific data Public Access Plan?*

Findings:

A draft plan for Public Data Access was presented and looks to be well done and based on other experiments' experiences and practices (e.g. microBooNE).

Comments:

Public data access was presented with data at 4 levels. The plan for level 1-3 data is solid, well founded and looks consistent with the FAIR principles and, likely, new DOE policy. It is also good practice for good science. Level 4 is a bit different in that it is hard to understand the need for the public to have data at the rawest level. The example from microBooNE seems a good way to satisfy the requirements, without too much wasted collaboration effort.

A near term opportunity for exploring the draft plan in practice would be the ProtoDUNE data taking and analysis.

Recommendations:

This data access plan should be further pursued to refine and to finalize with the DUNE collaboration.

The collaboration should further check that this policy is consistent with DOE guidance on the Nelson memo on public data access prior to February 2024.

Thank you

The review panel would like to thank the US DUNE Software and Computing Consortium for all their hard work in preparing for this review, and for the enlightening discussion!