

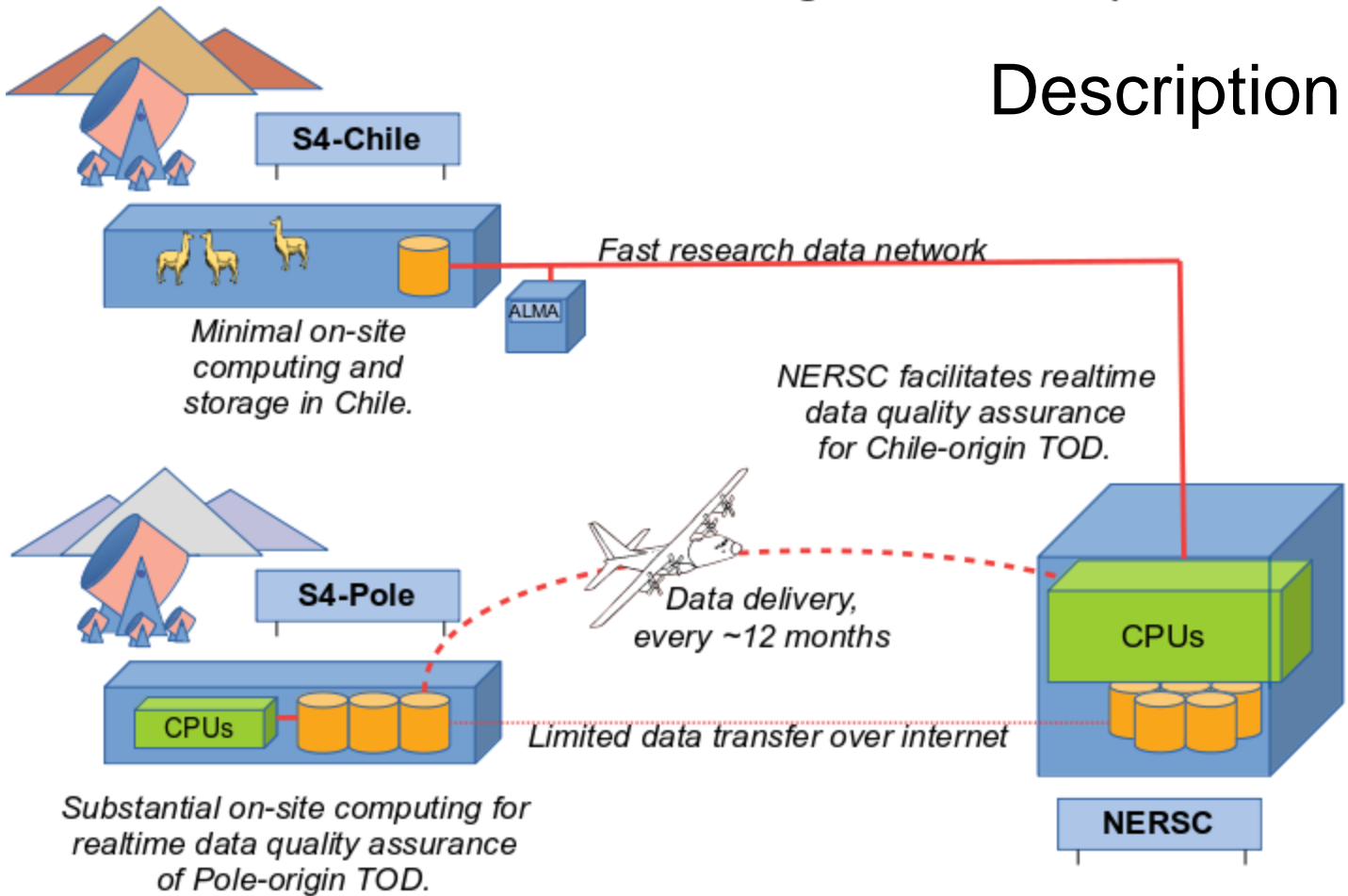
# 1.08 Data Management

T. Crawford, M. Hasselfield,  
M. Gilchriese, J. Borrill

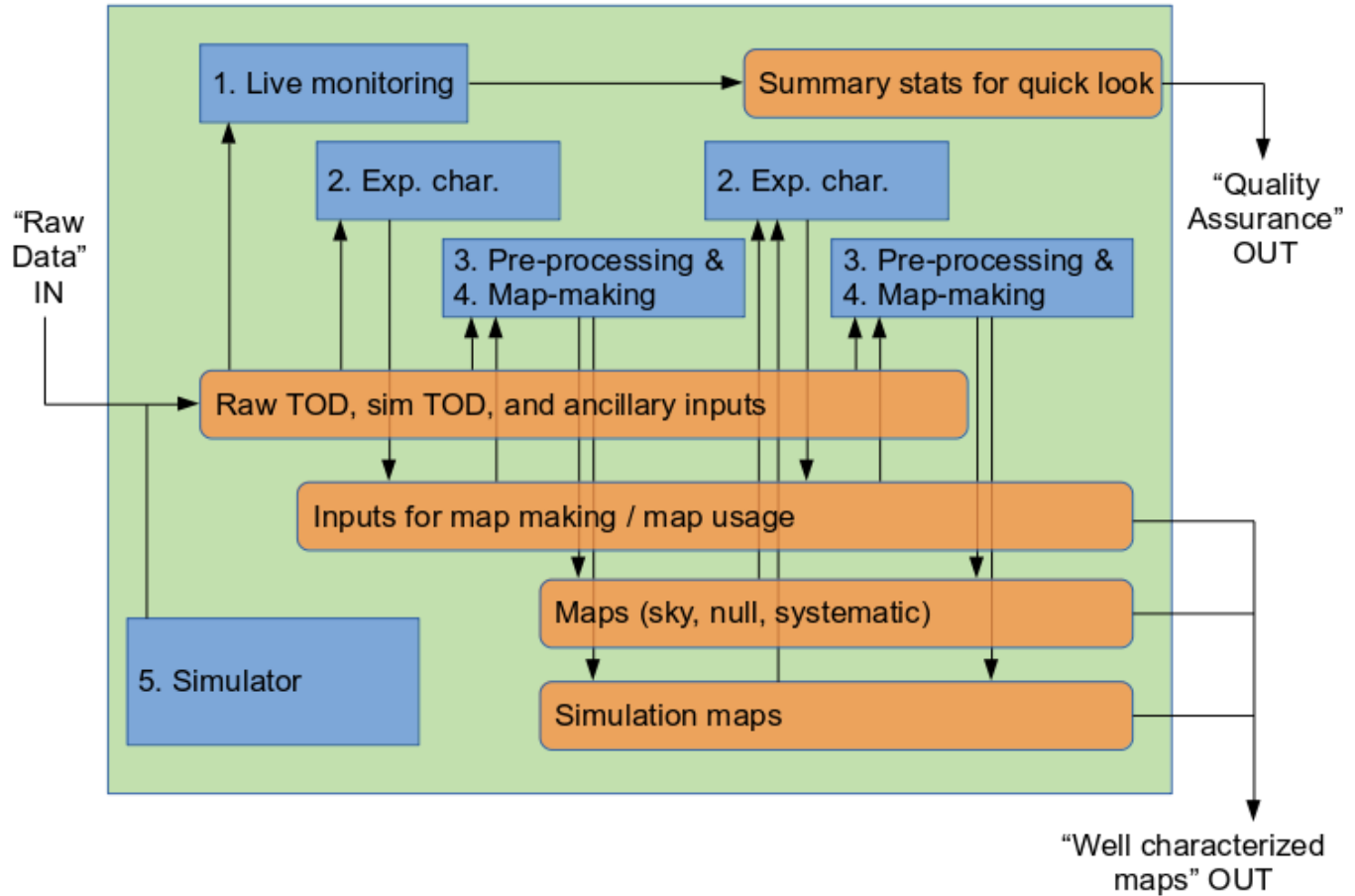
September 8, 2018  
Princeton University

# 1. DM & the sites

# Description



## 2. DM & data reduction



# Key Interfaces

## With Data Acquisition

- Data rates and storage. Storage primarily in operations. Check data rates and networking bandwidth assumptions
- Networking (assumption of good net from Chile, weak net from SP)
- Realtime Data Quality - DM responsible for analysis and hardware. Quick turn-around and display; DM hardware in SP for SP-origin data and on NERSC for Chile-origin data.

## With Science Teams

- DM purview includes bulk of TOD processing, ending in "well characterized maps". Downstream science analysis needs dictate the types of maps and ancillary products required.

# WBS 1.08 Data Management - Level 3

## 1.08.01 Management

Subsystem management, software configuration control, oversight of computing resources and interface definition

## 1.08.02 Transmission, Storage, Distribution & Publication

Transmission, storage, archiving, distribution and publication - personnel

## 1.08.03 Software Infrastructure

Software framework and pipelines - personnel

## 1.08.04 Data Synthesis

Experiment modeling and generation of simulated data - personnel

## 1.08.05 Data Reduction

Live monitoring, experiment characterization, data pre-processing and map making - personnel

## 1.08.06 Hardware Acquisition

Acquisition of CPU/disk located Pole, disk for data transmission from Pole, disk in US, networking - all hardware



01-05  
Almost all labor

# Major Cost Elements and Key Assumptions

CPU in US - provided at no cost to Project by allocation. Standard for DOE, possible for NSF.

- Non-US CPU for data analysis only, opportunity only - not part of Project plan
- Centralized - funding small resources at collaborating institutions not in Project

Disk Storage in US - currently same model as CPU, no Project cost. This is a new paradigm.  
Still under discussion - see risks

Other hardware - modest costs

- CPU/disk at Pole for data quality monitoring
- Physical disk transport from Pole (small on Project, mostly Operations)
- Networking - mostly no cost to Project (ESNET, etc). Assumes leverage other fiber (LSST, SO...). Needs more work to define cost.

Software personnel - the dominant Project cost by far: see next page

# WBS 1.08 Personnel FTEs

Estimate by WBS at L4

By year

By the following job types

- Senior software engineer
- Junior software engineer
- Senior scientist
- Junior scientist (costed)
- Junior scientist (not costed)

All these rates not in rate table so I mapped to match as best possible

# WBS 1.08 Data Management - Summary Costs

BOE TOTALS		LABOR			NON-LABOR			TOTALS	
ID	WBS/Activity	DIRECT	OVERHEAD	TOTALS	DIRECT	OVERHEAD	TOTALS	Labor + Non-Labor	Labor + Non-Labor + Estimate Uncertainty
	CMBs4.1.08 Data Management	\$15,251,538.40	\$6,710,676.90	\$21,962,215.30	\$1,075,458.00	\$172,073.28	\$1,247,531.28	\$23,209,746.58	\$32,562,394.79
	CMBs4.1.08.01 Subsystem Management	\$1,888,633.40	\$830,998.70	\$2,719,632.10	\$385,458.00	\$61,673.28	\$447,131.28	\$3,166,763.38	\$3,608,519.99
1	Management	\$995,465.00	\$438,004.60	\$1,433,469.60	\$37,458.00	\$5,993.28	\$43,451.28	\$1,476,920.88	\$1,704,781.92
2	Software Control	\$387,160.00	\$170,346.00	\$557,496.00	\$0.00	\$0.00	\$0.00	\$557,496.00	\$641,120.40
3	Computational Resources	\$253,009.20	\$111,324.05	\$364,333.25	\$0.00	\$0.00	\$0.00	\$364,333.25	\$418,983.24
4	Interfaces	\$253,009.20	\$111,324.05	\$364,333.25	\$0.00	\$0.00	\$0.00	\$364,333.25	\$418,983.24
5	Reviews	\$0.00	\$0.00	\$0.00	\$48,000.00	\$7,680.00	\$55,680.00	\$55,680.00	\$63,811.20
6	Travel	\$0.00	\$0.00	\$0.00	\$300,000.00	\$48,000.00	\$348,000.00	\$348,000.00	\$450,840.00
	CMBs4.1.08.02 Transmissin, storage distribution & publication	\$555,360.00	\$244,358.40	\$799,718.40	\$0.00	\$0.00	\$0.00	\$799,718.40	\$1,199,577.60
7	Transmission	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
8	Registration	\$122,820.00	\$54,040.80	\$176,860.80	\$0.00	\$0.00	\$0.00	\$176,860.80	\$265,291.20
9	Archiving	\$122,820.00	\$54,040.80	\$176,860.80	\$0.00	\$0.00	\$0.00	\$176,860.80	\$265,291.20
10	Internal	\$154,860.00	\$68,138.40	\$222,998.40	\$0.00	\$0.00	\$0.00	\$222,998.40	\$334,497.60
11	External	\$154,860.00	\$68,138.40	\$222,998.40	\$0.00	\$0.00	\$0.00	\$222,998.40	\$334,497.60
	CMBs4.1.08.03 Software Infrastructure	\$3,903,540.00	\$1,717,557.60	\$5,621,097.60	\$0.00	\$0.00	\$0.00	\$5,621,097.60	\$8,431,646.40
12	Framework	\$2,716,280.00	\$1,195,163.20	\$3,911,443.20	\$0.00	\$0.00	\$0.00	\$3,911,443.20	\$5,867,164.80
13	Pipelines	\$1,187,260.00	\$522,394.40	\$1,709,654.40	\$0.00	\$0.00	\$0.00	\$1,709,654.40	\$2,564,461.60
	CMBs4.1.08.04 Data Synthesis	\$3,519,960.00	\$1,548,778.00	\$5,068,728.00	\$0.00	\$0.00	\$0.00	\$5,068,728.00	\$7,603,092.00
17	Experimental modeling	\$1,759,975.00	\$774,389.00	\$2,534,364.00	\$0.00	\$0.00	\$0.00	\$2,534,364.00	\$3,801,546.00
18	Component generation & combination	\$1,759,975.00	\$774,389.00	\$2,534,364.00	\$0.00	\$0.00	\$0.00	\$2,534,364.00	\$3,801,546.00
	CMBs4.1.08.05 Data Reduction	\$5,384,055.00	\$2,368,984.20	\$7,753,039.20	\$0.00	\$0.00	\$0.00	\$7,753,039.20	\$11,629,558.80
22	Live Monitoring	\$570,045.00	\$250,819.80	\$820,864.80	\$0.00	\$0.00	\$0.00	\$820,864.80	\$1,231,297.20
23	Experiment characterization	\$1,440,910.00	\$634,000.40	\$2,074,910.40	\$0.00	\$0.00	\$0.00	\$2,074,910.40	\$3,112,365.60
24	Pre-processing	\$1,688,550.00	\$742,082.00	\$2,428,632.00	\$0.00	\$0.00	\$0.00	\$2,428,632.00	\$3,642,948.00
25	Map-making	\$1,688,550.00	\$742,082.00	\$2,428,632.00	\$0.00	\$0.00	\$0.00	\$2,428,632.00	\$3,642,948.00
	CMBs4.1.08.06 Hardware Acquisition	\$0.00	\$0.00	\$0.00	\$690,000.00	\$110,400.00	\$800,400.00	\$800,400.00	\$1,041,556.00
27	Disk Acquisition in the US	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
28	Pole Hardware	\$0.00	\$0.00	\$0.00	\$690,000.00	\$110,400.00	\$800,400.00	\$800,400.00	\$1,041,556.00

Base Cost: \$23.2M

Contingency:\$9.4M(40%)

Total:\$32.6M



# Uncertainties and Contingency

## Major uncertainties

- CPU by allocation only (DOE NERSC, ALCF; NSF XSEDE)
- Disk by allocation. NERSC moving in this direction. Significant risk: about \$3M
- How much software labor is needed?
- What software labor is costed i.e. scientists and NSF-supported role - see next slide

## Minor uncertainties

- Networking, hardware at Pole and Chile

## Contingency

- CDT contingency for DM was 66%
- We believe 40-45% is reasonable now.

# Total Cost and Comparables

## With Most Scientific Labor Costed - As Rolled Up

Base Cost: \$23.2M                      Contingency:\$9.4M(40%)                      Total:\$32.6M

## Without Scientific Labor Costed

Base about \$14M and total about \$20M

CDT estimate: \$13.5M(base) and \$22.5M(with contingency, 66%)

Comparables - compare FTEs since this is cost driver. Compare at peak FTEs/yr

- CMBS4 - 26 FTE(costed, including sci labor), about 32 total
- LSST - about 40-60(upper includes science operations) costed
- DESI - about 13 all funding sources

# Schedule and Decision Points

Major decision points are really “what are the assumptions” as already noted

Need to decide about costing disk, some or all scientists before DSR submitted

Otherwise schedule is synched to subsystem design reviews (CDR, PDR, FDR) and whatever mock data challenges you want (usually completed before DRs).

Schedule less coupled than other systems to CD gates or NSF equivalent

# Issues and Next Steps

Check our critical assumptions and update

- Free US disk
- No NSF involvement in US hardware.
- Are scientists costed or not or mixed(if so how)?

The cost driver is software labor so would be beneficial to have a review of labor plan, including experts from other collaborations, prior to December. Could be done by phone.

Higher level decision - is WBS 1.08 Data Management solely a DOE responsibility or does the NSF provide funding (e.g. university software personnel)?

# Backup

# CPU and Disk vs Time

CPU vs year does not exist but will before December

Disk vs time. Not yet well justified. Does not yet take into account retirement (e.g. after 5 years). Note this table assumes 8 year project or take FY26 as transition year

	<b>FY19</b>	<b>FY20</b>	<b>FY21</b>	<b>FY22</b>	<b>FY23</b>	<b>FY24</b>	<b>FY25</b>	<b>FY26</b>
	CD-0/CDR		CD-1+3a/PDR	CD-2/FDR	CD-3b			CD-4
		NERSC-9	ALCF		NERSC-10			NERSC-11
Disk/yr(TB)	0	250	500	750	3000	5000	5500	5000
Cumulative(TB)	0	250	750	1500	4500	9500	15000	20000