



LSSTへの貢献と データアクセス権

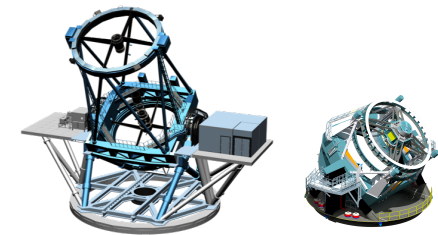
宮崎聡

2022-09-21

光天連シンポジウム

LSSTのデータアクセス

- ・ LSST: 可視広視野専用望遠鏡 (チリ)
- ・ HSCの20倍の探査速度
- ・ 2024年に運用開始予定。専用望遠鏡のため、2年以下でHSC-Wide(1200平方度)の深さのデータが、20,000平方度に達する。10年程度で、それらが全てHSC-Deepの深さに達する
- ・ データ保護期間2年 (ただし、その後どのような形で公開されるかは不明)
- ・ 米国外から優先データアクセス権を得る方法
 - ・ 1 PI (4人のAssociate(学生・ポスドク)がつけられる)あたり、約US\$300K相当のIn-Kind-Contributionをすることにより、13年間に渡るオペレーション期間中の優先データアクセス権を得る
- ・ 日本からのアクセス
 - ・ 光赤外コミュニティーからの意見を得つつ、代表であるすばる科学諮問委員会(SAC)において、2019-20に貢献方法が議論された。それを受け、国立天文台とKIPMUが協力して貢献案を取りまとめ、LSST側と協議中。平行して国立天文台における貢献体制(予算・人員)を台内で協議中



	HSC	LSST
Survey Period	6 yr	10 yr
Nights	300	~3000
Rel. Inst. Survey Speed	1	3.4
Rel. Yearly Pace	1	20 !!!

$$3.4 \times \frac{3000}{10} / \frac{300}{6}$$

日本からのIn Kind Proposal (協議中)

			PI
1	Telescope Time on Subaru	13年間で合計50晩	20
2	Help with Commissioning	2.5名を2年間のコミッショニングに派遣	5?
5	Serving Rubin Science Platform (Development and Support)	RSPの開発とユーザーサポートに貢献	8?
7	Serving LSST Catalog from Lite IDAC	日本人がデータアクセスするときにも必要。データセンターと協力が必須。	5 (NAOJ) ? 2 (IPMU)
N	Dark Energy Science Tools	IPMU/Nagoya/U-Tokyo	5
N	PFSによるPhoto-Z calibration Data	PFS-SSPから抽出、PFS Open UseのFiller	3
N	PFSによるTransient Follow-up	Open Use Programを提出予定	4?
9	Contribution Team Management	Contribution Packageの取りまとめもカウントされることになった	1?
			30 + 23?

(XRISM Timeや日本の中小口径望遠鏡時間も提案したが、貢献のRecipientが米側で特定できなかった)

LSSTデータアクセス権を得る意義

- ・ LSSTはHyper Suprime-Cam(HSC)の大型版にすぎず、HSCで研究成果を挙げつつある日本の研究者は、LSST観測開始初期から、それぞれの分野で研究をリードできる。
- ・ LSSTのデータ解析環境の導入により、HSCを用いた研究環境も現代化され、生産性の向上が期待できる。

1粒で2度おいしい

公募するPI

- 日本の天文コミュニティの共有財産であるすばる望遠鏡の観測時間、天文台の計算機資源などを使った貢献で得られるPIは公募で決める。
 - 通常のTACプロセスを経て獲得するすばる時間で得られるデータを LSSTに提供するin-kind contributionもあるが、それらによる PI は除く。
- In-kind contributionに実働する人には PI の権利を与える。
- 共同利用機関ではない IPMU の貢献(計算機資源、人的貢献など)によるPIは、IPMUが選択する。
 - IPMUでの必要数を超えるPIが得られた場合は、コミュニティに還元する。
 - IPMUの必要なPI数が得られれば、IPMUから公募枠のPIに応募することはない。
- 公募するPI数は最大30人前後(すばる時間20人＋計算機関連など10人)が見込まれる。

- 基本は LSST のデータを使って行う研究を提案してもらう。
- 日本コミュニティに対する貢献もプロポーザルに含めてもらう。
- 審査はすばるTACで行う。
- インテンシブプログラムの審査のように他分野の審査員も加わった形の審査になると考えられる。
- 年齢、分野なども考慮した審査になる可能性がある。
- 3-5年での入れ替え(再公募)が望ましい。
 - 継続中の研究がある場合に data right を失うのは好ましくない。
 - Science Collaboration と協力している場合は、短期に人が変わるのは好ましくない。
 -
- 以下の derived data products と target lists は data right を持たない研究者とシェアできる(詳細は <http://ls.st/RDO-013>)。日本コミュニティへの還元としてプロポーザルに書ける。
 - Derived data products : derived from LSST proprietary data but cannot be used to recreate any proprietary LSST data products (color-color diagram、color-magnitude diagram、shear measurement による mass map など)
 - Target lists : limited to 1000 objects and may include coordinates and magnitudes from the proprietary catalog.

PI の義務

- 日本コミュニティが将来 LSST データを有効活用できるように活動を行ってもらう。
- たとえば、すばるのSSPやインテンシブのPIがすばるUMで報告する義務があるように、天文台の大規模広視野観測準備室(設置準備中)が毎年報告会兼研究会を主催し、PIはその開催に協力し、そこで活動報告をしてもらう。
- また、データ解析勉強会等も同時に行う。
- すばるUMや光赤天連シンポにくっつけて開催することが考えられる。

HSCのデータ解析(現状)

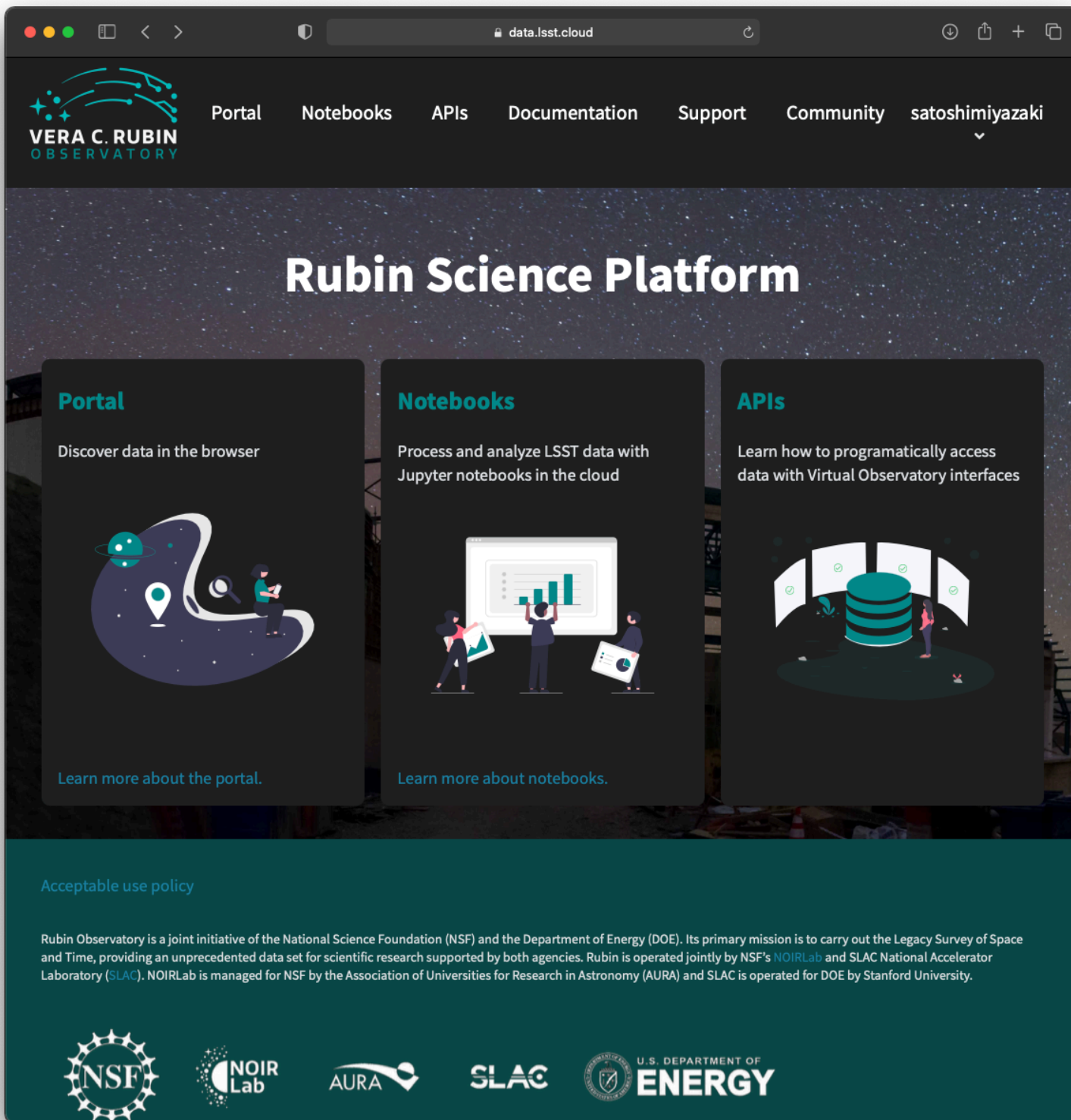
- ・ 共同利用観測者
 - ・ **lsst stack** (データ解析パッケージ)を用いて、**raw data (0.8 Gpixel/exposure)**を**stack**し、その最終画像から天体カタログを作成、研究素材とする
 - ・ 一晩 多くて**200 exposure** 年間数晩観測？
 - ・ メモリ多めの普通程度のPC (**10 core CPU, メモリ64GB, 数TB**)で十分解析可能
 - ・ ただし、極端に天体が混んだ領域、大型の天体が存在する領域では、天体**Deblend**のために、**> 100 GB**のメモリーが必要 → データセンター大規模解析システムへ
- ・ **HSC-SSP (大型サーベイ)**
 - ・ 専門チームが作成した天体カタログを**SQL Server**から取り出し、研究素材とする
 - ・ 場合によっては、解析途中の画像までもどり、検証 (**e.g. transient/variable**などを探する場合)

HSCのデータ解析 問題点

- ・ HSC-SSP (大型サーベイ)
 - ・ 専門チームが作成した天体カタログを**SQL Server**から取り出し、研究素材とする
 - ・ 場合によっては、解析途中の画像をダウンロードし、検証 (e.g. **transient/variable**などを探す場合)
 - ・ データ量が増大すると・・・
 - ・ データベース検索結果や画像をダウンロードするのに時間がかかる。研究における試行錯誤過程の効率が悪化。

これを解決するために、リモートにある計算機資源 (**CPU**、メモリー、**Data Base**)を利用して、ダウンロードすることなく、研究素材を吟味 (統計処理、グラフ化)する仕組みが整えられつつある。

Science Platform



Simulation Dataを
用いて、Proto type
の運用が最近開始された

300平方度

1.4億天体

200人程度で試使用中
(内日本人3人)

Portal

VERA C. RUBIN OBSERVATORY

Pixel Size: EQ-J2000: Image Pixel: Lock by click Background Monitor Logout

TAP Searches

- Select TAP Service**
- Select Query Type**
- Select Table**
 - IRSA** https://irsa.ipac.caltech.edu/TAP
 - Gaia** https://gea.esac.esa.int/tap-server/tap
 - CADC** https://www.cadc-ccda.hia-ihc.nrc-cnrc.gc.ca/tap
 - MAST Images** https://vao.stsci.edu/CAOMTAP/TapService.aspx
 - GAVO** http://dc.g-vo.org/tap
 - HSA** https://archives.esac.esa.int/hsa/whsa-tap-server/tap
 - NED** https://ned.ipac.caltech.edu/tap/
 - NASA Exoplanet Archive** https://exoplanetarchive.ipac.caltech.edu/TAP/
- Enter Constraints**
 - ☒ **Spatial**
 - Longitude Column:
 - Latitude Column:
 - Shape Type:
 - Coordinates or Object Name:
 - Abell 859 resolved by NED**
145.91543, 8.85368 Equ J2000 or 9h43m39.70s, +8d51m13.2s Equ J2000
 - Radius:
 - Valid range between: 1" and 360000"
 - ☐ **Temporal**

☒	ra		deg	pos.eq.ra
☒	dec		deg	pos.eq.dec
☒	blendedness			
☒	clean			
☒	cModelFlux_flag_g			
☒	cModelFlux_flag_i			
☒	cModelFlux_flag_r			
☒	cModelFlux_flag_u			
☒	cModelFlux_flag_y			
☒	cModelFlux_flag_z			
☐	cModelFlux_g			
☐	cModelFlux_i			
☐	cModelFlux_r			
☐	cModelFlux_u			
☐	cModelFlux_y			
☐	cModelFlux_z			
☐	cModelFluxErr_g			
☐	cModelFluxErr_i			
☐	cModelFluxErr_r			
☐	cModelFluxErr_u			
☐	cModelFluxErr_y			
☐	cModelFluxErr_z			
☒	extendedness			

Row Limit:

v2.3.3_Development Built On: 2021-06-22

世界各地のVO
TAP Serviceも
見える

米国以外では
カナダ、ドイツ
ESA
(Herschel/GAIA)

Notebook

The image shows a JupyterHub interface with two main windows. The left window, titled 'Server Options', displays a table with columns 'Image' and 'Options'. The 'Image' column lists various Ubuntu releases and weekly updates, with 'w_2021_35' selected. The 'Options' column lists hardware configurations (Small, Medium, Large) and checkboxes for 'Enable debug logs' and 'Clear .local directory (caution!)'. A large orange 'Start' button is at the bottom. The right window shows a Jupyter notebook titled 'Intermediate TAP Queries for DP0 catalogs'. It features the VERA C. RUBIN OBSERVATORY logo, a 'Credit' section, 'Learning Objectives', and a 'Set Up' section with a code cell for importing Python packages.

Image	Options
☒ Recommended (Weekly 2021_33)	☒ Small (1.0 CPU, 3072M RAM)
☐ Release r22.0.1	☐ Medium (2.0 CPU, 6144M RAM)
☐ Weekly 2021_35	☐ Large (4.0 CPU, 12288M RAM)
☐ Weekly 2021_34	
☐ Daily 2021_08_29	☐ Enable debug logs
☐ Daily 2021_08_28	☐ Clear .local directory (caution!)
☐ Daily 2021_08_27	
☐ Weekly 2021_25	
☐ Select uncached image (slower start):	
w_2021_35	

Start

Intermediate TAP Queries for DP0 catalogs
Last verified to run on 2021-08-16 with LSST Science Pipelines release w_2021_33
Contact authors: Leanne Guy
Target audience: All DP0 delegates.
Container Size: medium
Questions welcome at community.lsst.org/c/support/dp0
Find DP0 documentation and resources at dp0-1.lsst.io

Credit: Originally developed by Leanne Guy in the context of the Rubin DP0.1. Please consider acknowledging Leanne Guy if this notebook is used for the preparation of journal articles or software releases.

Learning Objectives

The Rubin Science Platform provides QUERY access to the DP0.1 catalogs via TAP from jupyter notebooks. TAP is a Virtual Observatory protocol for access to catalog data. In this tutorial, we will learn how to explore the DP0.1 archive via TAP and execute complex queries to retrieve data. Full TAP documentation can be found [here](#).

This notebook demonstrates how to:

1. Explore the DP0.1 schema and catalogs using the Rubin TAP service
2. Query the DP0.1 Object and Truth Match catalogs, and retrieve data for analysis
3. Visualize and interact with the retrieved data set
4. Work with asynchronous TAP queries

Set Up

```
[1]: # Import general python packages
import numpy as np
import re
import pandas
```

Tutorialは以下で参照可能

<https://github.com/rubin-dp0/tutorial-notebooks>

Notebook

data.lsst.cloud

Kernel Rubin Tabs Settings Help

02_Interm X 03_Image_X 04_Intro_t X 05_Intro_t X 06_Compe SQL-bencl X result-view X

LSST

```
[12]: # Convert the results to pandas data frames and assert that the
# contents of the two tables are identical
assert_frame_equal(results.to_table().to_pandas(),
                    results1.to_table().to_pandas())
```

2.2 Cone search around a point with specified radius

We will execute a cone search on the Object table centered on (RA, Dec) = (62.0, -37.0) with a radius of 0.1 degrees and applying a cut on magnitude. We expect to get 15,670 results.

```
[13]: # Define our reference position on the sky and cone radius in arcseconds
# to use in all following examples
coord = SkyCoord(ra=62.0*u.degree, dec=-37.0*u.degree, frame='icrs')
radius = 0.1 * u.deg
```

```
[14]: query = "SELECT ra, dec, mag_g, mag_i " \
            "mag_i, mag_g_cModel, mag_r_cModel, mag_i_cModel, " \
            "psFlux_g, psFlux_r, psFlux_i, " \
            "cModelFlux_g, cModelFlux_r, cModelFlux_i, " \
            "tract, patch, extendedness, good, clean " \
            "FROM dp01_dc2_catalogs.object " \
            "WHERE CONTAINS(POINT('ICRS', ra, dec),CIRCLE('ICRS', " \
            "+ str(coord.ra.value) + ", " + str(coord.dec.value) + ", " \
            "+ str(radius.value) + " )) = 1"
print(query)

SELECT ra, dec, mag_g, mag_i mag_i, mag_g_cModel, mag_r_cModel, mag_i_cModel, psFlux_g,
psFlux_r, psFlux_i, cModelFlux_g, cModelFlux_r, cModelFlux_i, tract, patch, extendedness
, good, clean FROM dp01_dc2_catalogs.object WHERE CONTAINS(POINT('ICRS', ra, dec),CIRCLE
('ICRS', 62.0, -37.0, 0.1 )) = 1
```

```
[15]: # For more detailed analysis of results, converting
# to a pandas dataframe is often very useful
results = service.search(query).to_table().to_pandas()

# Use an assertion to make sure we got the correct number of results.
assert len(results) == 15670
```

2.3. Cone search joining the results with the truth information

We will now join the results from the above query with the truth information. We will also add in some quality filters on the match. Objects in the truth-match table that do not have matches in the object

data.lsst.cloud

Rubin Tabs Settings Help

02_Interm X 03_Image_X 04_Intro_t X 05_Intro_t X 06_Compe SQL-bencl X

Markdown git

```
hover_tool_cmd = HoverTool(tooltips=tooltips)

[25]: # Create a Colour-Magnitude Diagram, color coding the different object typ
p = figure(title="Colour - Magnitude Diagram",
           x_axis_label='g-i', y_axis_label='g',
           x_range=(-1.8, 4.3), y_range=(29.5, 16),
           **plot_options)

# Define a palette for the truth types
truth_type_palette = ['darkred', 'lightgrey', 'blue']
p.add_tools(hover_tool_cmd)
p.circle(x='gmi', y='gmag', source=source,
         size=3, alpha=0.6,
         legend_field="truth_type",
         color=factor_cmap('truth_type',
                           palette=truth_type_palette,
                           factors=['star', 'galaxy', 'SNe']),
         hover_color="firebrick")
show(p)
```

Colour - Magnitude Diagram

48

Notebook

- ・ **Kubernetes**で計算機資源 (**CPU cluster**)を管理、ユーザーに仮想化環境を提供
- ・ ユーザーは**Jupyter Notebook**を通じて、データベースアクセス、統計解析、**Plotting**など、研究素材の吟味・加工を行う
- ・ データ転送の必要性はなくなる

ただし、、、

データベースアクセス速度が研究生産性のボトルネックであるのは変わらない

高速な**DB**が必要

現在、データセンター:古澤、すばる:小池で高速データベース付き**RSP**の試験運用の準備中である。

HSCのデータ解析効率向上、**RSP**開発への貢献を目指す

Rubin対応 To Do List

- ・ 天文台内の体制整備
 - ・ 2022/04に関係者がいろいろ動いてしまった。
- ・ 先方と協議し、すばる望遠鏡時間分を確定し、PIの選考を始め、PIが早期にScience Collaborationに参加できるようにする
- ・ quickDBを実装したScience Platformを稼働させ、HSCデータ解析の生産性を向上させる。これにより、やがてやってくるRubin dataに備える。