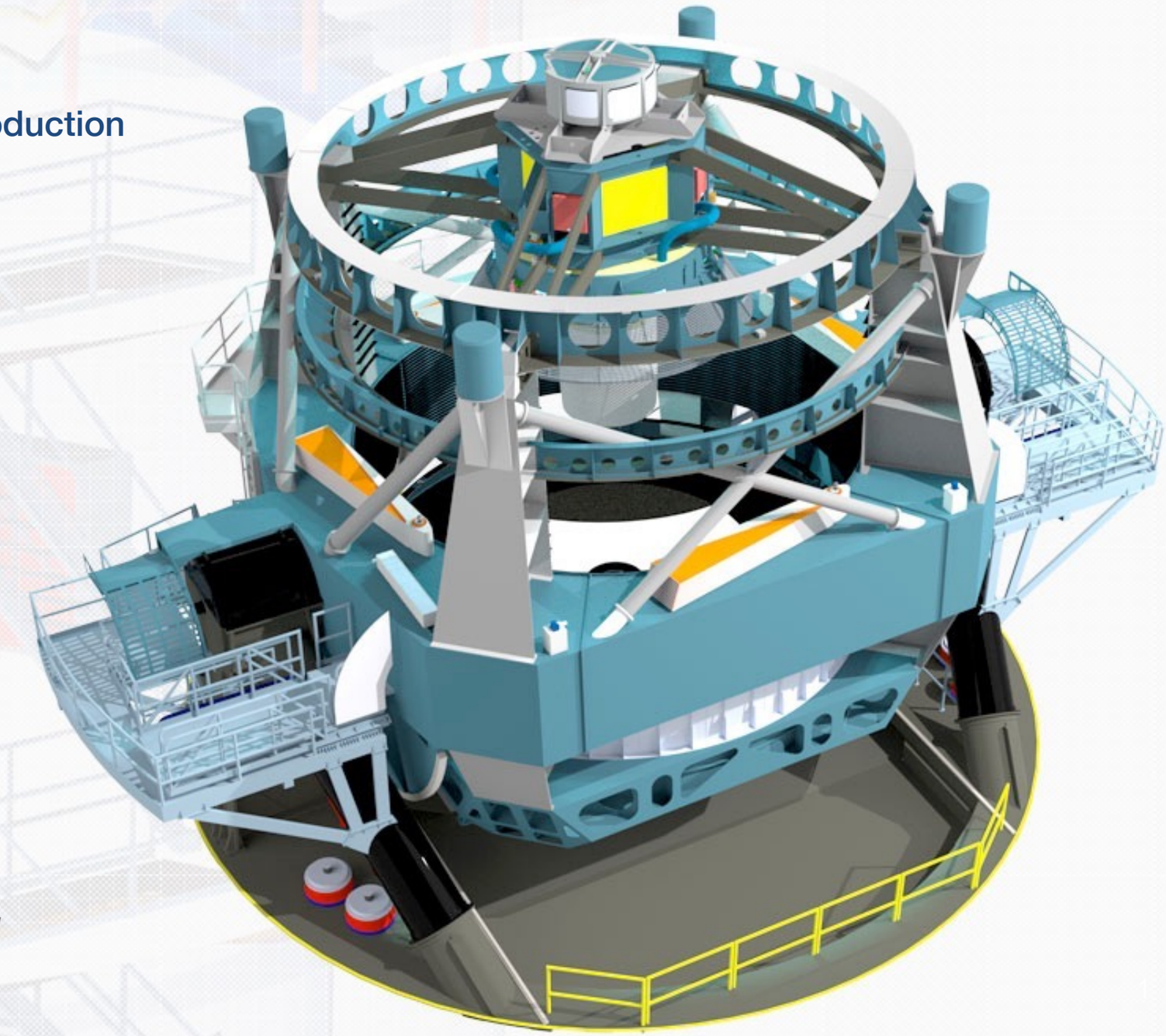
Large Synoptic Survey Telescope

Introduction & data management requirements

John Swinbank
Technical Manager, LSST Data Release Production
swinbank@princeton.edu



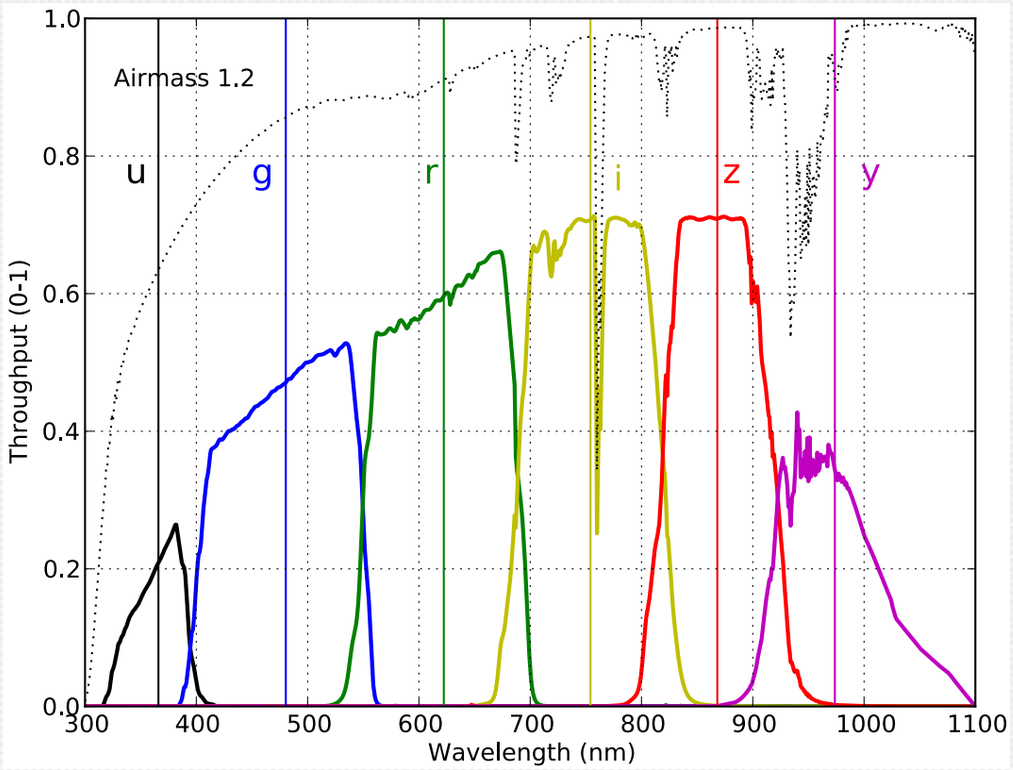
IVOA Interoperability Meeting
June 2015

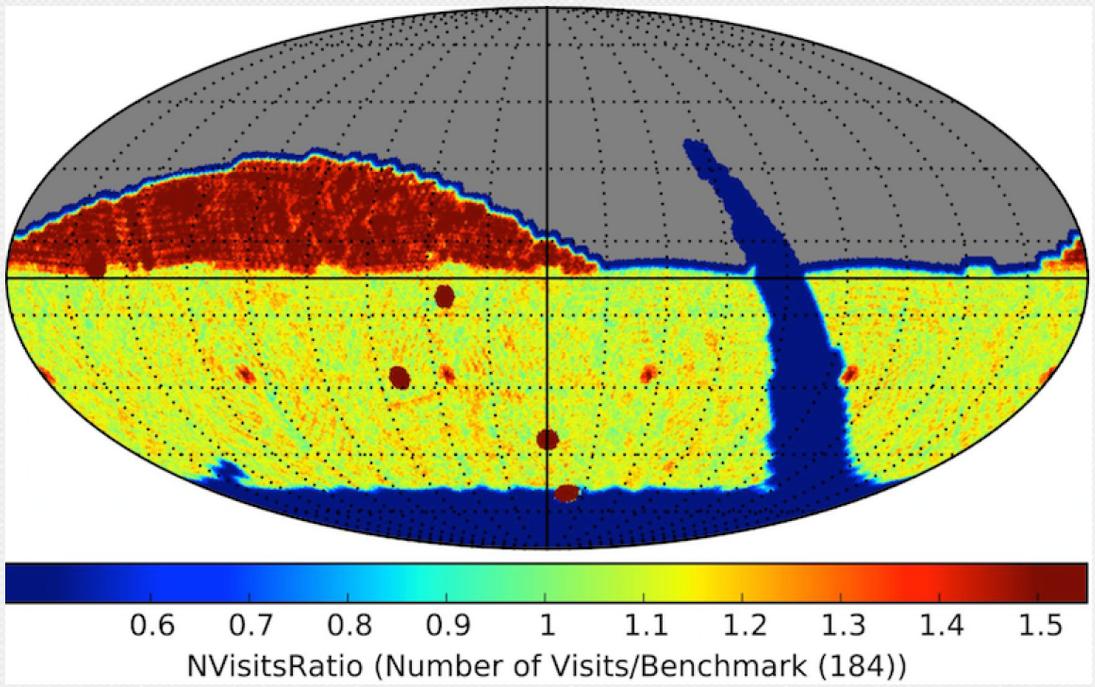
Outline

- The facility
  - Design & capabilities
  - Status & timeline
- Science goals
- Data products
  - Alerts
  - Catalogues & images
  - Community driven
- Software stack
- Data distribution & VO

Deep, Wide, Fast: pick any three

- The telescope:
  - 8.4m (~6.5m effective)
  - ~10 deg^2 field of view
  - 3.2 gigapixel camera
  - 6 bands (ugrizy; 320-1050nm)
  - 2s readout; 5s slew & settle
- The survey:
  - 18000+ deg^2
  - 10 years
  - 30s exposure/visit
  - ~825 visits
  - r~24.5/visit; r~27.5 total





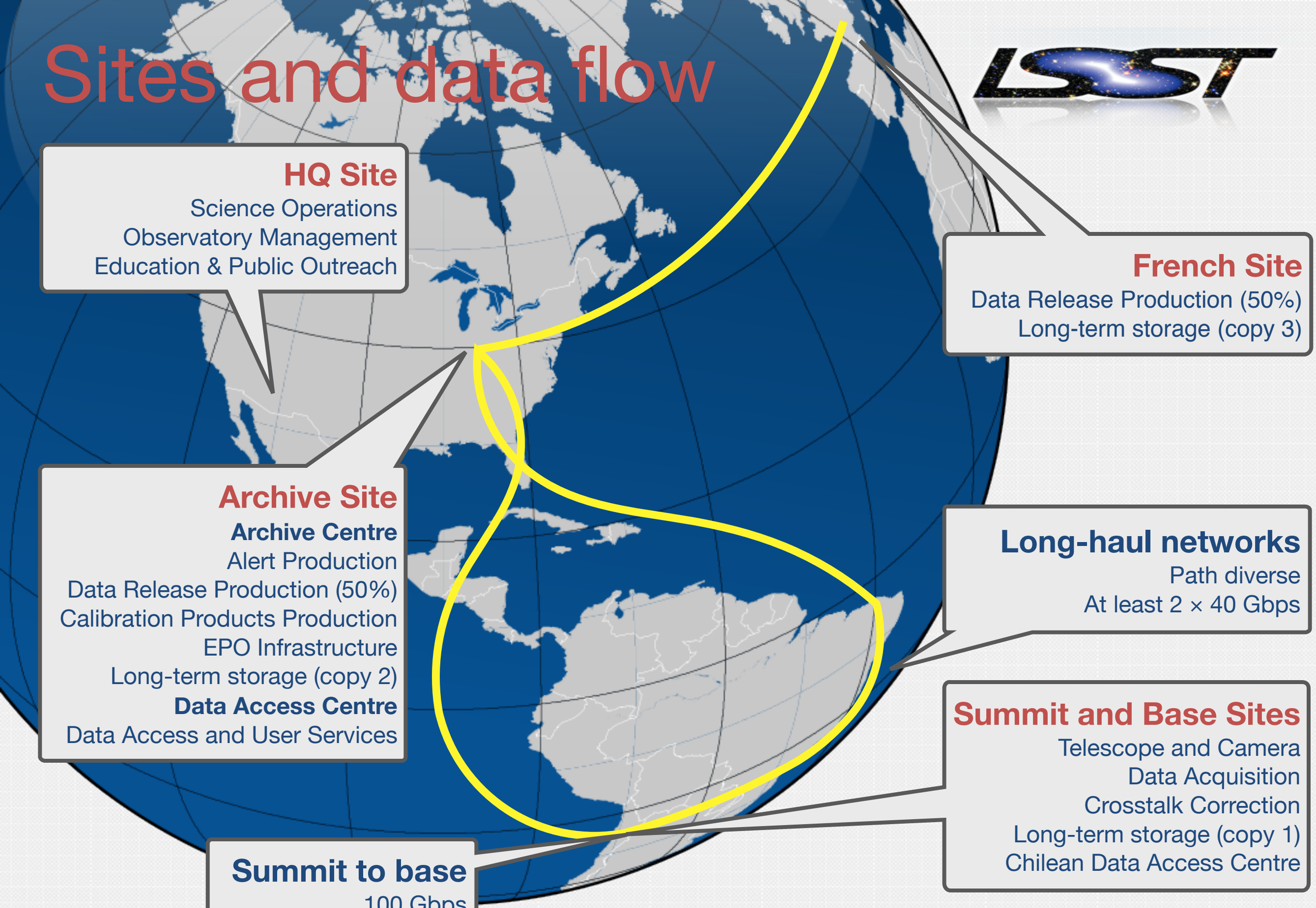
Figures: Ivezic et al

Location


LSST Site
La Serena
Santiago


Cerro Pachón – Future site of the LSST
LSST Rendering on El Peñón
SOAR
Gemini
Cerro Pachón ridge – view from northwest

Sites and data flow



HQ Site
Science Operations
Observatory Management
Education & Public Outreach

Archive Site
Archive Centre
Alert Production
Data Release Production (50%)
Calibration Products Production
EPO Infrastructure
Long-term storage (copy 2)
Data Access Centre
Data Access and User Services

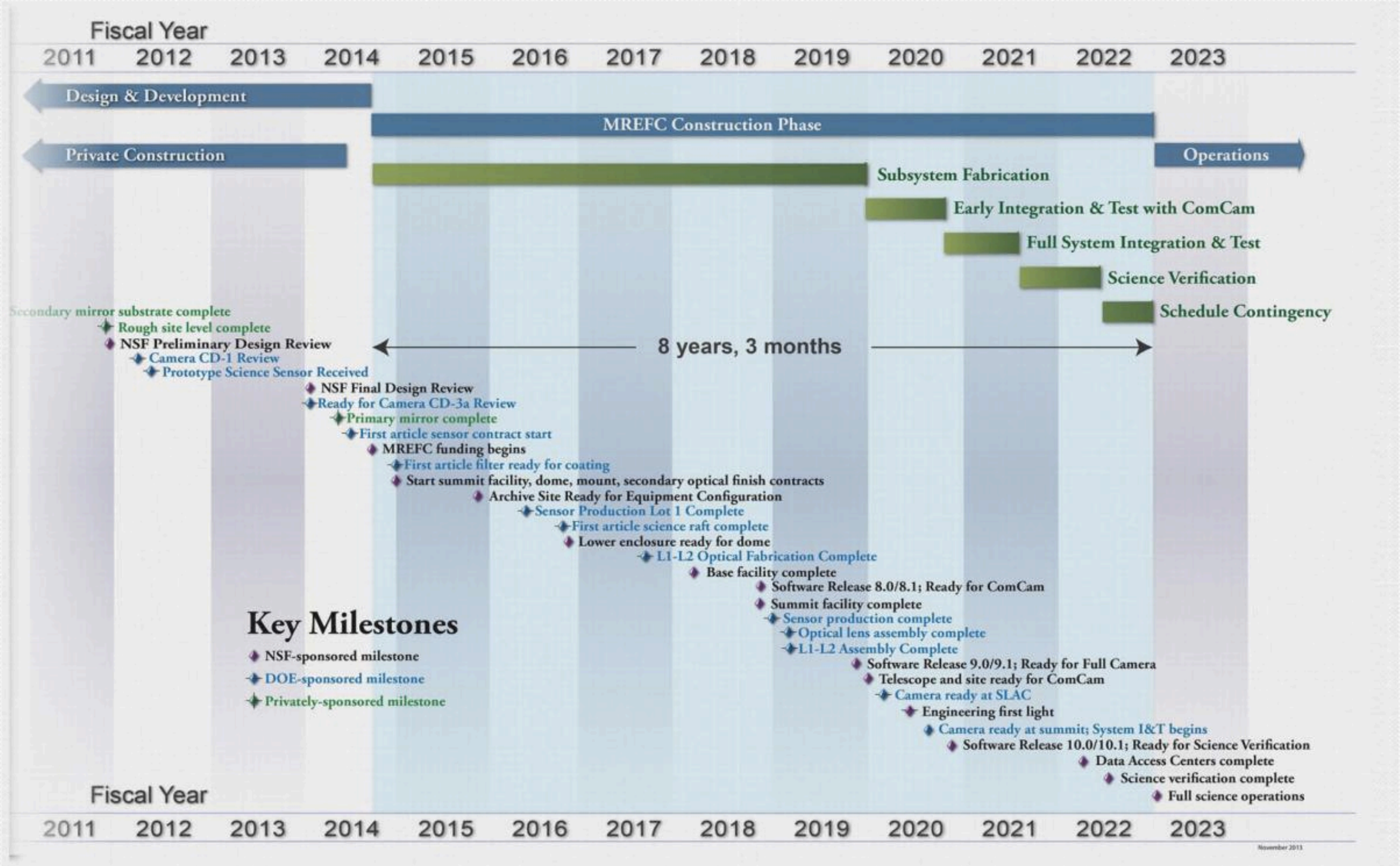
French Site
Data Release Production (50%)
Long-term storage (copy 3)

Long-haul networks
Path diverse
At least 2 × 40 Gbps

Summit and Base Sites
Telescope and Camera
Data Acquisition
Crosstalk Correction
Long-term storage (copy 1)
Chilean Data Access Centre

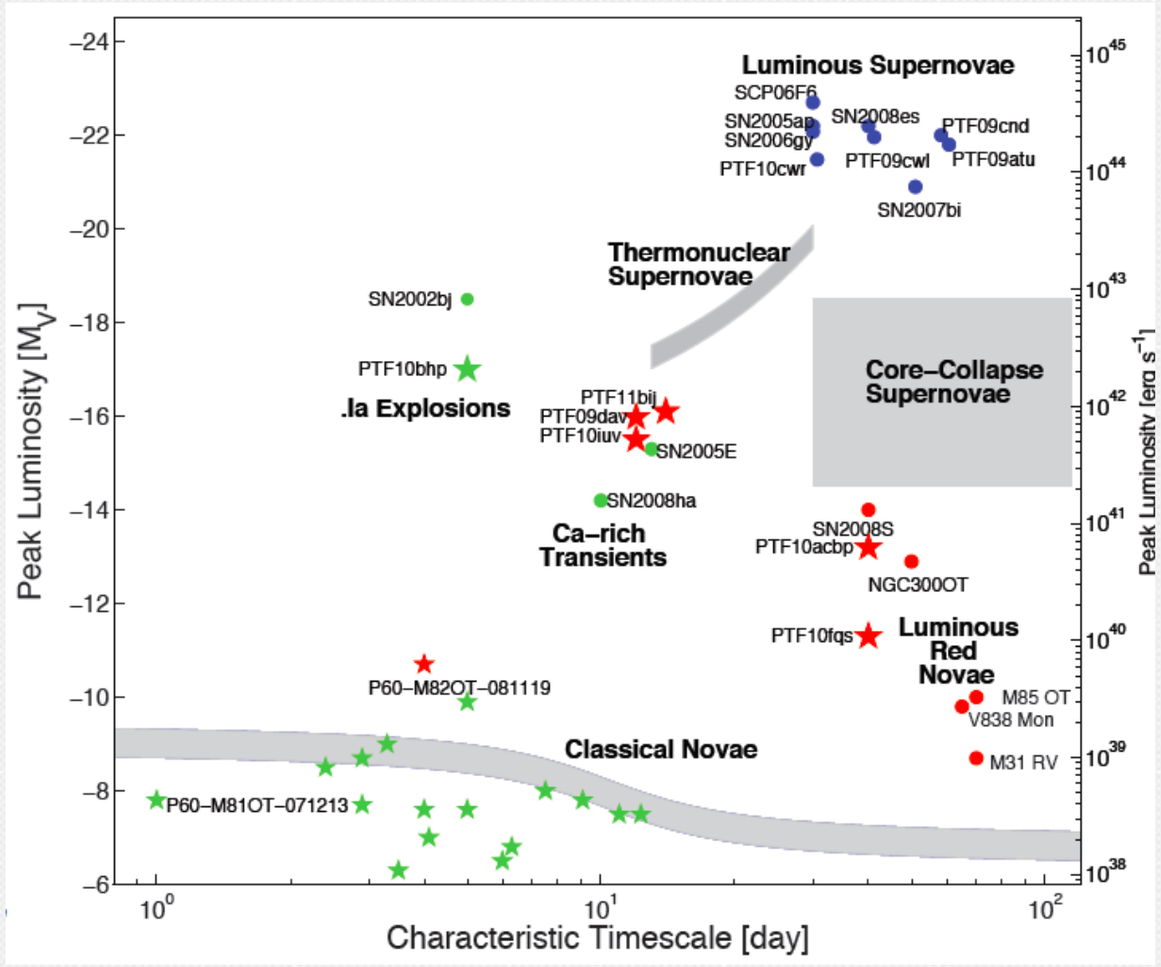
Summit to base
100 Gbps

Survey operations: 2022



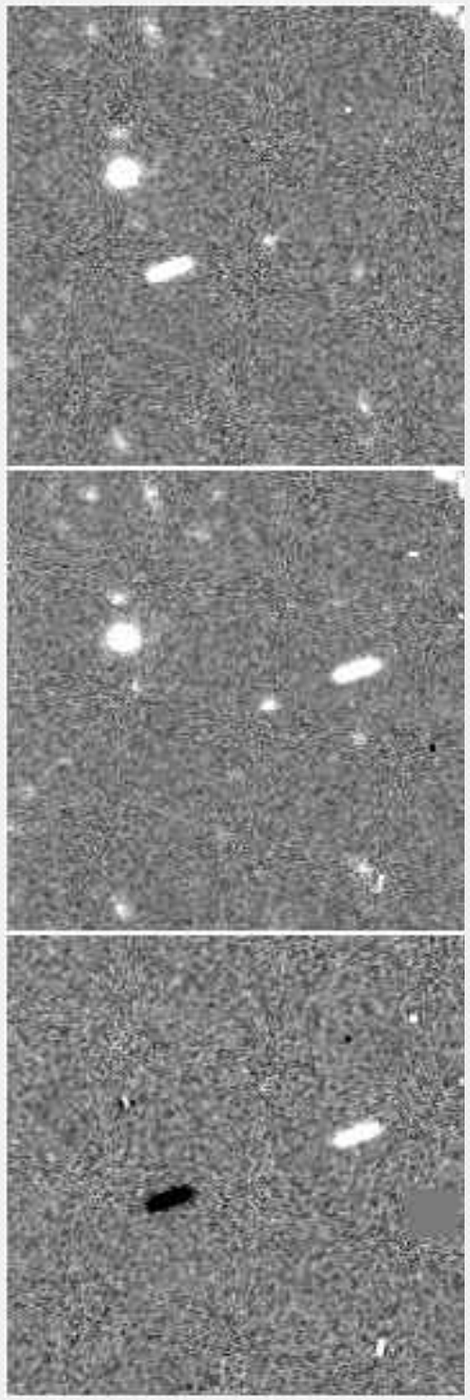
Frontiers of survey astronomy

- Time domain science:
  - Novae, black-hole binaries, GRBs …
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System:
  - NEOs, PHAs, moving objects
  - Solar system & planet formation
- Mapping the Milky Way:
  - Structure and accretion history
  - Properties of all stars within 300 pc
- Dark energy and dark matter:
  - Strong Lensing
  - Weak Lensing
  - Supernovae



Frontiers of survey astronomy

- Time domain science:
  - Novae, black-hole binaries, GRBs …
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System:
  - NEOs, PHAs, moving objects
  - Solar system & planet formation
- Mapping the Milky Way:
  - Structure and accretion history
  - Properties of all stars within 300 pc
- Dark energy and dark matter:
  - Strong Lensing
  - Weak Lensing
  - Supernovae



Exposure 1

Exposure 2

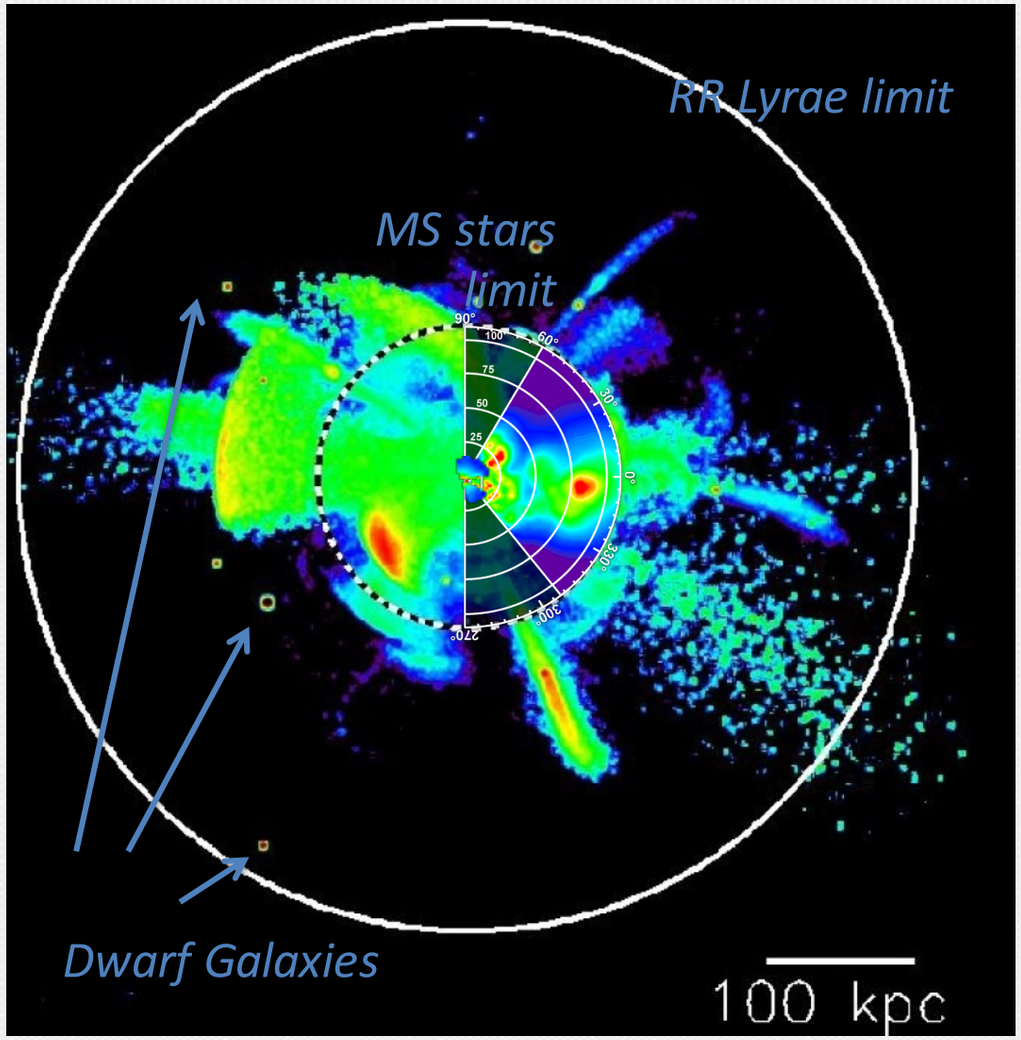
Exposure 1

-

Exposure 2

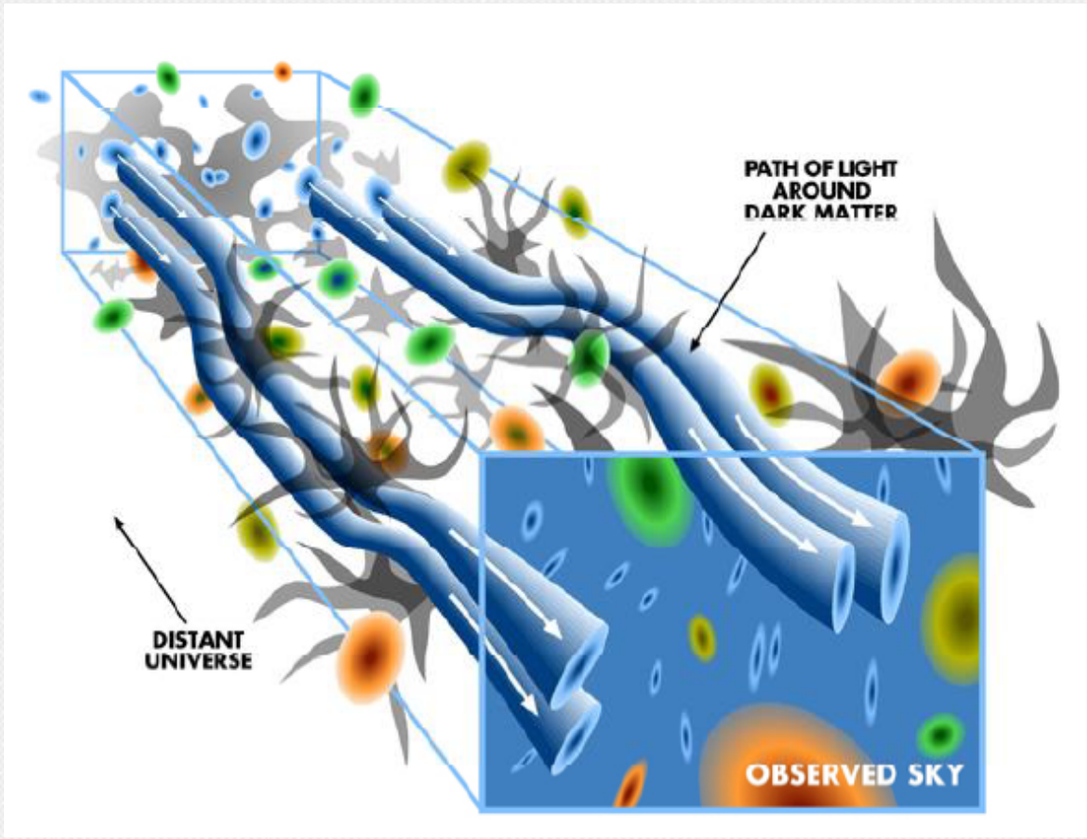
Frontiers of survey astronomy

- Time domain science:
  - Novae, black-hole binaries, GRBs …
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System:
  - NEOs, PHAs, moving objects
  - Solar system & planet formation
- Mapping the Milky Way:
  - Structure and accretion history
  - Properties of all stars within 300 pc
- Dark energy and dark matter:
  - Strong Lensing
  - Weak Lensing
  - Supernovae



Frontiers of survey astronomy

- Time domain science:
  - Novae, black-hole binaries, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System:
  - NEOs, PHAs, moving objects
  - Solar system & planet formation
- Mapping the Milky Way:
  - Structure and accretion history
  - Properties of all stars within 300 pc
- Dark energy and dark matter:
  - Strong Lensing
  - Weak Lensing
  - Supernovae



Data products

Level 1

- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

Level 2

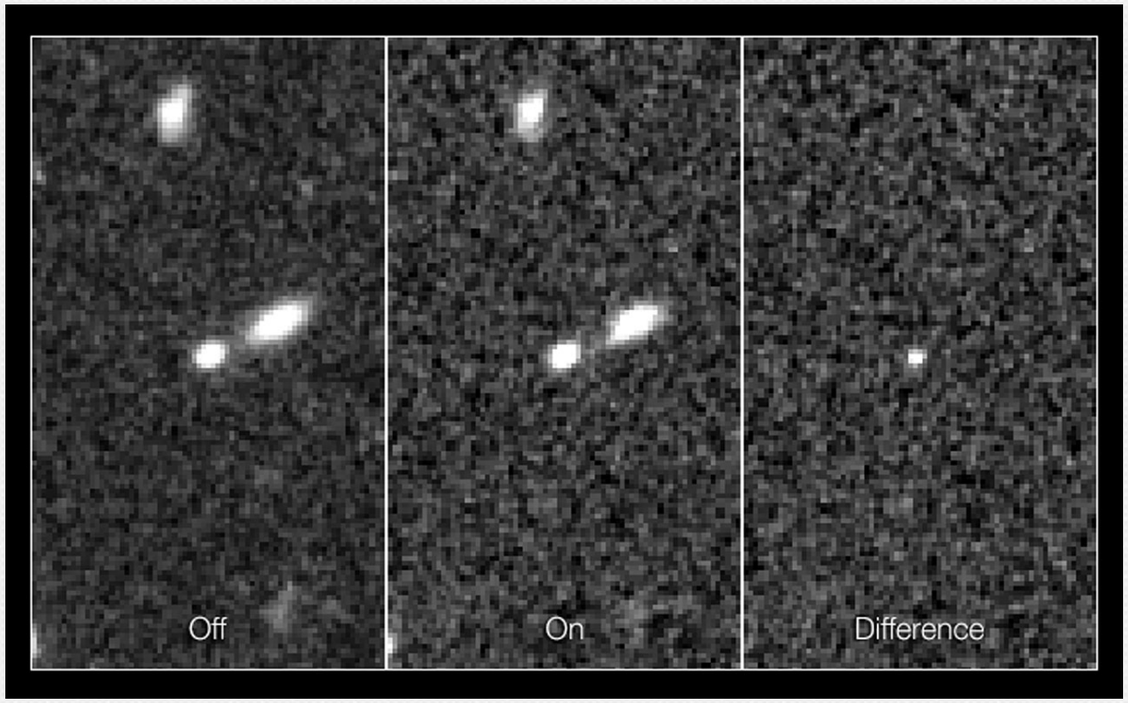
- A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations (“sources”), and ~30 trillion measurements (“forced sources”), produced annually, accessible through online databases.
- Deep co-added images.

Level 3

- Services and computing resources at the Data Access Centres to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Level 1: Alerts

- State-of-the-art image differencing pipeline
- Alerts issued within 60 seconds of observation
- 10M/night (average), 10k/visit (average), 40k/visit (peak)



- Each alert includes:
  - Position
  - Flux, size, and shape
  - Light curves in all bands (up to a ~year; stretch: all)
  - Variability characterization (eg., low-order light-curve moments, probability the object is variable)
  - Cut-outs centred on the object (template, difference image)

Alert distribution

- VOEvent
  - “Or some other format that is broadly accepted and used by the community at the start of LSST commissioning”
- LSST limited broker
  - Rate limited distribution directly to end users
  - Minimal filtering capability, based only on the contents of individual event packets; no classification
  - Early experiments now underway
    - e.g. https://github.com/SimonKrughoff/CometDemo
- “Fire hose” streams to selected public brokers
  - Likely operated under an MOU with LSST
  - Providing advanced filtering and event annotation services

Level 2: Annual Releases

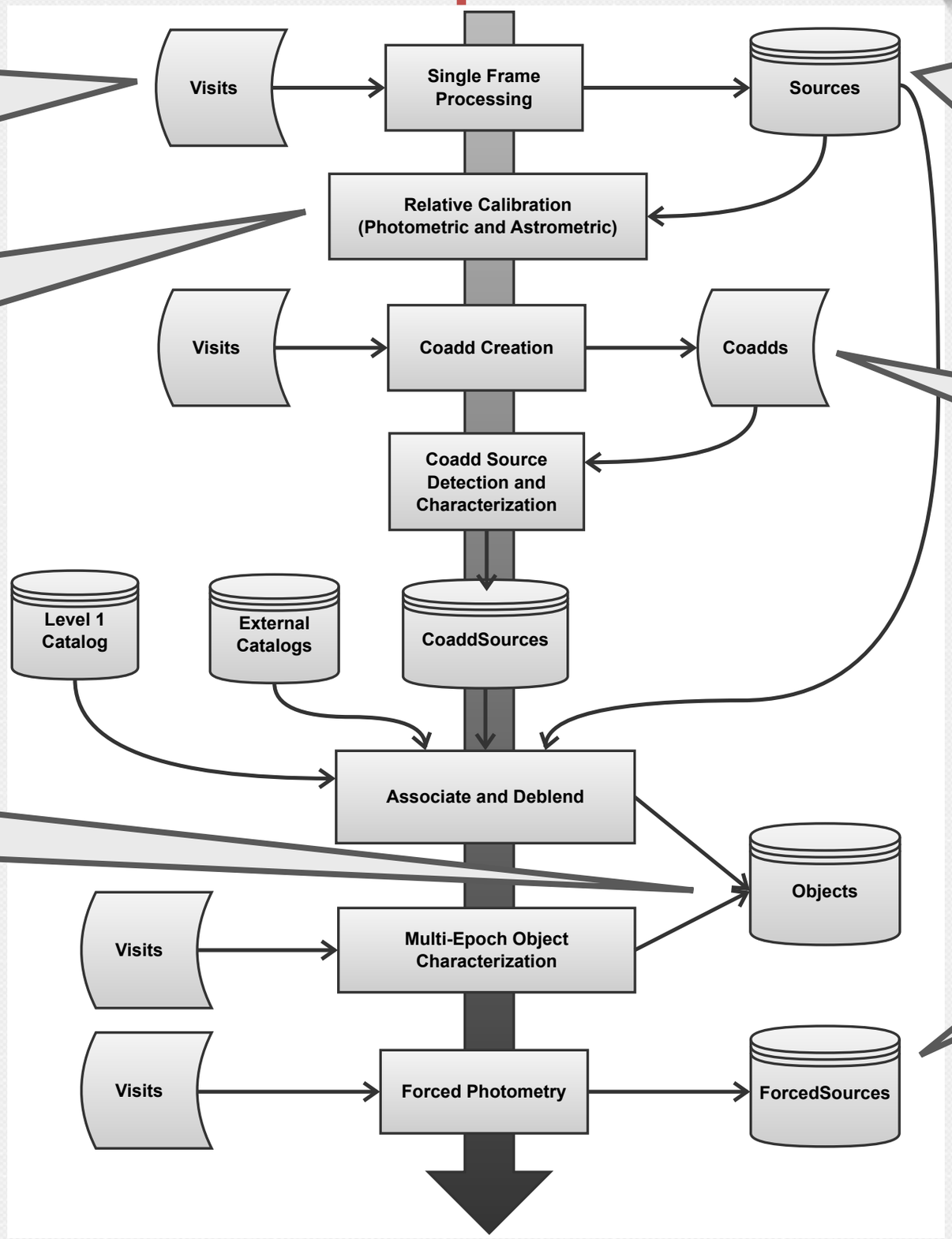
- Calibrated and consistent catalogues & images
  - Objects, detections, detections in difference images, coadds, etc
  - Enable static sky science and time-domain science which is not time sensitive (e.g. statistical investigations of variability)
- Made available in annual Data Releases
  - Two releases in the first year
- Complete reprocessing for each release
  - Every DR will reprocess all data taken up to the beginning of that DR
  - Including reprocessing of level 1 data
- Projected catalog sizes:
  - 18 billion (DR1) ➜ 37 billion (DR11) separate objects
  - 750 billion (DR1) ➜ 30 trillion (DR11) individual measurements
- Cumulative ~500 PB image and ~50 PB catalogue data

Level 2: Process & products

Single visit images
Raw exposures, processed frames

Calibration products
Darks, flats, biases, fringe, etc

Object catalogue
Detailed, multi-epoch characterization; model fits, colours, centroids, fluxes, surface brightness, etc. Include sampling of likelihood function for science cases requiring full posteriors



Source catalogue
Independent measurements on each exposure; high-SNR variables & moving objects

Coadded images
Deep, short-period, best seeing, PSF-matched; those not stored generated on request

Forced photometry
Measure flux of every source on every visit; characterize the light curve for every object in the survey

Level 3: User created

- Products created by the community and made available through an LSST Data Access Centre
- Use-cases not fully enabled by Level 1 and 2:
  - Reprocessing images to search for SNe light echos
  - Characterization of diffuse structures (e.g., ISM)
  - Extremely crowded field photometry (e.g., globular clusters)
  - Custom measurement algorithms
- Enabling Level 3:
  - User databases and workspaces (“mydb”)
  - Enabling user computing at the LSST data centre
    - For processing that will greatly benefit from co-location with the LSST data
    - Sized for ~10% of total compute budget
  - Making the LSST software stack available to end-users

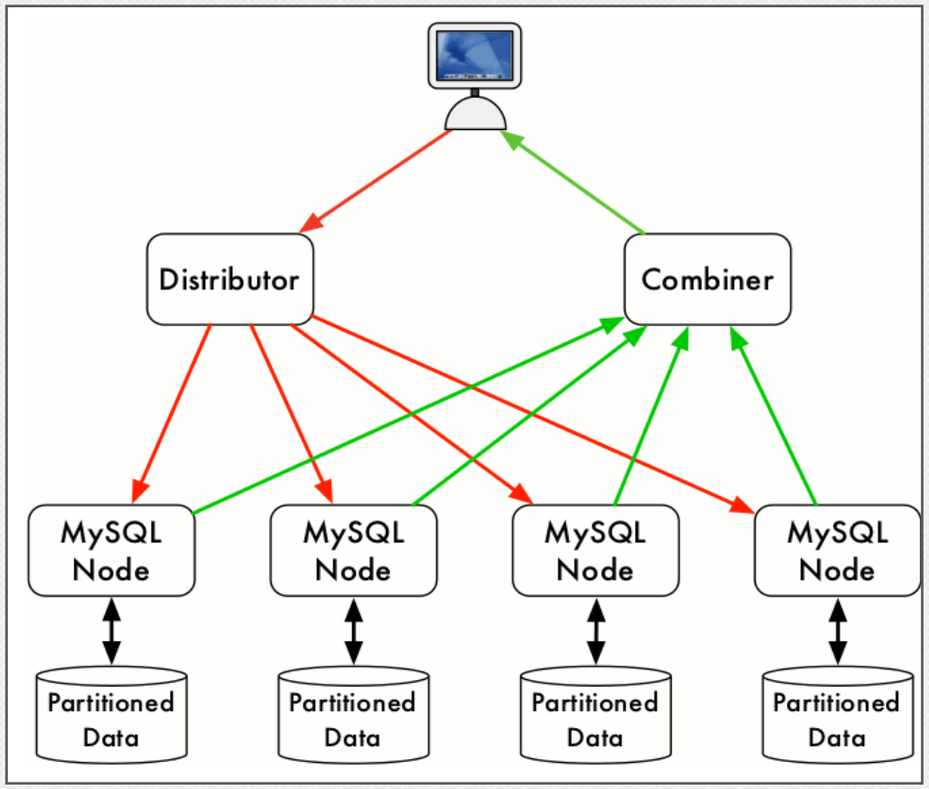
All-new software stack

- Difficulty adapting existing public codes to LSST requirements (AstroMatic suite, PHOTO, Elixir, IRAF-based pipelines, etc):
  - Run efficiently at scale
  - Flexible (plugging/unplugging of algorithms at runtime)
  - Developed by a large team (20+ scientists and programmers)
  - Maintainable over ~25 years of R&D, Construction, and Survey Operations
  - Support a variety of hardware and software platforms
  - Logging and provenance built into the design
- Early on (~2006), a decision was made to transfer scientific know-how, but not code.

Design & language choices

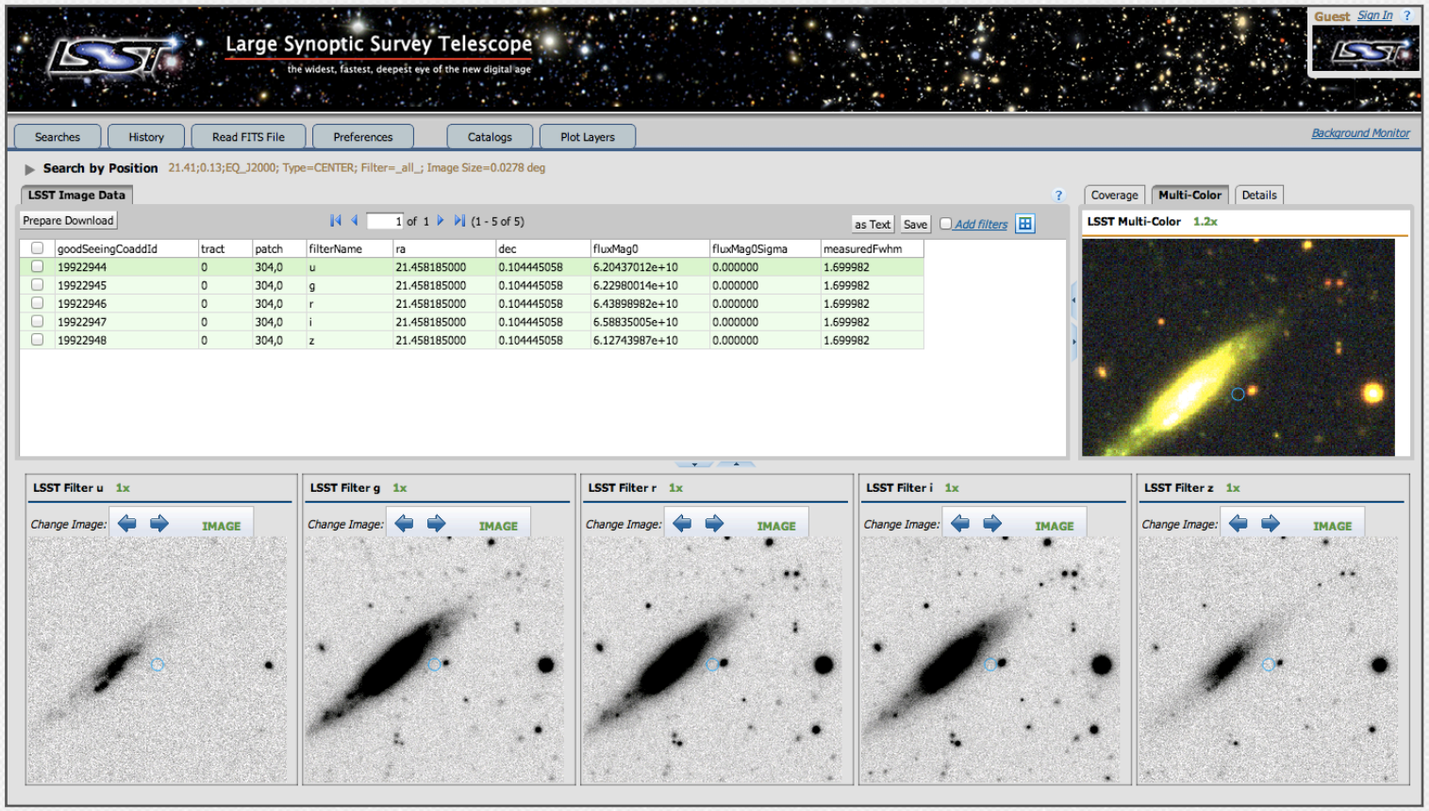
- Python
  - 2.7 currently; 3+ later
  - All high-level code
  - Whenever performance demands allow
- C++
  - Limited subset of C++11 currently; more as compilers permit
  - Computationally intensive code
  - Exposed to Python through SWIG
- Modular
  - (Virtually) everything is a Python module
  - ~60 separate packages
- Open source, transparent development
  - GPL v3+
  - http://github.com/lsst • http://dm.lsst.org • ⚠ Under construction! ⚠

Database & Science UI



- Massively parallel, distributed, fault tolerant, spatially sharded, relational database
- Built on existing, well understood technologies (MySQL, xrootd)
- Commodity hardware, open source
- Advanced prototype available (qserv)

- Science User Interface provides access to and analysis of LSST data
- Web and machine interfaces to database
- Visualization capabilities
- User workspace



Data distribution & the VO

- Actively assessing distribution methods:
  - Access protocols
    - Internal: native RDBMS protocol, HTTP(S)
    - Public: SCS, TAP, SIAPv2, VOSpace, DataLink, OAuth/OpenID, …
  - Catalogs
    - Internal: native RDBMS storage
    - Query: native SQL, ADQL
    - Bulk: compressed FITS binary tables (or HDF5)
  - Images:
    - Internal: internal format (may be FITS)
    - Public: compressed FITS (MEF)
  - Time Domain Events
    - Internal: native RBMS storage
    - External: VOEvent + VOEvent Transport Protocol (*evolved*)

Libraries and toolkits in languages we can use are vital to drive adoption

Adopt community standards where practical; drive development of standards where possible; build from scratch only when unavoidable.

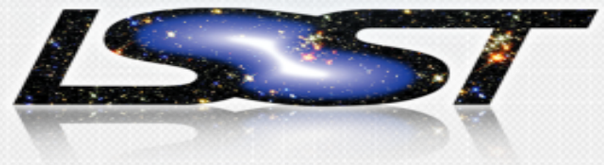
Conclusion

LSST will:

- Commence survey operations in ~7 years
- Produce an unprecedented volume of transient alerts
  - Published to the worldwide community with low latency
- Generate annual data releases providing trillions of source measurements and petabytes of image data
  - Available to the US, Chile and international partners with no proprietary period
- Use and develop community standards for making data available wherever possible
  - Virtual Observatory standards are expected to play a major role

Extra slides

Example: VOEvent evolution



- Existing VOEvent format imposes significant overheads
- Example:

  From http://wiki.ivoa.net/internal/IVOA/IvoaVOEvent/example1-v1.0.xml

  - Information content:

    ~40 bytes

  - Data on the wire:

    787 bytes

```
<VOEvent xmlns="http://www.ivoa.net/xml/VOEvent/v1.0"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 id="ivo://raptor.lanl/VOEvent#23564" role="observation" version="1.0"
 xsi:schemaLocation="http://www.ivoa.net/xml/VOEvent/v1.0
 http://www.ivoa.net/internal/IVOA/IvoaVOEvent/VOEvent-v1.0.xsd">
  <Who>
    <PublisherID>ivo://raptor.lanl/organization</PublisherID>
    <Date>2005-04-15T14:34:16</Date>
  </Who>
  <What>
    <Param name="RA" ucd="pos.eq.ra" unit="deg" value="185.0"/>
    <Param name="Dec" ucd="pos.eq.dec" unit="deg" value="13.2"/>
    <Param name="magnitude" ucd="phot.mag;em.opt.R" unit="mag"
    value="18.2"/>
  </What>
  <Why>
    <Concept>Fast Orphan Optical Transient</Concept>
  </Why>
</VOEvent>
```

Need compact serialization,
provisions for bulk delivery.