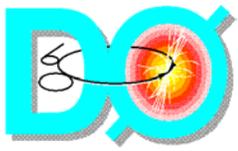


Bird Review Close Out

Amber Boehnlein

Oct 8, 2003



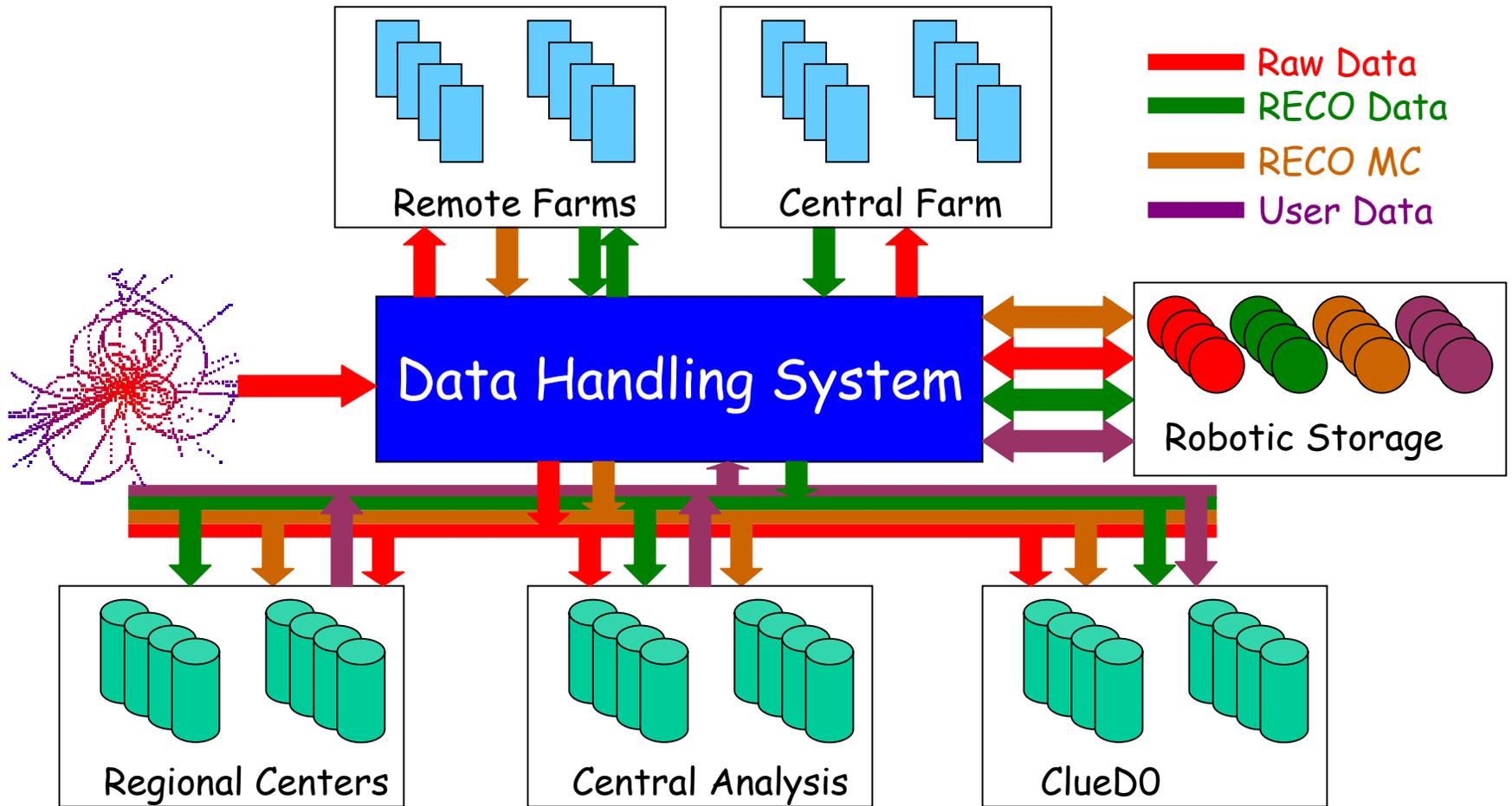
Charge To Bird Committee

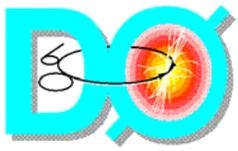
The CDF and D0 experiments are asked to together propose and arrange a series of talks from experiment and computing division personnel that present

- a) the current status of the computing systems and how they operate, both at Fermilab and worldwide, to enable the experiments to collect, store and analyze the Run II data.
- b) the experiment requirements and proposed computing model for the next 3 years, together with the estimated costs at Fermilab in terms of both equipment and manpower.
- c) the agreements in place by collaborating institutions to provide either manpower or services that the experiment relies on for some part of the processing and analysis of data.

D0 gave talks on Remote computing, D0Reco and a talk that used current operational understanding to make estimates

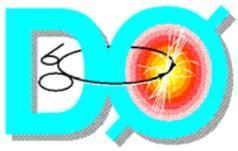
Data Flow





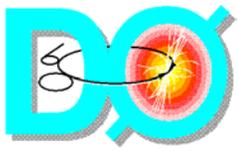
Towards a Global Model

- Two planning efforts—Regional Analysis-> Offsite Analysis Task Force
 - ◆ Computing resources and infrastructure are only one aspect of effective offsite analysis
- Effectively integrating hardware resources requires structural changes and additional effort (management and technical)
 - ◆ The Computing Planning Board now has different composition and is charged with focusing on global issues.
 - ◆ We must develop a support model in which developer and operations effort is supplied in conjunction with the hardware to make the global model a success—with support also covered by MOU.
 - ◆ We are increasing focus on aligned activities within GRID projects.
 - ◆ We are track available resources and determine the deployment as best meets strategic and tactical needs
- Financial considerations have to be addressed—use computing contributions to offset common fund contributions, start from BaBar Model, details are still being settled
- Track total estimated needs and value of contributions—use and extend planning spreadsheet.
- Operational contributions also important and valued.



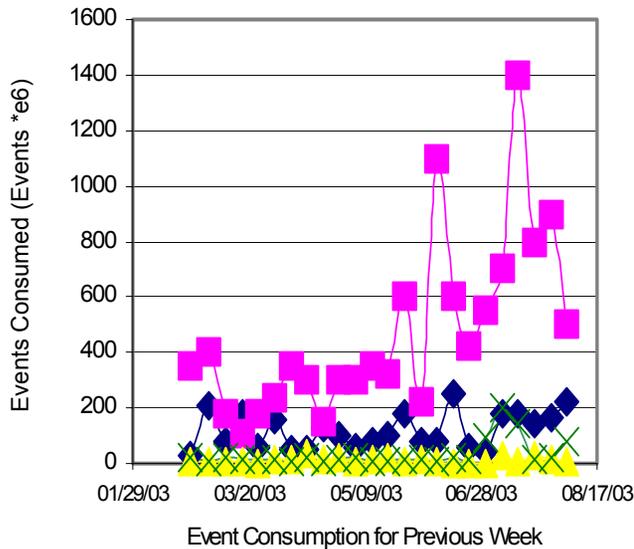
The Virtual Center

- For cost basis, determine the cost of the full computing system to met all needs as if the center were located at FNAL, plus equipment required to support offsite
- Presented today as draft
 - ◆ Disk and servers and CPU for analysis
 - ◆ Production activities such as MC generation, processing and reprocessing.
 - ◆ Infrastructure such as gateway machines and code servers
 - ◆ database machines and servers
 - ◆ Mass storage
 - ◆ Cache machines and drives to support extensive data export
- Not included as a cost estimate, but vital
 - ◆ Wide Area Networking
 - ◆ Desktop computing



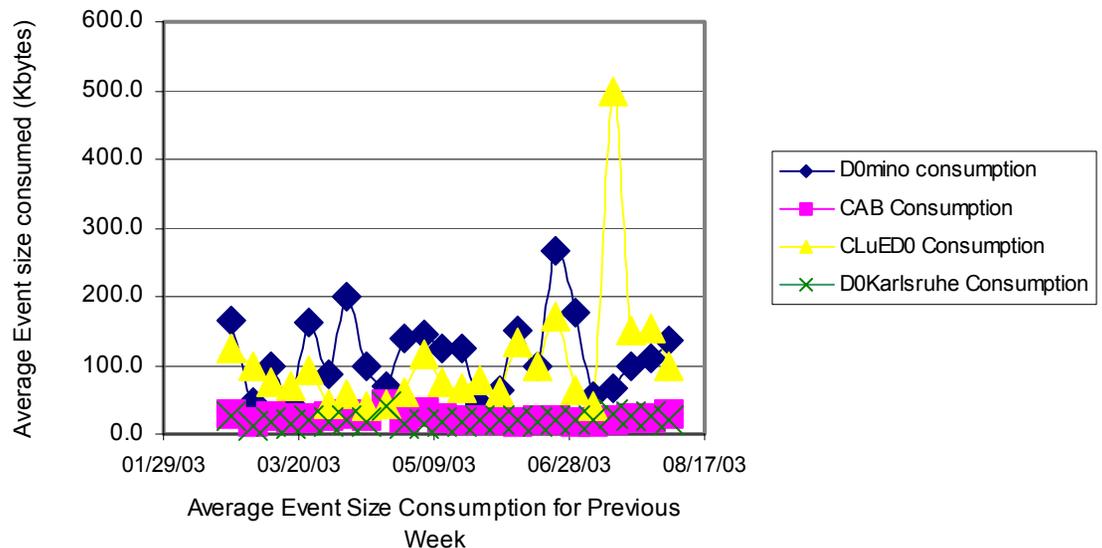
Data Handling Metrics

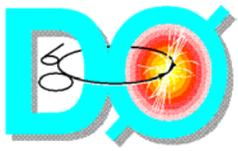
Event Consumption on Analysis Stations



Consumption plot gives information
On analysis uses cases.
Use more varied on desktop, Domino
Than on CAB or DOKarlsruhe

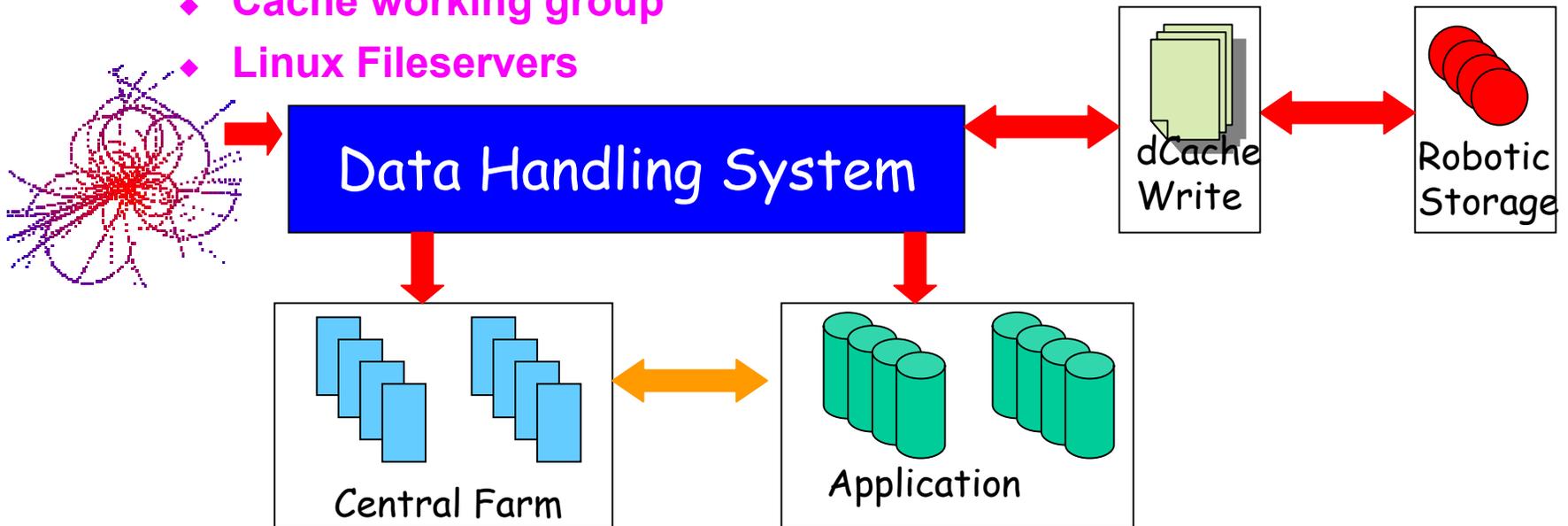
Average Event Size Consumed on Analysis Stations

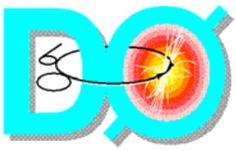




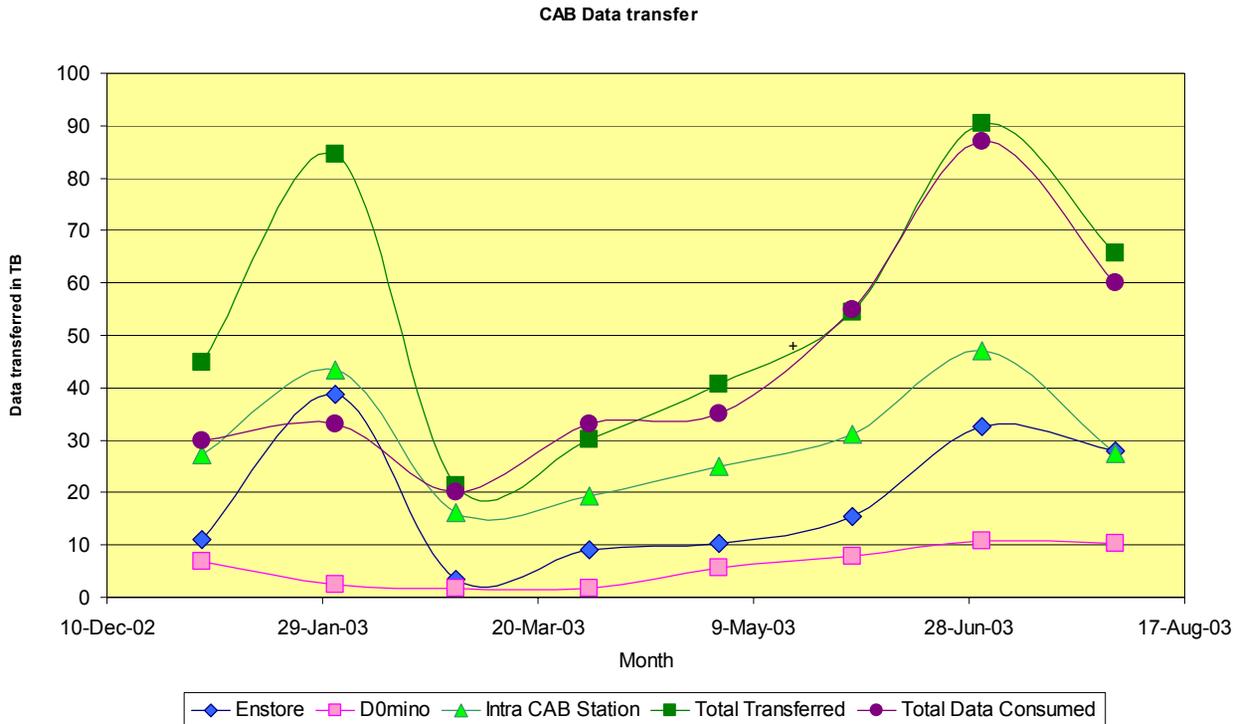
Data Handling, cont.

- Improved operations
 - ◆ Investigate slow transfers
 - ◆ Continue to improve tools, documentation metrics
- Extensions
 - ◆ Grid
 - ◆ Support for Remote Systems
 - ◆ Integrate dCache-first in online for monitor data
 - ◆ Cache working group
 - ◆ Linux Fileservers

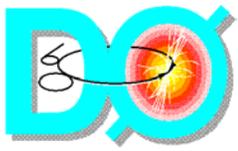




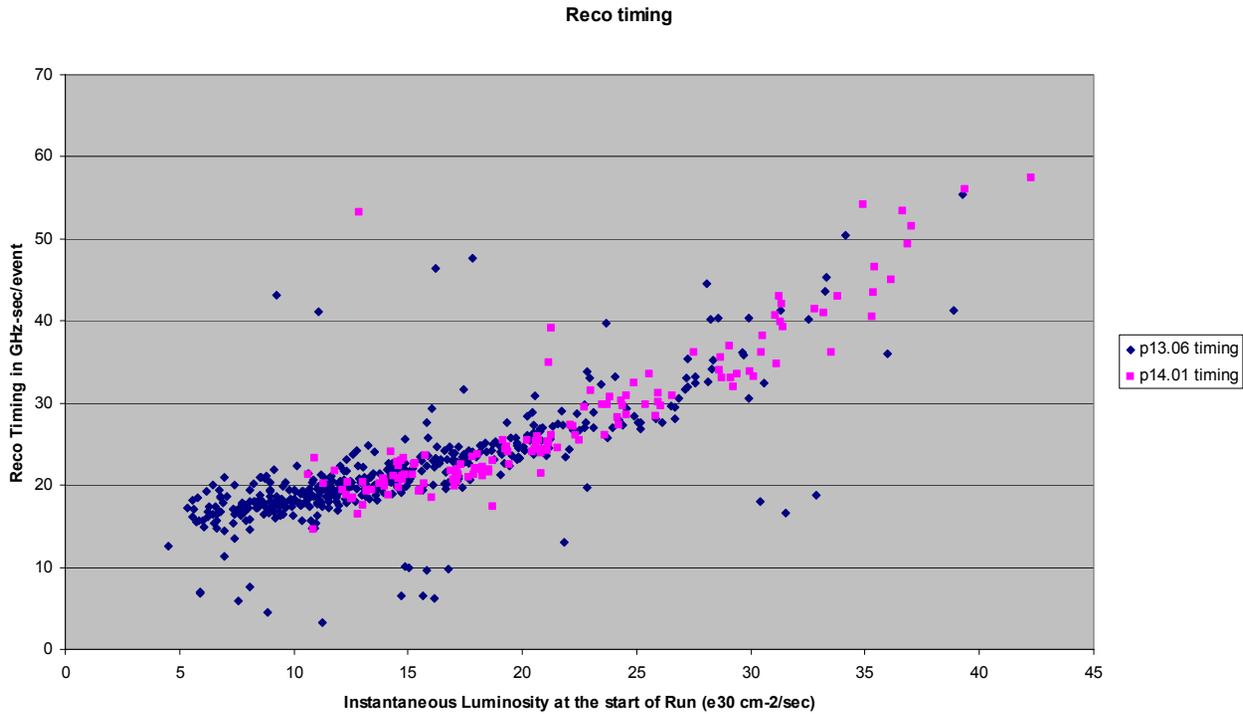
CAB File Transfers

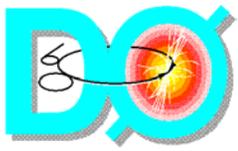


New Plot!—Major components of data transfer on CAB
Slight excess transfers is fine—jobs crash while SAM delivers
Station problem in Feb now fixed.



Reco Timing





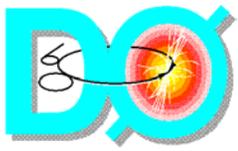
Contributions

Use the FNAL equipment budget to provide very basic level of functionality

- ◆ Database and other infrastructure
- ◆ Primary Reconstruction farm
- ◆ Robotic storage and tape drives
- ◆ Disk cache
- ◆ Basic analysis computing
- ◆ Support for data access to enable offsite computing

Institutional Contributions

- ◆ All Monte Carlo production takes place at remote centers
- ◆ Secondary reprocessing
- ◆ Analysis at home institutions
- ◆ Contributions at FNAL to project disk and to CLuED0
- ◆ Eventually collaboration wide analysis

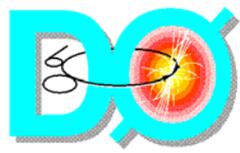


Infrastructure Estimates

Infrastructure				
Year		2004	2005	2006
databases				
	servers	\$30,000	\$30,000	\$30,000
	disk	\$30,000	\$30,000	\$30,000
Home Areas		\$50,000	\$10,000	\$10,000
Networking		\$120,000	\$80,000	\$100,000
Machines		\$60,000	\$60,000	\$60,000
Totals		\$290,000	\$210,000	\$230,000

Home areas—either keep SGI and buy faster disk or
Buy replacement system: took average cost.

Networking cost under-estimated-Phil ~\$260K in 2004



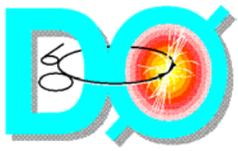
Cost Estimate-Sept 2003

	2003	2004	2005	2006
FNAL Analysis CPU	\$505,400	\$339,000	\$522,000	\$337,000
Primary Reconstruction	\$200,000	\$407,507	\$423,464	\$303,573
Re-Reco	NA	\$1,969,574	\$1,767,617	\$1,565,659
Monte Carlo	NA	\$446,940	\$105,485	\$70,323
File Servers/disk	\$262,000	\$360,000	\$230,000	\$140,000
Mass Storage	\$280,000	\$230,000	\$100,000	\$500,000
<i>Remote Analysis</i>				
Infrastructure	\$244,000	\$290,000	\$210,000	\$230,000
FNAL Basic	\$1,491,400	\$1,626,507	\$1,485,464	\$1,510,573
Virtual Center Total		\$2,454,105	\$2,006,482	\$1,954,730

Reconstruction is a cost driver—selective reprocessing, speeding up Reco
File servers and farms are not generous—no reprocessing at FNAL in most
Basic plan.

Global Remote Analysis in preparation

Very Little flexibility in this plan.



My Conclusions

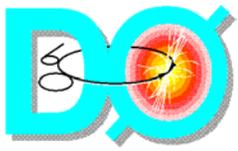
- **The DO computing model is successful**

Having an integrated data handling system enables flexibility in the allocation of resources and effective use of disk and robotic storage and is our path into the GRID era

Most performance tracking metrics shown today come from the SAM database

TMB format extremely valuable

- **Use Virtual Center Concept to calculate all costs.**
- **DO is shifting our thinking towards a more global model—and making structural changes and plans accordingly.**
- **We will need increased effort in order to make good use of all available hardware resources**



Committee's conclusions

- **Quite complimentary to DO-called out**
 - ◆ Remote Computing/Virtual Center accounting
 - ◆ CLuEDO and CAB
 - ◆ TMBS
 - ◆ SAM
 - ◆ Cost estimates sensible given assumptions
- **Complimentary to CD projects (enstore, dCache, databases...)**
- **CDF should move aggressively to SAM**
- **DO should aggressively move off D0mino**
- **Concern expressed about reco timing**
- **Pursue combining hardware resources**
- **Oh, by the way, FCC is out of space, power and cooling—Wideband facility**