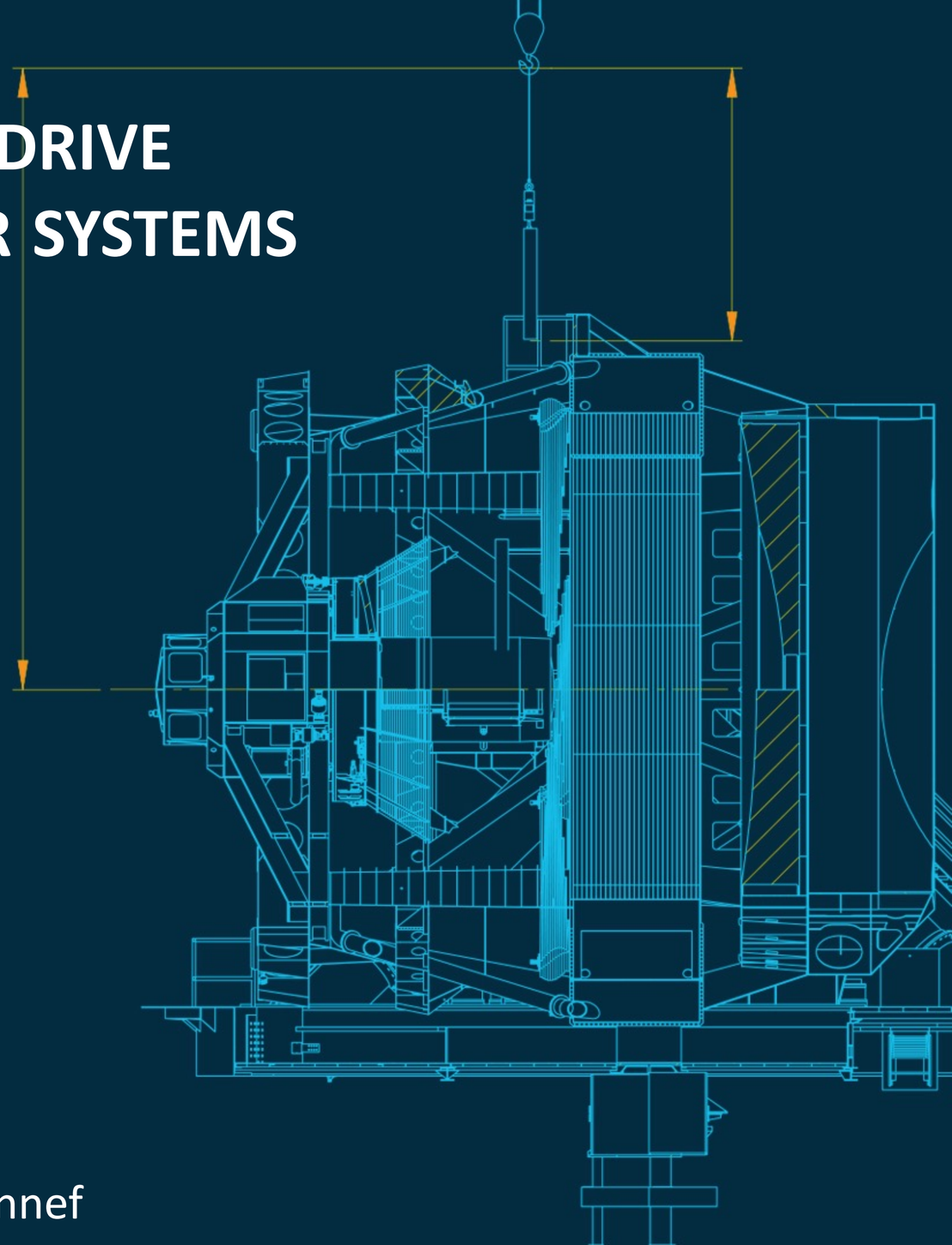
How Big Data missions like LSST drive new models of how we build our systems - and our teams

Frossie Economou

frossie@lsst.org
Technical Manager, SQuaRE, a Data Management Team
Project Manager, LSST Science Platform



Large Synoptic Survey Telescope

2020-01-13 Bad Honnef

Hallo! 🇩🇪🇪🇺

Tucson, AZ

22°C

- 🙏 WE-Heraeus / Organisers / Martina Albert
- **Young people please come and talk to me**
  - You don’t have to think of a clever question first!
  - We can be socially awkward together!
  - Your English can’t be worse than my German
  - Also: Je parle Français 🇫🇷 / Μιλάω Ελληνικά 🇺🇾 / I speak geek 👾
- I talk too fast even for Americans! Ask me to slow down.
- If you see something in green I can tell you more - A LOT LOT MORE - about it!

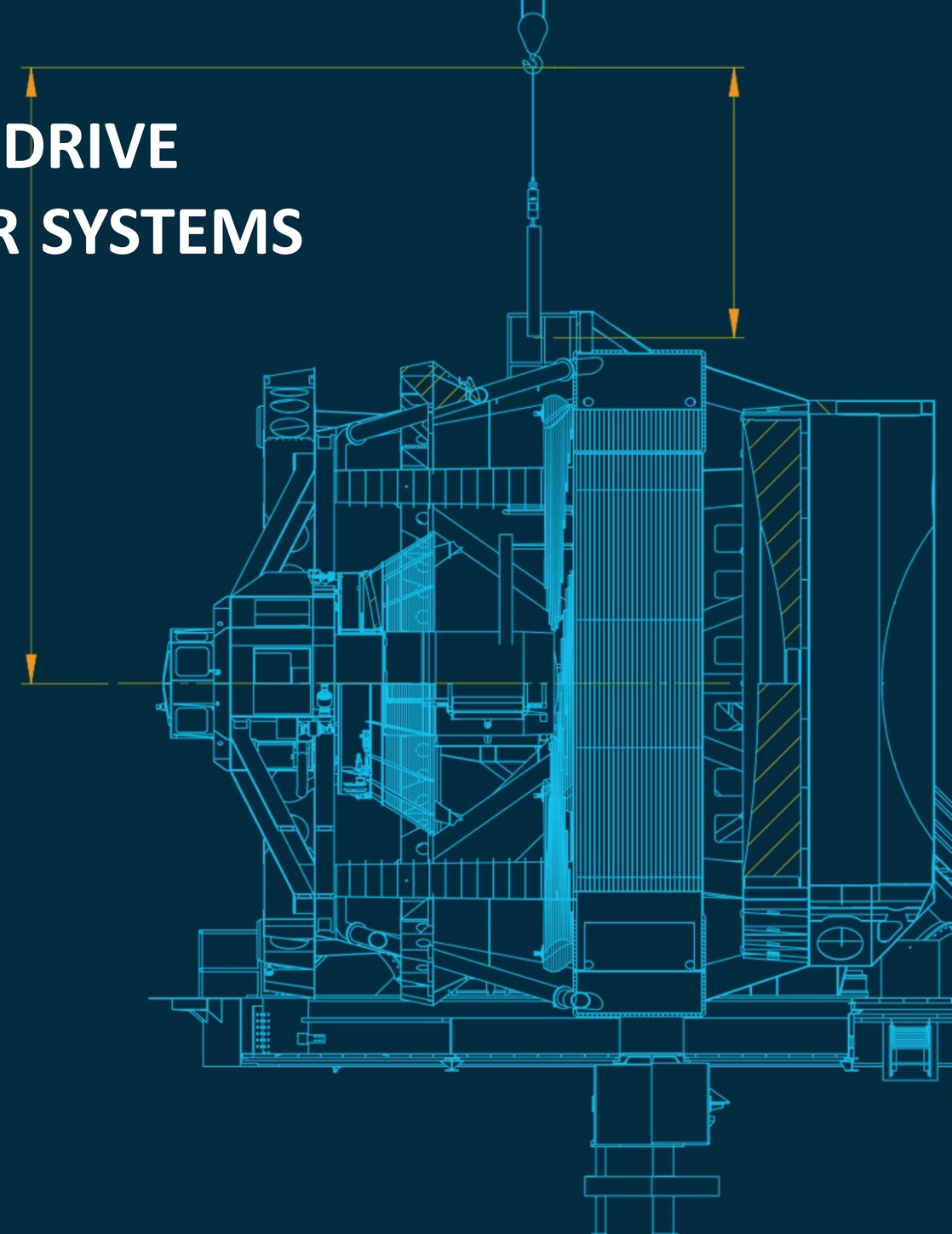
frossie@lsst.org

How Big Data missio[illegible]e LSST drive new models of [illegible]ld our systems - and our t[illegible]

But first some BREAKING NEWS

Frossie Ecor[illegible]

frossie@lsst.org
Technical Manager, [illegible] Management Team
Project Manager, LSST [illegible] Platform



Large Synoptic Survey Telescope

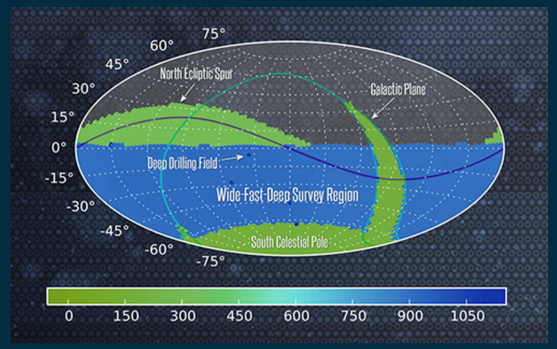
RIP Large Synoptic Survey Telescope

- LSST - the project/organisation -> NSF Vera C Rubin Observatory (vro.org)

- LSST - the actual astronomical telescope -> Simonyi Survey Telescope

This is all only a few days old, brain re-wiring in progress !

- LSST - the 10 year data campaign -> Legacy Survey of Space & Time (also LSSTCam etc)

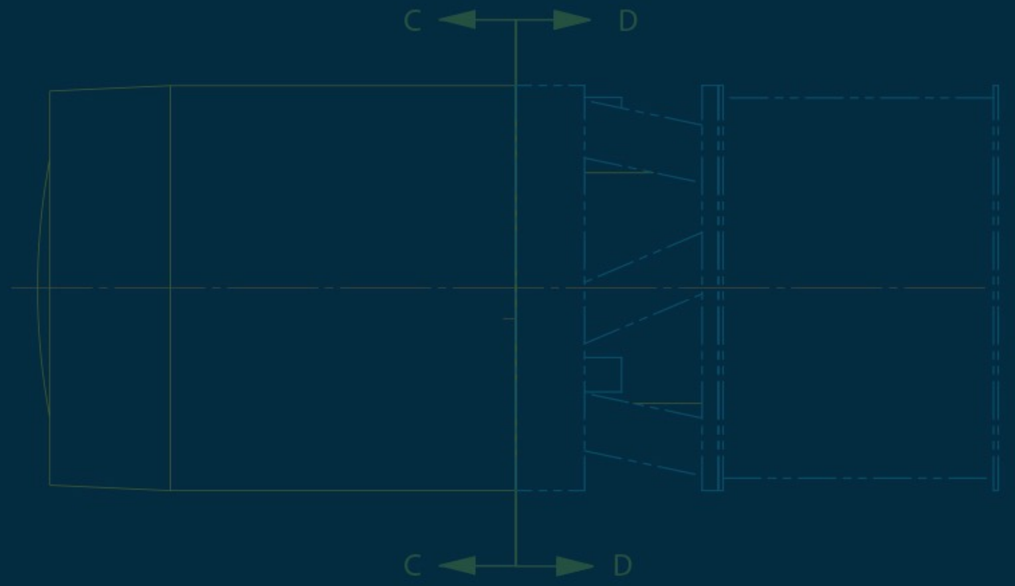


frossie@lsst.org

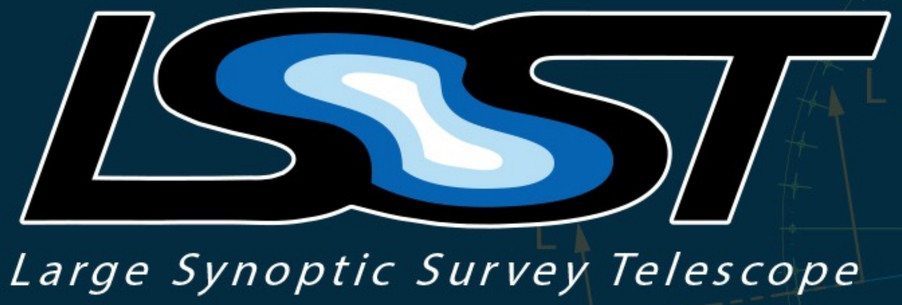
This Talk Is Three Talks

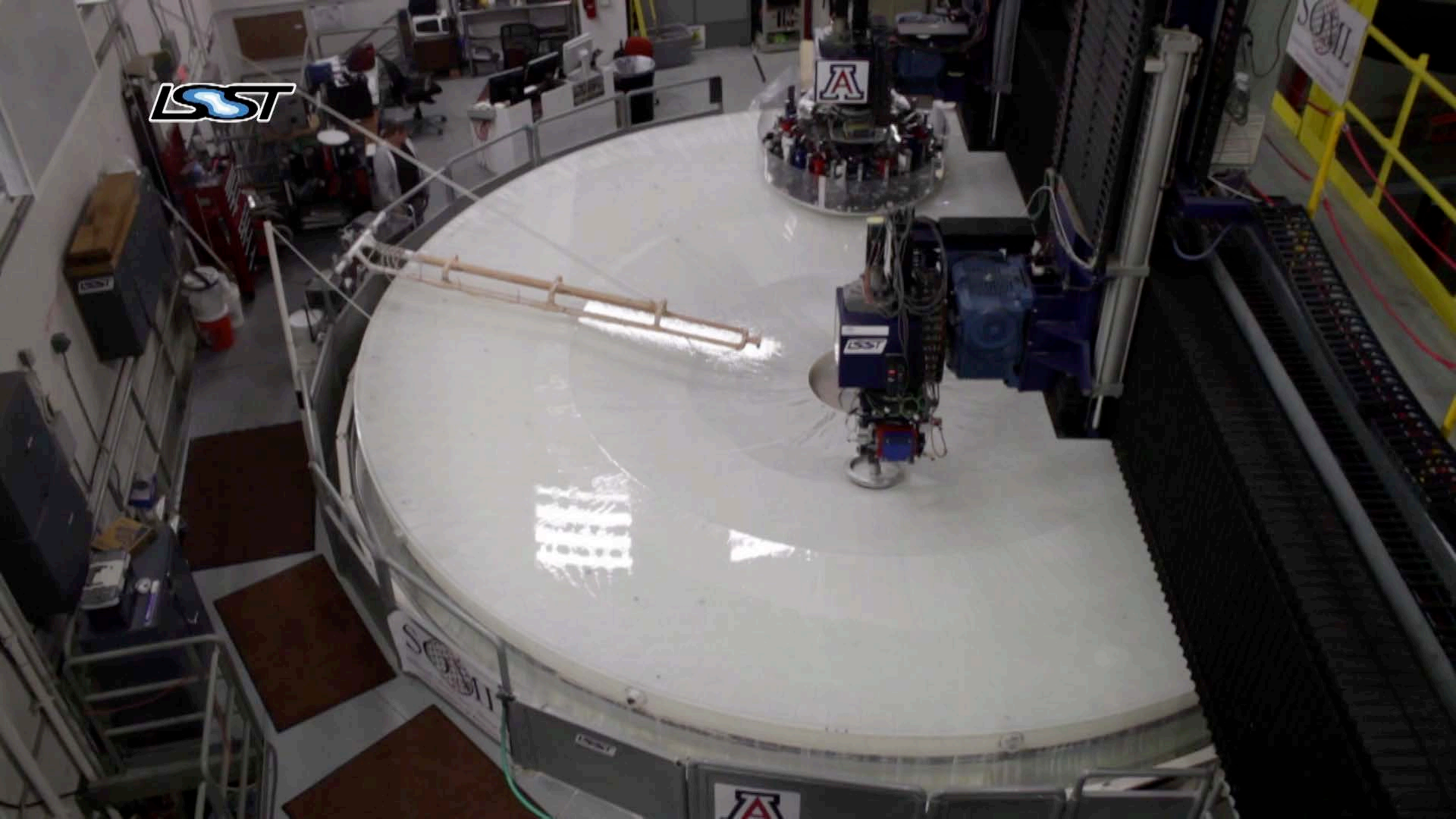
- Part I: Educational
- Part II: Moralising
- Part III: Technical

Let’s Go!



Big Thing On A Mountain




LSST

~ Dec 2019

← ~ Oct 2019

photo: Wil O'Mullane


TRANSP. JAVIER CORTES S.A.
VOLVO


+Y

Coating Chamber
Made In 🇩🇪


LSST

Little White Puffy Science
Clouds In The Data Sky

Large Synoptic Survey Telescope

Everything Is Big About This Project

- Goal is 60 seconds from shutter close to the broadcast of transient alerts
- Half an Exabyte (~500 PB) of data holdings over the 10-year survey
- ~ 7 trillion single epoch sources (ie “rows in the database”)
- ~ 100 individuals in Data Management alone (~ 68 FTE)
- ~ 7,500 astronomers with data rights

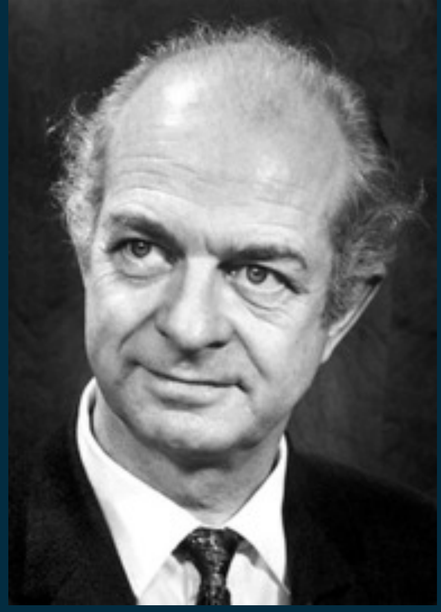
  … the vast majority of which can’t possibly download the data

  … and probably want it at the same time

  … in order to do whatever they used to do with data they COULD download (ie anything and everything)

  … OMG WHAT ARE WE GOING TO DO 😱

Here is what we are NOT going to do: a morality tale



Linus Pauling:

- One of the founders of quantum chemistry
- Devised the first accurate scale of element electronegativities
- Inspired Watson Crick & Franklin (DNA)
- One of only two people ever to get two Nobel prizes in different subjects
- Peace activist during the Vietnam War

Also Linus Pauling:

- Promoted mega-doses of Vitamin C as a cure for arteriosclerosis, cancer, the common cold
- Proposed that people with genetic diseases get forehead tattoos to avoid mating accidentally with people with the same genetic disease

Being Smart And Good At Physics != Being Good At Everything

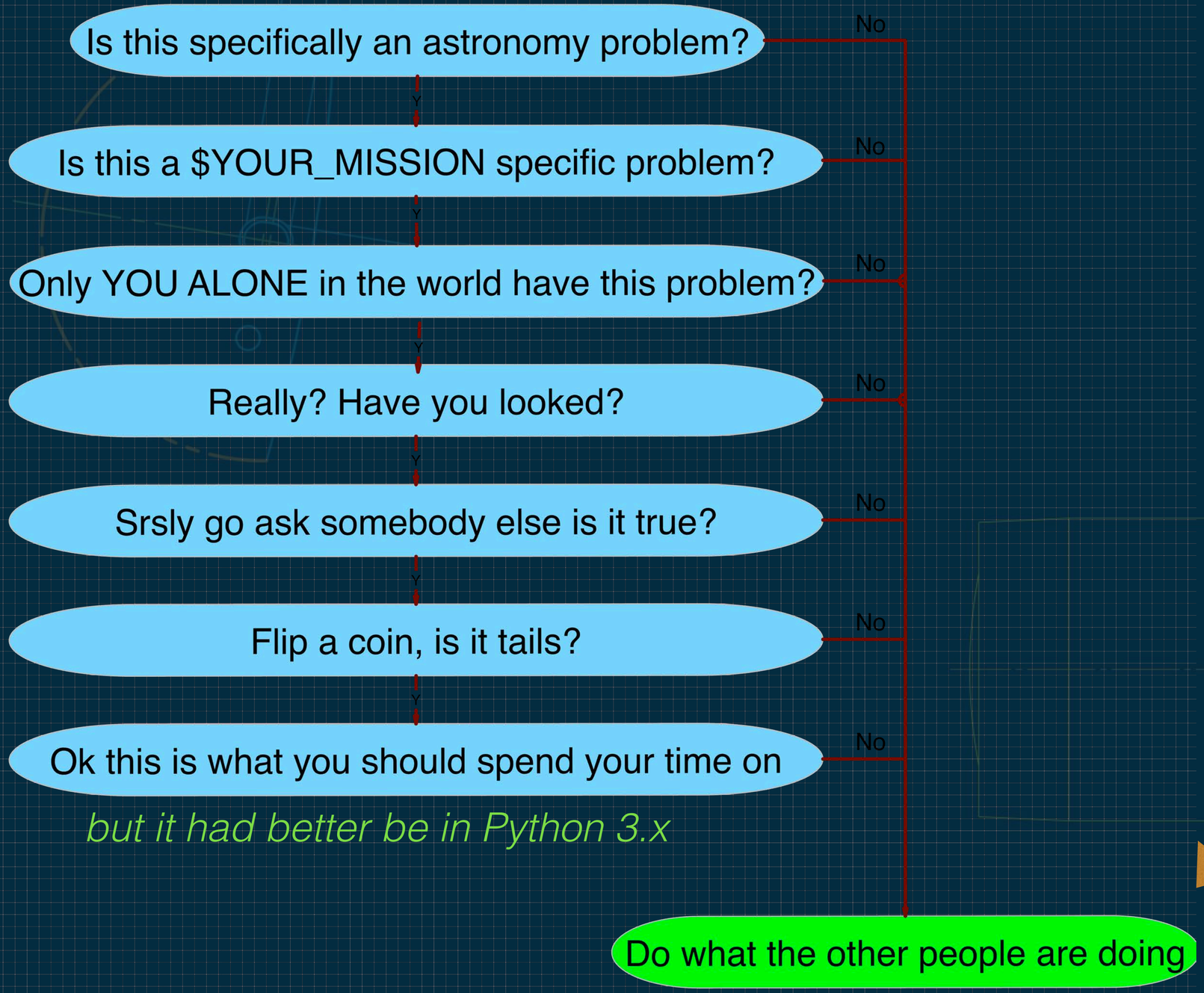
- Phsycists are really bad at knowing this
- Astronomers are even worse at knowing this
- And when it comes to software, they are absolutely terrible at knowing this

Astronomers suffer from the narcissism of omni-competence

– Matt Turk

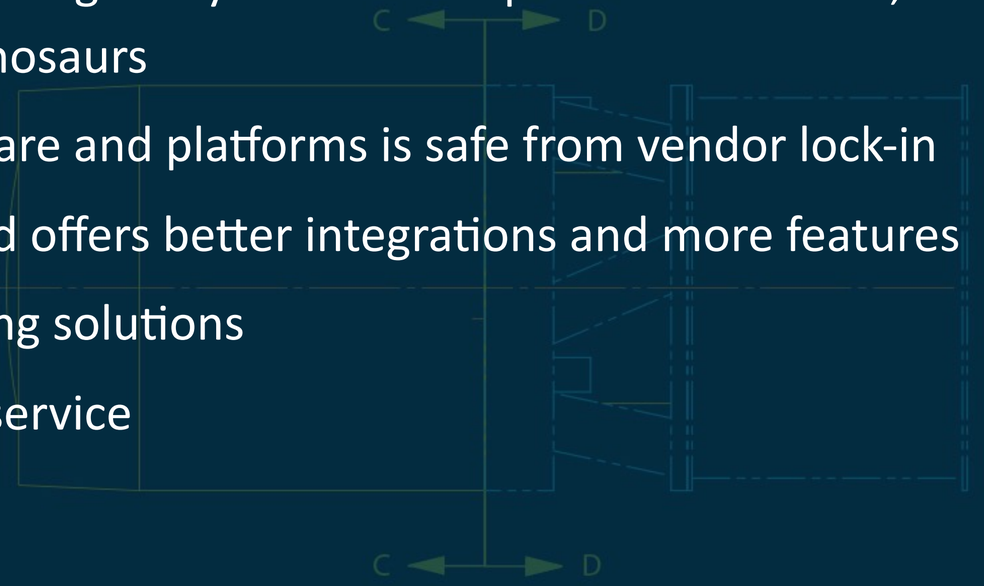
Example 1: half an Exabyte of data holdings sounds like a very special problem, but Facebook had that problem back in 2014 - this is not a novel problem
Example 2: … but neither is dealing with such problems “trivial” or “just software”

So You Want To Write Some Software?

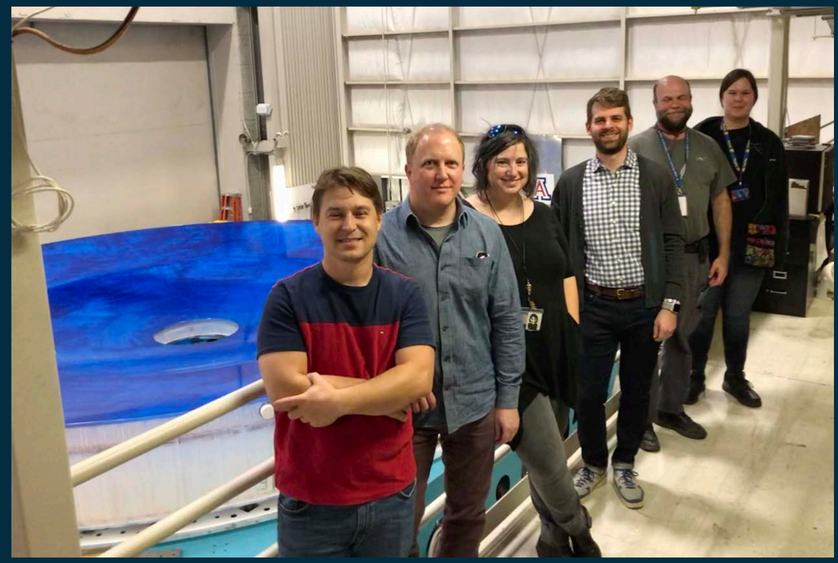


Why Does It Matter?

- In the Big Data era we need scalable, growing systems w/small hybrid teams (good luck hiring though)
- Software issues are now often in the critical path to scientific discovery
- The dotcom world has solved A LOT of our problems; we need cross-fertilisation and we're not going to get it if we are know-it-all snobs. Modern astronomy is an software as much as a scientific discipline.
- In astronomy we have always over-produced students knowing many would end up outside the field; we owe it to them to make sure academia projects are not dinosaurs
- Open source triumph means using non-homegrown software and platforms is safe from vendor lock-in
- Using newer techniques makes it easier to stay current and offers better integrations and more features
- Adopting new paradigms naturally drives better engineering solutions
- General SaaS experience sets expectations of the level of service



Generalist, High-Performing Teams = ❤ ❤ ❤



Software development services (continuous integration at ci.lsst.codes, software distribution, container builds, science pipelines release infrastructure, managing Gitub orgs, git-lfs service). The Notebook aspect of the Science Platform (aka nublado). Notebook service for test stands/summit (single-machine nublado). The QA infrastructure including a metrics framework (lsst.verify), metrics curation (SQuaSH), interactive and API access to the data (squash.lsst.codes), metric alerts (simple for now). LSST-the-Docs, our documentation build-publish system (lsst.io). Template-based reporting (data quality reports, end-of-night) via nbreport. Slackbots that perform common tasks from JIRA ticket expansion to information retrieval and document creation. The Engineering Facilities Database at the summit, LDF, base if required. A microservices platform (api.lsst.codes). The kafka infrastructure used by the EFD, the microservices platform and (soon) more. community.lsst.org management. JIRA project administration for DM, EPO, PUB, anybody who asks. Mailing lists. Support for tutorials and workshops (ad-hoc LSP deployments etc). We do all our own IT: certificate management, secret management, AWS and Google cluster administration and management, security updates, monitoring for all out services, managing the MacOS build cluster in Tucson

Cloud technologies are designed with the intent of allowing a few devs to do a lot; and a confluence of converging technologies allows teams to do more without exploding the size of the technical stack

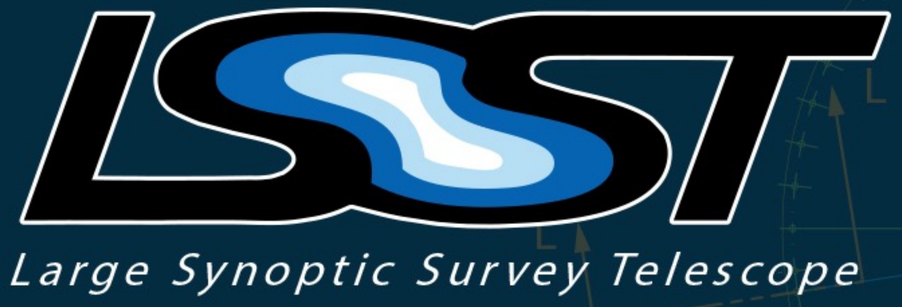
frossie@lsst.org

The Challenge (Opportunity?) To Getting More Serious About Your IT

There are some things we don't want to imitate Silicon Valley at…

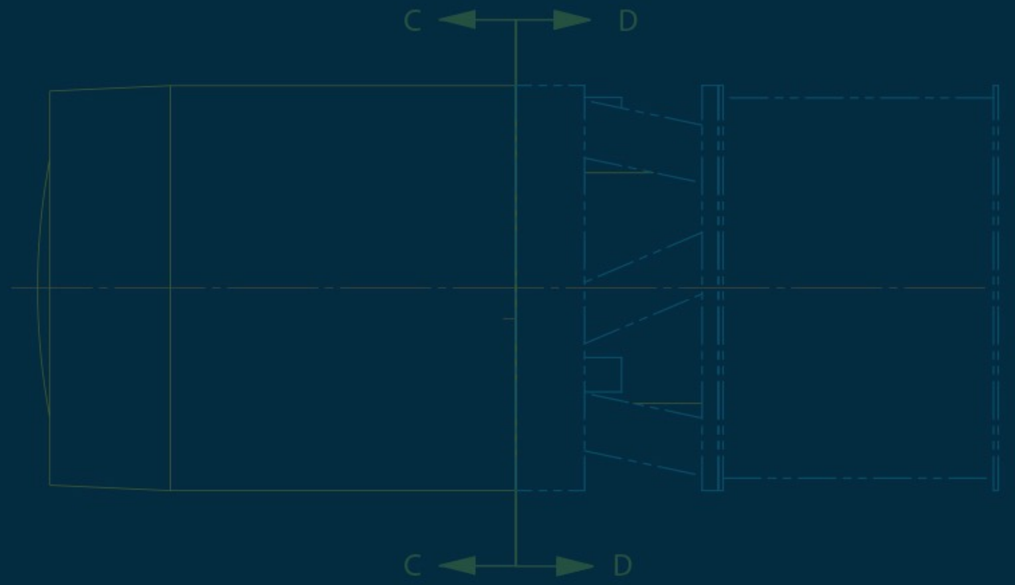
frossie@lsst.org

Specific Examples From LSST

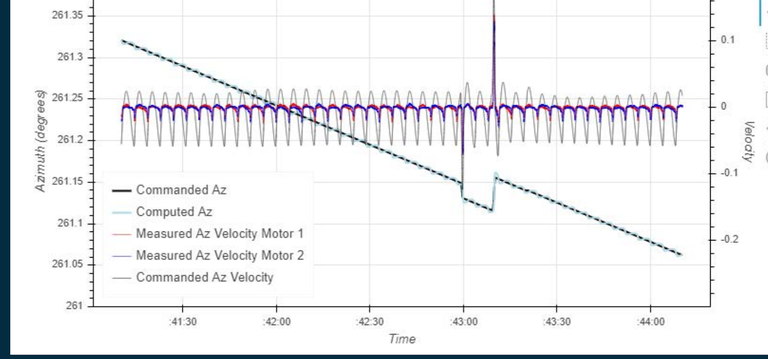


Github

- We switched from in-house git repository management to Github at the beginning of construction
- Huge success that signalled and solidified open source and good developer culture
- But the real gain has been the features, services and integrations ecosystem that did not exist when we made that decision => this is the dividend for doing what other people are doing
  - code reviews
  - OAuth provider
  - security alerts
  - etc.
- Ask me if I am worried if Github goes bankrupt



Engineering Facilities Database

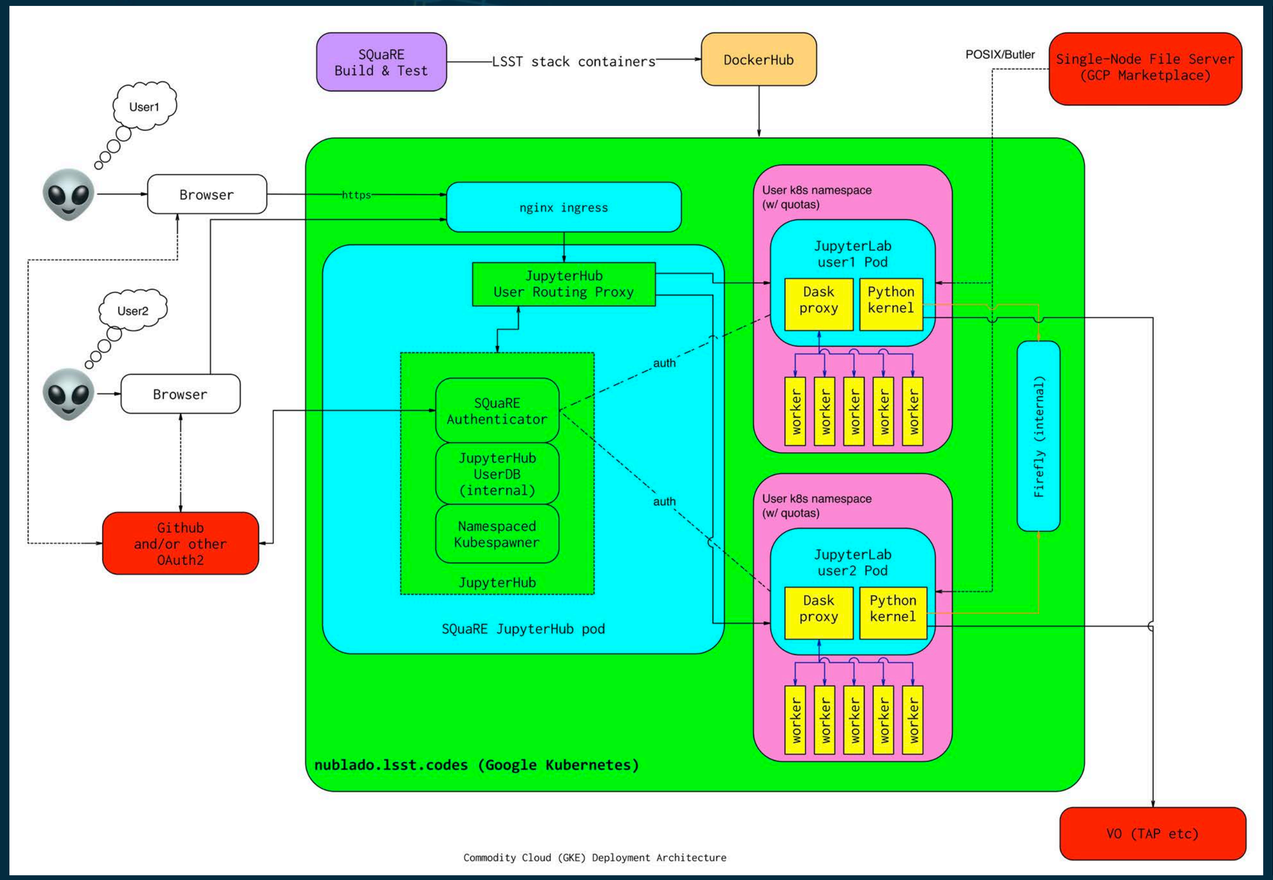


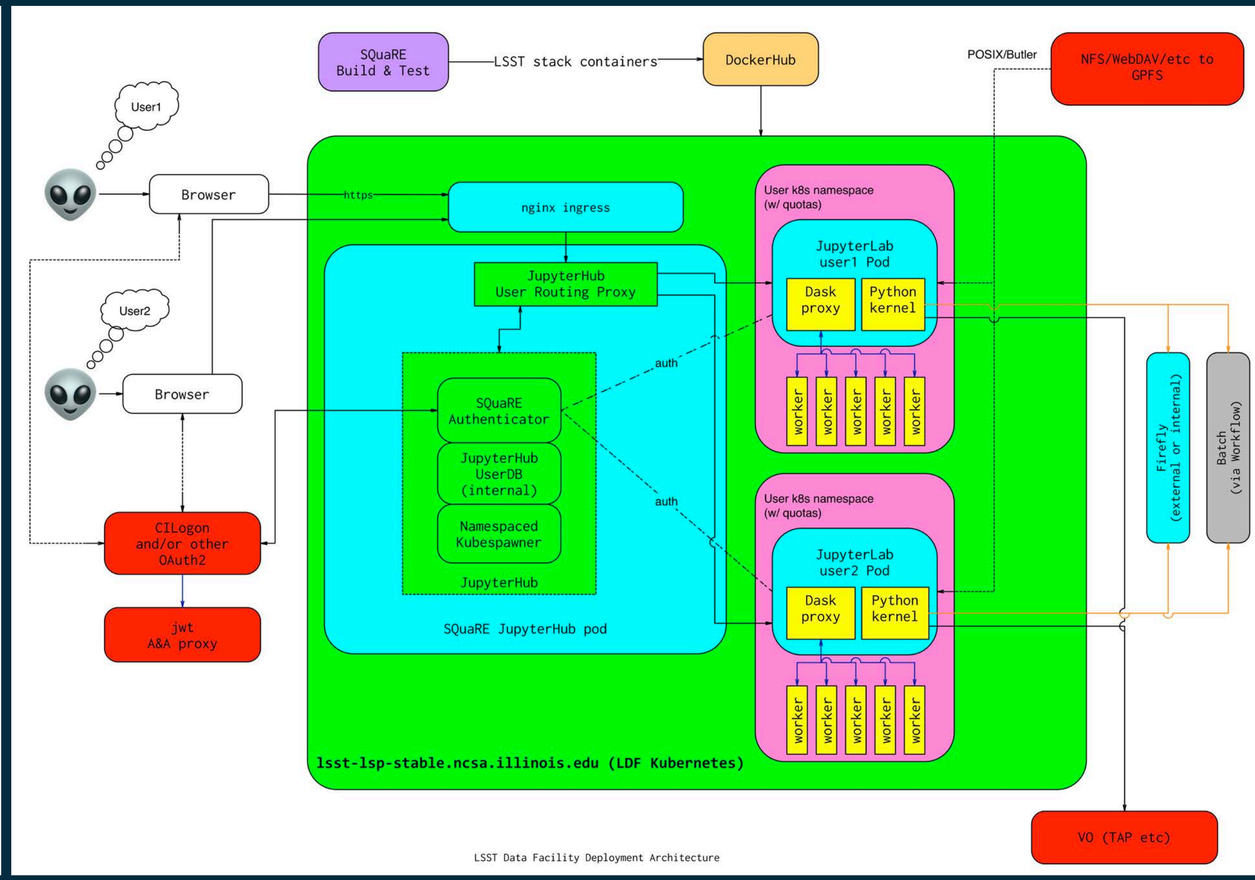
- A traditionally “Telescope group” problem, needs to store telemetry at ~ 50Hz
- At LSST we replaced an in-house system with an InfluxDB based ecosystem on k8s with great success
- This is not only highly performant but it brought a powerful feature ecosystem (chronograf, kapacitor, etc) and exposed the data to “scientist” workflows

frossie@lsst.org

Nublado & the LSST Science Platform

- How do you give people the flexibility they had on their laptops - but server/data side?
- How does your code **and services** go where the data is?
- Time for a demo?

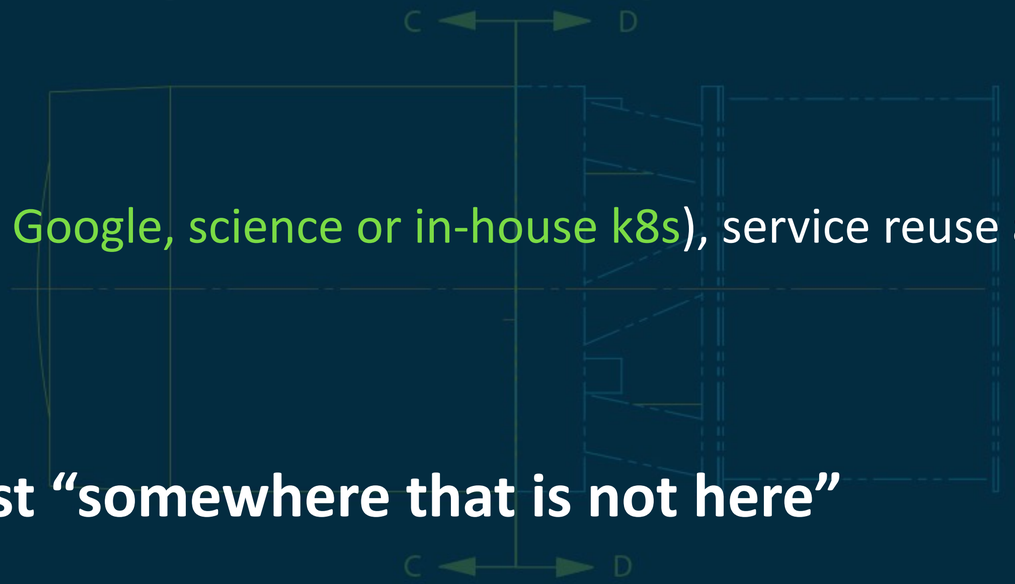




frossie@lsst.org

Want to be cloud-ready? Deploy your services on kubernetes

- Auto-scaling
- Elasticity
- Reproducibility
- Service discovery, rolling updates, load balancing, lifecycle management, volume management, resource monitoring, ingress, healthchecks, ponies…
- Best practices service deployment (gitops etc)
- Unprecedented opportunity for service portability (AWS, Google, science or in-house k8s), service reuse and services preservation (👋 Alex)
- Kubernetes Will Save Astronomy ™
- **=> “Cloud” is an engineering paradigm not just “somewhere that is not here”**



STRONGER TOGETHER

Live long and upstream

Also don’t be Bad Linus Pauling