INTEGRATING ONLINE ANALYSIS WITH EXPERIMENTS TO IMPROVE X-RAY LIGHT SOURCE OPERATIONS [*]

N.M. Cook[†], E. Carlin, J. Einstein-Curtis, P. Moeller, R. Nagler, R. O'Rourke,
RadiaSoft LLC Boulder, CO 80301, USA
A. Barbour, M. Rakitin, L. Wiegart, H. Wijesinghe,
National Synchrotron Light Source II, Brookhaven National Laboratory, NY, 11973, USA

Abstract

The design, execution, and analysis of light source experiments requires the use of sophisticated simulation, controls and data management tools. Existing workflows require significant specialization to accommodate specific beamline operations and data pre-processing steps necessary for more intensive analysis. Recent efforts to address these needs at the National Synchrotron Light Source II (NSLS-II) have resulted in the creation of the Bluesky data collection framework, an open-source library for coordinating experimental control and data collection. Bluesky provides high level abstraction of experimental procedures and instrument readouts to encapsulate generic workflows. We present a prototype data analysis platform for integrating data collection with real time analysis at the beamline. Our application leverages Bluesky in combination with a flexible run engine to execute user configurable Python-based analyses with customizable queueing and resource management. We discuss initial demonstrations to support X-ray photon correlation spectroscopy experiments and future efforts to expand the platform's features.

INTRODUCTION

X-ray light sources are prominent drivers of scientific discovery across a range of disciplines. These facilities serve a diverse user community, often providing concurrent beam time and user support to tens of domain scientists with unique backgrounds. Increasing demand for beam time, coupled with the increasing sophistication of experiments, places constraints on the software infrastructure required to successfully carry out experiments within available time and resource allocations. Recently, significant development efforts have been made towards improving experimental planning and execution; however, significant challenges remain to integrating real-time analysis tools within the experimental workflow. In this proceedings, we discuss a strategy for incorporating analysis pipelines within common experimental workflows, focusing on applications at the NSLS-II light source. We present a schematic workflow for orchestrating analysis in concert with experimental execution. We then demonstrate this workflow via an open source, browser-based interface furnishing beamline agnostic analysis pipelines.

AN INTEGRATED FRAMEWORK FOR EXPERIMENT AND ANALYSIS

Our platform supports customizable analysis coupled with pre-existing experimental and data management frameworks. The Bluesky Data Collection Framework is used to coordinate data and metadata access and processing [1]. Bluesky aims to provide end-to-end experimental planning, execution, and data acquisition tools through a set of interoperable Python libraries; it is currently used across all active beamlines at NSLS-II. Bluesky is comprised of a set of distinctive libraries, and here we highlight a few critical ones. First, the eponymous `bluesky` library implements a run engine and event model to permit experimental control and data collection through the execution of high level plans. The `ophyd` library provides hardware abstraction to communicate plans to devices along the beamline. The `databroker` library implements an API for structured access to experimental data and metadata generated during an experiment executed by Bluesky. These libraries are under active development, with new capabilities added frequently.

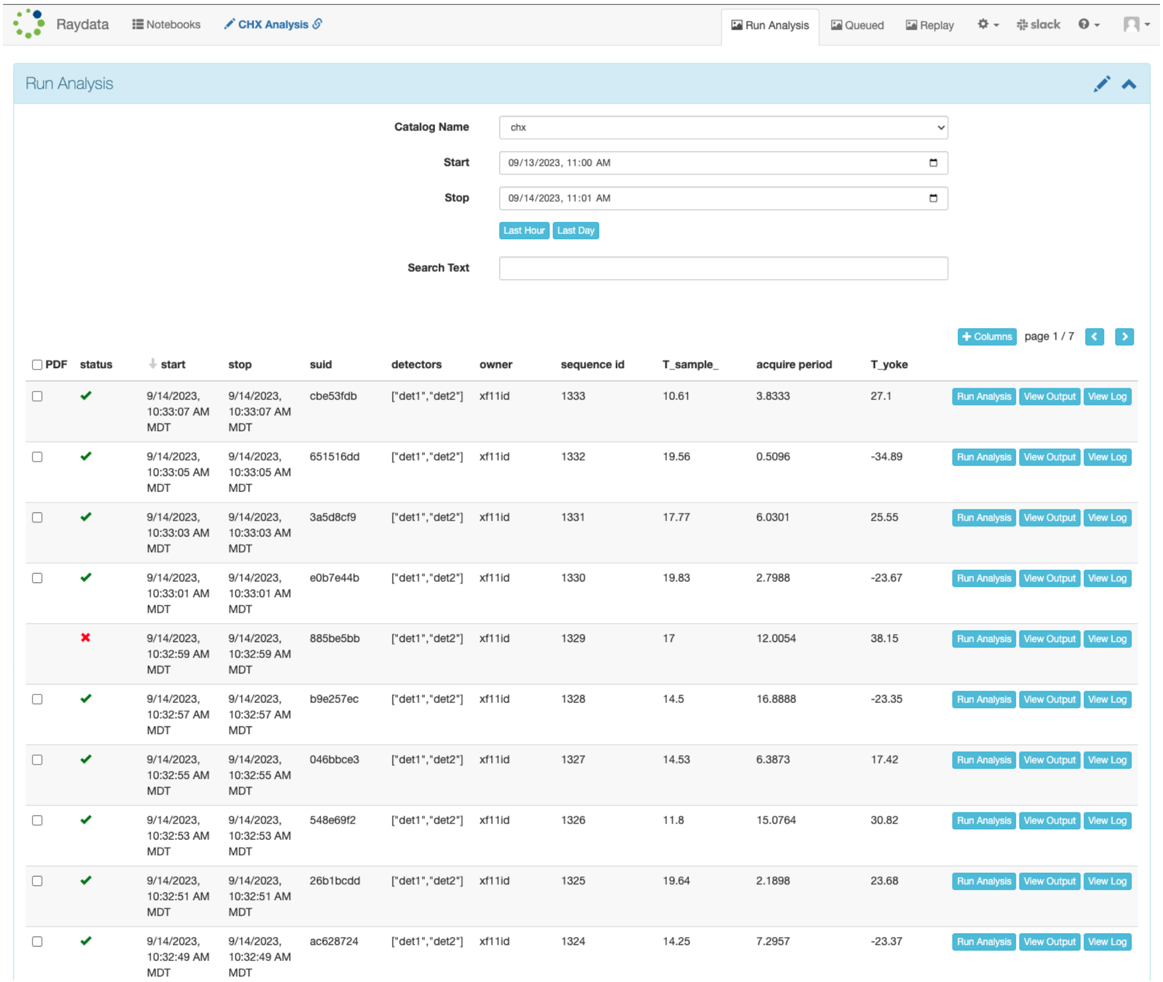
To orchestrate analyses, we have developed a custom scan monitor to provide asynchronous execution of Python-based analyses, operating in tandem with a Sirepo front end user interface. Sirepo is an open-source scientific computing gateway that provides access to community codes through custom, browser-based interfaces and an embedded JupyterHub instance. Sirepo is designed to be hardware agnostic; simulation environments are deployed via Docker containers, and can be executed across a range of computing systems, ranging from a laptop to a GPU cluster an high performance computing facility. Sirepo provides support for numerous accelerator modeling and related tracking codes; existing applications have been employed to provide customized simulations of X-ray beamlines at the NSLS-II using the Synchrotron Radiation Workshop code [2]. Sirepo has also been integrated with Bluesky to enable the asynchronous execution of long-running SRW simulations to support multiparametric optimizations of beamlines [3].

ANALYSIS SERVICE AT NSLS-II

We have deployed a prototype analysis web-service at NSLS-II. The application is hosted on the NSLS-II science network, accessible from beamline workstations or via remote desktop applications such as Apache Guacamole or VMWare Horizon by all credentialed users. The application is deployed via a Podman container providing the environ-

[*] This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC00215553.

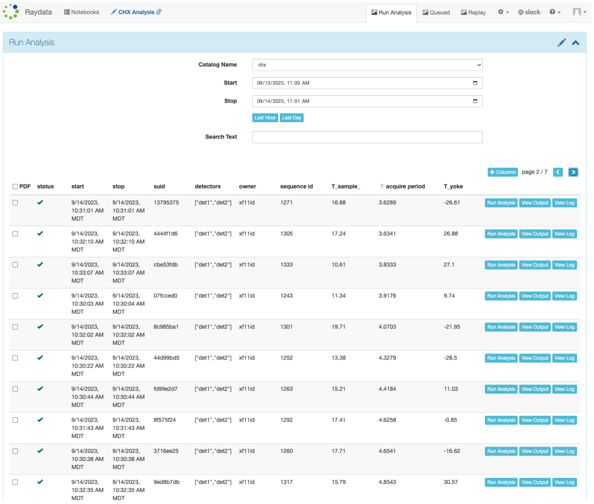
[†] ncook@radiasoft.net



Catalog Name: chx
Start: 09/13/2023, 11:00 AM
Stop: 09/14/2023, 11:01 AM
Search Text:

□ PDF	status	start	stop	suid	detectors	owner	sequence id	T_sample_	acquire period	T_yoke
□	✓	9/14/2023, 10:33:07 AM MDT	9/14/2023, 10:33:07 AM MDT	cbe53fdb	["det1","det2"]	xf11id	1333	10.61	3.8333	27.1
□	✓	9/14/2023, 10:33:05 AM MDT	9/14/2023, 10:33:05 AM MDT	651516dd	["det1","det2"]	xf11id	1332	19.56	0.5096	-34.89
□	✓	9/14/2023, 10:33:03 AM MDT	9/14/2023, 10:33:03 AM MDT	3a5d8cf9	["det1","det2"]	xf11id	1331	17.77	6.0301	25.55
□	✓	9/14/2023, 10:33:01 AM MDT	9/14/2023, 10:33:01 AM MDT	e0b7e44b	["det1","det2"]	xf11id	1330	19.83	2.7988	-23.67
	✗	9/14/2023, 10:32:59 AM MDT	9/14/2023, 10:32:59 AM MDT	885be5bb	["det1","det2"]	xf11id	1329	17	12.0054	38.15
□	✓	9/14/2023, 10:32:57 AM MDT	9/14/2023, 10:32:57 AM MDT	b9e257ec	["det1","det2"]	xf11id	1328	14.5	16.8888	-23.35
□	✓	9/14/2023, 10:32:55 AM MDT	9/14/2023, 10:32:55 AM MDT	046bbce3	["det1","det2"]	xf11id	1327	14.53	6.3873	17.42
□	✓	9/14/2023, 10:32:53 AM MDT	9/14/2023, 10:32:53 AM MDT	548e69f2	["det1","det2"]	xf11id	1326	11.8	15.0764	30.82
□	✓	9/14/2023, 10:32:51 AM MDT	9/14/2023, 10:32:51 AM MDT	26b1bcdd	["det1","det2"]	xf11id	1325	19.64	2.1898	23.68
□	✓	9/14/2023, 10:32:49 AM MDT	9/14/2023, 10:32:49 AM MDT	ac628724	["det1","det2"]	xf11id	1324	14.25	7.2957	-23.37

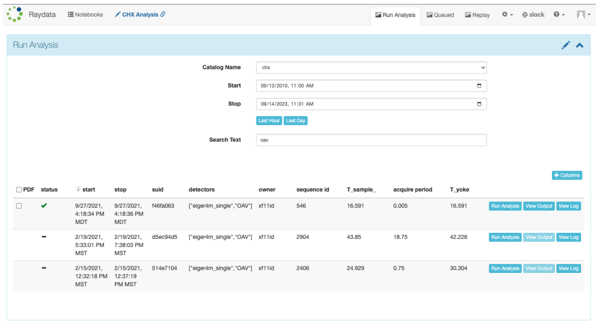
Figure 1: The run analysis tab permits users to search, sort, and select from available runs in a prescribed catalog, using high level metadata to guide decisions about selections.



Catalog Name: chx
Start: 09/13/2023, 11:00 AM
Stop: 09/14/2023, 11:01 AM
Search Text:

□ PDF	status	start	stop	suid	detectors	owner	sequence id	T_sample_	acquire period	T_yoke
□	✓	9/14/2023, 10:31:01 AM MDT	9/14/2023, 10:31:01 AM MDT	13795375	["det1","det2"]	xf11id	1271	16.88	3.6289	-26.61
□	✓	9/14/2023, 10:32:10 AM MDT	9/14/2023, 10:32:10 AM MDT	4444f1d6	["det1","det2"]	xf11id	1305	17.24	3.6341	26.88
□	✓	9/14/2023, 10:33:07 AM MDT	9/14/2023, 10:33:07 AM MDT	cbe53fdb	["det1","det2"]	xf11id	1333	10.61	3.8333	27.1
□	✓	9/14/2023, 10:30:03 AM MDT	9/14/2023, 10:30:04 AM MDT	07fcced0	["det1","det2"]	xf11id	1243	11.34	3.9176	9.74
□	✓	9/14/2023, 10:32:02 AM MDT	9/14/2023, 10:32:02 AM MDT	8c985ba1	["det1","det2"]	xf11id	1301	19.71	4.0703	-21.95
□	✓	9/14/2023, 10:30:22 AM MDT	9/14/2023, 10:30:22 AM MDT	44d99bd5	["det1","det2"]	xf11id	1252	13.38	4.3279	-28.5
□	✓	9/14/2023, 10:30:44 AM MDT	9/14/2023, 10:30:44 AM MDT	fd99e2d7	["det1","det2"]	xf11id	1263	15.21	4.4184	11.03
□	✓	9/14/2023, 10:31:43 AM MDT	9/14/2023, 10:31:43 AM MDT	8f575f24	["det1","det2"]	xf11id	1292	17.41	4.6258	-0.85
□	✓	9/14/2023, 10:30:38 AM MDT	9/14/2023, 10:30:38 AM MDT	3716ee25	["det1","det2"]	xf11id	1260	17.71	4.6541	-16.62
□	✓	9/14/2023, 10:32:35 AM MDT	9/14/2023, 10:32:35 AM MDT	9ed8b7db	["det1","det2"]	xf11id	1317	15.79	4.8543	30.57

Figure 2: Runs may be sorted by unique ID, start and stop times, and high level metadata entries from Bluesky's primary document stream.



Catalog Name: chx
Start: 09/13/2010, 11:00 AM
Stop: 09/14/2023, 11:01 AM
Search Text: oav

□ PDF	status	start	stop	suid	detectors	owner	sequence id	T_sample_	acquire period	T_yoke
□	✓	9/27/2021, 4:18:34 PM MDT	9/27/2021, 4:18:36 PM MDT	f46fa063	["eiger4m_single","OAV"]	xf11id	546	16.591	0.005	16.591
	–	2/19/2021, 5:33:01 PM MST	2/19/2021, 7:38:03 PM MST	d5ec94d5	["eiger4m_single","OAV"]	xf11id	2904	43.85	18.75	42.226
	–	2/15/2021, 12:32:18 PM MST	2/15/2021, 12:37:19 PM MST	514e7104	["eiger4m_single","OAV"]	xf11id	2406	24.929	0.75	30.304

Figure 3: The search feature of the run analysis tab enables queries of high level metadata entries from Bluesky's primary document stream. Additional efforts are in place to support baseline stream queries.

ment and file system support necessary for typical end user operations. The container configures access to a facility-wide LUSTRE filesystem via specific mounts, enabling data access consistent with how beamline scientists organize their data and analysis outputs.

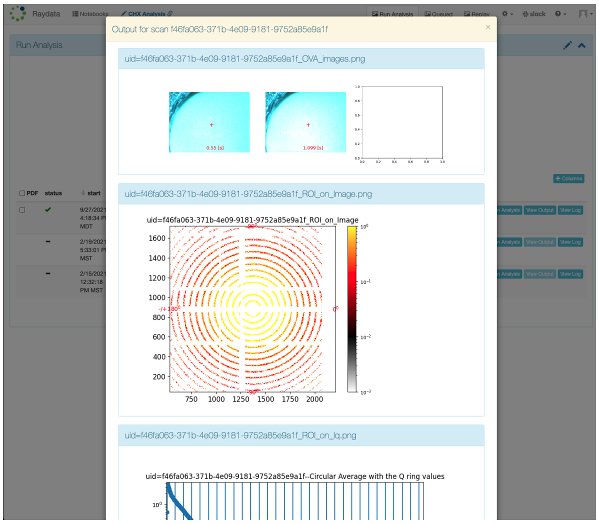

Raydata
Notebooks
CHX Analysis
Run Analysis
Output for scan f46fa063-371b-4e09-9181-9752a85e9a1f
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_OVA_images.png
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_ROI_on_Image.png
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_ROI_on_Image
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_ROI_on_Iq.png
uid=f46fa063-371b-4e09-9181-9752a85e9a1f--Circular Average with the Q ring values
PDF
status
start
+ Columns
Analysis
View Output
View Log

Figure 4: Figures generated from XPCS analysis during an experiment at the CHX beamline show the masked intensity profile from a scattering image for varying Q rings.

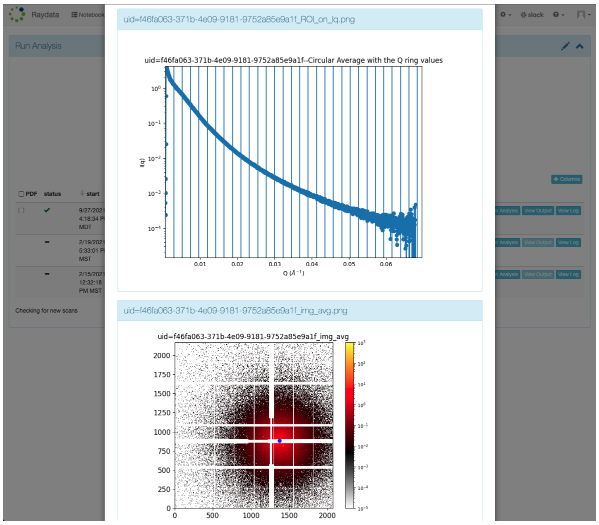

Raydata
Notebooks
Run Analysis
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_ROI_on_Iq.png
uid=f46fa063-371b-4e09-9181-9752a85e9a1f--Circular Average with the Q ring values
I(q)
Q (Å⁻¹)
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_img_avg.png
uid=f46fa063-371b-4e09-9181-9752a85e9a1f_img_avg
PDF
status
start
Checking for new scans
+ Columns
Analysis
View Output
View Log

Figure 5: Figures generated from XPCS analysis during an experiment at the CHX beamline depict the cylindrically-averaged intensity values are displayed as a function of Q, with vertical markers indicating ring separation.

Existing analysis capabilities are supported by a custom scan monitor that provides asynchronous resource and queue management for script execution. At CHX and CSX beamlines of NSLS-II, many users leverage Jupyter notebooks running IPython kernels [4] customized to provide near real-time analysis for a subset of the data generated during runs. Executing these notebooks necessitates dedicated support for required dependencies, as well as sufficient computing resources to run notebooks for each of the hundreds to thousands of datasets generated during a given experiment. Notebook execution may take 1 – 10 minutes depending upon the analysis, while data capture may only require a few seconds. As such, the analysis step constitutes a bottleneck in completing the pipeline in real-time. The deployed service leverages local beamline resources to identify, queue, and execute beamline- and run-specific Jupyter notebook workflows. Intermediary and final outputs are saved in local directories and made accessible through the web interface along with typical logging and file access tools.

A PROTOTYPE SIREPO INTERFACE FOR INTEGRATED ANALYSIS

Users of the service can navigate a browser-based interface for carrying out analysis at NSLS-II. The initial prototype provides capabilities to support a few custom workflows at the CHX and CSX beamnlines. At CHX, an XPCS analysis workflow is supported, while for CSX an image pre-processing workflow and prototype XPCS workflow are provided. Currently, these analyses are pre-selected and static for all runs; however, a new infrastructure is in testing that will identify and execute different analyses on the basis of metadata identified in the run's document stream.

Figure 1 depicts the run analysis tab, which permits users to browse data generated by Bluesky runs at a particular beamline. By leveraging the `databroker` library's catalog structures, runs can be searched and filtered according to primary metadata such as start and stop time, unique identifier (UID), user, and other relevant flags. The search results can be subsequently sorted, and users may select a subset of remaining runs for further processing. Figure 2 depicts a sorting of the runs according to displayed metadata entries, while Fig. 3 depicts search functionality, which permits the search of additional metdata entries within the document stream.

Analyses are executed by dispatching containerized environments to carryout the prescribed calculations for the requested datasets. These analyses primarily consist of Python scripts and corresponding Jupyter notebooks; this workflow permits flexibility and visibility for endusers, and its common for beamline scientists to reconfigure these notebooks on a regular basis to adapt for different experimental campaigns and machine cycles. Products of these analyses are stored on shared filesystems and made available from the interface. Users may explore executed analyses and examine output figures and logs from each run. Figures 4 and 5 depict produced images during XPCS analysis of runs at CHX. Masked intensity profiles are displayed for varying Q rings, alongside averaged intensity values, as computed according to the prescribed routines.

CONCLUSION

We report on the development of an interactive online platform for the analysis of X-ray light source experiments, built atop the Sirepo platform and Bluesky framework. The interface leverages shared metadata structures provided by the Bluesky, making use of the `databroker` library to enable active searching, sorting, and retrieval of runs from catalog structures generated during experimental execution. Custom analysis routines are supported by deploying templated Jupyter notebooks via Docker images containing the necessary dependencies. Job management, resource allocation, and queueing are provided by Sirepo, along with real-time feedback through automated figure previews and reports. Future developments will enhance queueing capabilities to support dynamic analysis selection alongside screening and prioritization of runs, with the goal of providing an integrated analysis workflow to complement Blueksy's experimental procedures on active beamlines at NSLS-II and elsewhere.

REFERENCES

[1] Bluesky Project, https://blueskyproject.io

[2] M. S. Rakitin *et al.*, "Sirepo: an open-source cloud-based software interface for X-ray source and optics simulations", *J. Synchrotron Radiat.*, vol. 25, pp. 1877–1892, 2018. doi:10.1107/S1600577518010986

[3] M. S. Rakitin *et al.*, "Introduction of the Sirepo-Bluesky interface and its application to the optimization problems", in *Proc, SPIE 11493 Adv. Comput. Methods X-Ray Opt. V*, 2020, p. 1149311. doi:10.1117/12.2569000

[4] F. Pérez and B. Granger, "IPython: A System for Interactive Scientific Computing", *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 21–29, 2007. doi:10.1109/MCSE.2007.53