

# The Capability Release: A Process to Coordinate the Integration and Deployment of ASCI Supercomputing and Network Technologies

James A. Ang, Martha J. Ernest, John P. Noe  
Sandia National Laboratories<sup>†</sup>

Michel G. McCoy, Dave Wiltzius  
Lawrence Livermore National Laboratory<sup>‡</sup>

Richard A. Light, Stephen E. Turpin  
Los Alamos National Laboratory<sup>§</sup>

## I. Objectives and Definition of the ASCI/DisCom<sup>2</sup> Capability Release

The Accelerated Strategic Computing Initiative (ASCI) is a U.S. Department of Energy program among Lawrence Livermore National Laboratory (LLNL), Los Alamos National Laboratory (LANL) and Sandia National Laboratories (SNL). The ASCI program has worked with the High-Performance Computing and Communications community to develop terascale computing technologies and systems to help ensure the safety and reliability of the U.S. nuclear weapons stockpile without full-scale testing. The Distance and Distributed Computing and Communications (DisCom<sup>2</sup>) program element of ASCI has two key goals: (1) extend the environments required to support high-end computing to remote sites, and (2) develop a complex-wide integrated supercomputing environment to support stockpile stewardship. A major challenge associated with these DisCom<sup>2</sup> goals is coordinating and integrating technology deployments among the existing Tri-Lab production computing centers and maintaining their interoperability.

The primary objective of the Capability Release is to ensure ASCI developed technologies are integrated and deployed into the ASCI user environment with the support and commitment from the Tri-Lab production computing organizations. Closely aligned with this objective is the need to have an operational baseline of integrated capabilities to support meaningful discussions with the user community about the prioritization of improvements to that environment. While many ASCI technology development projects work (by necessity) on stand-alone activities, the real impact of these technology development projects is only seen after they are successfully integrated into and supported by the user's production computing environment.

---

<sup>†</sup> Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

<sup>‡</sup> This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

<sup>§</sup> Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

The definition of the capability release has been the subject of much Tri-Lab discussion. An initial strawman definition was, "A Capability Release (CR) is an Integrated Set of Technical Resources with "Guaranteed" Level of Reliability, Availability & Supportability (RAS)." There has been much discussion over the use of the term "guaranteed". Some of the production computing managers are uncomfortable with the notion that we can really provide guaranteed levels of RAS. On the other hand it is this very notion that helps define the scope of what the CR is going to be able to provide. It conveys an important message to the ASCI technology development community. But maybe it limits the scope of the CR too much. The second option was to replace "guaranteed" with "known". However this term seems too soft. The third option was to indicate a commitment on the part of the production computing organizations. Another option is to replace "guaranteed" with the term "production quality". But the argument has been made that ASCI systems never attain true production quality status, the best we can achieve is "general availability" levels of RAS. If we can equate this level with "ASCI production quality", the following is the current CR working definition,

*A Capability Release (CR) is an Integrated Set of Technical Resources with a Commitment to Support "ASCI-Production Quality" Levels of Reliability, Availability & Supportability*

**II. CR Scope: Technology Maturity Levels and Tri-Lab Focus**

Closely related to the CR Definition is the scope of technology maturity levels that the Capability Release will work with. The following availability definitions are based on Livermore Computing's treatment of different levels of availability and support for the ASCI White platform. For consistency the CR process is adopting and promoting these same definitions for other technology deployments into the Tri-Lab ASCI user environment.

<b>Livermore Computing Definitions of Technology Availability</b>	<b>CR Scope</b>
<p><b>General Availability (GA)</b>                      The Program (working with the Center) has a sufficient level of confidence in the functionality, reliability and security of the computer system and its operating environment to allow access to all users who have a programmatic need. In such an environment, it is possible, in general, to schedule downtimes with adequate lead time for users to plan their work schedule with confidence. However, given the leading edge nature of the computing environment, there is an expectation of significantly more downtime due to software and hardware upgrades and debug shots than is traditional in a production environment. It is also likely that certain system features necessary for a balanced, production quality environment, for example interconnect bandwidth, MTBF, resource management and scheduling software, and I/O bandwidth may not be at final specifications, but rather will evolve in time according to development schedules.</p>	<p>GA is the goal for ASCI CR-deployed technologies. At this level of technology maturity the CR is primarily concerned with fostering interactions with the end-user community, with the object of identifying and prioritizing improvements to the ASCI user environment for future CRs.</p>
<p><b>Limited Availability (LA)</b>                      The Program (working with the Center) has a sufficient level of confidence in the functionality, reliability and security of the computer system and its operating environment to allow access to a limited (small) set of users and projects working towards programmatic deliverables. This would probably mean two ASCI code groups per lab accessing the system remotely. There would be weekly teleconference calls scheduled to discuss prioritizing the</p>	<p>LA deployments are an important part of CRs. These may be an initial deployment that requires implementation in the production environment. Promotion to GA pending</p>

<p>jobs. While attempts will be made to schedule downtimes, it is <u>not</u> possible, with high probability, to schedule these with adequate lead time for users to plan their work schedule with confidence. In such an environment, it may be necessary that users and teams coordinate their major calculations with the computer center in order to enhance chances of a successful outcome. A number of tools that users expect to be routinely available may not yet be in place, and workarounds will need to be coordinated with the Center.</p>	<p>acceptance tests, documentation, training materials, and orientation of consultants, help desk and user support organizations to these new capabilities.</p>
<p><b>Beta Availability (BA)</b>  The Center does not have sufficient confidence in the functionality and reliability of the computer system and its operating environment to allow teams on the machine focused on programmatic deliverables. However, it is possible to allow a few friendly users on the system who can run problems and do some work with the understanding that the environment is extremely volatile and that their presence is encouraged primarily to aid the Center in bringing the machine to Limited Availability. A number of tools that users expect to be routinely available may not yet be in place, and workarounds will need to be coordinated with the Center. Remote access is allowed.</p>	<p>These technologies are available as extensions to the GA/LA CR foundation. At this technology maturity level the CR is primarily concerned with fostering communication between the technology developers and the production computing community.</p>
<p><b>Alpha Availability (AA)</b>  The System is undergoing continuous development and integration. In this environment users run primarily to gain some limited knowledge about system performance and to aid the Center in diagnosing problems. Very few ASCII accounts are allowed per lab and no remote access is allowed (users will have to be on site to access the machine).</p>	<p>Generally not part of the CR, AA technologies should reside back in the development environment.</p>

The Capability Release is intended to facilitate the interactions between the three labs' production computing centers to ensure integration of capabilities. The primary focus for CR's is for Tri-Lab capabilities. Of secondary importance are those capabilities that are deployed only between two labs. Site-specific capabilities are only important for CR's to the extent that they set the boundary conditions for Tri-Lab integration, or may be affected by the need for Tri-Lab integration. To the extent possible, the CR process seeks to work through the Configuration Management Board (CMB), and the existing organizations and programs that are responsible for the administration, operations, and maintenance of the local-site computing resources.

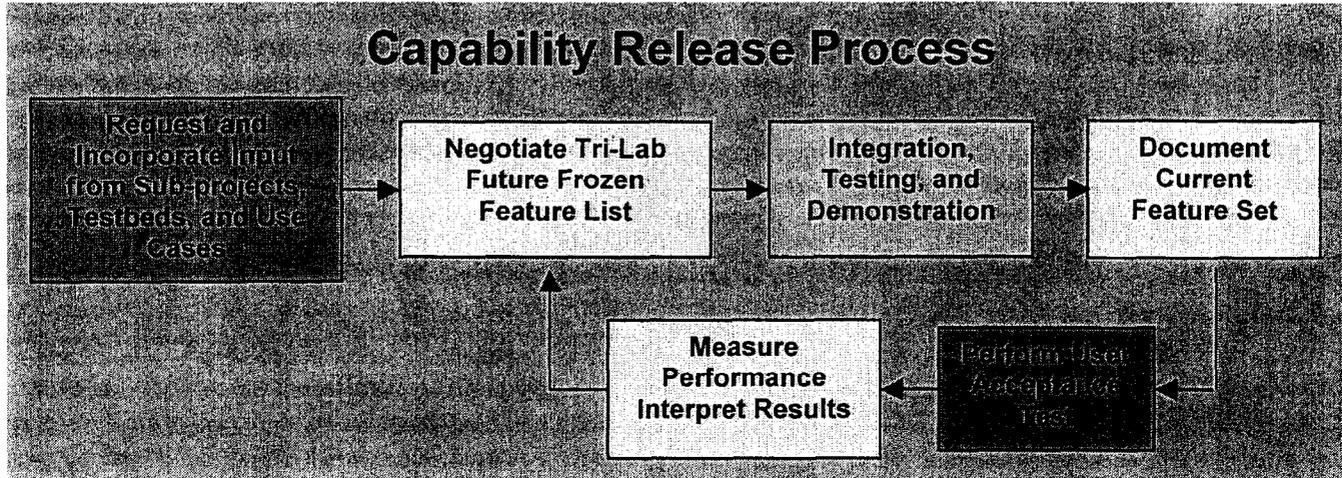
### III. Support for DisCom<sup>2</sup> Objectives

Each of DisCom<sup>2</sup>'s major projects develops and deploys capabilities that must be integrated into a coherent system. To ensure the integration and to focus the projects and subprojects of DisCom<sup>2</sup>, we have instituted a capability release cycle that started in FY00 that spans the entire DisCom<sup>2</sup> program. Every six months the ready capabilities are being aggregated into a system release. The act of defining and documenting the features and support for the release provides a higher level of formalism to the program. The CR features list provides the user community with a concrete specification of what is available. Each release also provides a baseline of features from which future improvements can be negotiated and prioritized among the technology developer and user communities. Prior to the capability release, a feature freeze occurs, followed by integration testing and then transition to operations. Each of the DisCom<sup>2</sup> projects contains milestones consistent with this release cycle. The capability releases provide a distributed computing system with defined functionality and features that will, in their full

deployment to meet the DC-3.1 milestone and other high-level milestones, span the weapons laboratories and production plants.

#### IV. Capability Release Process

Each CR is the product of a systematic process of requirements definition, incorporation of input from technology development projects, testbeds, use cases and user feedback, negotiation of frozen feature list, system integration, integration testing, CR feature demonstrations, user acceptance testing, documentation, user support, and deployment. A simplified schematic diagram of the CR Process is shown below.



#### High Level CR Process Elements

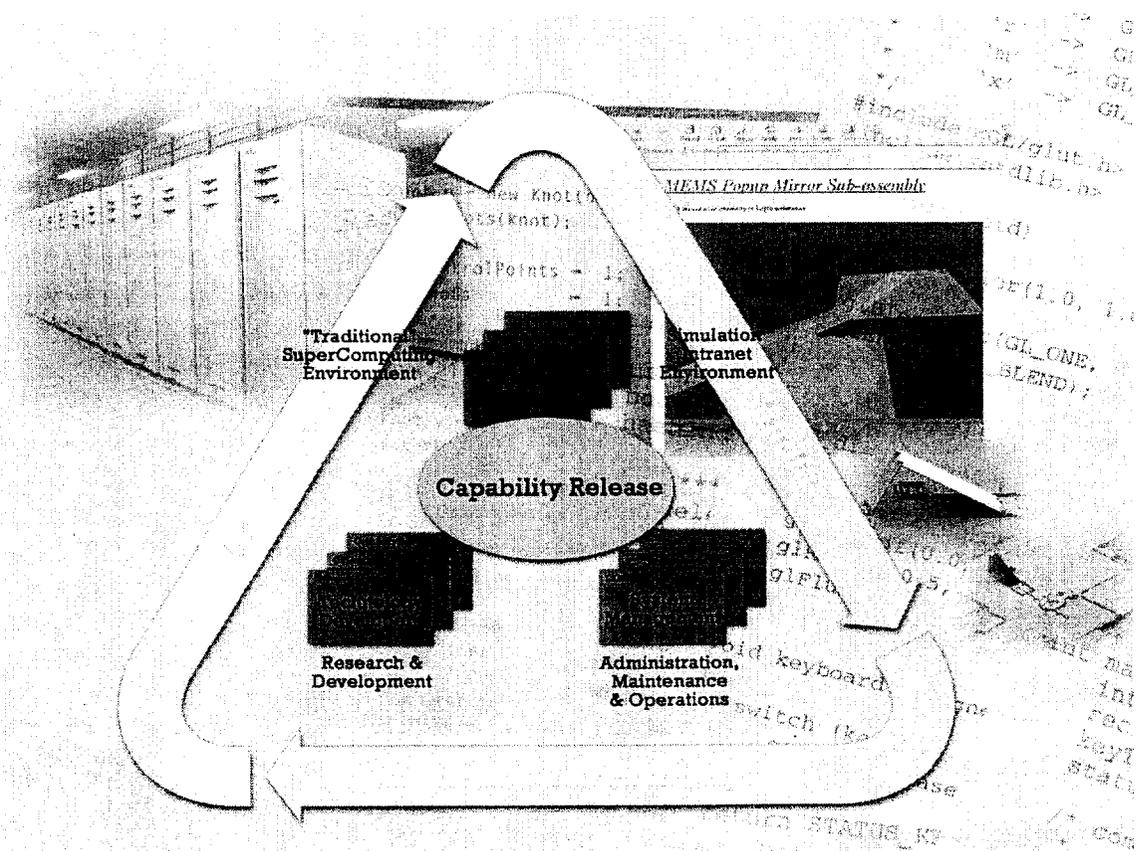
1. Request and incorporate input from sub-projects, testbeds, and use cases
  - a. Input from Technology Development Projects for Capability Feature Set
  - b. Flow-down of system requirements from Program Milestones to technology development projects
  - c. Input from Technology Development Projects for Integration into and overarching System Design/Architecture
  - d. Integration/coordination with existing production computing processes
  - e. Feedback from User Community
2. Negotiate Tri-Lab future frozen feature list
  - a. User Ranking of Features
  - b. Proposed Feature Assessment
  - c. Negotiation of Frozen Feature List
  - d. Tri-Lab Frozen Feature Agreements
3. Integration, testing and demonstration
  - a. Integration Phase
  - b. Test Suite Development
  - c. System Test Phase
  - d. Transition to Production
4. Document current feature set
  - a. Description of new features and extensions to the previous baseline
  - b. Include targeted feature availability level: BA, LA, GA, pending results of UAT

5. Perform User Acceptance Test (UAT)
  - a. Analyze New Features
  - b. Establish Requirements & Use Cases
  - c. Develop Test Scenarios & Test Plan
  - d. Conduct UAT (including deficiency corrections and retesting as needed)
6. Measure performance interpret and publish results
  - a. Document and Review Test Results
  - b. Including promotion of feature availability levels (BA→LA, LA→GA), or
  - c. Possible demotion of feature availability levels (GA→LA, LA→BA)

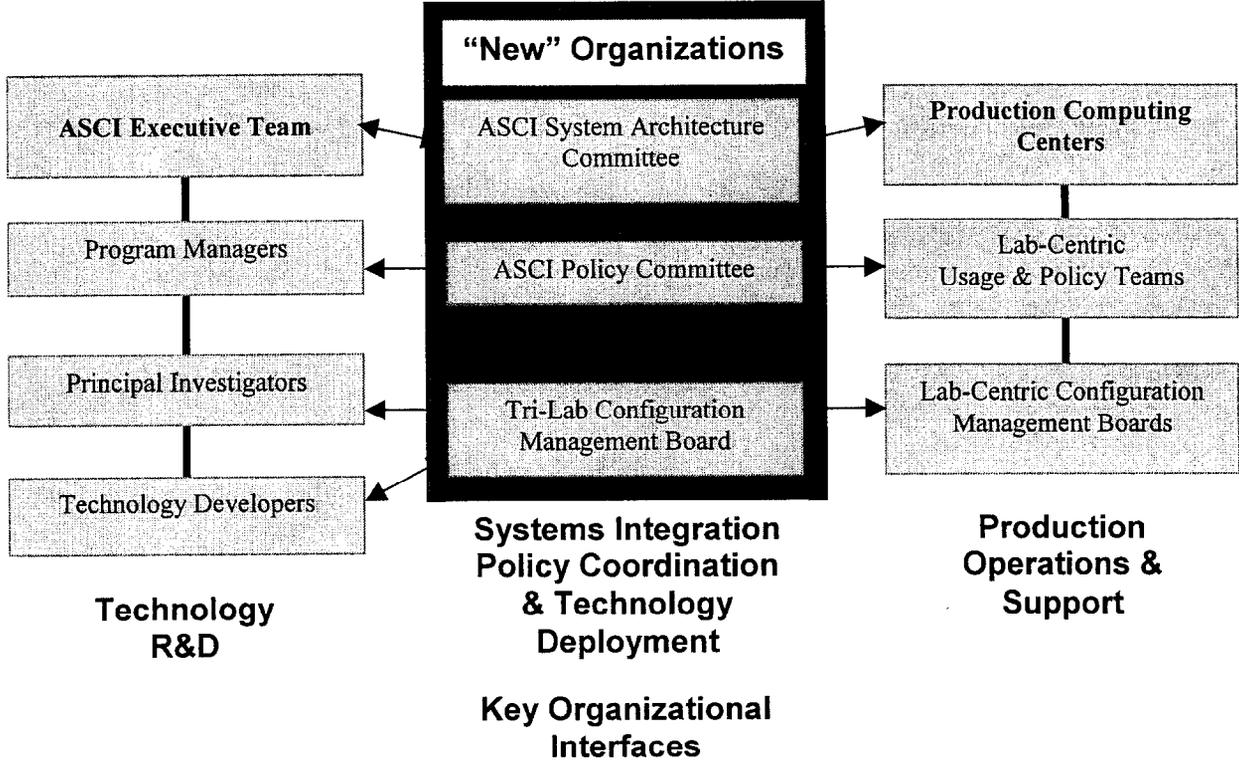
#### IV. CR Process Roles, Responsibilities, and Communication Interfaces

Within the DisCom<sup>2</sup> scope the Capability Release process is expected to be implemented by an integration team that is composed of participants from within DisCom<sup>2</sup> (the Tri-Lab CR process owners) as well as participants from within the Tri-Lab production computing centers, primarily CMB members and PI's from relevant ASCI technology development teams. The primary interface between the CR process and the Tri-Lab production computing centers will be provided through the Tri-Lab Configuration Management Board.

The primary role for the CR Process is to ensure new ASCI technologies are integrated with existing resources into the nuclear weapons complex user environment. This integration requires communication and coordination among the technology development, system management and user communities across the three labs. A notional schematic of this communication and coordination is shown in the figure below.



The effort to integrate and deploy ASCI resources to the Tri-Lab user community has led to the need to define a number of new teams to provide coordination and resolution of Tri-Lab policy and operational issues. The new Tri-Lab governance teams and a rough summary of key interfaces is illustrated in the schematic below.



**V. Planning, Evolution & Integration of New CR Features**

An important role for the CR process is to provide some systematic review for how new features are proposed and introduced into the ASCI user environment. There are at least three ways needs for new features are identified. Feedback from the user community for new features or modifications of existing features should be of top priority. New features are also introduced by the various ASCI technology development projects. Finally, new features may be introduced from a top down need based on assessment of technology requirements to meet ASCI Program Milestones.

To facilitate the gathering of needs and requirements from the user community the CR process includes holding user forums, coordinating the inclusion of information on new features into existing training materials and on-line documentation. Members of the CR integration team also monitor trouble tickets and user support requests to identify needs and user environment issues.

The CR process effort to understand of new feature requirements that flow down from ASCI milestones began with the FY01 ASCI Implementation Plan. The initial effort was focused on laying out the DisCom<sup>2</sup>, PSE and VIEWS milestones against a common timeline for the next few years to help us identify inter-program linkages and synchronizations that will be required to deploy integrated capability to the ASCI user community. The next step will be to extend this mapping to the ASCI Applications milestones.