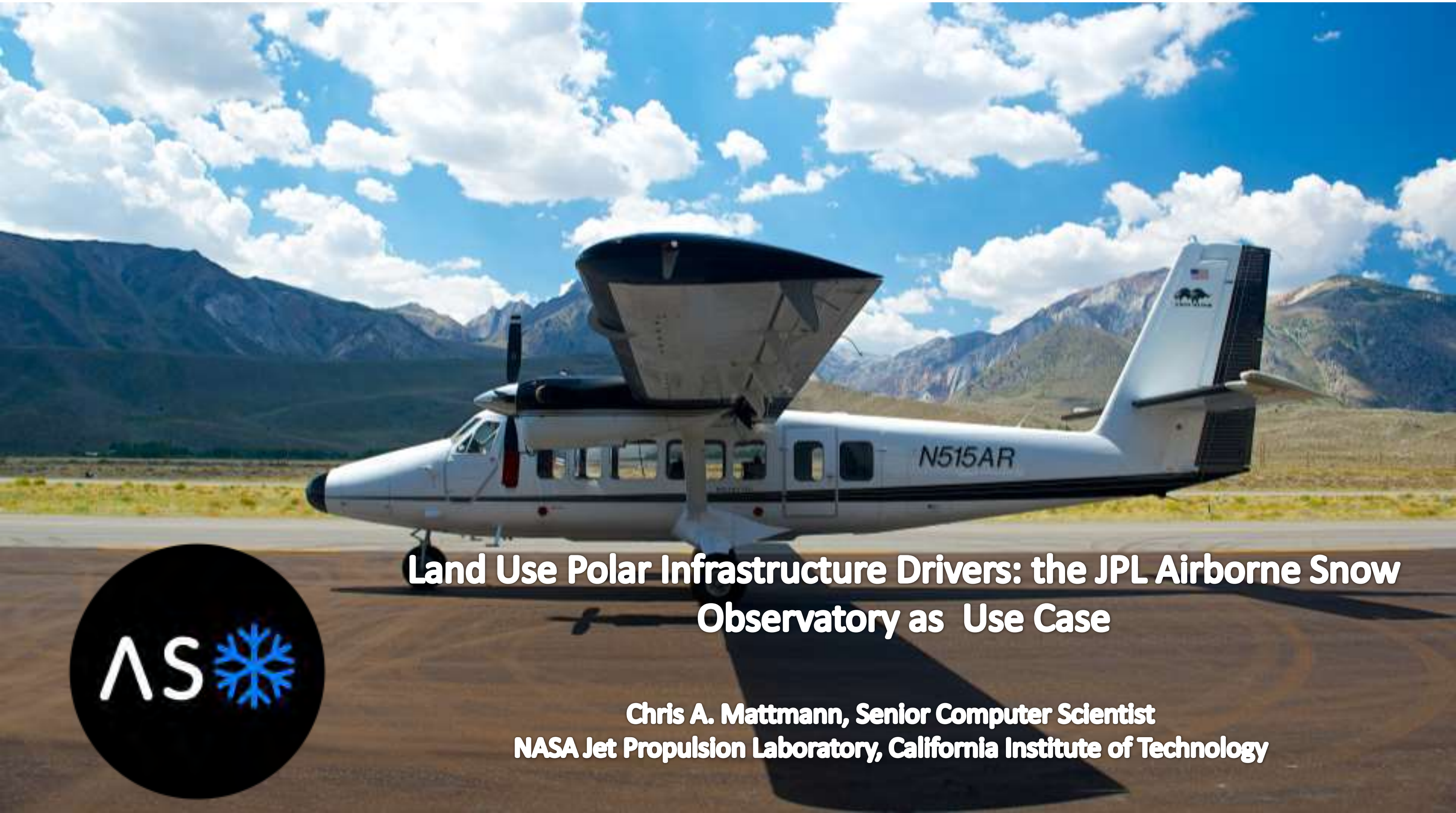
Jet Propulsion Laboratory
California Institute of Technology

Western Water Assessment



Land Use Polar Infrastructure Drivers: the JPL Airborne Snow Observatory as Use Case

Chris A. Mattmann, Senior Computer Scientist
NASA Jet Propulsion Laboratory, California Institute of Technology

And you are?

- Senior Computer Scientist at NASA JPL in Pasadena, CA USA
- Software Architecture/Engineering Prof at Univ. of Southern California
- **Compute Team Lead, JPL Airborne Snow Observatory**

- Apache Board Member involved in
  - OODT (PMC), Tika (PMC/progenitor), Nutch (PMC), Incubator (PMC), SIS (PMC), Gora (PMC), Airavata (PMC)

Jet Propulsion Laboratory
California Institute of Technology

Polar Data: Land Use

- Polar Data
- Land use related to
    - Carbon/Ecosystems
    - Glaciology
    - Water runoff
    - **Snowpack measurement**
- Land data can be generated
    - In-situ
    - Satellite
    - **Airborne instrumentation**
- NASA, JPL
    - S: MODIS, VIRS
    - A: CARVE, **ASO**

Snowmelt Runoff Forecasting

Jet Propulsion Laboratory
California Institute of Technology

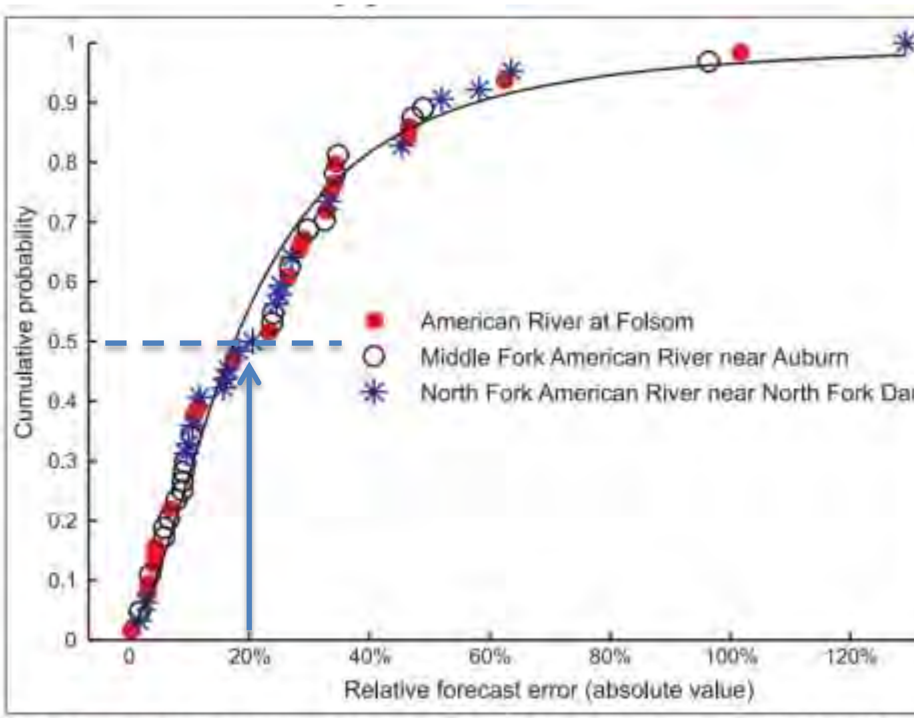


Fig. 1. Errors in the 1 April forecast for April–July runoff in the American River, 1990–2011, bas on gauges at Auburn and Folsom, in California. Note that the median error is 18% and the 80 percentile (1 year in 5) error is 39%. The plot was generated from information from the Califo Data Exchange Center.

In 1 of 5 years, forecast errors are greater than 40%. Half the time, they are greater than 20%. These come from poor data and poorly constrained science.

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | U.S. Edition

The New York Times

U.S.

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINI

POLITICS | EDUCATION | TEXAS

In California, Reading the Snow to Tell the Future for the Water Supply

Max Whittaker for The New York T

Frank Gehrke, center, has led snowpack surveys in California for a quarter-century. The state's multibillion-dollar agricultural industry pays close attention.

By NORIMITSU ONISHI
Published: February 7, 2013

Credit: Tom Painter, JPL

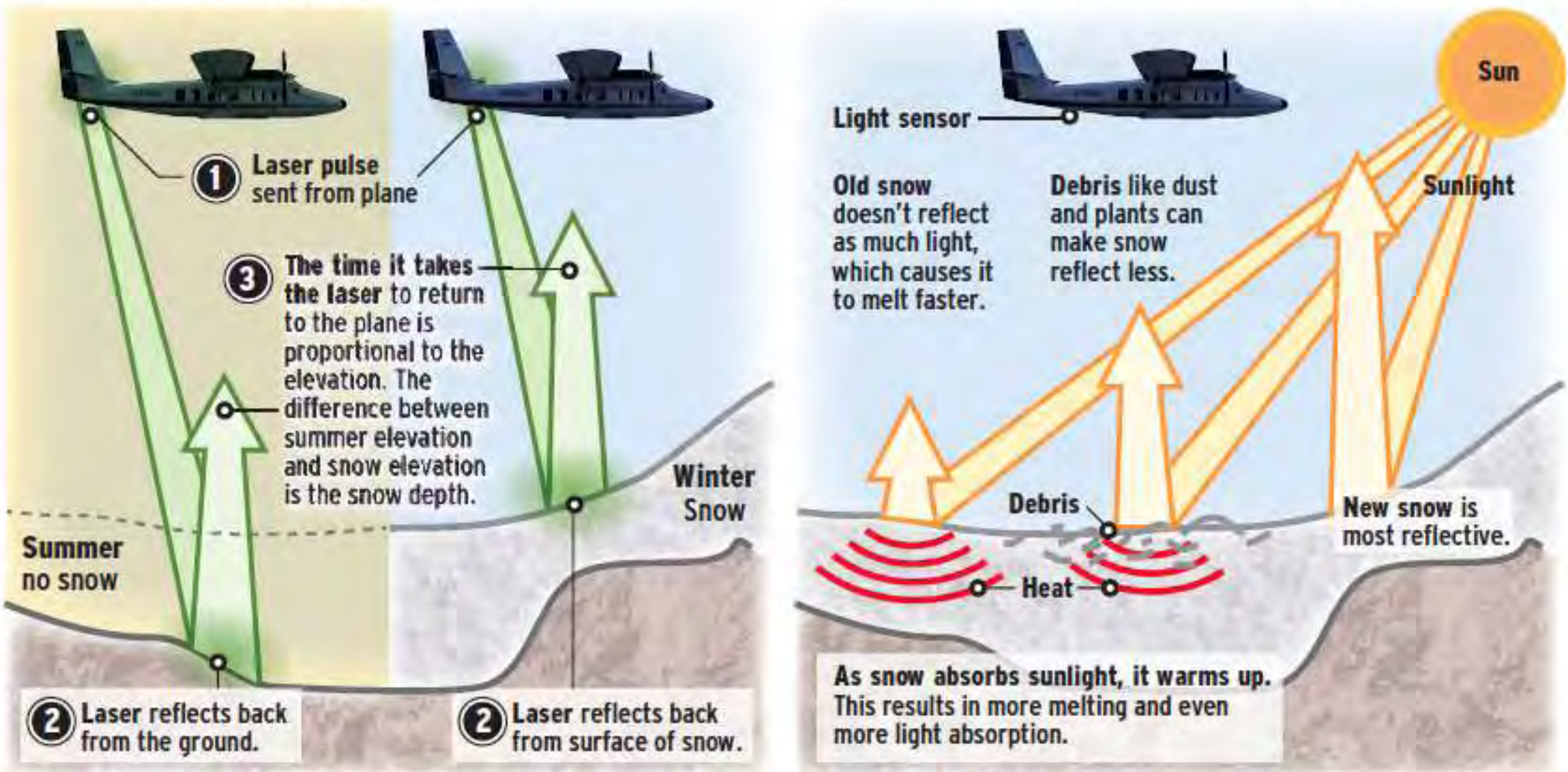
Jet Propulsion Laboratory
California Institute of Technology

How much snow?

Using laser radar, known as Lidar, researchers measure the depth of snowpack in California.

How will it melt?

With an advanced light sensor, scientists measure snow's reflectivity – an indicator of how it will melt.



Sources: Thomas Painter, Frank Gehrke, Optech Inc.

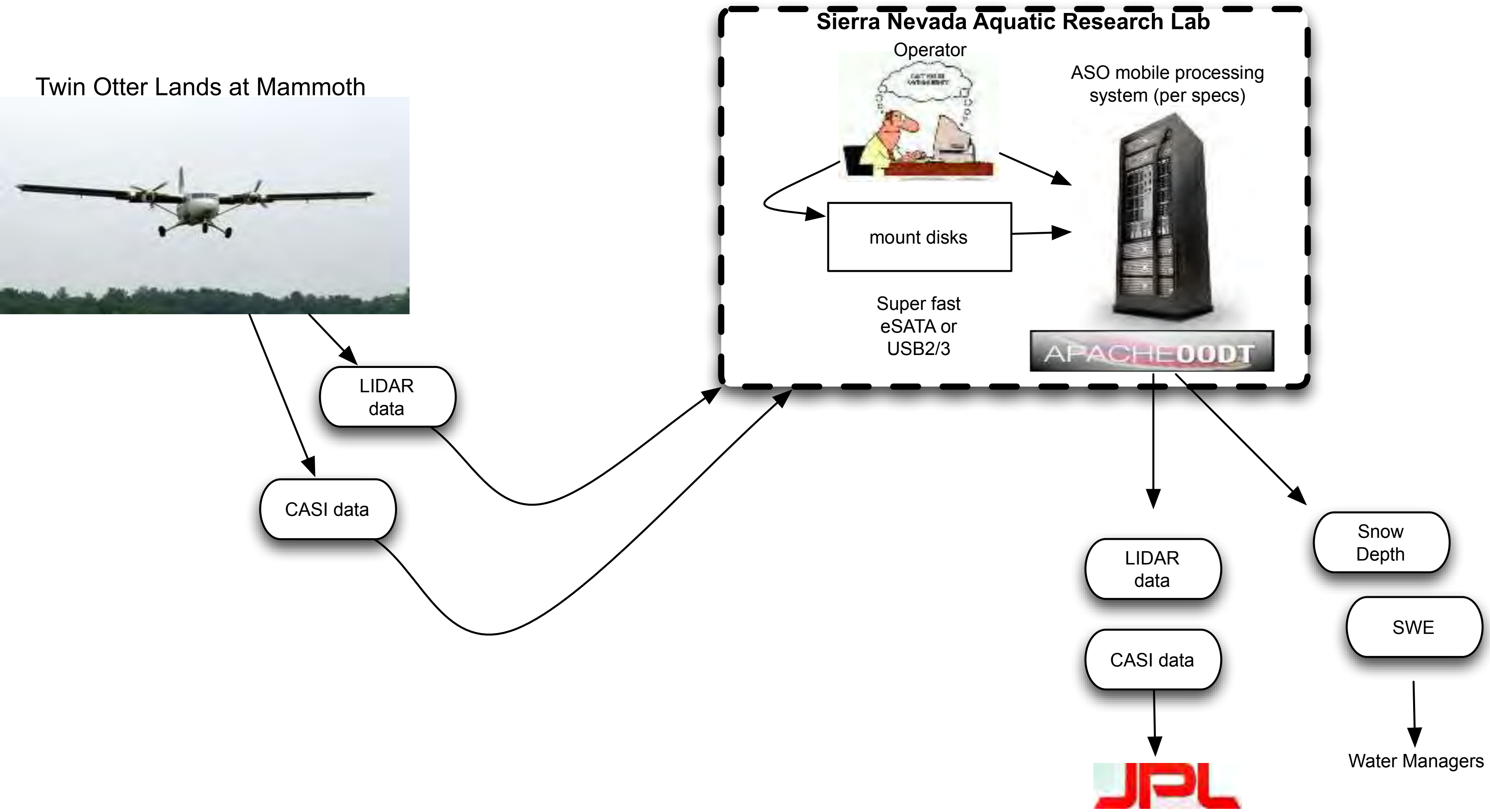
Maxwell Henderson / The Register

Credit: Tom Painter, JPL

Jet Propulsion Laboratory
California Institute of Technology

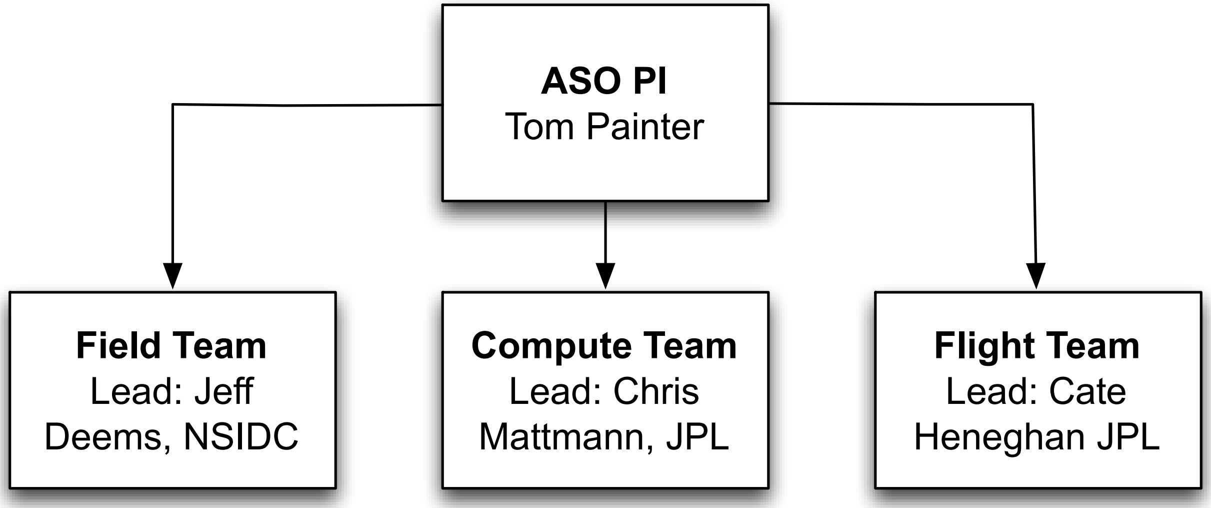
How did ASO go from acquired data to…improving water estimates?

The ASO Compute Team



Jet Propulsion Laboratory
California Institute of Technology

Who is the ASO Compute Team?



Paul Ramirez, Andrew Hart, Cameron Goodale, Felix Seidel, Paul Zimdars, Susan Neely, Jason Horn, Rishi Verma, Maziyar Boustani, Shakeh Khudikyan, Joseph Boardman, Amy Trangsrud, Cate Heneghan

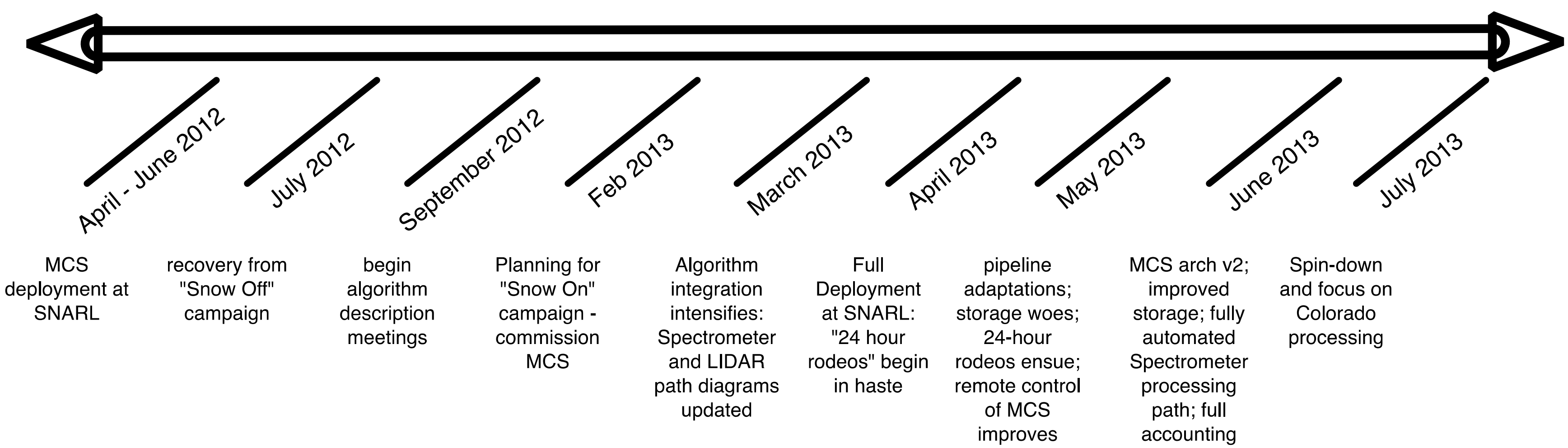
Jet Propulsion Laboratory
California Institute of Technology

What do we do?

- Job #1
  - Don’t lose the bits
- Job #2
  - Rapidly, and automatically process algorithms delivered by ASO scientists
    - Spectrometer (raw radiance data through basin maps of albedo)
    - LIDAR (raw data through snow depth/SWE)
- Job #3
  - Ensure that executed algorithms can easily be rerun, and that we catalog and archive the inputs, and outputs
- Job #4
  - Deliver the outputs of the algorithms (“move data around”)
- Job #5
  - Reformat the data, and convert it, and deliver maps, movies, higher level EPO
- Job #6
  - Entertain the rest of the team

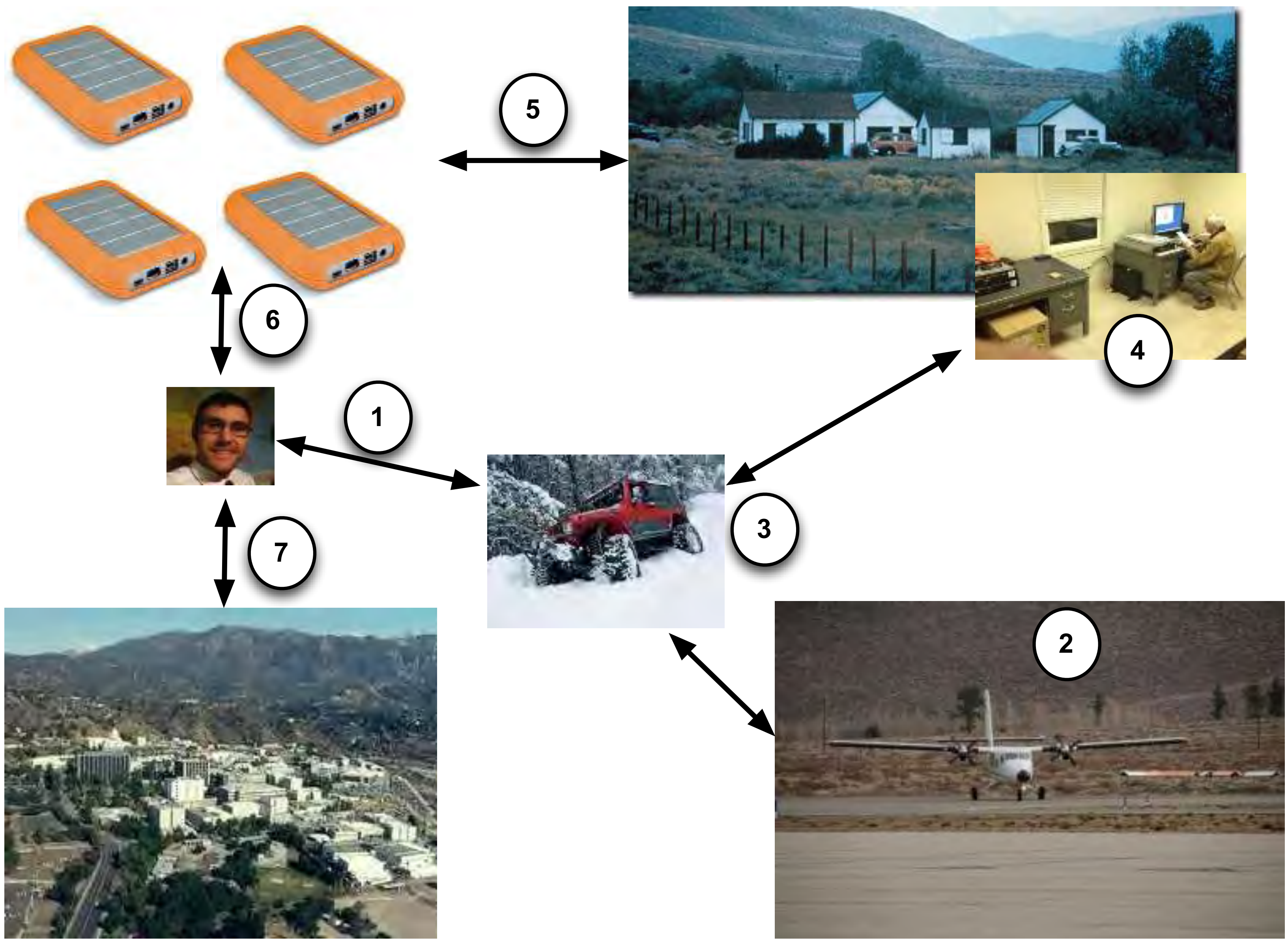
Compute Team Timeline

- Two Campaigns
  - "Snow Off" Campaign (Summer 2012)
    - Goal was Job #1 (don't lose the bits)
    - Won't spend much time covering this
  - "Snow On" Campaign (Feb/Mar 2013 – June 2013)
    - Full demonstration of 24 hour latency processing (we didn't start out there)
    - Snow in the mountains, and measure until full snowmelt
- Snow On Detailed Timeline



Jet Propulsion Laboratory
California Institute of Technology

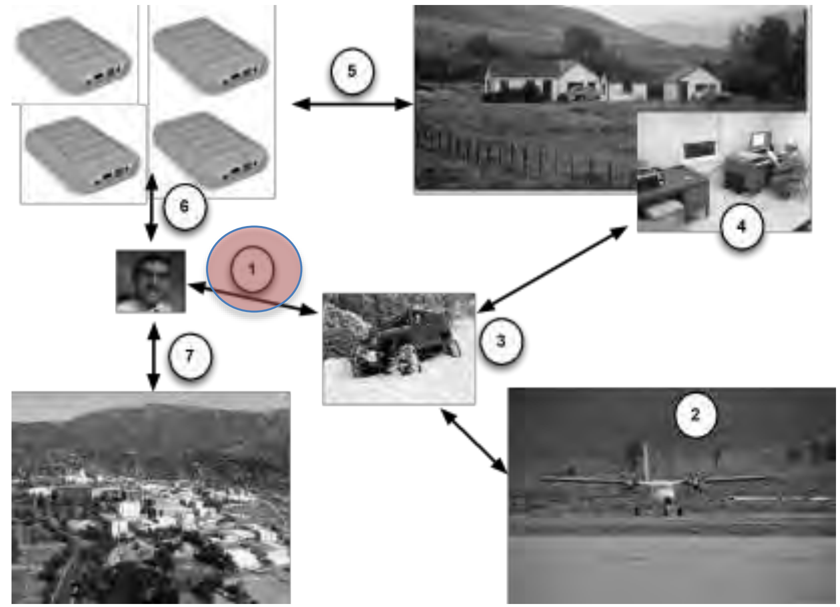
Weekly Detailed Compute Flow



Jet Propulsion Laboratory
California Institute of Technology

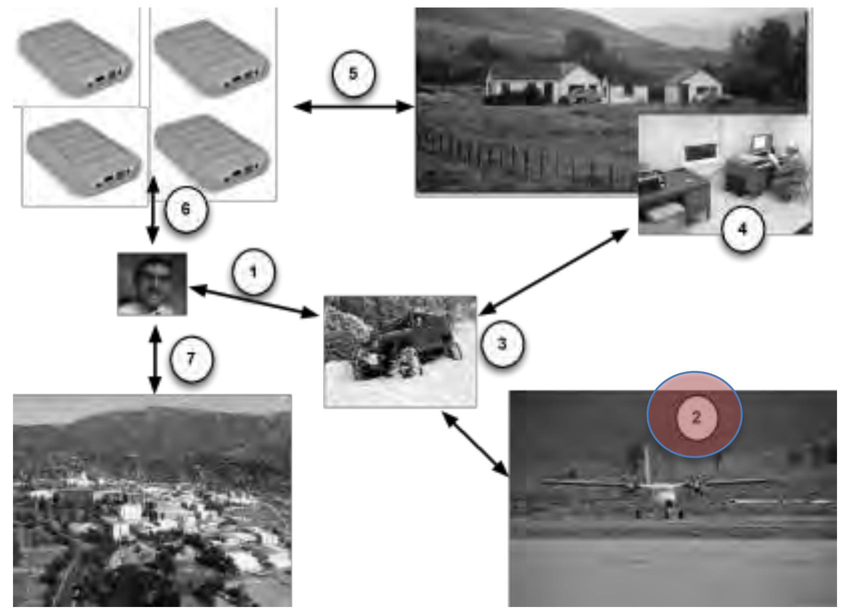
Weekly Detailed Compute Flow: 1/7

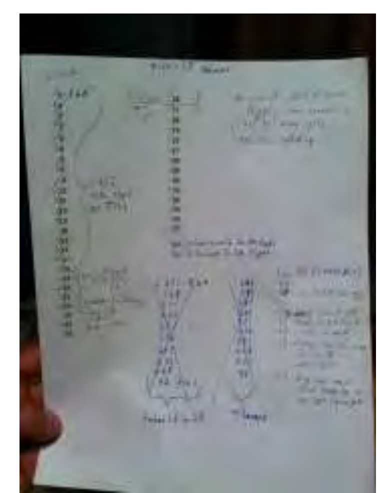
- Flight team member (e.g., Dan Berisford) heads up to SNARL
  - May take LaCie Bricks with him
  - Bricks are tagged with QR codes that are scannable by iPhone
  - Handoff of bricks coordinated with SA/devops portion of Compute Team



Weekly Detailed Compute Flow: 2/7

- Flights Happen
  - Could take multiple flights to cover the entire basin
  - Flights are labeled with a “Flight Id”, e.g.., 20130410f1 is April 10, 2013’s 1st flight
  - Flight captures LIDAR and spectrometer raw data, camera data, and other ancillary info
  - A “flight log” indicates notes about flight lines (issues encountered), and is delivered, along with flight summary
  - A “survey” is a set of flights (e.g., 20130410f1 and 20130410f2) that cover the whole basin
    - All info is captured on ASO wiki

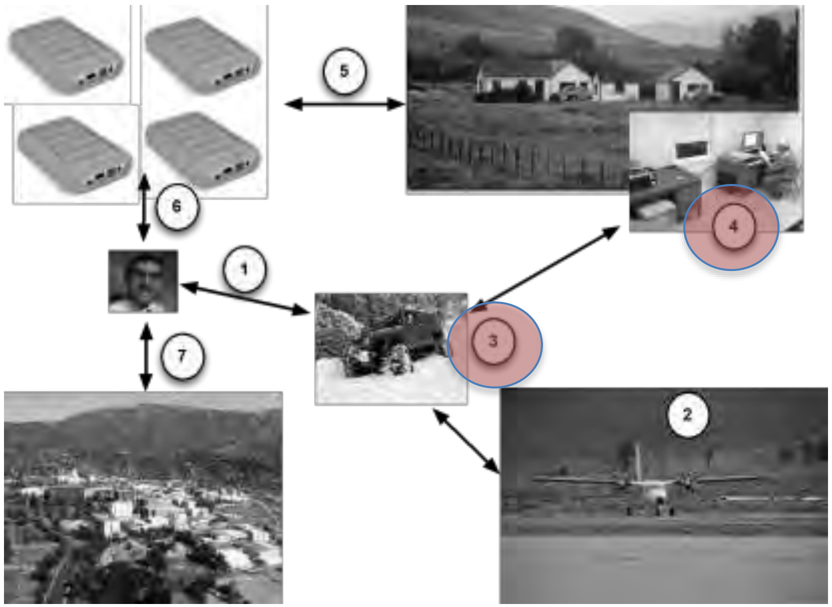




Jet Propulsion Laboratory
California Institute of Technology

Weekly Detailed Compute Flow: 3 and 4/7

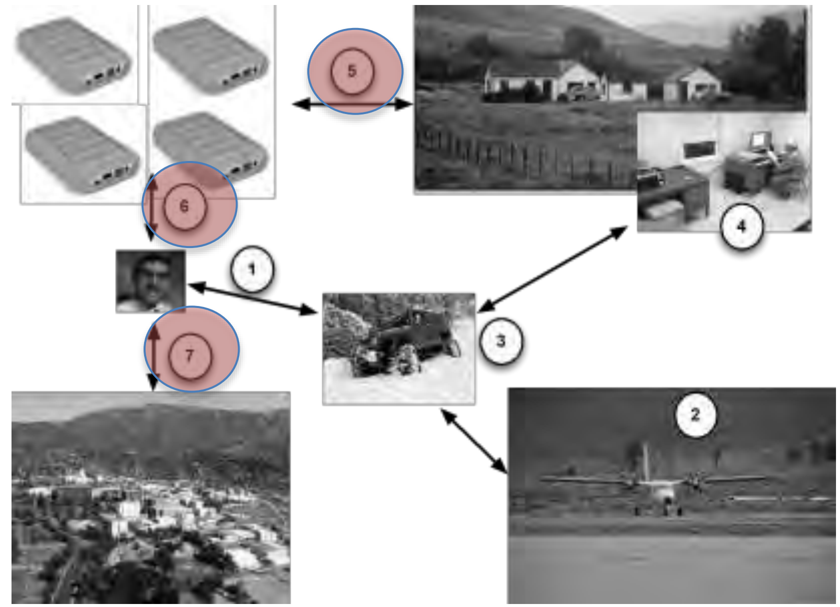
- When Flight lands
  - Data is driven from Mammoth Lakes Airport to SNARL by Flight Team member(s)
  - Data is plugged in to Mobile Compute System using eSATA and USB3 hub
  - As data is being transferred, Compute Team is notified
  - Data processing begins
    - "the rodeo"
    - Will explain in detail later



Jet Propulsion Laboratory
California Institute of Technology

Weekly Detailed Compute Flow: 5, 6 and 7

- Near end of week, after rodeo
  - Flight team works with Compute Team to define what raw LIDAR and spectrometer and camera and anc data can be copied
  - Data is automatically checksummed, and then copied to free LaCie bricks
  - LaCie bricks are taken by Flight team member (QR scanned via iPhone) and then transited to JPL for deep storage



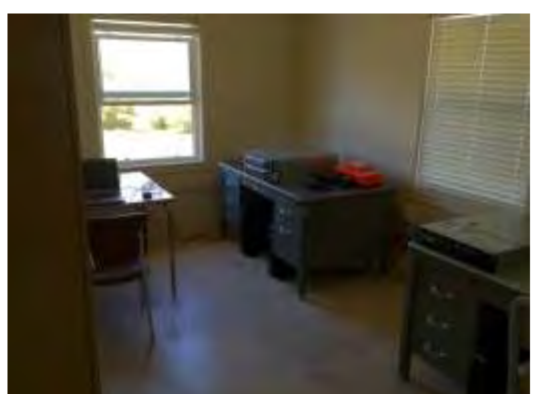



Image credit: Andrew Hart

Jet Propulsion Laboratory
California Institute of Technology

Lots of Complexity Hidden in there

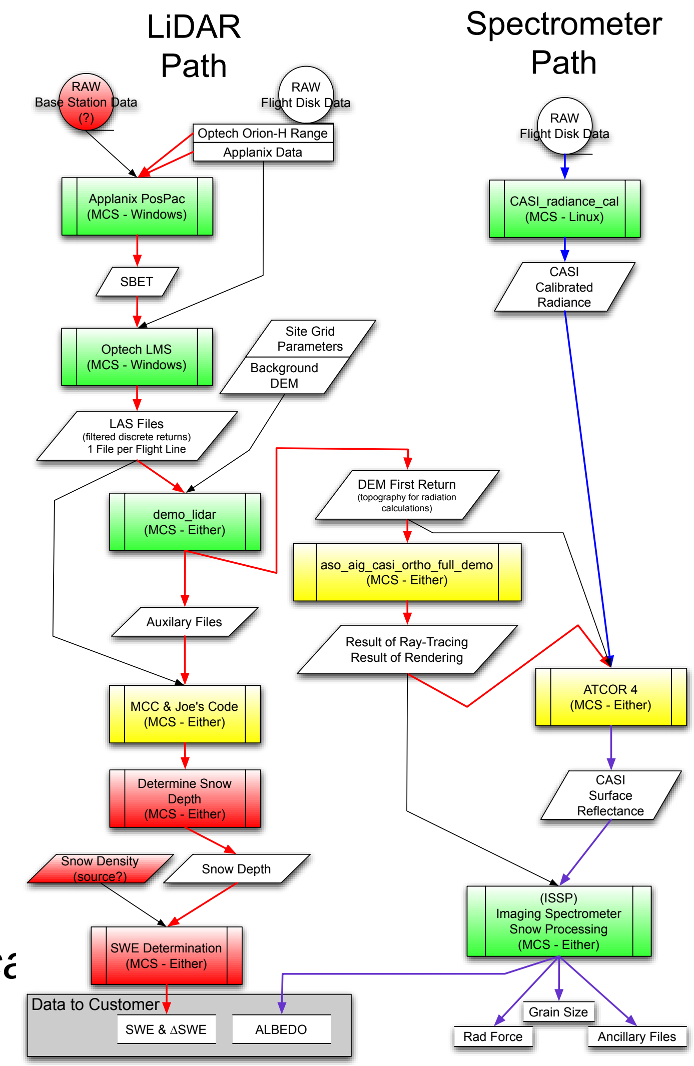
- LIDAR path

Human in the loop

  - Start from .las files in a directory
  - Run through *demo_lidar*
  - Run output of *demo_lidar* through *mcc_part_a*
  - Run output of *mcc_part_a* through *mcc_part_b*
  - Run output of *mcc_part_b* through *mcc_part_c*
  - Run output of *mcc_part_c* through *snow depth*
- Spectrometer path

Made this fully automated

  - Start from raw spectra, run through *radcorr/ATCOR*
  - Run *radcorr* output through *rebin*
  - Run *rebin* output through *orthorectifica*
  - Run *ortho* output through *issp*
  - Run *issp* output through *mosaic*



Jet Propulsion Laboratory
California Institute of Technology

Yee-haw, "Rodeo Time"

- With weather information and proper flight planning, the team and PI were able to figure out ahead of time what the outlook was in terms of acquiring the entire basin (full survey) or even for the single flight, or subset of flights
- Using the above, the 24 hour "rodeos" were planned
  - From the point of inception of data driven to SNARL, and data available on the MCS
    - Rip through LIDAR (runs for many hours on the *whole LIDAR dataset* to begin with/predictable)
    - Separately, grab CASI data, run through
      - Radcorr (minutes, *per flight line* can run many/all in *parallel*)
      - Rebin (10s of minutes, *per flight line* cannot run in *parallel* must run *sequential*)
      - Ortho (hours, *per flight line* can run 3-4 in *parallel*)
      - ISSP (hours, *per flight line* can run 3-4 in *parallel*)
      - Mosaic (1/2 hour per *entire Flight Id* or *Survey*, can run *in parallel*)
    - Quickest to Snow Depth/Height (LIDAR) and Snow Albedo (CASI) and Mosaic (CASI) wins!

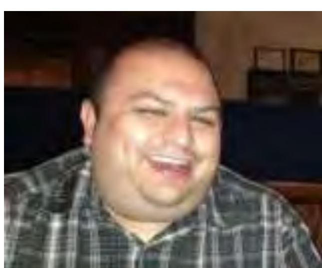
Jet Propulsion Laboratory
California Institute of Technology

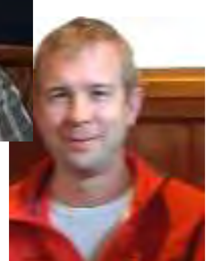
"Rodeo as an Illustration"

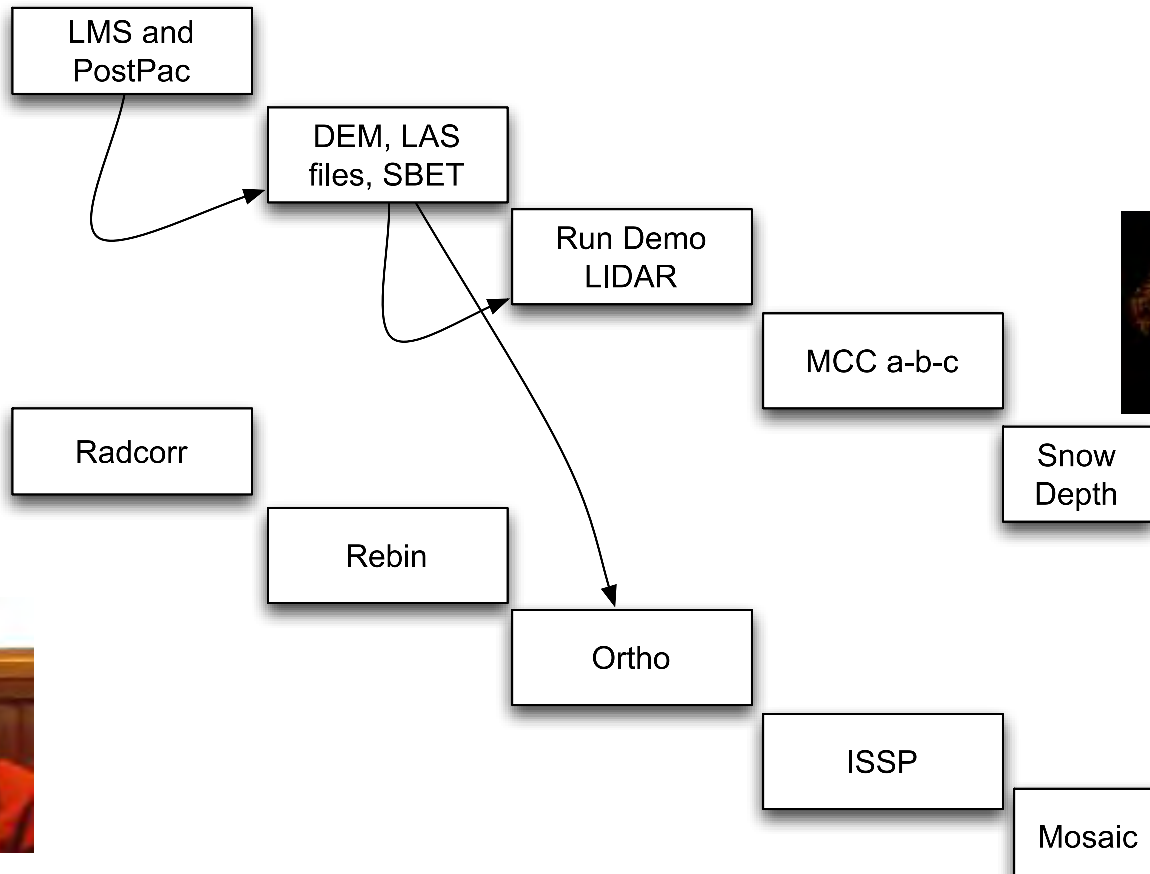
Start!

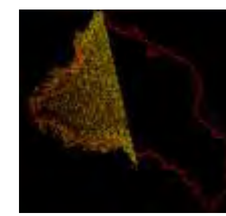
T=24 hours

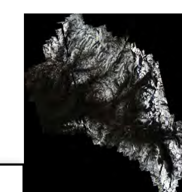
L0 products → L4 products









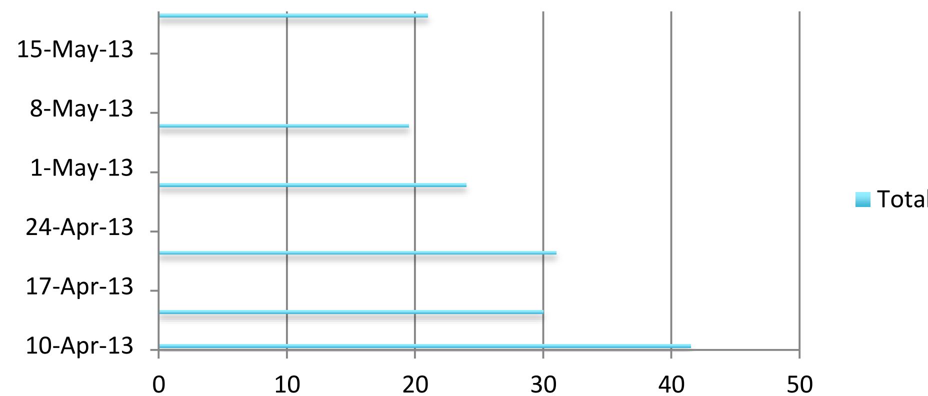


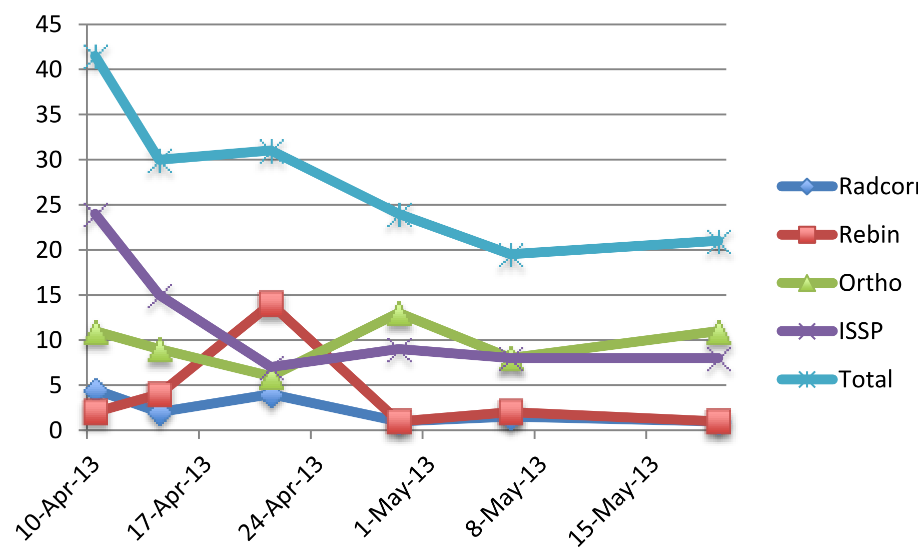
Rodeo improvements over time (CASI/spectrometer)

Jet Propulsion Laboratory
California Institute of Technology

- Earlier, ISSP was dominant processing time in rodeo
  - Eventually Ortho became a problem too due to issues like flying off DEM; and/or discovery of resource contention at alg. Level
- Radcorr and Rebin processing time were equated to nil through paralllelism and automation
- Within a month of near automation, we were making 24 hrs on CASI side
- Updates to algs to make deadline

Total CASI 24 rodeo processing time in hours: 4/10/2013 - 05/15/2013





Communication “Winning”

- So, what was going on during these rodeos?

Beirc

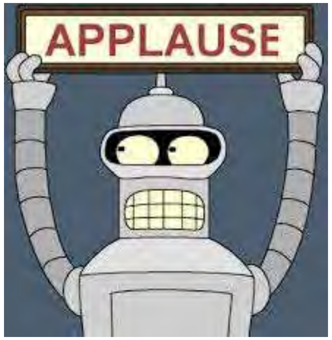
#lisp #beirc *Server*

<TITHAF> So…
<TITHAF> Where are we with CASI?
<crankshaft_> LIDAR first surface return done
<paulramirez> He asked about CASI? TITHAF we just finished radcorr and are working on rebin for 20130410f1
<chrismattmann> Am I late to the party?
<TITHAF> chrismattmann ping!
<TITHAF> chrismattmann!
<c0dified> hi all! Just to let you know, md5erator is done for flight 20130503
…

16:58 splittist on #lisp 33 messages

Used Internet Relay Chat (#aso-ops)

- Effective communication for a mission
  - Originally focused on use of cell phones, and/or text, and even social media to connect with one another (Facebook chat; Skype; etc.)
  - Found that all of these message platforms are async, 1-way
- Compute Team moves to IRC room for faster, located, and async and sync broadcast communication
  - It's easy for us nerds to adopt it, but we really succeeded *when*
    - Initially we were able to get some of the Flight team members (mountainair and wingsuiter) and other Compute team members (joeb) to join
    - *REALLY succeeded* when TITHAF showed up (aka "Tom Painter")
- With #aso-ops communication, we then moved to see what else we could localize into #aso-ops
  - Not-001 "the robot" was born



Not-001 and automation

Jet Propulsion Laboratory
California Institute of Technology

Beirc

#lisp #beirc *Server*

<chrismattmann> Not-001 latest email from mountainair

<Not-001> SBET production delayed on flight 20130501f1 – Amy

<Not-001> you are welcome master chrismattmann

<chrismattmann> Thanks Not-001, please pat yourself on the back

..some time later

<Not-001> Processing status update for flight 20130501f1: CASI: 10/31 lines processed: current running PGEs (radcorr, rebin) 2 nodes thorodin, 2 nodes trango: LIDAR: mcc-part-b running for 12 hrs

<Not-001> svn commit r929888 /aso/pge/rebin by pramirez log: Update rebin IDL code from mammut

- Email status
- Processing updates
- SVN updates
- JIRA updates
- Automatic monitoring and emailing

Jet Propulsion Laboratory
California Institute of Technology

How did Not-001 work?

Notifico Home Login Register

NOTIFICO

Notifico is an open-source replacement for the now-defunct cia.vc service. Notifico receives messages from services and scripts and delivers them to IRC channels.

+ Sign Up

Already have an account? Login instead.

NEWEST PROJECTS

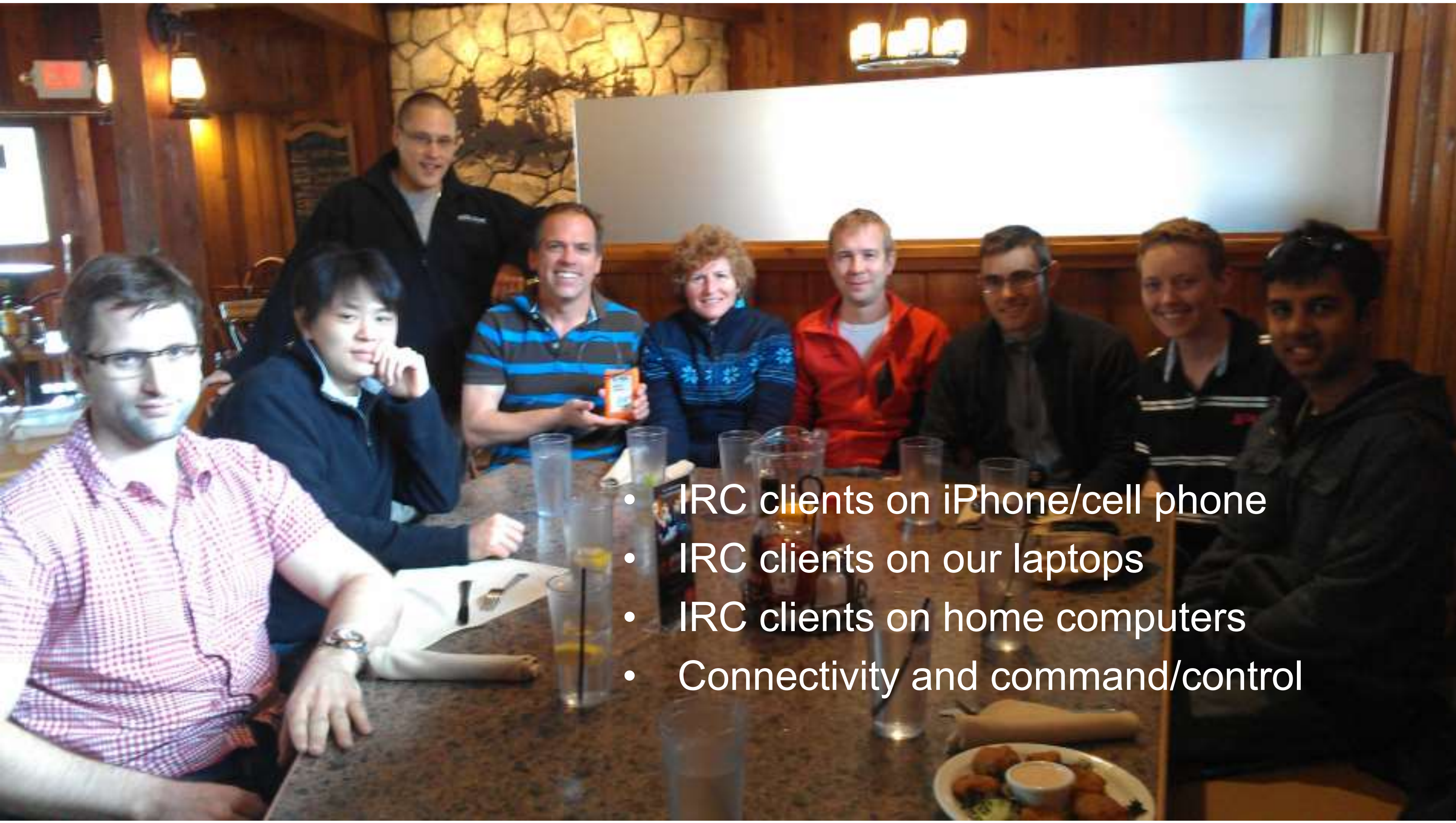
Created	Name	Messages
9 hours ago	VegaDark / Wikia	1
23 hours ago	Minetweak / Minetweak-IO	0
23 hours ago	Minetweak / Minetweak	1
3 days ago	SirCmpwn / GifQuick	74
4 days ago	SinZ / NotEnoughMods	7

TOP NETWORKS

Host	Channels
irc.freenode.net	97
irc.esper.net	30
irc.ponychat.net	8
irc.rizon.net	7
irc.gamesurge.net	5

Jet Propulsion Laboratory
California Institute of Technology

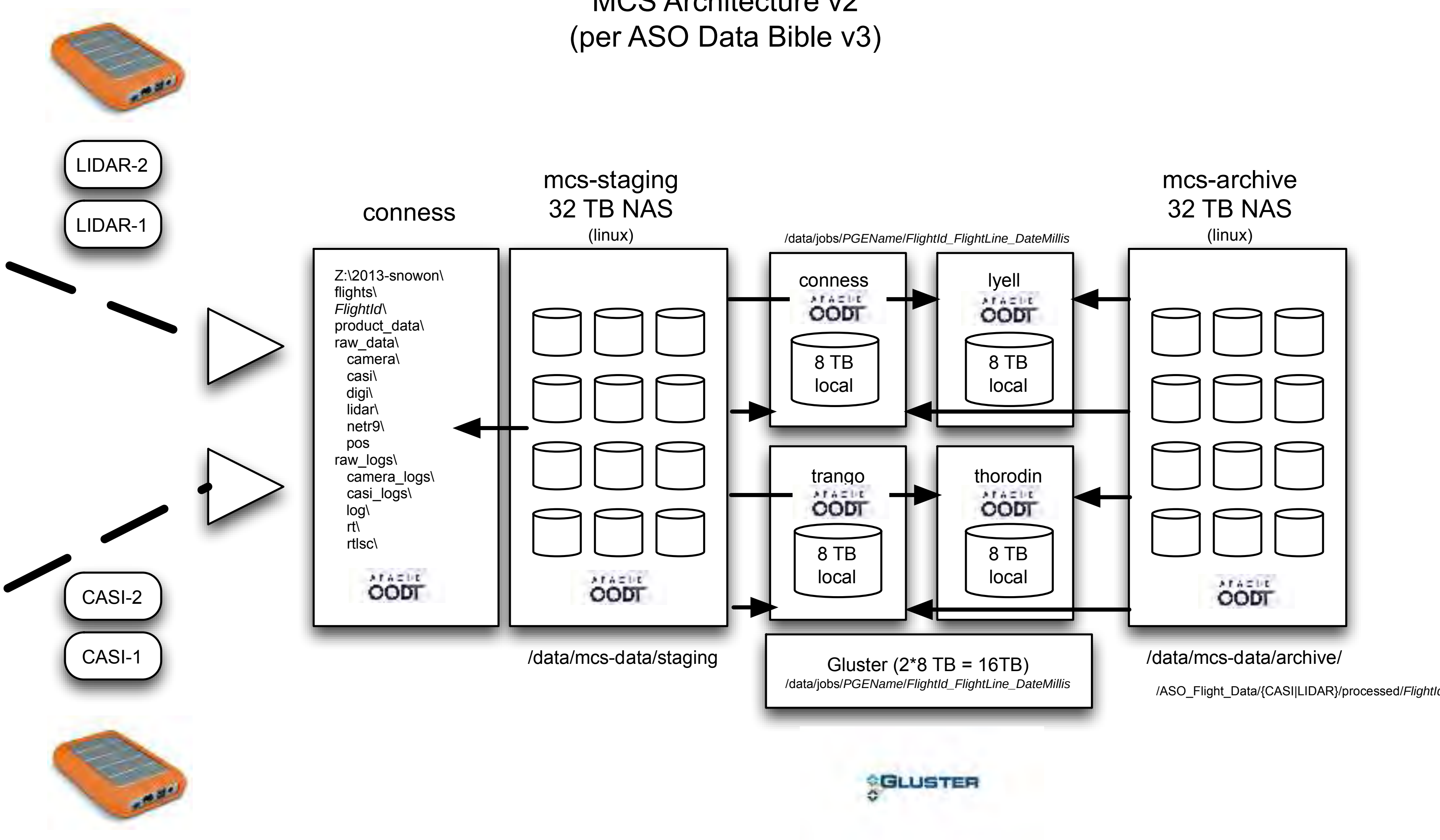
Full “situational awareness”



- IRC clients on iPhone/cell phone
- IRC clients on our laptops
- IRC clients on home computers
- Connectivity and command/control

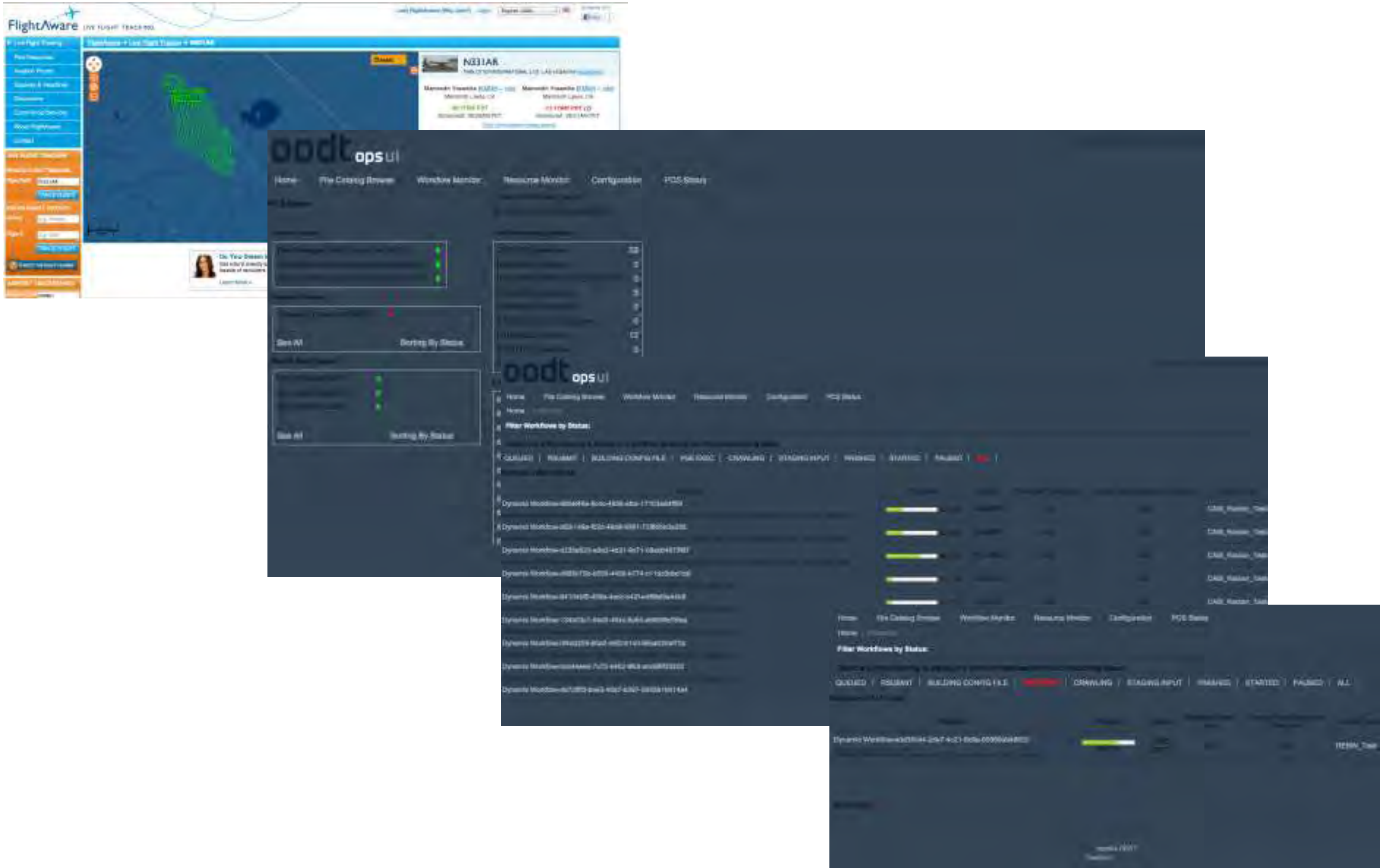
The Mobile Compute System Architecture

MCS Architecture v2
(per ASO Data Bible v3)



Jet Propulsion Laboratory
California Institute of Technology

System in Action



Lessons Learned

- *Algorithm Integration*
  - Be able to ramp up as a Compute Team when actual test data is produced
    - Efforts before at scale are futile before initial end to end, or even piece-wise science code is available; generating output; and required inputs are specified
  - Be adaptable to change
- *Flexibility in home/location of Compute System*
  - Went from very little understanding or control of remote site and monitoring to *fully automated 24 hour system*
  - Without ability to deploy Airborne Data processing software and system "at the vein", could have never made 24 latency
  - Data movement alone would have cost us 24-48 hours
  - Potential for multiple rodeos (or full week schedules) in a single week
  - The next rodeo comes at you a lot quicker in an airborne mission compared to a space mission so you have less time to adapt

Jet Propulsion Laboratory
California Institute of Technology

Lessons Learned

- *Use of Wiki as a Mission document, architecture and procedures repository*
  - All traditional mission documentation is available in the wiki
  - Full situational awareness on outputs, since wiki can email and notify us
  - If it can email, it can be automated, so Not-001 is plugged in and thus #aso-ops is plugged in, and thus *we* are plugged in
  - All Compute Operator Logs are captured in wiki, with commands/runtimes
- *Handoff between Flight Team and Compute Team*
  - Increasing communication abilities and use of IRC effective
  - Exchange in terms of the flight planning versus science goals could be better captured and automated
  - If automated and better captured, then Compute Team could respond more effectively
  - Need to figure out how to increase automation of LIDAR step
  - PDAL (Python LIDAR toolkit) a potential option, along with macros and scripts

Lessons Learned

Operator on Duty: Cameron E. Goodale
"CRANKshaft_"

- *Operator Rotation Strategy*
  - Lead operator versus backup
  - Instrument based breakdown (CASI lead; LIDAR lead; overall compute lead; SA/devops team lead)
  - Weekly rotations with daily "pow wow" with Tom and Chris
  - Lessons learned drive improvements in 24 hour latency and need for full situational awareness
- *Open Source Software*
  - Data processing system development was configurable
  - Cost less than just building scripts (we tried that and used it to flush out the control flow in the beginning, but didn't provide full sit. awareness)
  - Made a bunch of updates to upstream projects (OODT and Tika) because of ASO
  - Got a whole bunch more people interested in our project and disseminated the work more effectively

Jet Propulsion Laboratory
California Institute of Technology

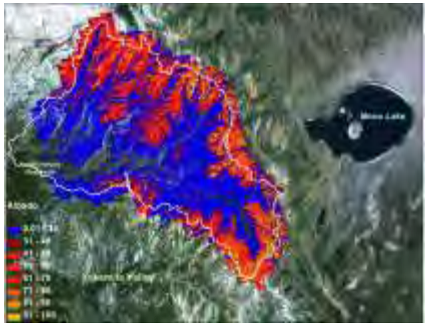
Lessons Learned

Operator on Duty: Shakeh Khudikyan "MathGeek"

- *Proprietary Algorithm Software*
  - Use of IDL and ATCOR are expensive
  - Even with per host license policy (with many cores and mem), and virtualization, we were still constantly purchasing IDL and ENVI licenses and ATCOR licenses to scale up
  - We had the progenitor of these technologies working for the team, so sunk costs in licenses were swept up by the "Boardman intellect"
    - And you can't put a cost on comedic relief and team camraderie and heart and all around winningness that is Joe Boardman
  - Had difficulties with centralized licensing server on scale out
  - Inheritance of code and algorithm knowledge made it too difficult to rewrite and swap with FOSS alternatives
- *Data backup strategy*
  - Rugged disks, and flight operator rotation can save a lot in expected FedEx costs
  - Buy lots of rugged disks, and upfront storage, and be ready to buy more

Jet Propulsion Laboratory
California Institute of Technology

Lessons Learned



- *Compute Team Deployment Strategy*
  - Bring leads in early on to optimize the architecture
  - In parallel, work algorithm integration, and system integration issues
- *Work towards automation as fast as possible*
  - Don’t build the perfect system (1 out of every 3 jobs failed, but if the mean time to fail is <5 min, then who cares and what have you lost)
  - Schedule Multiple copies of the same job
  - Wipe products that you can generate again
  - Realize that you are an SCF and not an archive but if you need to become an archive..
- *Data science strategy*
  - Get algorithm scientists that understand computer scientists, and computer scientists that are willing to read a paper on radiative forcing and understand it
- *Accountability*
  - Understand the archive structure and build scripts to say what’s done and what’s not done

Contributing to space and future airborne missions

- EVS work being baselined off this project
- Uniqueness between ASO and CARVE
  - Pushing the compute system “to the vein” and processing for latency
- SMAP and NPP already back and forth technology and use case flow
- SWOT, HYSPIRI and future missions will benefit from the Apache OODT and other innovations from this project
- Represents an “in the field” rugged deployment of Apache OODT, unlike traditional on large computing devices and storage networks where OODT is traditionally deployed
- Strategies and lessons learned will inform science computing facilities, that more naturally lend themselves towards archive transitions

Jet Propulsion Laboratory
California Institute of Technology

Getting to the entire Sierra Nevada

Jet Propulsion Laboratory
California Institute of Technology

Summary

- ASO provides a more complete picture of snow water storage for water management
- Compute Team supports ASO hardware, software, data products, backup, and storage
- Many lessons learned constructing system
- Preparing us for future airborne missions including EVS and also space-based missions