

Response to Questions and Comments

USQCD/LQCD team

LQCD extension III Review
Cambria Rockville Hotel | July 9–10, 2019



Project Management

- *Project management structure, esp. the execution plan, seem to be based on the dedicated cluster model. How will PM evolve as institutional clusters mature? Is a "project" necessary?*
- *Response:*
 - the project management is 0.5 + 0.1 + 0.1 FTE (the two small parts for site management); expect 0.5 → 0.4 in ~1 year;
 - there are many tasks best suited to professionals: negotiating MoUs w/ labs, monthly tracking of IC usage, budget planning, annual survey;
 - instead of fully designing and procuring, the site architect is now part of a design team—LQCD will pay its share; (procurement centralized);
 - the work of procurement, *etc.*, is replaced with the MoU process.

- *Response* (con't):
 - That said, the details will certainly evolve.
 - We're choosing to discard PM procedures when they are obsolete:
 - *e.g.*, we retired the "Change Control Board" that was in place for the dedicated-cluster era.
 - the project team is aware of some aspects of the execution plan that are candidates for "early retirement".

Funding for Storage

- *How much more storage makes sense? How much would it cost?*
- **Response:** We address archival tape storage, which is an important growing challenge for USQCD and the topic raised during yesterday's discussion.
 - Two examples of RBC (compressed) eigenvector data that would be reused multiple times over 5 years:

Lattice	# eigenvectors	Computing cost	Storage
$64^3 \times 128$	2000 \times 200 configs	105 M BG/Q core-hrs	0.8 PB
$96^3 \times 192$	5000 \times 50 configs	300 k Summit node-hrs	1.0 PB

- MILC requirements are expected to be similar.
- 4 PB storage cost would be \$206K/year (average of BNL and FNAL costs).
- This storage resource would be allocated annually by SPC.

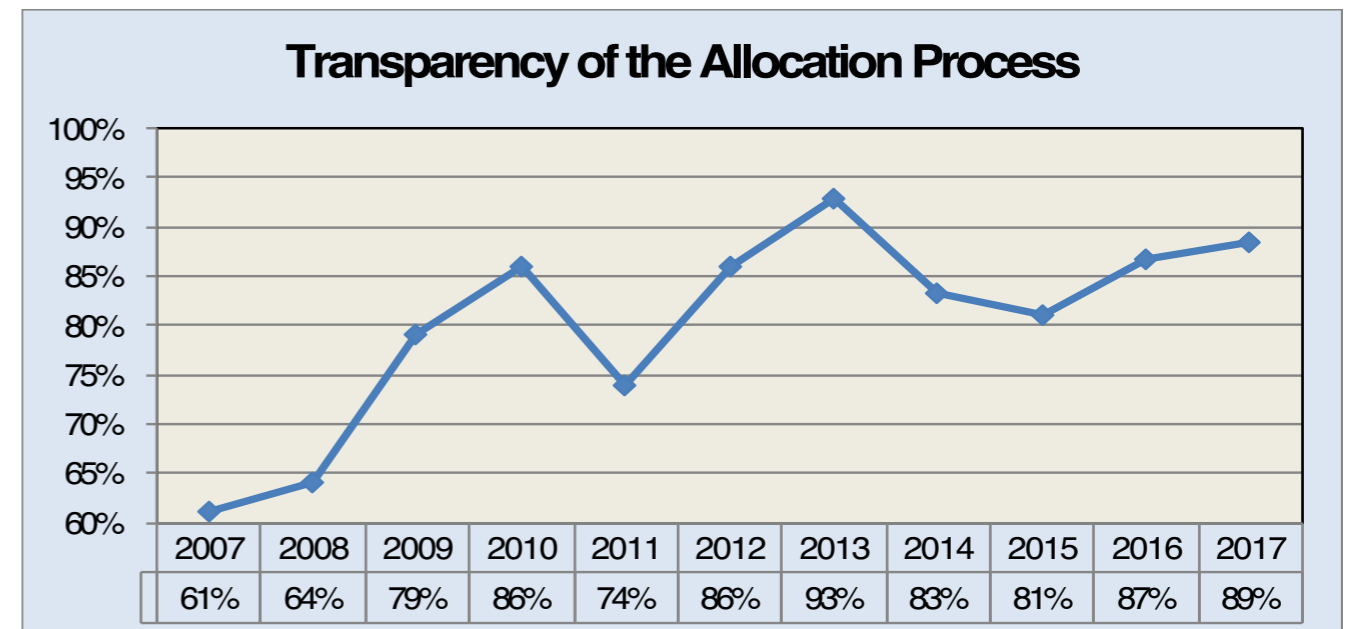
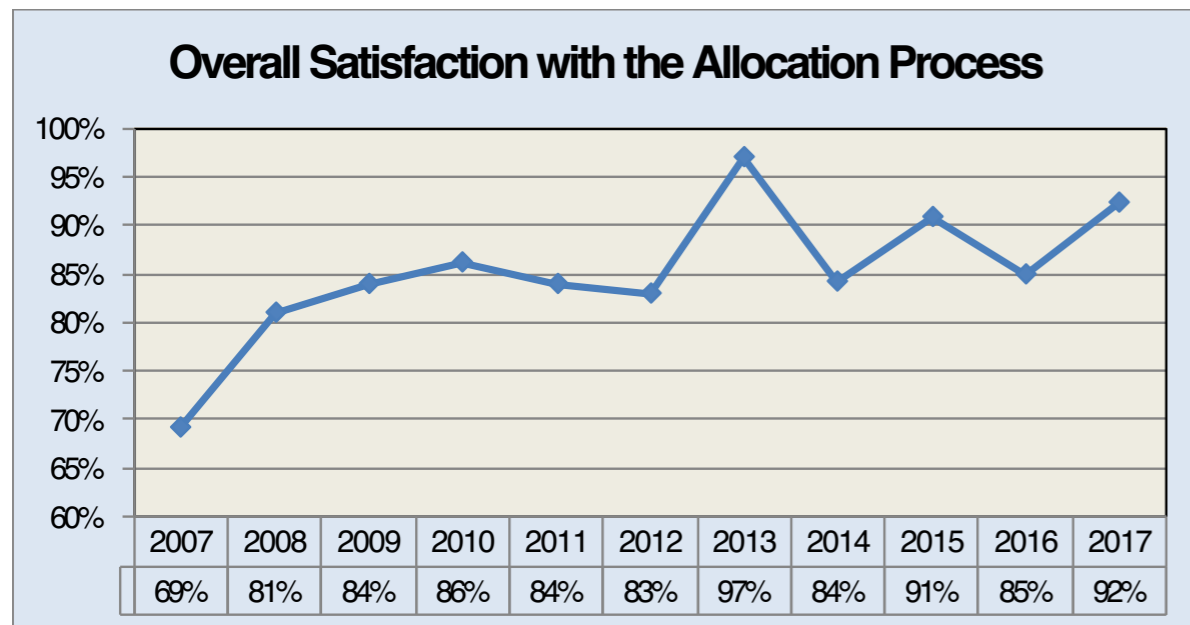
Data Sharing

- *How is data shared now? How will new data be shared? Who outside of USQCD uses data produced within the collaboration?*
- *Response:*
 - USQCD and the labs (FNAL+BNL+JLab) are developing a data management framework. It starts with “best practices” for managing data. The facilities will provide a unique namespace for data (e.g., directory structure) that will also be usable as a unique URI for the data. The data structure will match that on tape. The users will register the URI matched to the datasets.
 - Once datasets have a URI, they can be registered as a publication under SPIRES and citations will be tracked. Phys. Rev. now supports citing a URI and will list the URI as part of the bibliography.

- Data currently are transferred with Globus. The typical pattern is that data is pulled from tape (if stored there) and held on disk. A consumer will need an account at the facility holding the data, and an account at the target destination for the transfer. We expect new data will also be transferred via Globus.
- Who outside of USQCD uses data produced by the collaboration?
 - USQCD has a policy that configurations generated using collaboration resources should be available promptly to members of the collaboration, in particular at most 1 year after first journal publication using the data, the configurations will be made public
 - For more complicated data, such as propagators, there is no standard policy on sharing. These data are much more specialized. Discussions between the producers and consumers are carried out on a case-by-case basis.
 - Details of new types of data (e.g., Dirac eigenvectors) on a new storage facility will fall somewhere between the two.

User satisfaction with allocation process

- *Why are some users dissatisfied with the allocation process?*
- **Response:** Results from user survey regarding allocations:



- "2017" survey results were obtained from March-May 2018, looking backwards. Decided to delay next survey to make it more prompt: next survey planned in August 2019 for FY18.
- Specific comments made in the last survey requested more transparency in terms of disclosing scientific priorities.
- Another comment requested more clarity in terms of assignment of allocations to sites; in 2019 a new web form for proposals was released, including explicit rank-ordering of site preference.

Allocation procedure

- Call for Proposals sent to membership e-mail list beginning of February 2019; deadline of March 4.
- One round of questions sent to all proponents from the SPC, to allow clarification of proposal before final decisions.
- *New*: SPC reports publicly on the decision process at the USQCD All Hands' Meeting (AHM), as well as reporting on proposals received and science priorities by subfield (flavor physics, nuclear, thermodynamics, BSM...)
- The AHM is live-streamed for remote participation by all members of USQCD (at least since 2015).
- All proposals & allocations are released to members of the collaboration; overlooked for a couple years (2015–2018) and partly fixed; hasn't been done yet for 2019.
- Open call for suggestions to improve the process was made this year at the AHM; no responses were given.

Design Control

- *USQCD will have less control over the design of institutional than dedicated clusters. What will USQCD/LQCD do if a lab's IC no longer meets needs for lattice QCD?*
- *Response:*
 - Such divergence seems unlikely—BNL and Fermilab's philosophy behind the institutional clusters is that the hardware should encompass all elements of HPC: in particular GPUs and other accelerators.
 - Nevertheless, if that were to change at one of the labs, USQCD could send all LQCD ext. III funding to the other one.
 - If both labs were to abandon ICs, then we could go back to the dedicated model.

Questions and Discussion