The Rosalind Franklin Institute

# SciCat at The Rosalind Franklin Institute

Dr Laura Shemilt
SciCatCon25

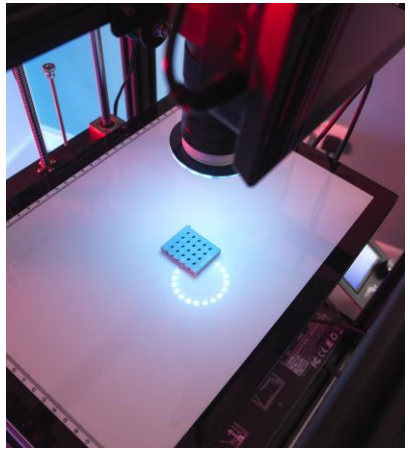
# The Rosalind Franklin Institute

technology innovation
transforming life science
improving human health




The Rosalind Franklin Institute

# A different data management problem…



Off the shelf instruments

No access to control machines

No access to control software

Instruments Developed with Industrial partners

Development under NDA

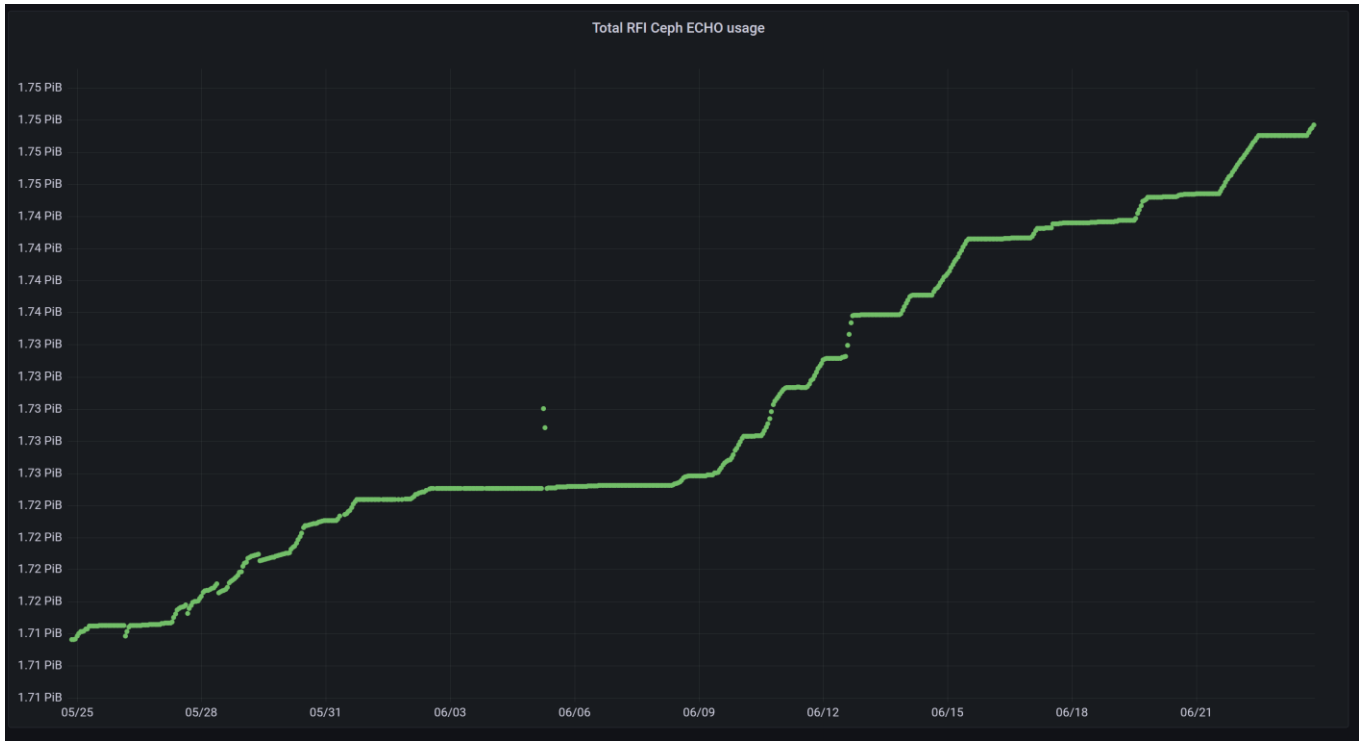
Industrial partner develops controls software

# Statistics

**1.75 PB** data and growing

**415 k** records in SciCat

**39** number of instruments in our data infrastructure, **23** pushing metadata to SciCat



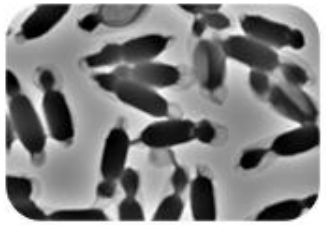



The Rosalind Franklin Institute

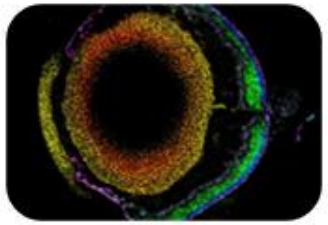
# SciCat Migration

# About the Rosalind Franklin Institute

## Technology Innovation Challenges

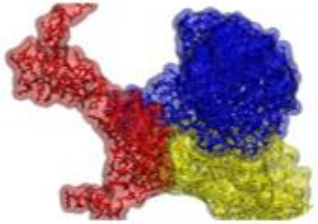


Multidimensional Molecular Imaging



Integrated Chemical Imaging in Cells and Tissues

Molecular Perturbations: Chemistry Engineering Biology

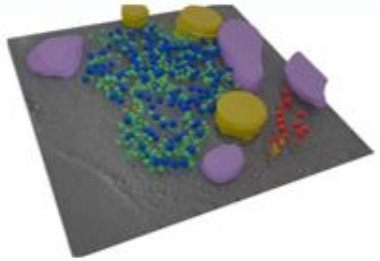


AI and Informatics for Predictive Biology

Application *pull*

Technology *push*

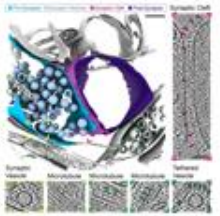
## Life Science Challenges



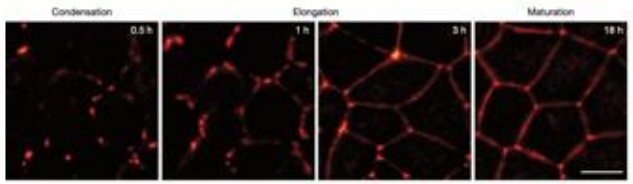
Biology across scales

Host-pathogen interactions

## Emerging interest areas



Cell compartmentation

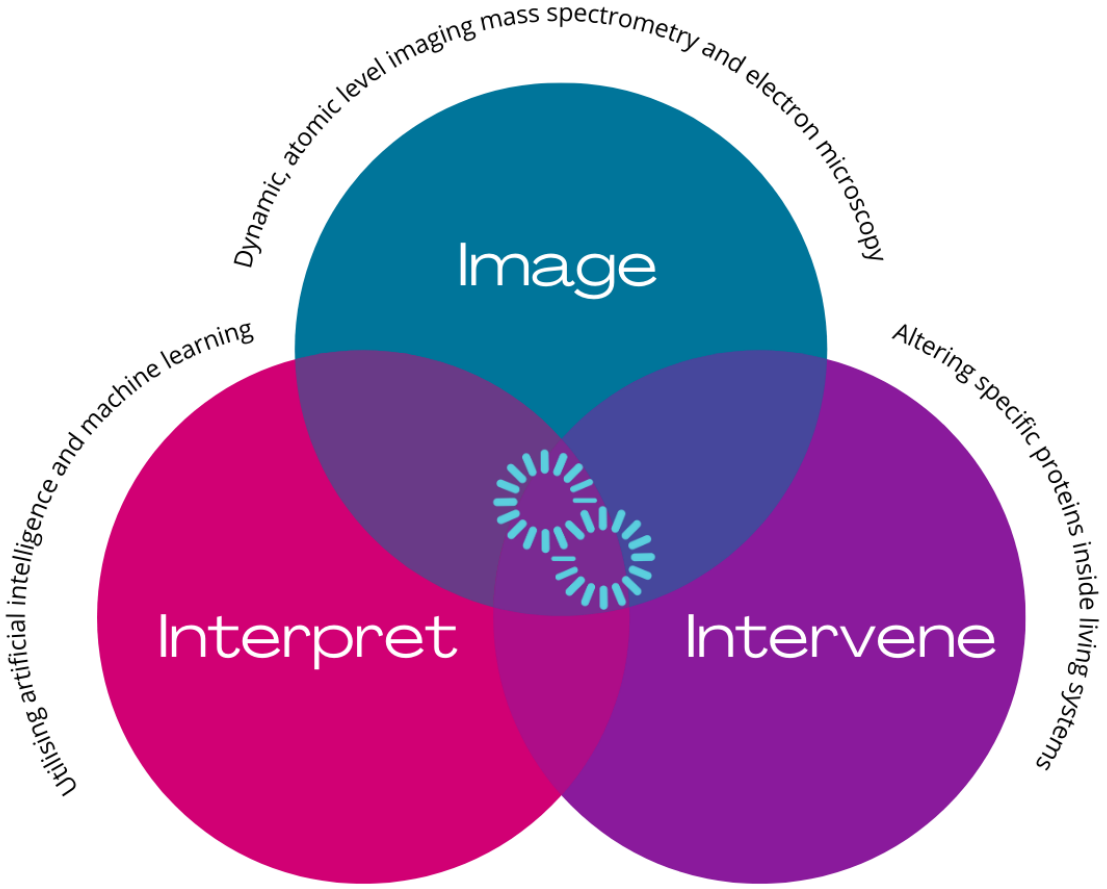



Cell-cell interaction



The Rosalind Franklin Institute

# About The Rosalind Franklin Institute Jan 2024






The Rosalind Franklin Institute

# Challenges of migrating: Business Logic Changes

Franklin in the early days

- Business logic is not always the same as ways of working
- Demand for flexibility lead to no standardisation

Franklin now

- Age and size of institute requires necessary management changes
- this requires us to change data management

# Challenges of migrating: Business Logic Changes

Franklin in the early days

- Business logic is not always the same as ways of working
- Demand for flexibility lead to no standardisation

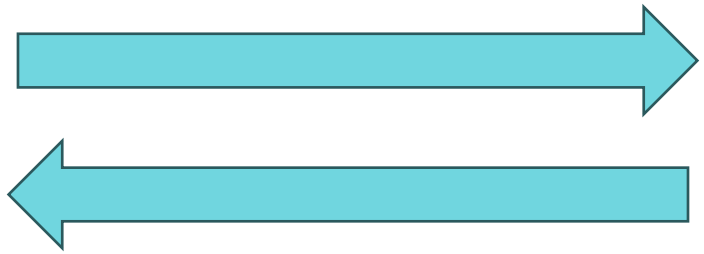
Franklin now

- Age and size of institute requires necessary management changes
- this requires us to change data management

**Isn’t metadata supposed to be persistent?**

# Challenges of migrating: Business Logic Changes

## Defining a Working System



## Demonstrating a Working System

Strong definition leads to easier automation and management

Longer time to product

Difficult to maintain interest whilst you work everything out

Develop from feedback

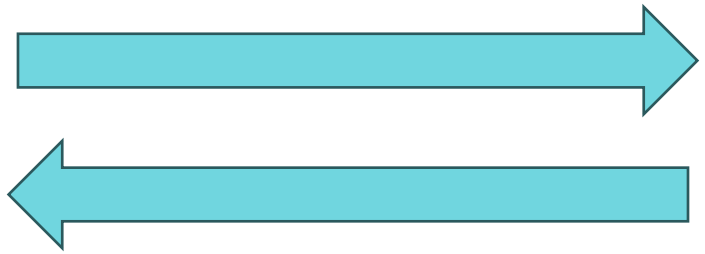
Iterative cycle

Easier to keep your colleagues' interest

Lots of edge cases that are difficult to maintain

# Challenges of migrating: Business Logic Changes

Defining a Working System



Demonstrating a Working System

Strong defini
easier autom
managemen

Longer time

Difficult to m
interest whils
everything ou

m feedback

cle

ep your colleagues'

e cases that are
aintain

It doesn't matter which approach you take, when making business logic changes you need to make large structural changes

# Challenges of migrating: Metadata Provenance

- Keeping metadata provenance
  - Mapping group names to deal with new structure? Or keep old ones even though they will eventually be meaningless
- Dealing with leavers
  - In our current version of SciCat you either have full access or no access
  - Use new features of SciCat to control what a user sees
- Ingestion of historic data
- Clearing up mistakes
  - We do not let users delete records, where there have been mistakes with the file-monitor should we fix these?

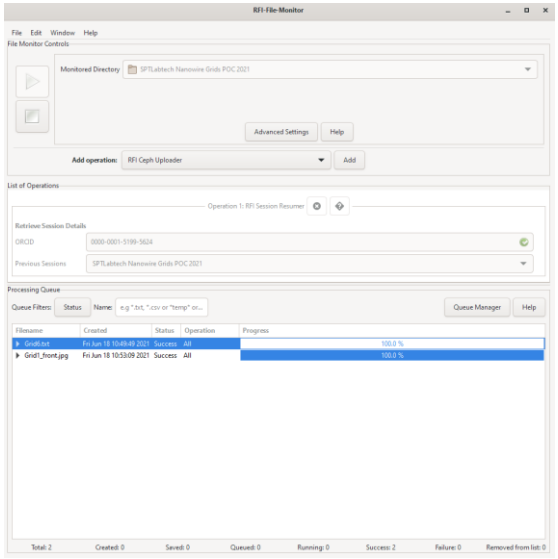
# Challenges of Migration: Maintaining Functionality

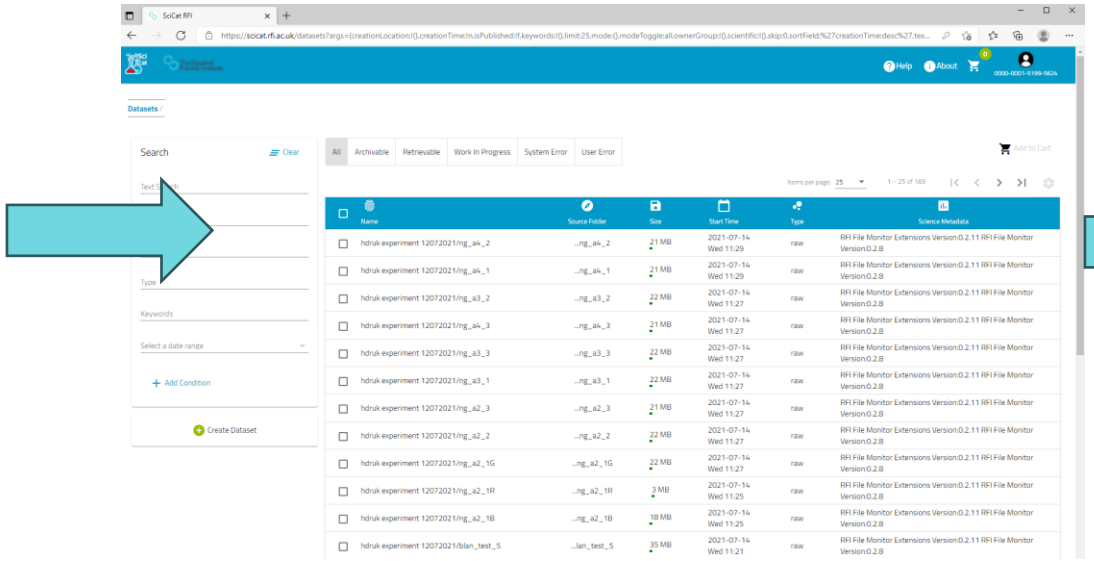
Maintaining user experience

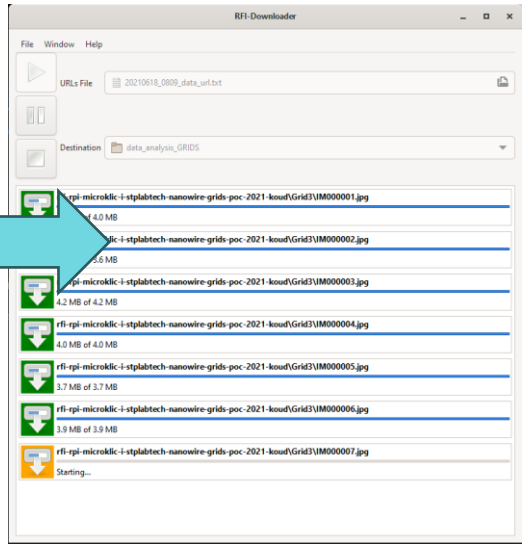
➔ A user needs to be able to upload analysed data to SciCat
➔ Download their data from SciCat
➔ Have access to the same metadata as before
➔ Have access to the correct metadata going forward

# Phase 1

## RFI-File-Monitor -> SciCat -> RFI-Downloader



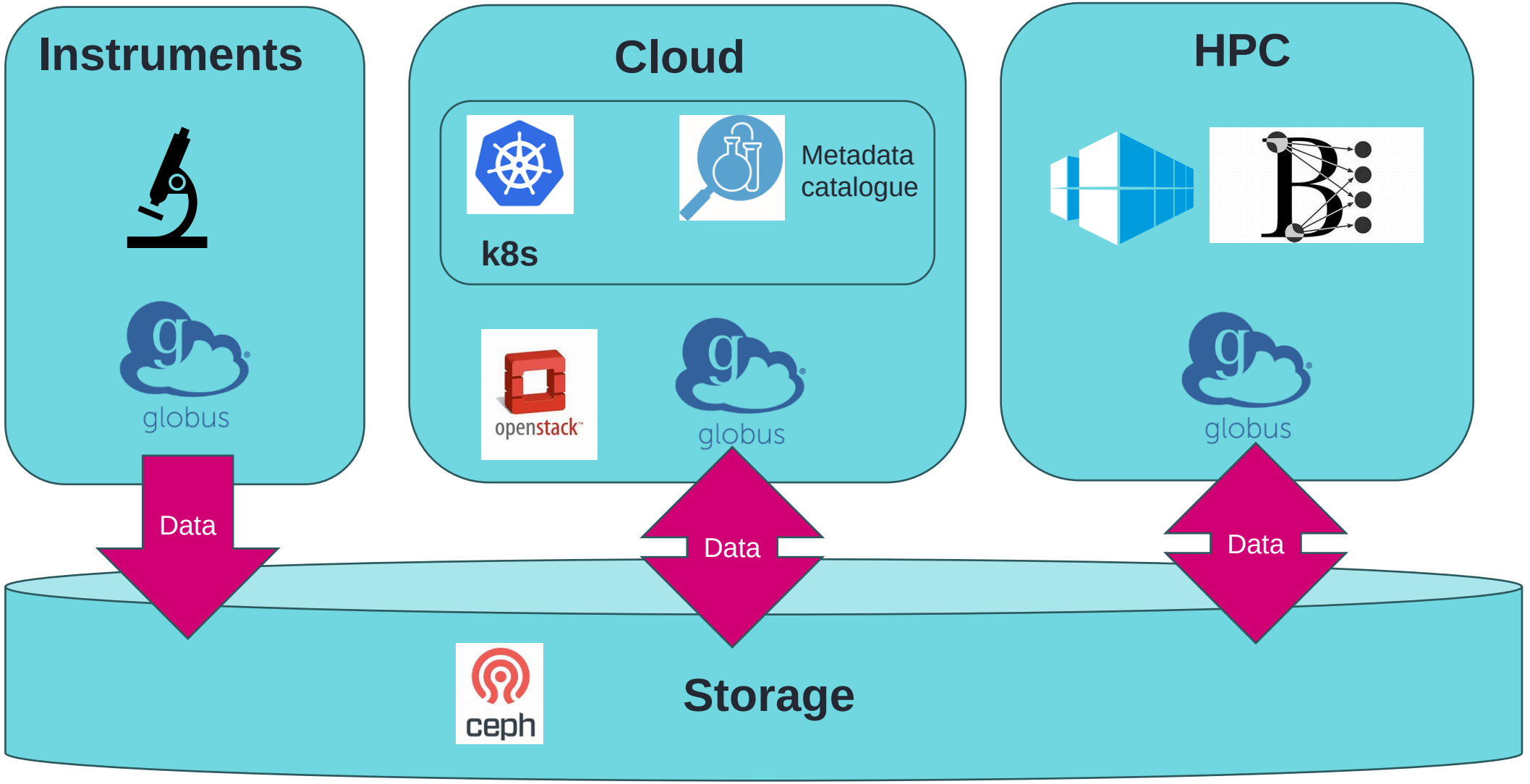






The Rosalind Franklin Institute

# Phase 2

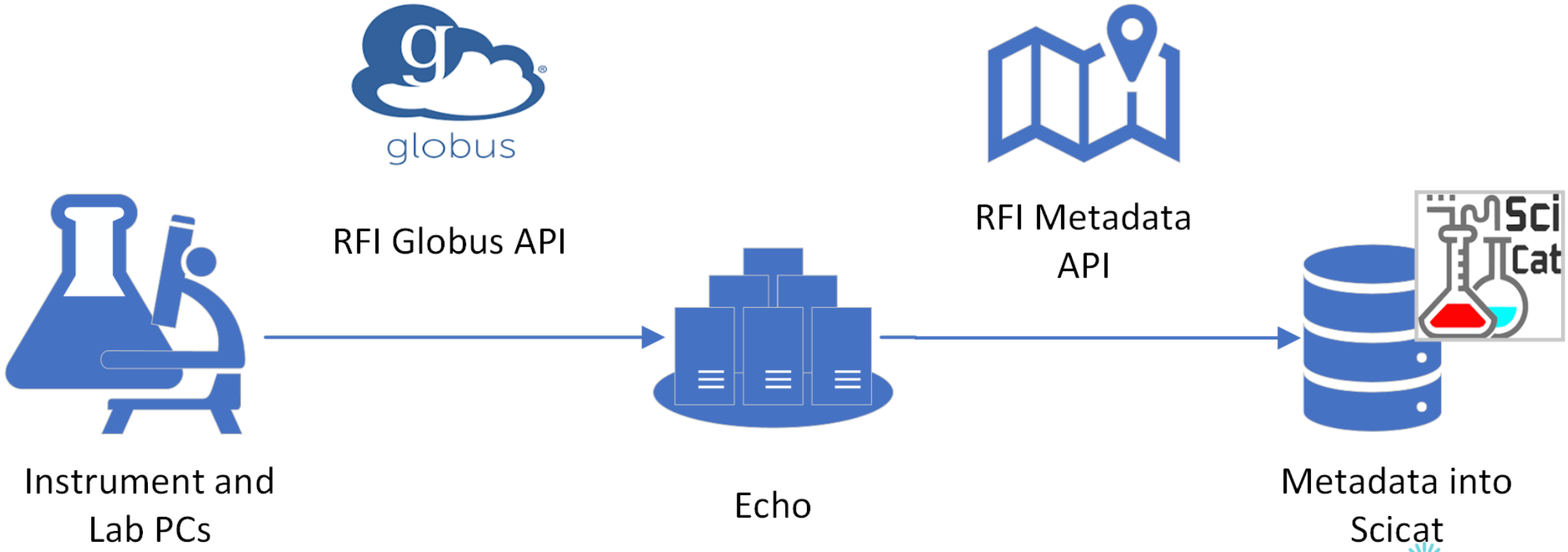





The Rosalind Franklin Institute

# RFI-Metadata-API and RFI-Scicatalogger-API

- Not RESTful
- Python CLI interfaces in functional programming style

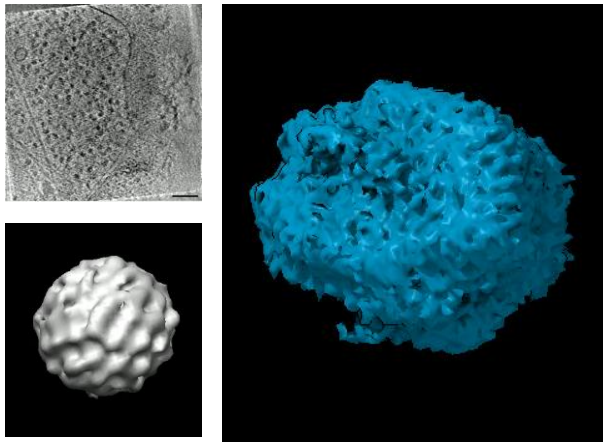


# Kubernetes in The Franklin

- Running Kubernetes for 3 years
- Training the team
- Changing the way we think
- Building for the new platform
- SciCat migration will have a large change in architecture
- Using ArgoWF for our pipeline

# Expanding SciCat

# Cryo Electron Microscopy Pipeline



Microscope

SciCat
advance your data to advance your science

argo

kubernetes

**Visualisation**
View images in the browser (Neuroglancer)
View reports on quality metrics

**Subset selection**
Choose most promising data for further analysis
Curate training data for machine learning

**Publish**
Scicat records contain information needed for community databases, e.g., EMDB, EMPIAR

**CPU jobs**
motion correction, CTF estimation

**GPU jobs**
tilt series alignment, tomogram reconstruction, particle picking

ceph-echo
longer term storage



The Rosalind Franklin Institute

# Growing a UK community (SciCatUK)

UK Research and Innovation

## UKRI Digital Research Infrastructure

## Interconnected DRI

fostering community engagement

promoting the creation of federated infrastructure

## Human DRI

promoting careers, and supporting  professional development

foster a supportive culture and ensure a skilled and diverse DRI workforce.

## FAIR DRI

driving adoption of the FAIR principles

## Sustainable DRI

addressing efficiency, security, environmental and financial challenges to create a sustainable DRI ecosystem.

# Research Software Maintenance Fund

- Call from the Software Sustainability Institute
- Application to **improve UX/UI** and do **essential frontend maintenance**
- Franklin leading EOI with Max Novelli (ESS) as co-investigator
- **£500k** over two years

# Thanks to

**The Rosalind Franklin Institute**
Mark Basham
Alex Lubbock
Dimitrios Bellos
Laura Crawford
Elaine Ho
Nick Crawford
Piper Fowler-Wright
Tibor Auer (Former)
Silvia Ramos (Former)
Joss Whittle (Former)

**SciCat**
The whole community …
..and for their advice and discussion
Max Novelli (ESS)
Dylan McReynolds (ALS)
Carlo Minotti and Frederic Poitier (PSI)
Spencer Bliven and Despina Adamopoulou (Swiss OpenEM)
Daphne Van Dijken and Max Burian (DECTRIS)
Amir Tosson (Uni. Seigen)
Regina Hinzmann (DESY)

**STFC**
Martin Summers
Tom Byrne
Jacob Ward
John Good
Aidan Mc Coomb