LINCC Frameworks

Rachel Mandelbaum (Carnegie Mellon University)
June 9, 2022

Outline

1. What is LINCC Frameworks?
2. Recent and ongoing LINCC Frameworks activities
3. How can SSSC connect to and benefit from LINCC Frameworks?

What is LINCC Frameworks?

LSST Interdisciplinary Network For Collaboration And Computing (LINCC)

LINCC is an LSSTC initiative to prepare for and support science with LSST in collaboration with University of Arizona, Carnegie Mellon, Northwestern University and the University of Washington

Programs LINCC supports include:

- **LINCC Frameworks.** A program to develop software infrastructure to support LSST science at scale
- **The LSSTC Catalyst Fellowship Program.** A prize postdoctoral fellowship program to support LSST science and analysis infrastructure
- **LSSTC Data Science Fellowship Program.** A graduate training program to prepare researchers with the skills to work with LSST data
- **More to come....**

The LINCC Frameworks Project

LSST Interdisciplinary Network For Collaboration And Computing

A collaboration between the University of Washington, Carnegie Mellon, LSSTC, the University of Pittsburgh, and NOIRLab to build software systems for key LSST science, especially infrastructure that supports multiple science use cases.

PIs: RM (CMU) and Andy Connolly (UW)
And a new team of ~14 Software developers, data scientists, cloud engineers, front-end developers, devops engineers, full stack scientists

We need **science** software infrastructure: combining user algorithms and code, astronomy packages, and industry tools.

This is what LINCC wants to enable

LSST Science Pipelines

Science Platform Research

Inference

Algorithms

Overall LINCC Frameworks Objectives

Helping the LSST/Rubin Obs User Community Prepare for LSST science

- Sharing end-to-end analyses to make them reproducible, scalable, personalizable - starting with existing data sets
- Encouraging engagement through dedicated cyberinfrastructure group, funded **workshops**, **incubator** projects, visiting fellowships, **hackweeks**
- Enabling access by a larger community, in turn generating more discoveries

Our goal is to build and support an inclusive community so that we can have a democratizing effect through community software.

https://www.lsstcorporation.org/lincc/frameworks

Recent and ongoing LINCC Frameworks activities

LINCC Frameworks Hiring

- **Software Engineering Team**
  - Jeremy Kubica (CMU): Software Manager
  - Hiring 8 software engineers at UW and CMU
- **Project Scientists** (starting in the next few months)
  - Neven Caplar (UW, 50%)
  - Colin Chandler (UW)
  - Alex Malz (CMU)
  - Sam Wyatt (UW, 50%)
  - 2 more to come from the University of Pittsburgh

Workshop: From Data to Software to Science with the Rubin Observatory LSST

Workshop goals:

1. Enabling *interactive development* of exciting scientific use cases for early LSST data, and identifying the common computational/technical challenges and enabling technologies associated with them.
2. Promoting the development of a broad and inclusive community of res engaged with LINCC Frameworks.

Program design, plenary talk content, and communication channels for the meeting were developed with both goals in mind.

https://indico.flatironinstitute.org/event/2777/

Science use cases

Divided the science into 7 research areas (not tied to the Science Collaborations)

- Solar System Science: 6 cases (active asteroids, TNOs)
- Local Universe Static Science: 5 cases (IMF, accreted stellar pops, dwarf gals)
- Local Universe Variable and Transient Science: 9 cases (YSO, microlensing)
- Extragalactic Static Science: 7 cases (morphologies, extinction, LSB dwarfs)
- Extragalactic Variable Science: 8 cases (AGN, lensing)
- Extragalactic Transient Science: 7 cases (SNe, TDEs, classification)
- Cosmology: 6 cases (weak lensing, SNe classification, spectroscopic followup)

~50 use cases for science in the first 2 years of Rubin

	Cross-matching	Photo-z	Selection functions	Reprocessing images	Time series	Image analysis or classification
Cosmology	■				■	
Solar System		■				
Extragal transient			■			
Extragal variable						
Extragal static			■			
Local static		■				
Local trans+var		■		■		■

A matrix of connections between science and technical software development. This reflects the subset of use cases discussed at the workshop, not all possible early LSST science cases.

What did we get out of the experience?

Scientists in the LSST community fleshed out their high-priority and high-urgency LSST science cases, technical challenges, and enabling technology – and made connections with and learn from others working in different scientific areas.

Members of the Rubin Observatory team, NOIRLab, and teams developing Independent Data Access Centers (IDACs) learned in detail about intended community software and data usage patterns.

Members of the LINCC team identified opportunities to develop cross-cutting technology that broadly enables the LSST science community.

Some takeaways from this exercise

There are clear technical areas where developing software and infrastructure could accelerate LSST science

Many science programs share common challenges (e.g. working with time series)

As LINCC Frameworks is starting out we should work to make the project inclusive (one breakout session and plenary was devoted to how to create inclusive collaborations)

There is a lot of existing software that we can build from and we don’t need to start from scratch.

White paper

We’re using the workshop material to develop a white paper (open to the LSST science community) with the following:

- ~50 science cases from across the LSST science community
- 6 computational challenges that cross multiple science cases & require new software
- Inclusive collaboration strategies for software development, scientific collaborations

Cross-cutting content in this white paper should form the basis for future collaborative efforts across the community. Great interactions with Rubin Observatory team and Astropy team members!

So what's next?

- We keep hiring and finish the white paper, but also...
- Summer 2022: initiate starter projects, with a goal of establishing the team and initial collaborations while delivering impactful software
- Summer-fall 2022: work on longer term development plan, in collaboration with groups across the LSST scientific community

How can you get involved?

1. Join #lincc-workshop-sw-data-science to get involved in and provide feedback on the white paper.
2. Attend a planned Rubin Project & Community Workshop session on collaborative software development

How can SSSC members connect to & benefit from LINCC Frameworks?

SSSC - LINCC Frameworks connections

Just a few ways SSSC members can benefit from LINCC Frameworks:

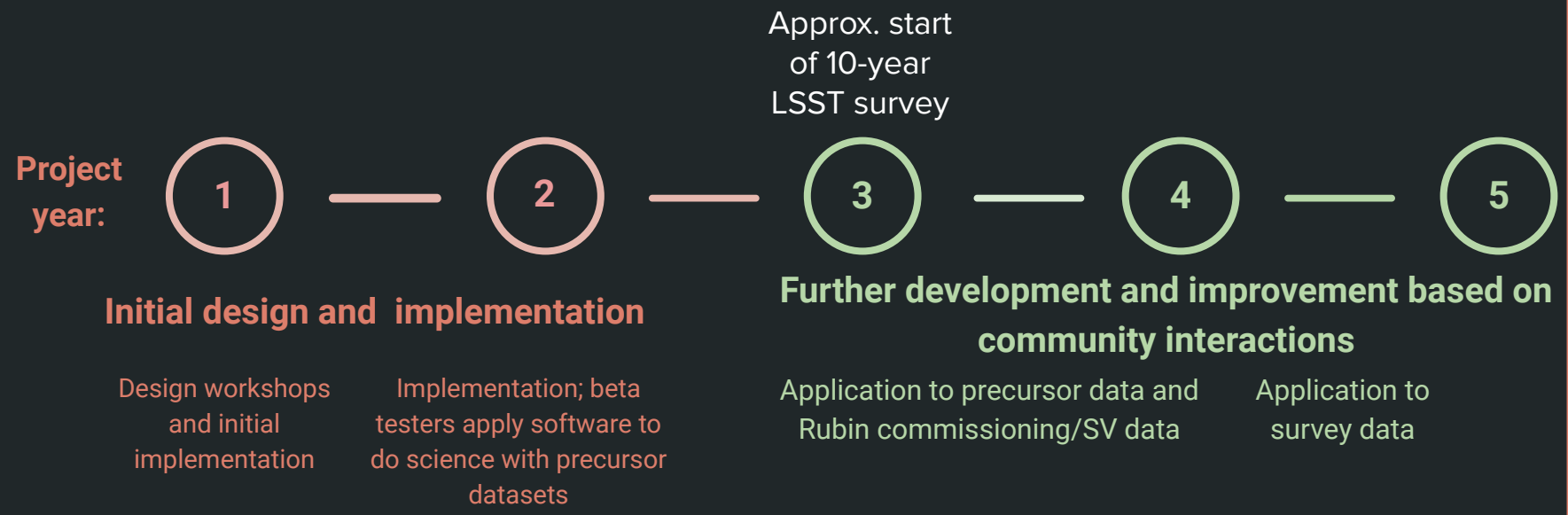
1. Short term (now!): computing resources – LINCC Hub, including solar system simulation data
2. Longer term: provide a clear definition of software needs requiring significant software engineering effort.
    a. Examples: getting tools for your science to work at scale (THOR, KBMOD, ...), identifying how SSSC would want to interact with more general-purpose tools, e.g., light curve analysis
    b. This is useful input that we can fold into the LINCC Frameworks development plan
3. Longer term: engage with incubator program and hack weeks (which come with funding and computing resources for participants) to do science with prototype software and drive improvements to get to LSST scale/precision

Conclusions

- The LINCC Frameworks team is coming together.
  - Several areas for potential collaboration on software development with SSSC
  - Workshop has seeded ideas for impactful early projects, and we are happy to discuss collaboration opportunities – reach out to the project PIs
  - Once the team is on board, we will solidify development plans
- “From Data to Software to Science with the Rubin Observatory LSST” was an opportunity to discuss software needs across the LSST science community
  - Revealed many collaboration opportunities
  - LINCC Frameworks team hopes to enable some of them, in collaboration with Rubin Observatory, the astropy team, LSST SCs, and others
- You’re all welcome to engage with white paper development by joining #lincc-workshop-data-sw-science on Slack, and a Rubin PCW session

Backup slides

5-year project timeline



After that, the landscape will have evolved substantially and there may be room for expansion into other scientific or technical capabilities, given additional funding.

LINCC Frameworks: From Data to Software to Science

LSST Science Pipelines

Real-time Streaming Technologies

Science Platform Research

Inference

Community Contributions and Support

Algorithms

We need **science** software infrastructure: combining user algorithms and code, astronomy packages, and industry tools. This is what LINCC wants to enable

The LINCC Frameworks Project

LSST Interdisciplinary Network For Collaboration And Computing

A collaboration between the University of Washington, Carnegie Mellon, LSSTC, the University of Pittsburgh, and NOIRLab to build the software systems for key LSST science.

A team of ~14 software engineers and research scientists building software for:

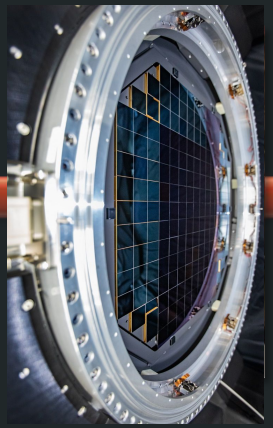
- Data infrastructure
- Solar System exploration
- Time domain science
- Extragalactic astronomy

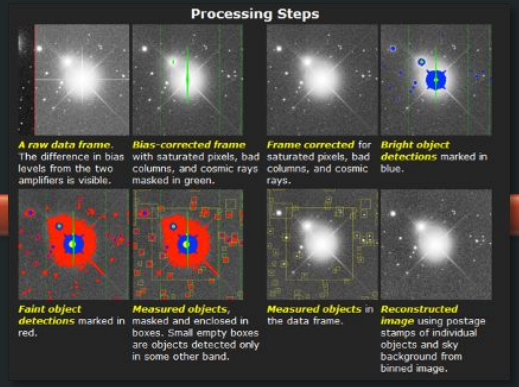
Software developers, data scientists, cloud engineers, front-end developers, devops engineers, full stack scientists

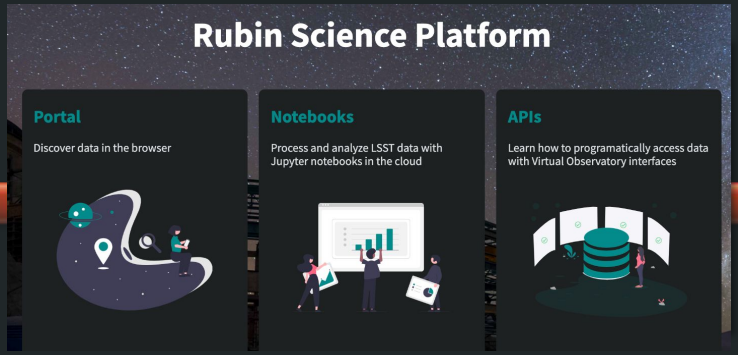
The Legacy Survey of Space and Time

Repeated imaging of the visible sky to ~24th mag.
10 years of operation.
60 PB of raw data.
40 billion stars, galaxies, asteroids.
30 trillion observations.

Automated Data Processing Pipelines







A database of object positions and properties available online for the astronomers to query.