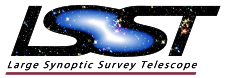
Federated Identity Needs for the Large Synoptic Survey Telescope (LSST)

Jim Basney
jbasney@ncsa.illinois.edu
12th FIM4R Workshop (Feb 5 2018)



This material is based upon work supported by the National Science Foundation under grant numbers 1258333 and 1547268. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government or any agency thereof.

Large Synoptic Survey Telescope

- Wide-field reflecting telescope on a peak of Cerro Pachón, a mountain outside La Serena, Chile
- Beginning in 2022: Imaging the visible sky once every 3 days, for 10 years
- Over 15TB of data per night collected
  - Initial requirements: 100 tflops of computing, 15PB of storage
- Ultimate deliverable of LSST is the fully reduced data
  - All science will come from survey catalogs and images

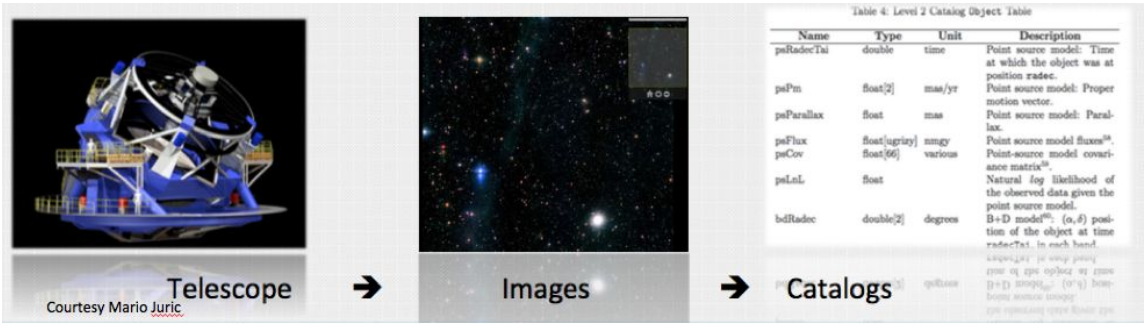


Table 4: Level 2 Catalog Object Table

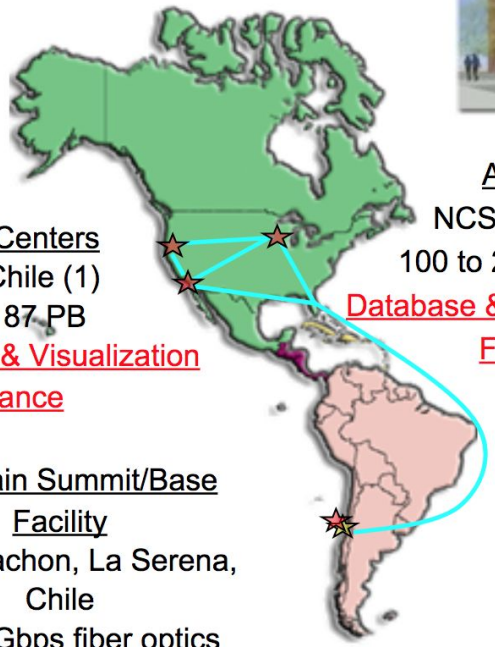
Name	Type	Unit	Description
psRadecTai	double	time	Point source model: Time at which the object was at position **radec**.
psPm	float[2]	mas/yr	Point source model: Proper motion vector.
psParallax	float	mas	Point source model: Parallax.
psFlux	float[ugrizy]	nmgy	Point source model fluxes[58].
psCov	float[66]	various	Point-source model covariance matrix[58].
psLnL	float		Natural *log* likelihood of the observed data given the point source model.
bdRadec	double[2]	degrees	B+D model[60]: (α, δ) position of the object at time radecTai, in each band.

Large Synoptic Survey Telescope

- Scientific goals:
  - Probe the nature of dark matter and dark energy
  - Cataloging the Solar System, particularly near-Earth asteroids and Kuiper belt objects
  - Observing transient optical events
  - Mapping the Milky Way: exploring structure and formation
- More information: www.lsst.org

LSST
Large Synoptic Survey Telescope

Archive Center
NCSA, Champaign, IL
100 to 250 TFLOPS, 75 PB
Database & Pipeline Parallelization
Fault Tolerance



Data Access Centers
U.S. (2) and Chile (1)
45 TFLOPS, 87 PB
Data Access, Mining & Visualization
Fault Tolerance

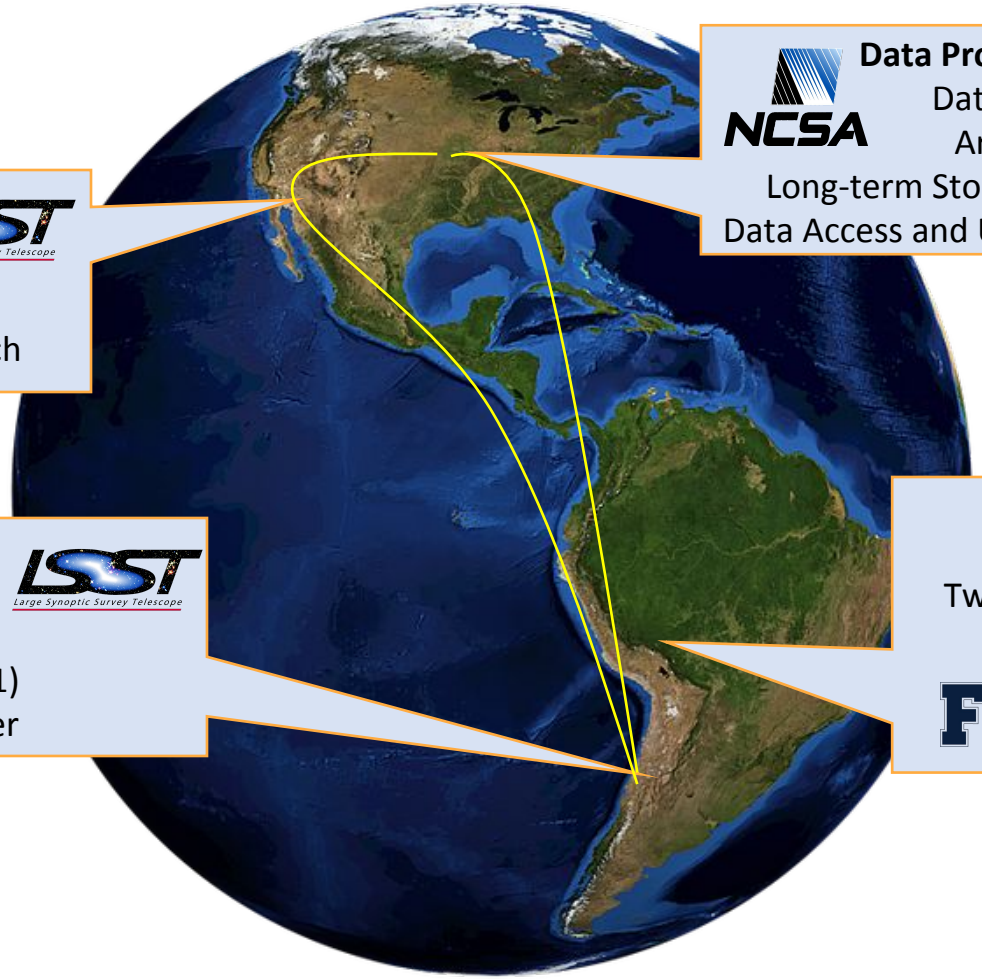
Image Courtesy of K. Gilmore SLAC

Mountain Summit/Base Facility
Cerro Pachon, La Serena, Chile
10x10 Gbps fiber optics
25 TFLOPS, 150 TB
Transient Alerts, Pipeline Parallelization
Fault Tolerance

Long-Haul Communications
Chile - U.S. & w/in U.S.
2.5 Gbps avg, 10 Gbps peak
High-speed transfer
Fault Tolerance

1 TFLOPS = 10^12 floating point operations/second
1 PB = 2^50 bytes or ~10^15 bytes

LSST
Large Synoptic Survey Telescope



HQ Site
Science Operations
Observatory Management
Education and Public Outreach

Summit and Base Sites
Telescope and Camera
Data Acquisition
Long-term Storage (copy 1)
Chilean Data Access Center

Data Production Site
Data Production
Archive Center
Long-term Storage (copy 2)
Data Access and User Services

Dedicated Long Haul Networks
Two redundant 100 Gb links from La Serena to NCSA

LSST Data

- Recall that LSST data is the deliverable…
- Data classification and access:
  - Lowest classification is data released to the public, this is where most data ends up
  - Higher levels are made available authorized users
- LSST's Information Classification Policy outlines the information categories and gives examples.
- Sites that provide access to LSST data (i.e. NCSA) need to follow LSST's security policy w.r.t. to that data.
  - *Identity management plays a very important role here.*

LSST From the User's Perspective

Raw data. Support timely detection and follow-up of time-domain events (variable and transient sources). Not produced for release.

Level 1

Products are generated as part of a Data Release, generally performed yearly, with an additional data release for the first 6 months of survey data. **These data releases are available for data rights holders only and made public after 2 years.**

Level 2

Derived data. Maybe restricted to subgroups within LSST.

Level 3

LSST Large Synoptic Survey Telescope

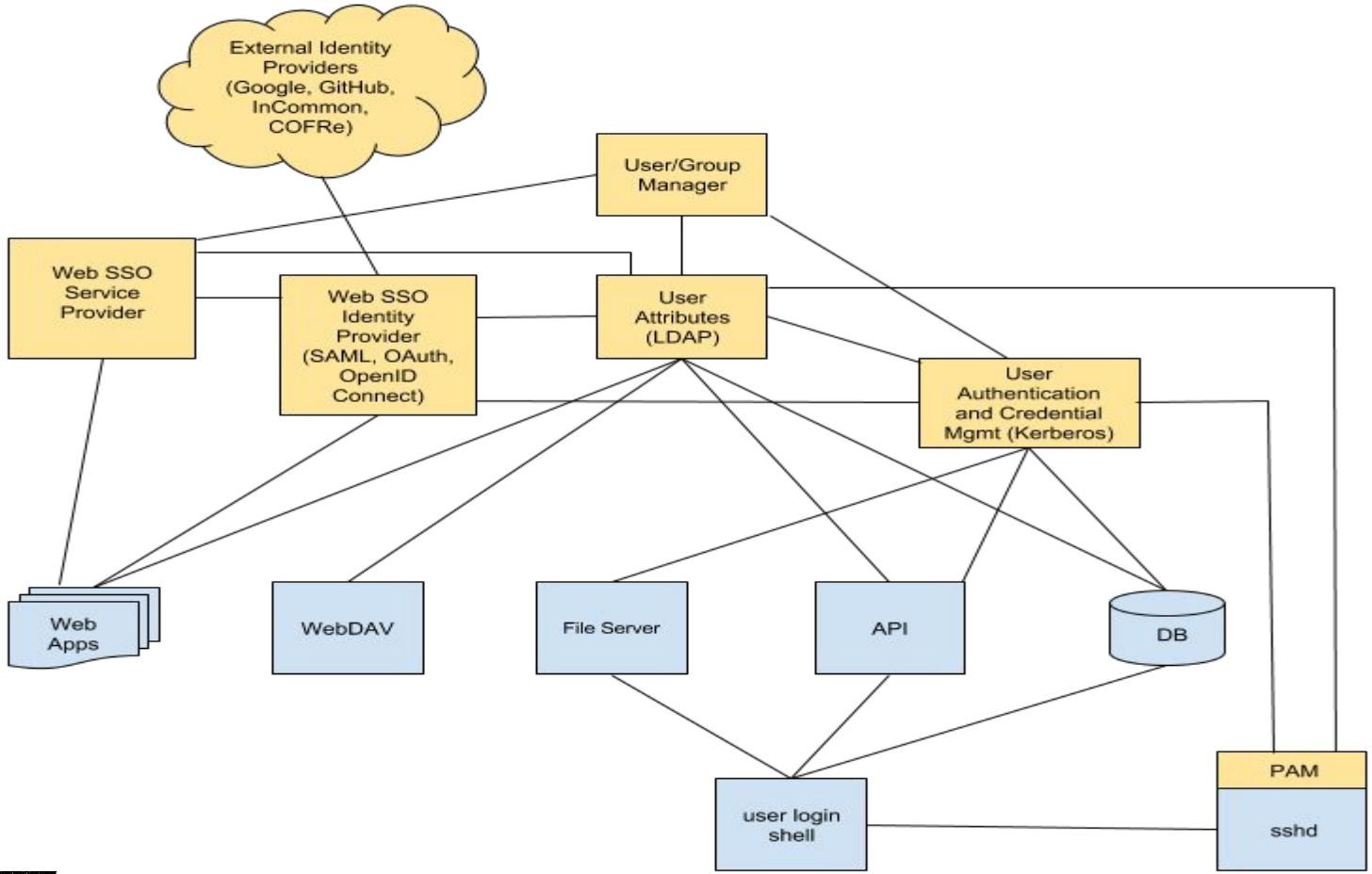
Identity and Access Management

- AuthN and AuthZ to LSST data and services
- Scalable, secure management to data access rights

FEDERATED IDENTITIES TO THE RESCUE!

Identity Linking

- External identities (University, GitHub, etc.) linked to individual's LSST identity
  - Established during initial enrollment and managed by user
- Group memberships based on LSST identity
  - LDAP queries using LSST IDs and external IDs



Large Synoptic Survey Telescope

Authorization

- L2 data rights
- L3 collaboration groups
- Access to applications/services
- Admin/staff roles

Access Control Components

- User/Group Manager
  - Implements the logic and workflows to determine who has L2 Data Access Rights and who is involved in L3 collaborations. These workflows set/unset User Attributes (i.e., group memberships).

Access Control Components

- User Attribute Store
  - Receives information from User/Group Manager and publishes the resulting User Attributes via a standard LDAP interface.

Access Control Components

- Service Level Authorization
  - Services implement authorization (access control) based on access control lists (ACLs) or database GRANT statements or other service-specific methods, based on the User Attributes.

L2 Data Rights (Proposed)

- National professional astronomical community
  - Use eduPersonAffiliation when available
    - No "astronomy department" affiliation
    - "Member" is close enough? inacademia.org?
  - Use American Astronomical Society membership directory?
    - https://aas.org/individual-membership/classes-membership
    - https://aas.org/posts/news/2016/08/do-you-have-orcid-id
  - Otherwise will require manual review/approval

L2 Data Rights (Proposed)

- Named individuals from international partners
  - Lookup existing LSST accounts
  - Email-based invitations
- A limited number of designated additional individuals (post-docs, grad students) per named individual
  - Named individuals can invite/grant others (from same institution)
- Periodic re-validation / review

Managing an L3 Group (Proposed)

- Via ORACLE
  - ORACLE (Observatory Resource Allocation Committee for Level Elevation) process defines a group indicating the users (group members) who can use the resource allocation. Also create an associated L3 data workspace private to that group.
- Via User/Group Manager
  - […]

Managing an L3 Group (Proposed)

- Via User/Group Manager
  - Any user with Data Access Rights can click "Create L3 Data Product Group" in the User/Group Manager web interface to create an L3 group and define its initial members. That user will be the initial owner of the group.
  - Users who own L3 groups will also see a "Manage My L3 Data Product Group(s)" button/link that allows them to add/remove members and add owners / transfer ownership.
  - Users with Data Access Rights will see a "Manage My L3 Data Product Group Memberships" button/link that allows them to request to join L3 groups or leave L3 groups they are currently a member of.

LSST FIM: Assurance and Incident Response

- Questions about eduGAIN identity assurance and federated incident response during design review
- https://refeds.org/assurance/profile/cappuccino & https://refeds.org/sirtfi likely satisfy LSST's FIM requirements (if they are supported/adopted)

LSST FIM Needs Summarized

- More IdPs in eduGAIN (especially in US and Chile)
- Persistent NameID for authentication
- Affiliation attribute for authorization
- Assurance and Incident Response

Large Synoptic Survey Telescope

Thanks!

- Contacts:
  - jbasney@ncsa.illinois.edu
  - https://confluence.lsstcorp.org/display/LAAIM

- Acknowledgements
  - Alex Withers (LSST Information Security Officer)
  - LSST Authentication, Authorization, and Identity Management (LAAIM) Group