NWSC Allocations Plans

David Hart
User Services Section Manager
April 21, 2011

Big changes in store…

- NWSC timelines

- NWSC resource division

- Accelerated Scientific Discovery

- Allocable resources and CHAP impact

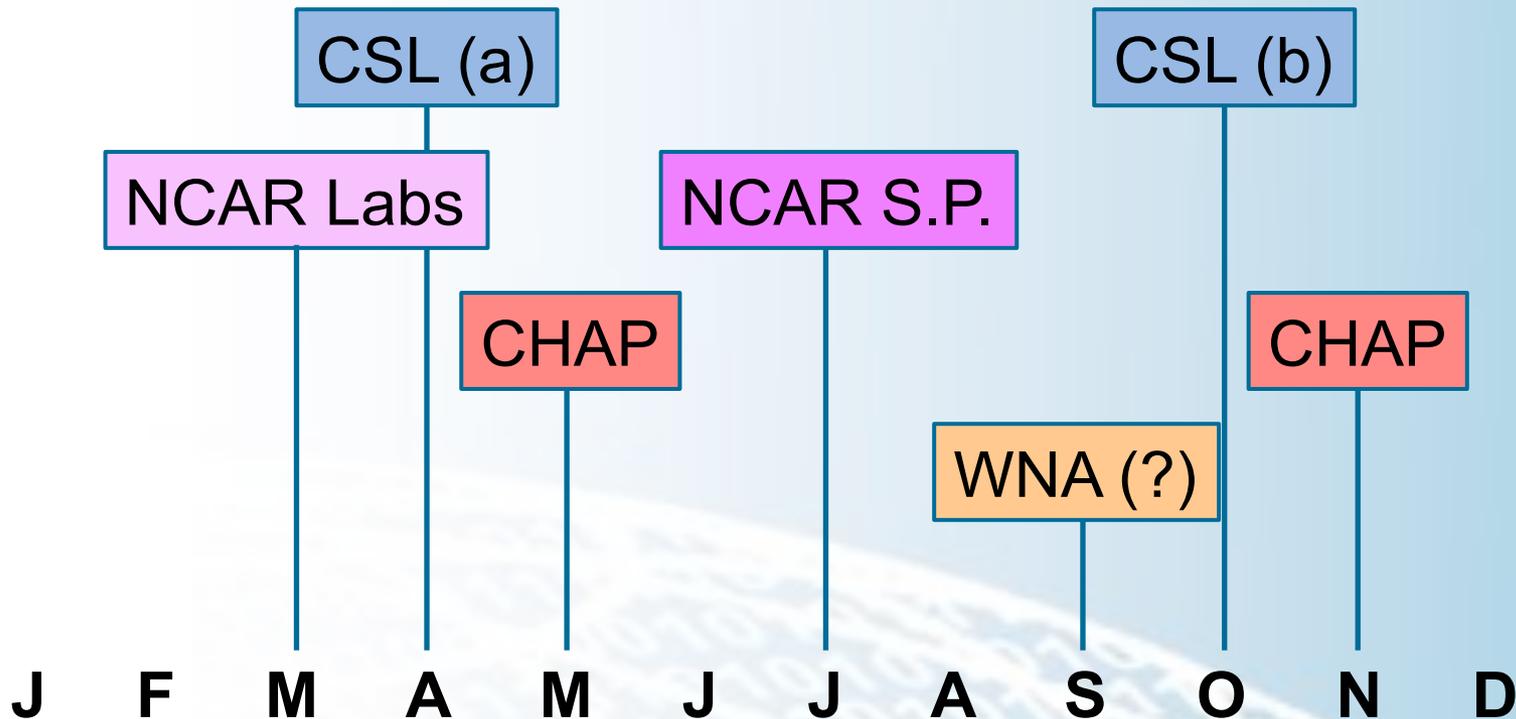- Data management plans and allocations

NWSC timelines

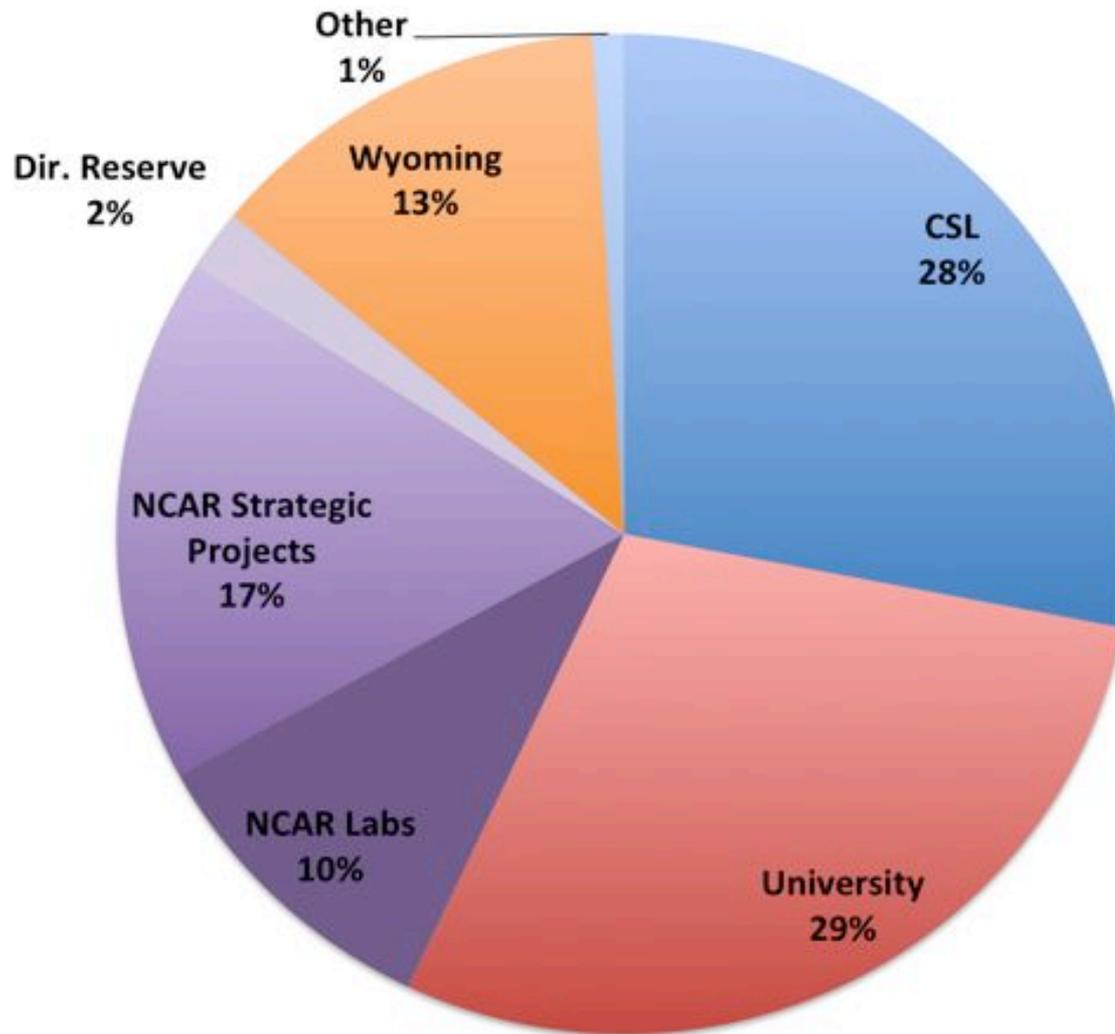| Date	NWSC milestone	Allocations event
ca. Sept. 2011	Vendor subcontract award	
Oct. 2011	NCAR/vendor announcement	Fall CHAP meeting
Jan. 2012	First hardware deliveries	CSL meeting (at AGU)
Feb. 2012?		Univ. ASD request review; NCAR Strategic Projects (NSP) review
Mar.-Apr. 2012	System acceptance	Spring CHAP meeting; NCAR Lab Allocation review; U Wyoming alloc. review (tbd)
Apr.-June 2012	Pre-production use	NSP, ASD, CSL(?) projects begin
June-July 2012	Production use	University, NCAR labs, CSL(?) projects begin

NWSC timeline issues

- Public announcement will happen same time as CHAP submission deadline in Sept 2011.
  - Recommend that CHAP requests be instructed to offer a "multi-year plan" with Year 1 = Bluefire (Nov.-Jun.) and Years 2+ = NWSC.
  - CHAP may need to balance two sets of numbers.
- Production date remains tentative, but conservative
  - CISL will be working to manage expectations and educate users in advance about changes or delays.
- After NWSC start, ultimately would like to stagger start dates of the 6 different allocation facilities.

Possible future allocation timeline

CSL (a)

CSL (b)

NCAR Labs

NCAR S.P.

CHAP

CHAP

WNA (?)

J F M A M J J A S O N D

Prelim. NWSC resource division

Other 1%

Dir. Reserve 2%

Wyoming 13%

CSL 28%

NCAR Strategic Projects 17%

NCAR Labs 10%

University 29%

Precise percentages subject to change.

Transition to NWSC-1

- CHAP projects awarded October 2011 will be given their "Year 2+" allocations to start on NWSC-1 production.
- CHAP projects awarded pre-October 2011 with (substantial) available GAUs will be carried over to NWSC-1.
- Small University projects may only be carried over upon request by PI or project lead.
- Current CSL awards "right-sized" to end of Bluefire.
  - If Bluefire retired early, CSL projects will be transitioned to NWSC-1.
- Early use period, ramp-up time to be leveraged for…

Accelerated Scientific Discovery

- Early access for small number of large-scale projects on post-acceptance NWSC resources.
  - DAV and GLADE space also available.
- ***How to manage process? Criteria for review?***
  - ***Need CHAP input here.***
  - ***Anticipate small number of awards.***
  - ***Announce in October-November, submit in January, review in February, begin in April.***
- On NCAR side, ASD rolled into NCAR Strategic Project (NSP) requests.
  - Reviewed by NCAR panel, approved by NCAR Executive Council.

ASD leverages early use, ramp up
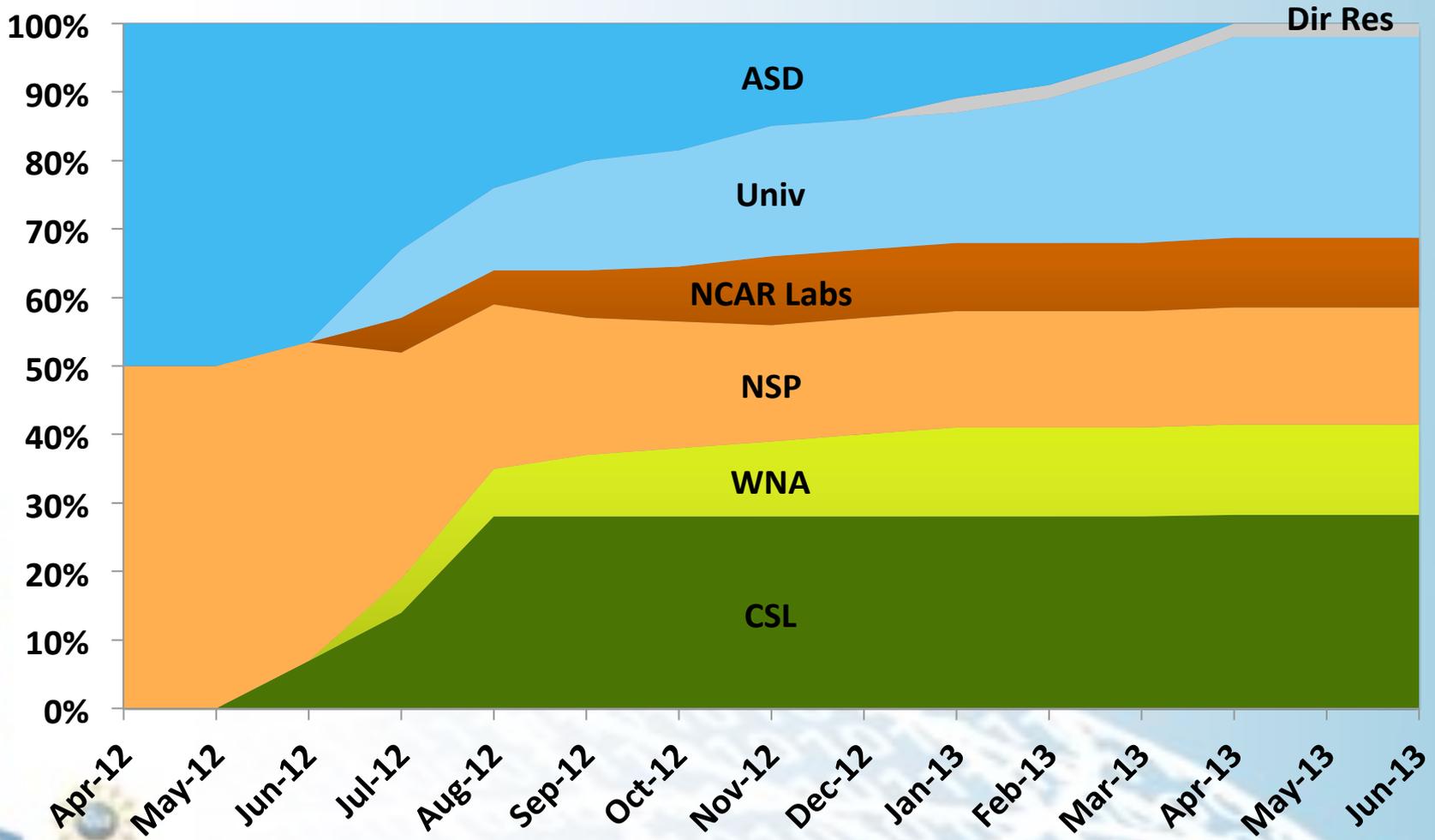
NWSC allocable resources

- With NWSC production, we will expand allocation processes to encompass the non-HPC resources

  - *Disk, tape and DAV resources represent more than 1/3 of CISL's financial investment.*

- The CISL objective is to ensure the full resource environment is used most effectively by the most meritorious projects.

- We want to balance optimal resource use with minimal reviewer and submitter burden.

- Submitters will need better guidance on what's needed in their data management plans.
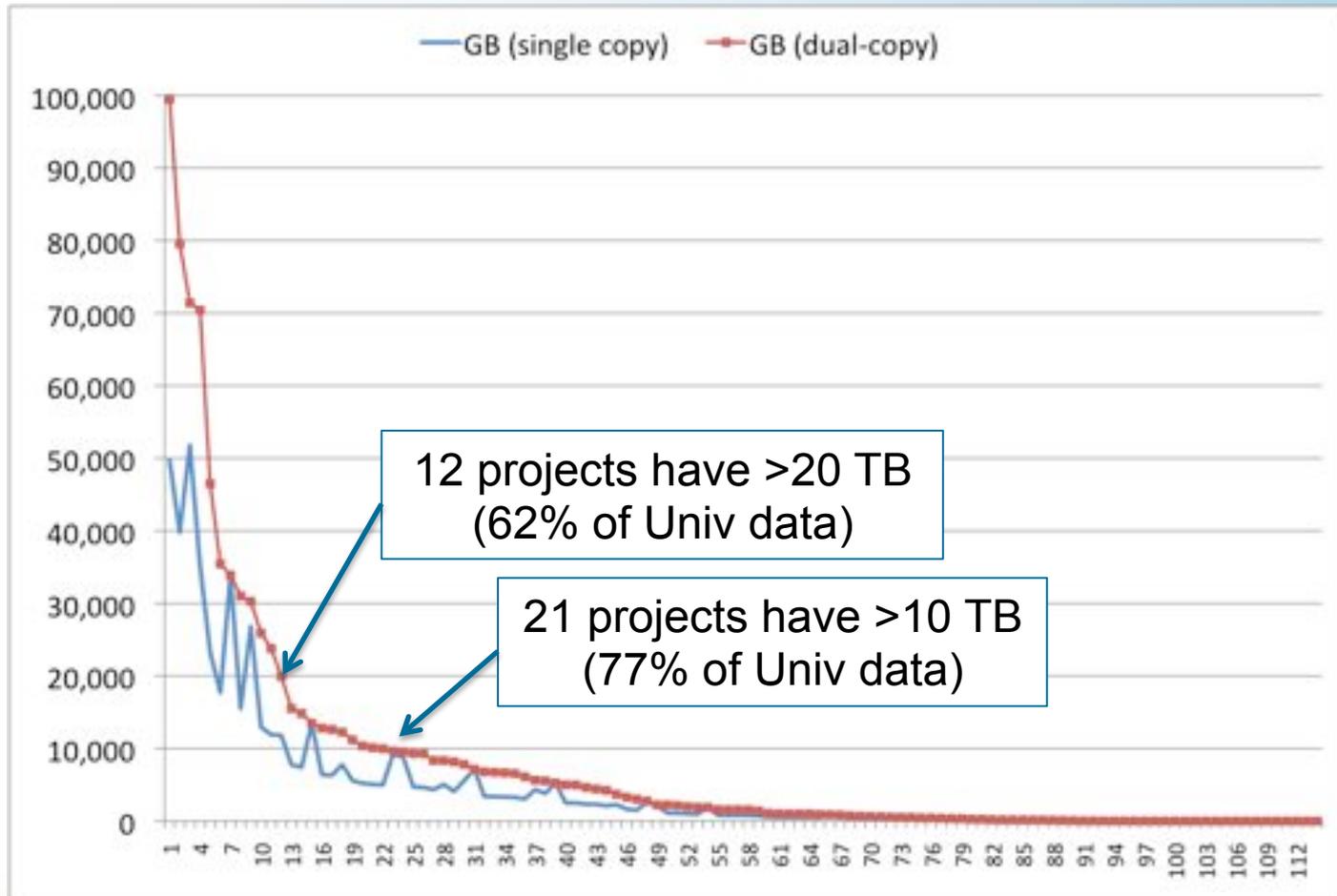
NWSC allocable resources

- **_HPC_** — similar to current practice
- **_DAV_** — will be allocated, similar to HPC practice
- **_HPSS_** — will be allocated, but requests only reviewed above some threshold (tbd).
  - Threshold will be set to balance reviewer burden with demand on resources. E.g., >20 TB.
- **_GLADE_** — project (long-term) spaces will be reviewed and allocated.

Impact on CHAP review burden

- **GLADE**—At this meeting 7 of 23 (1/3) requests asked for project space.
  - 1 definitely inappropriate
  - 2 very small, may be suitable for /glade/user
  - 2 unclearly justified, may not be needed
  - 2 appropriate

- **HPSS** — 19 of 23 identified some HPSS need
  - 6 requests (1/4) were for greater than 20 TB
 - Justifications range from "barely any" to "good."

Current university HPSS use

Legend: GB (single copy), GB (dual-copy)

12 projects have >20 TB (62% of Univ data)

21 projects have >10 TB (77% of Univ data)

Note: On MSS, dual-copy is the default. HPSS default is single-copy.

Data management plans, allocations

- NSF now requires all proposals to have a data management plan (DMP).

- CISL will soon be allocating long-term data storage resources.

- CISL's data generation capacity will soon be orders of magnitude larger.

- What are the implications and interactions?

CISL Usage Policies

Allocation Requests

NSF Data Plan Requirements
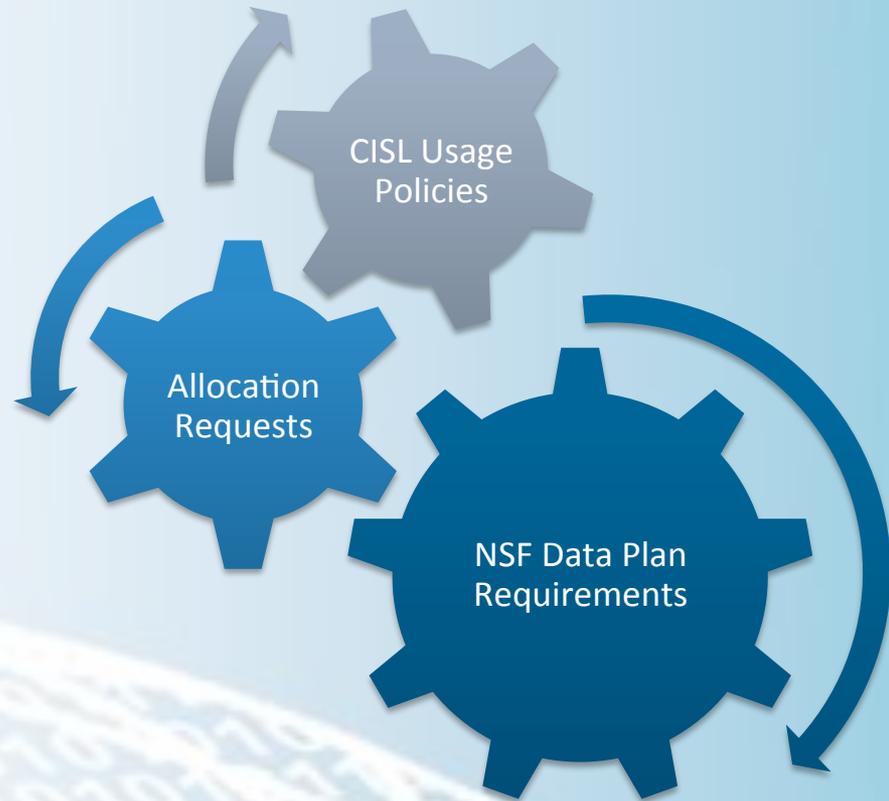
NSF Data Policies

- NSF ***data dissemination policies*** state that research data should be shared as soon as is practical or reasonable.

- NSF ***data retention policies*** state that data should be kept for three years after end of an award, or three years from first publication.

  - In some cases, longer retention periods are warranted, e.g. data supporting patents.

Current CISL Data Policies

- No limits on or review of archive use during NSF award period.
  - CSL awards granted "enough" GAUs for their HPSS use
  - Some CHAP requests include GAUs for HPSS, others don't.
- Data lifetime at CISL largely tied to NSF award period, with some leeway to extend project a year or so beyond that.
- User can request a Small Data allocation, no award needed, for small number of GAUs, to access data at CISL.
  - Generally (but not strictly) for accessing others' shared data
- Orphaned data in HPSS will be deleted after 12 months.
  - http://www2.cisl.ucar.edu/docs/hpss/policy
  - Policy for GLADE data after end of allocation is TBD, but will likely need to be a small number of weeks.

NCAR

Current CSL/CHAP Policies

- Allocation requests now being asked for Data Management Plan (April 2011 CHAP).
  - Can ask for GLADE project space. Otherwise, GLADE project space would be "first-come, first-served."
  - Asks for some detail for use over 10 TB.
  - Except for ability to request GLADE, this has been requested for at least since April 2010 CHAP meeting.
- Data plan is for CISL's information. Not reviewed by CHAP.
- CSL (Jan 2011) used a similar approach.

17

CISL's Potential Dilemma

- If NSF awardee tells NSF in his/her data management plan that data will be stored at CISL for *X* years, will CISL have to honor that user's plan?

  - We have first example of PI's NSF DMP appearing in CHAP request at this meeting.

- What are the user's responsibilities? CISL's responsibilities?

- CISL needs to set expectations.

Proposed CISL Approach

- Following NSF award completion, allow CHAP/Univ. awardees to request a *Data Extension,* with a term of up to 3 years.
  - Purpose: Use CISL resources to hold, access, and/or share data that result from computational work conducted at CISL.
  - Option: Can request "small" (tbd) DAV allocation to continue post-processing, data reduction, or analysis. More would require CHAP review.
- After 3 years, project lead can request up to 2 more years, with cause.
  - Cause may be tied to data volume, need to share data with a community.
  - CISL will set a minimum threshold for data volume.
  - No DAV access.
- After 3+2 years, project lead must assume ownership of data.
  - Delete or move to local resources.
  - Rare: User could pursue adding data as RDA/CDP/ESG data set.
- Note: Small Data Allocations would continue to be available to users for access to shared data in HPSS (or GLADE).

Elements of a CHAP/CSL DMP

1. Expected Data
   - Only for data destined to reside on GLADE project space or in the HPSS archive
   - Should be linked to proposed computational experiments *conducted at CISL*
   - CISL reserves the right to cap a project's HPSS use
2. Period of Data Retention
   - Should be X < 5 years
3. Data formats and dissemination
   - Standard formats, sharing, metadata—all good, preferred
4. Storage and preservation of access
   - I.e., need and plans for storing multiple copies, etc.
5. Additional requirements
   - Special needs or expectations, usually NSF mandated.
   - No guarantee that CISL can meet these needs.

QUESTIONS?