How IRIS Enables Science At Diamond Light Source

Chris Reynolds – Principal Platform Engineer

DLS Scientific Computing

IRIS Collaboration Meeting Jan 2026

Contents

- The science Diamond does on IRIS
- Some history
- Compute and storage architecture
- Queue configuration
- xChem & eBIC
- Summary

eBIC Titan Krios Electron Microscope (EM)

DLS has 8 EMs

Diamond on IRIS: Science Use Cases

- Macromolecular Crystallography (MX) Pipelines *
  - Automated, submitted by a workflow engine/middleware
  - Electron Density Maps built from beamline X-Ray Data
- eBIC (electron Bio-Imaging Centre) *
  - Tomogram reconstruction
- Spectroscopy
  - Pre-visit simulations, prototype application
- XChem
  - Fragment Screening Web Application:
  - https://fragalysis.diamond.ac.uk

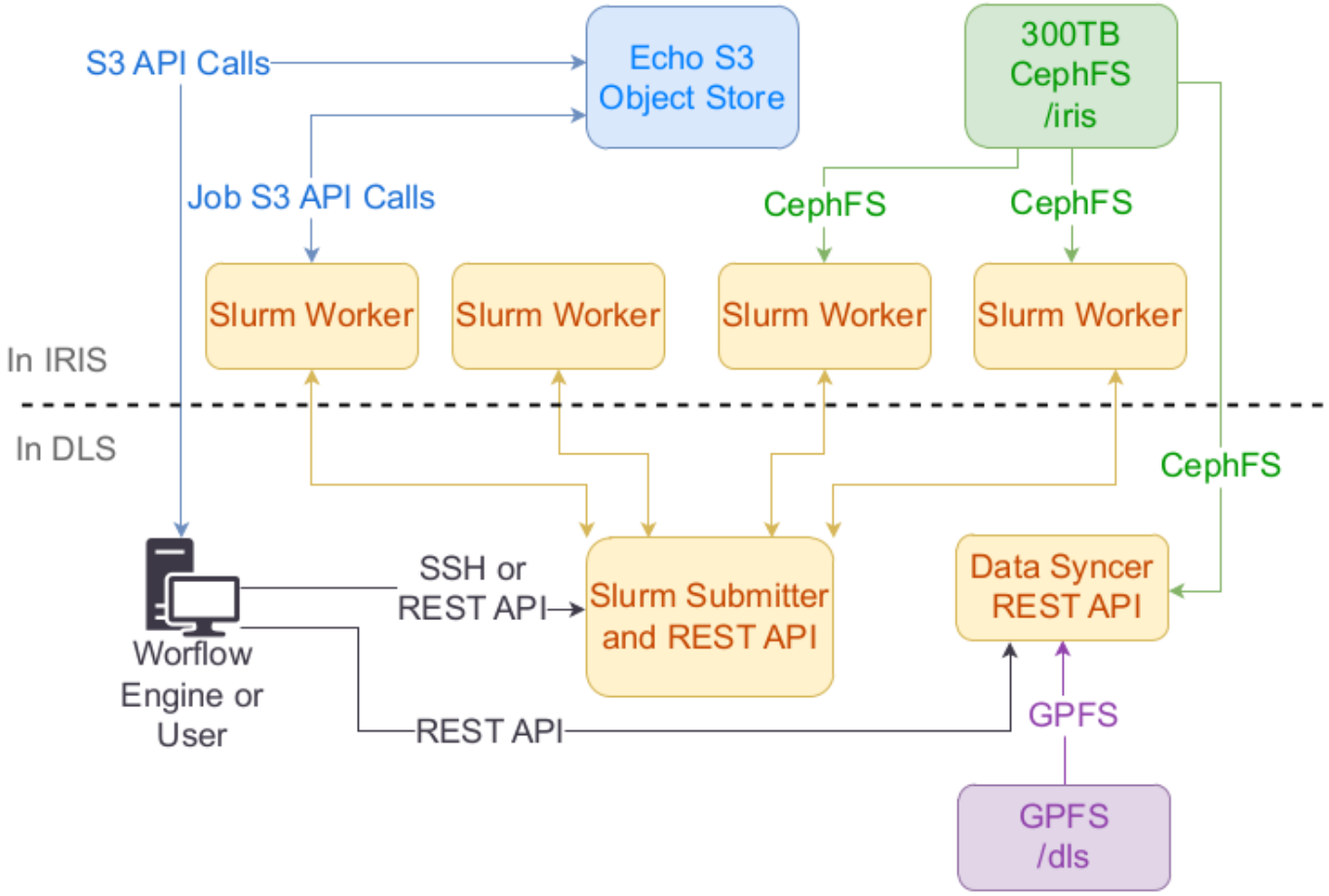
* Cloudbursting based on high demand

Some History...

- Diamond has been using IRIS for approx. 7 years
- STFC has been our IRIS provider. They do an awesome job!
- Various science use cases have come and gone
- Always "opportunistic" or "post-visit-processing"
- The compute scheduling infrastructure DLS runs on IRIS has been through several iterations
- Much experience gained about how to offer DLS Scientists reliable and performant computing infra based on IRIS components
- This experience guides our current choice of tools and architecture

Current Infrastructure

- Data syncer REST API moves data from DLS GPFS to Arided Flash CephFS (retains POSIX ACLs)
- Some data also held in Echo/S3
- Workflow Engine or User interacts with REST API or Slurm CLI to submit
- OpenVPN connects Slurm submitter to workers (allows interactive jobs, no direct login)
- DLS based AuthN/Z
- Singularity/Apptainer only for bins & libs
- Bulk of DLS IRIS allocation lives in Slurm (GPUs, ~7000 CPU Cores)



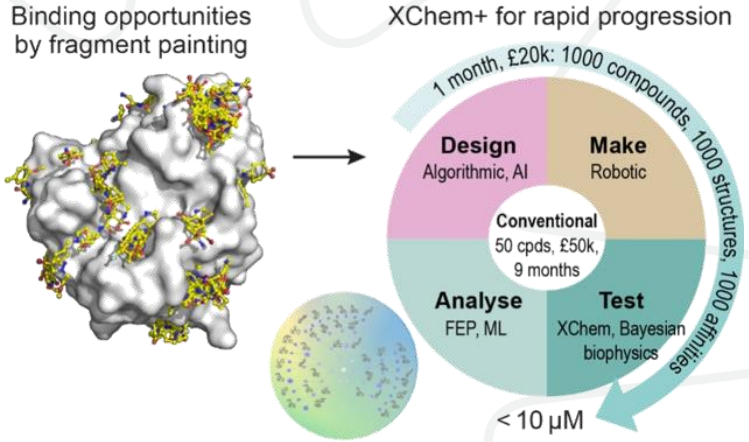
Queue Configurations

- A queue (Slurm partition) per science use case
- Some science use case partitions/queues only present on some instances
- A preemptible *main* partition/queue that is present on all instances
- A separate GPU partition/queue that is first come first served
- Effectively guarantees resource for the science use case, but offers the resource to anyone (via main partition/queue) if the science use case is not using it

Accelerating Drug Discovery Through Computational Speed – MX/xChem

The Challenge:

- Traditional drug discovery: Finding the right molecule = months or years of lab time
- We used **Fragment Screens** to speed up the process to give '**hot starts**' on target:drug interactions. Combining **Crystallography** and **chemical data**.
- **Design → Make → Test → Analyze"**(the DMTA cycle)
- Each step generates **massive amounts of data**

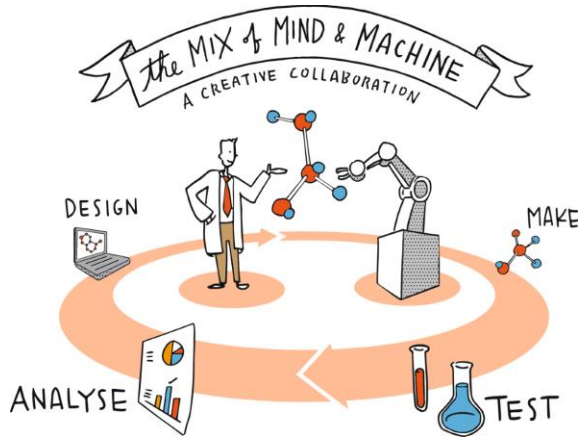


The IRIS Solution:

- **1000x Algorithmic Merging**: IRIS performed **docking** to computationally merge fragment hits into larger molecules
- **1000x Retrosynthetic Elaboration**: IRIS ran **enumeration** of building block combinations to design synthetic routes
- **Procurement Analysis**: IRIS identified which building blocks were **actually available** for purchase—making designs practical to synthesize

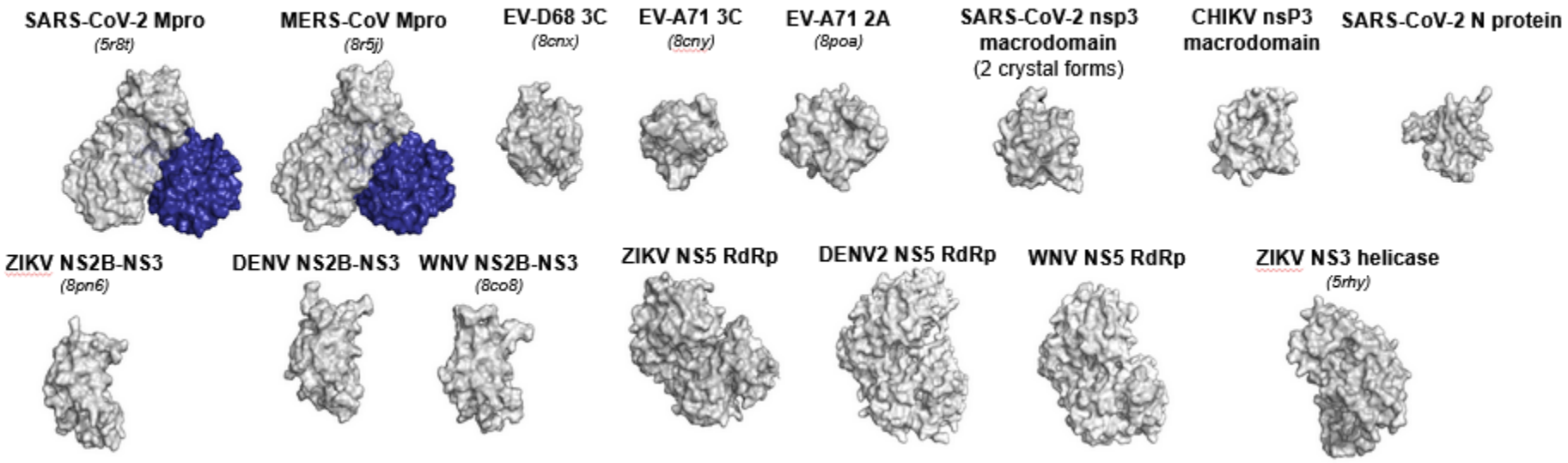
The Impact:

- From 10^1 fragment hits → 10^3 merges → 10^6 elaborations → 10^3 selected → 10^1 validated hits
- **Chemists could focus on making only the most promising compounds**, not guessing blindly
- Entire DMTA cycle compressed from months to weeks
- These cycles supported Moonshot and ASAP antiviral discovery projects. Many antiviral targets using IRIS cluster computing for drug discovery



Macromolecular Crystallography – IRIS Driven Protein Structure Determination

Project 2 (Target Enablement) developed 16 tractable crystal forms for 15 proteins (8 targeting opportunities)



Structural biology output is unprecedented:

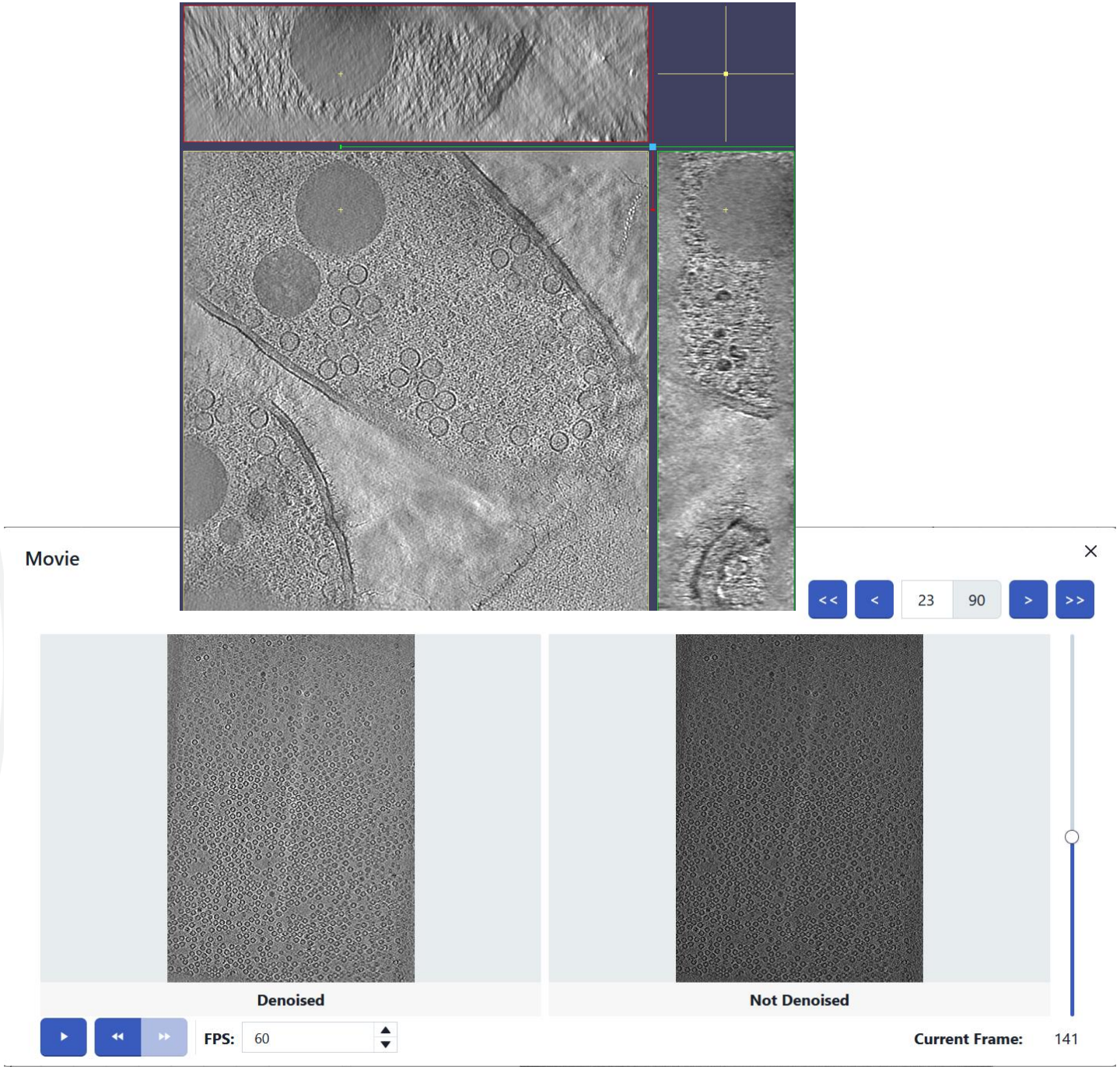
- Structural Biology Core **has collected 18,353 datasets** as of 1st of April 2024
- **601 structures deposited in Protein Data Bank** by ASAP project in 2023

  955 for all of Diamond, 916 for all American synchrotrons
- Open Science: 5 preprints, 63 Addgene, 24 protocols.io, 138 assays, 2104 compounds on ChEMBL

http://asapdiscovery.org https://covid.postera.ai/covid

eBIC

- We use IRIS to opportunistically reconstruct 3D tomogram volumes and then denoise them
- Both tasks are GPU processes taking around 5 minutes each
- Data collection of tomograms is completed in bursts, so IRIS allows us to overflow processing from the Argus cluster at times of high demand
- This means we give users feedback about the quality of their data quicker
- Fast feedback allows users to decide whether to continue collection or make changes to fix any problems



Summary

- Slurm + Apptainer built on the IaaS parts of IRIS
- Scientists mostly interact with Slurm and a data mover API only (no Openstack access)
- Hosts https://fragalysis.diamond.ac.uk web application allowing fragment hit visualisation and investigation
- IRIS is driving the "Open Source" contribution of high throughput drug discovery by xChem
- Real time experiment steering for eBIC

Thanks! Questions?

Acknowledgements Past and Present:

DLS Cloud Team:

Thomas Hartland, Sachin Gautam, Godson Alex, John Davies, Subin Saji, Richard Parke, Althaf Mohamed-Koya, Sam Diserens

- cloud.manager@diamond.ac.uk for future questions