



Science and  
Technology  
Facilities Council

# Introduction to CCP-EM

*Software and support for the  
cryoEM community*

Tom Burnley  
July 2025  
Padova, Italy



# What is CCP-EM?



*Collaborative Computational Project for Electron cryo-Microscopy*

Support users and developers in computational aspects of biological EM via **software & training**

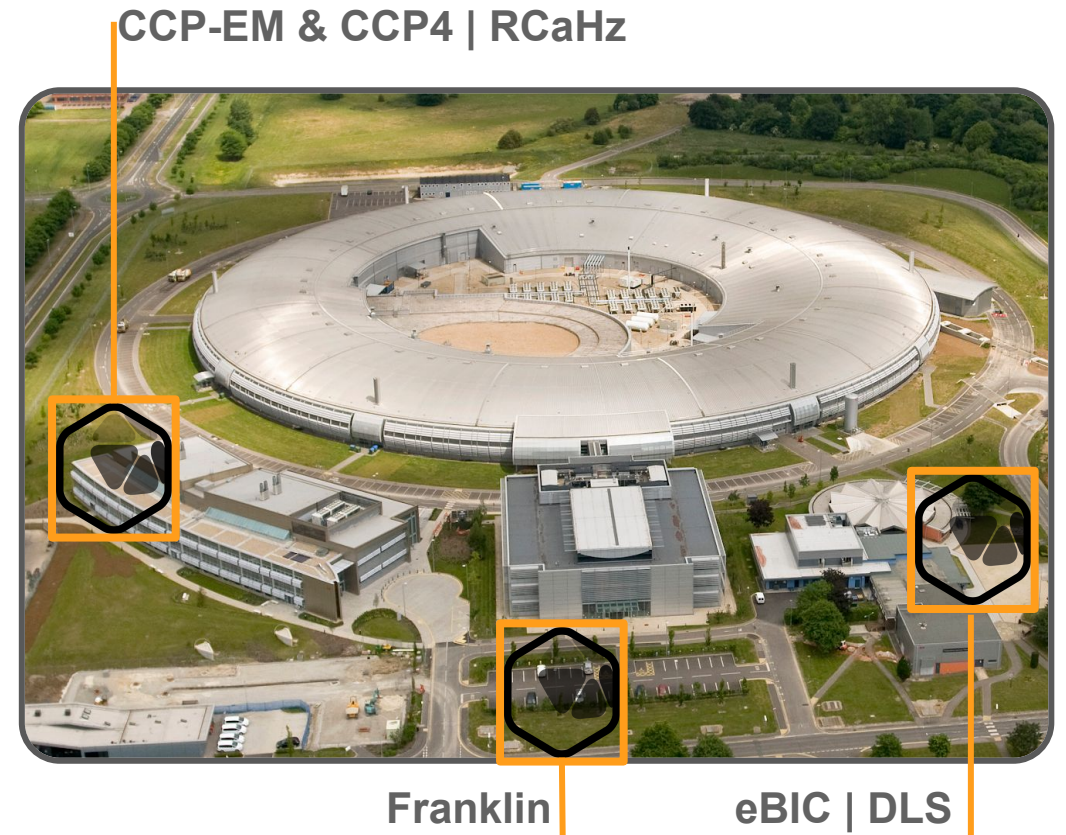
Based at STFC RAL national laboratory

EM community (>4k subscribers)

Software users (>1200 downloads, >40 industrial licences)

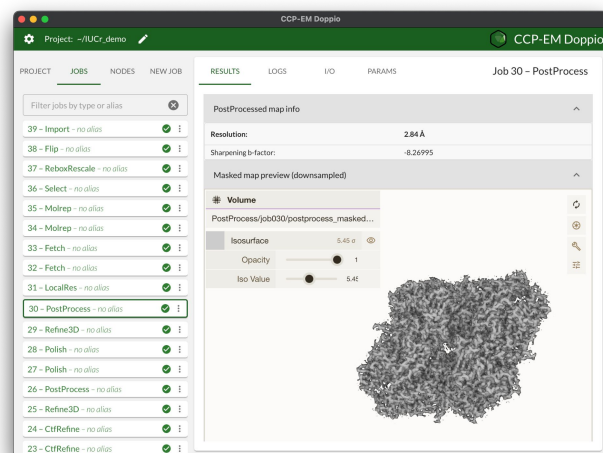
Support developers (>10 external groups)

*Core funding from MRC since 2012*





# Collaborative Computational Project



Tom  
Burnley



George  
Coldstream



Joel  
Greer



Lauren  
Giles



Matt  
Iadanza



Agnel  
Joseph



Sony  
Malhotra



Jola  
Mirecka



Colin  
Palmer



Rangana  
Warshamanage



Martyn  
Winn



Nick  
Whyatt

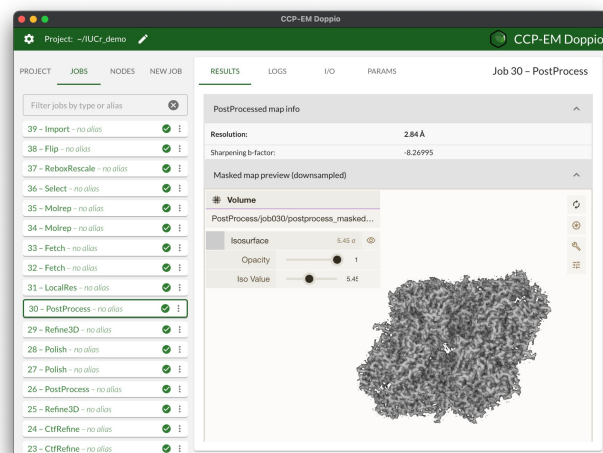


Science and  
Technology  
Facilities Council





# Collaborative Computational Project



## Executive Group

*Provides day-to-day management, sets policies, and approves new collaborations or software*



## Chair and Deputy

*Provide leadership and represent the project to external stakeholders*



## Working Group 1

*UK PIs, provide community oversight for the project and helps to set priorities*

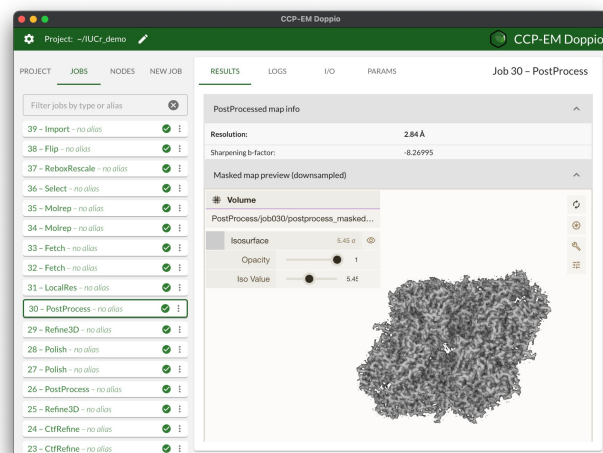


Science and  
Technology  
Facilities Council





# Collaborative Computational Project





Science and  
Technology  
Facilities Council

# CCP-EM Community

# CCP-EM Workshops & Mailing List

- Technical advice, scientific discussions, job and events announcements:

- <https://www.jiscmail.ac.uk/ccpem>

- Run and teach at many workshops:

...

- Nov 2023 CCP-EM/CCP4 Workshop, Spring8, Japan
- Jan 2024 CCP-EM & eBIC Industrial, RAL, UK
- Feb 2024 Intro to Doppio, EMBL, Germany
- Nov 2024 Icknield Workshop, RAL, UK
- June 2025 KMUTT EM Workshop, Bangkok, Thailand

...

**13-17th October 2025 Icknield Model Building Workshop**  
**RAL, Oxfordshire, UK**  
**Registration now open**



Science and  
Technology  
Facilities Council



# CCP-EM Spring Symposium

- Annual cryoEM conference
- Celebrating new method developments in cryo-EM
- YouTube lectures past 11 editions
  - [www.ccpem.ac.uk](http://www.ccpem.ac.uk)
- 11th Anniversary May 2025 Nottingham / Zoom
  - >360 in-person, >800 virtual
  - Free virtual attendance
- 12th 22-24 April 2026, Nottingham, UK
  - Bursaries available

## 2025 Talks

Alister Burt (Genentech)

Jose-Maria Carazo (Centro Nacional de Biotecnología)

Maud Dumoux (RFI)

Johannes Elferich (UMass Chan Medical School)

Andreas Engel (CryoWrite)

Arjen Jakobi (TU Delft)

Wanda Kukulski (University of Bern)

Taiana Maia de Oliveira (AstraZeneca)

Charlie Scarff (Leeds University)

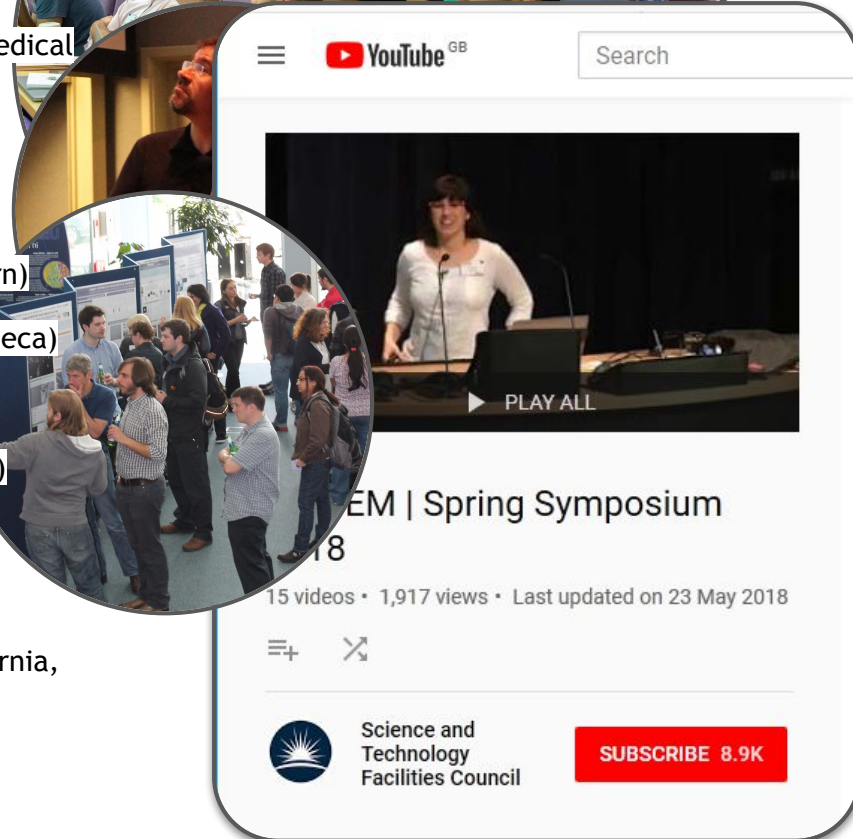
Thom Sharp (University of Bristol)

Alexander Shtyrov (MRC-LMB)

Amit Singer (Princeton)

Jessie Zhang (University of California, Berkeley)

Kai Zhang (Yale)





Science and  
Technology  
Facilities Council

# CCP-EM Software Suite

# CCP-EM software suite

30+ tasks in a common Python framework

Raw micrographs to validated atomic model

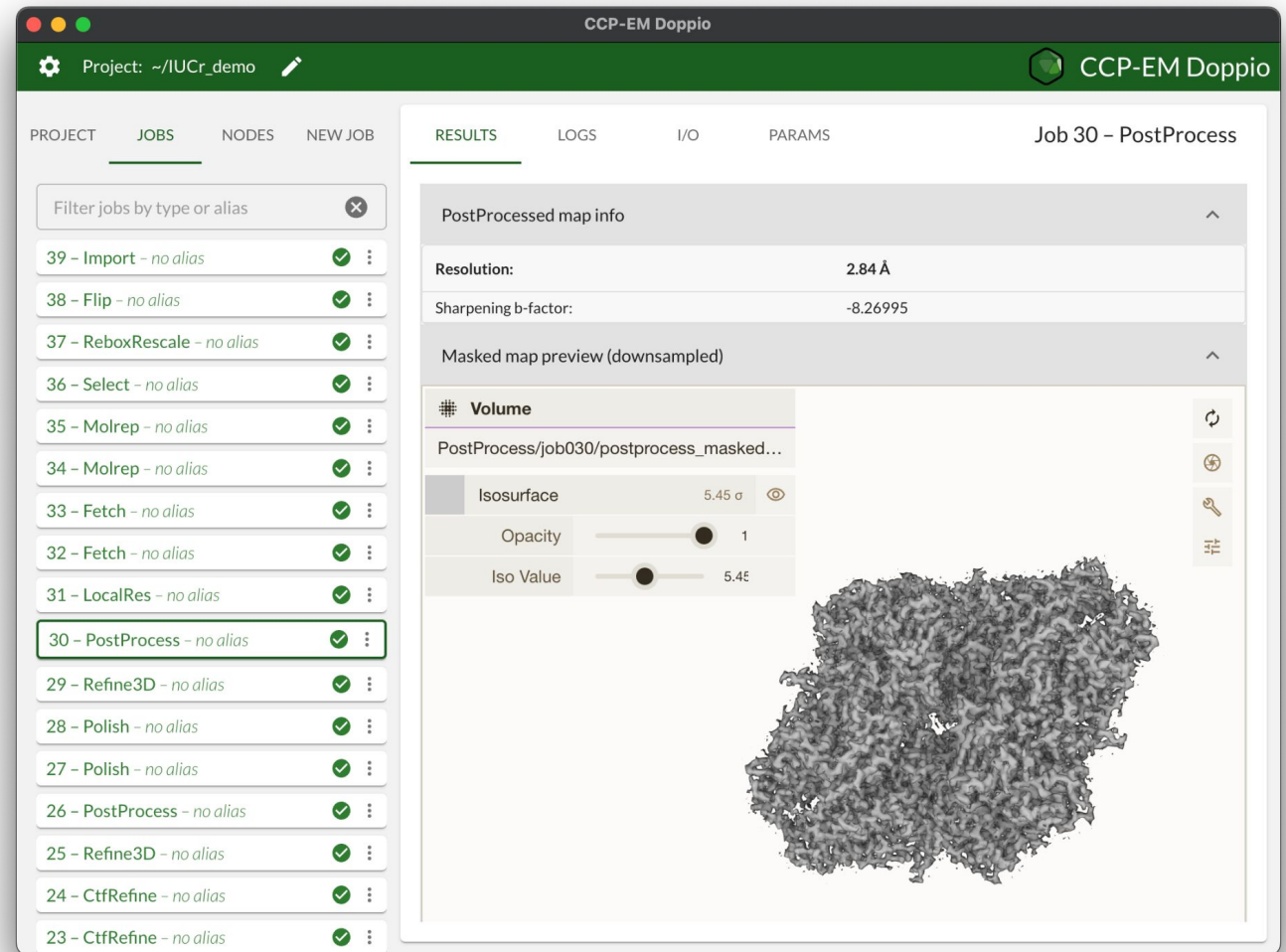
Download from [ccpem.ac.uk](http://ccpem.ac.uk)

Linux & Mac

Free for academic use, fee for commercial

Get in touch:

*[ccpem@stfc.ac.uk](mailto:ccpem@stfc.ac.uk)*





# CCP-EM workflow

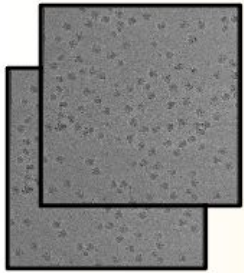
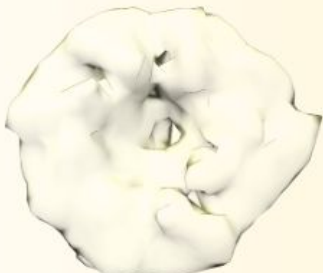
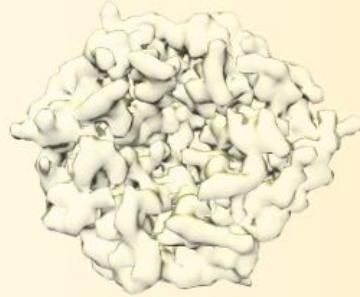


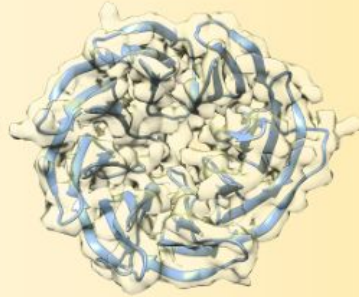
Image processing



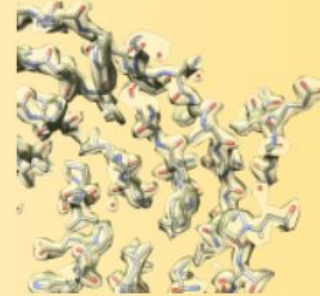
Reconstruction



Map optimization



Model building



Model refinement



Model validation

# SPA & atomic modelling workflow

## Single Particle Analysis

Relion

Topaz

CrYOLO

Icebreaker

## Map Optimisation

LocScale

Lafter

Confidence Map

## Model Building

ModelCraft

Model Angelo

Fetch AF2

## Model Docking

EM\_  
placement

MOLREP

## Automated Refinement

Refmac  
Servalvat

AceDRG

TEMPy  
-REFF

## Interactive Refinement

Coot

Moorhen

Chimera

## Map/Model Validation

CCP-EM  
Validation

Find/Check  
MySeq

CryoEF

## Model & Map Utilities

*e.g. Map  
trimming*

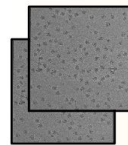
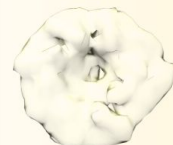
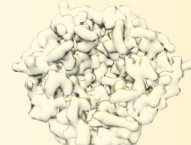


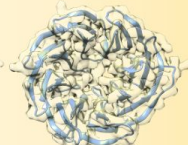
Image processing



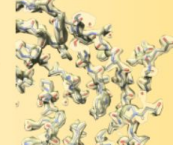
Reconstruction



Map optimization



Model building



Model refinement



Model validation

# SPA & atomic modelling workflow

## Single Particle Analysis

Relion

Topaz

CrYOLO

Icebreaker

## Map Optimisation

LocScale

Lafter

Confidence Map

## Model Building

ModelCraft

Model Angelo

Fetch AF2

## Model Docking

EM\_  
placement

MOLREP

## Automated Refinement

Refmac  
Servalvat

AceDRG

TEMPy  
-REFF

## Interactive Refinement

Coot

Moorhen

Chimera

## Map/Model Validation

CCP-EM  
Validation

Find/Check  
MySeq

CryoEF

## Model & Map Utilities

*e.g. Map  
trimming*

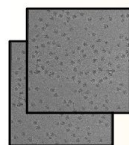
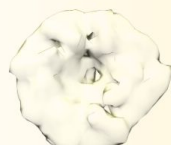
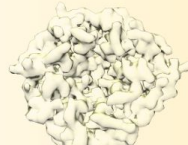


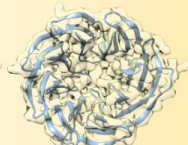
Image processing



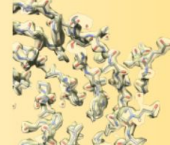
Reconstruction



Map optimization



Model building



Model refinement



Model validation



# Tutorial workflow

Single Particle  
Analysis

Map  
Optimisation

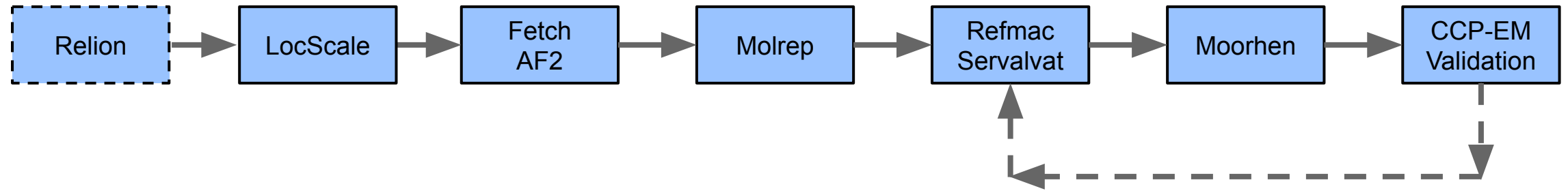
Model  
Building

Model  
Docking

Automated  
Refinement

Interactive  
Refinement

Map/Model  
Validation



*xMultiple human in the loop iterations*

CCP-EM

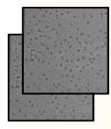
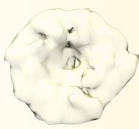
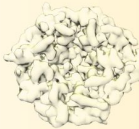


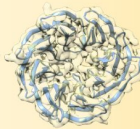
Image processing



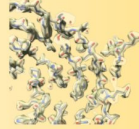
Reconstruction



Map optimization



Model building

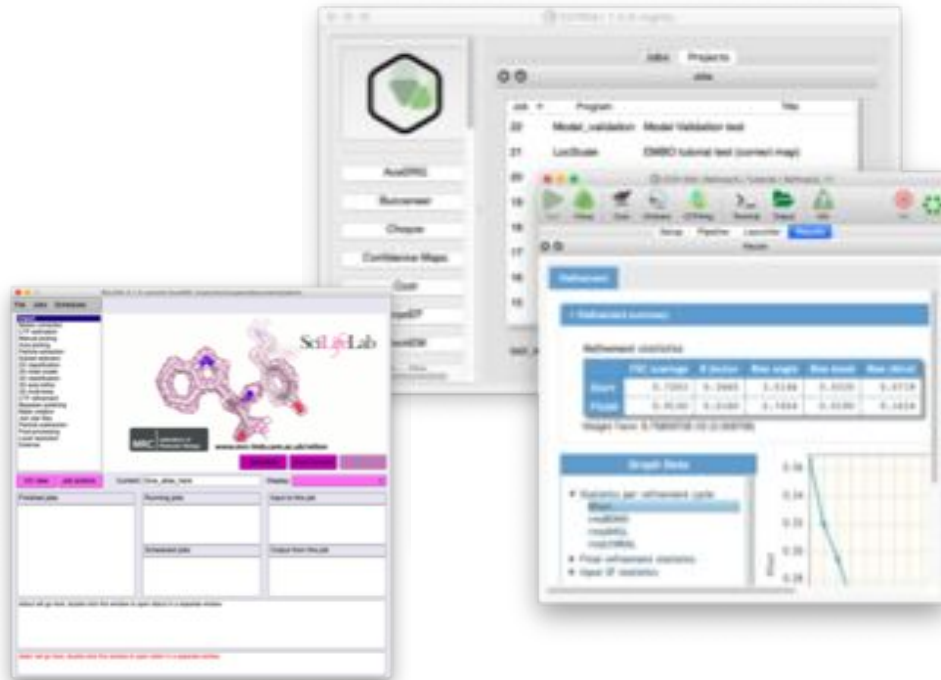


Model refinement



Model validation

# CCP-EM version 1.x



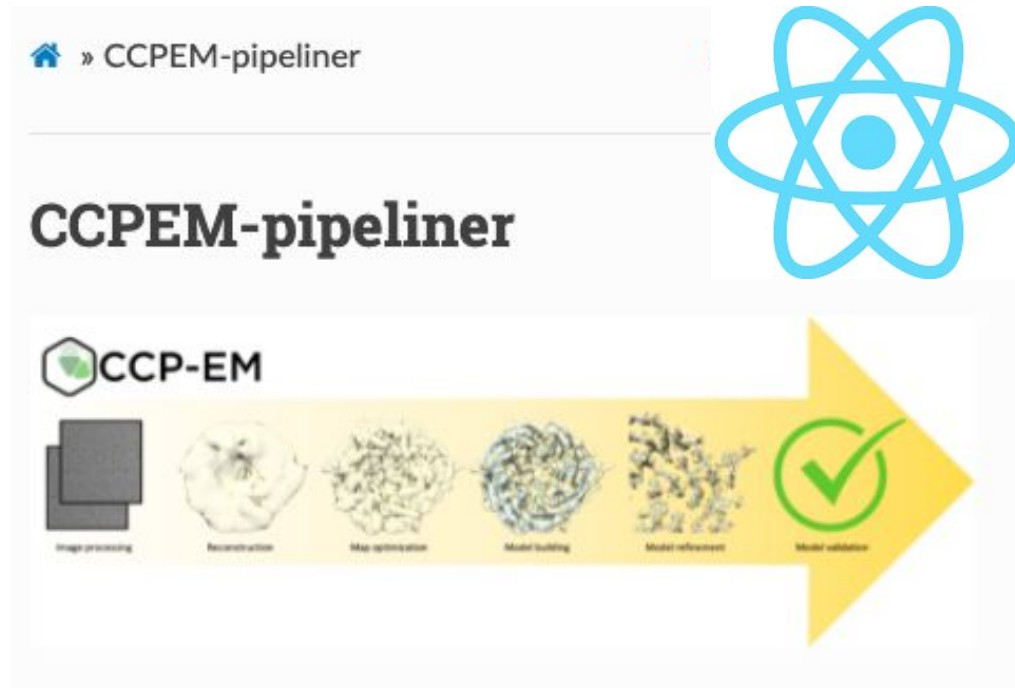
# CCP-EM

- Map post processing
- Atomic modelling
- PyQt5 GUI
- No data model
- Basic project management

**RELION**

- Single particle reconstruction
- FLTK GUI
- Data model
- DAG project management

# CCP-EM version 2.x



## *Images to structures*

- Python pipeliner python API
- JS react GUI
- Updated data model
- Project management
- Single particle
- Atomic modelling
- Tomography (coming soon)



# CCP-EM software plan

GUI



CCP-EM Doppio

Manager



CCP-EM Pipeliner

Tasks



Relion Refine



LocScale



EM  
Placement



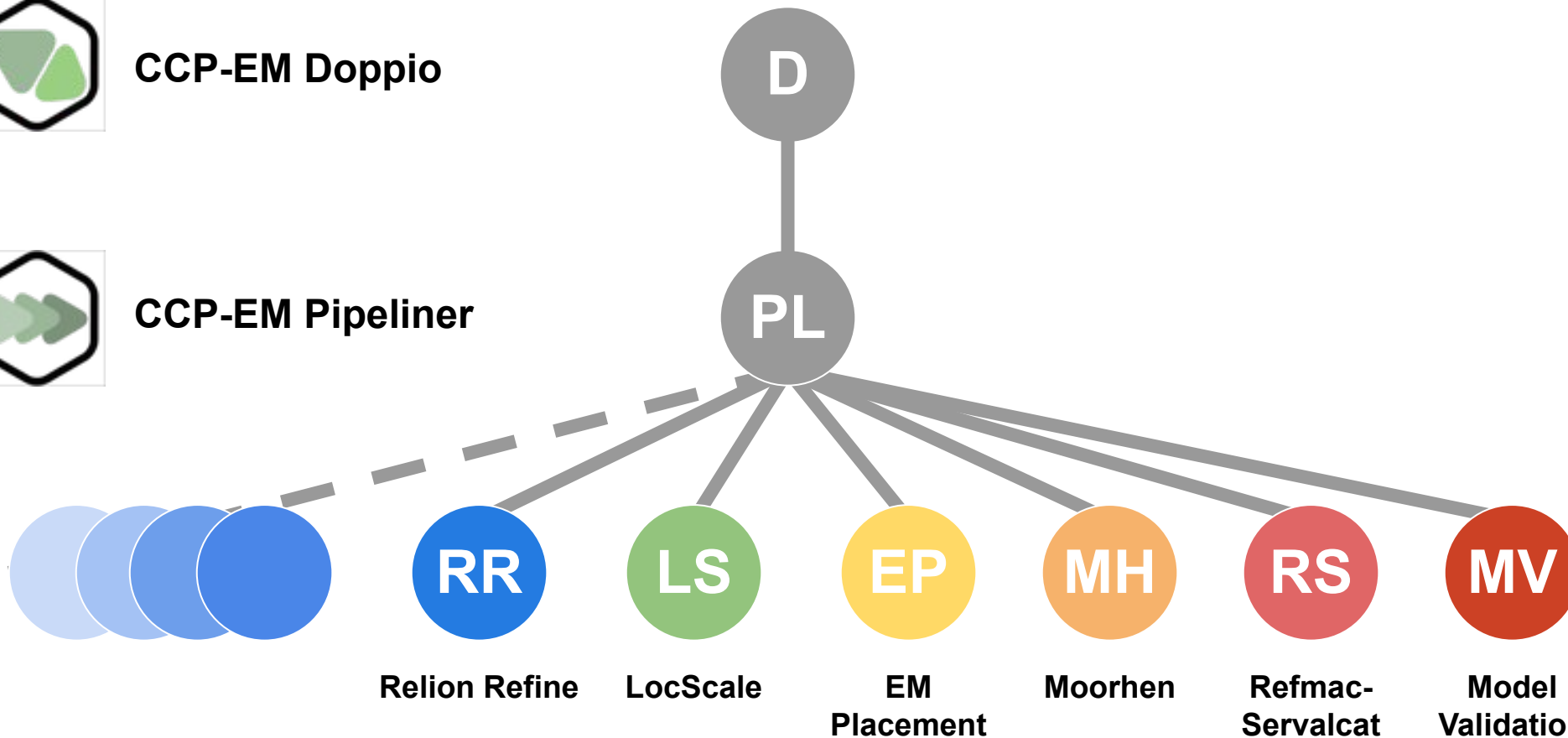
Moorhen



Refmac-  
Servalcat

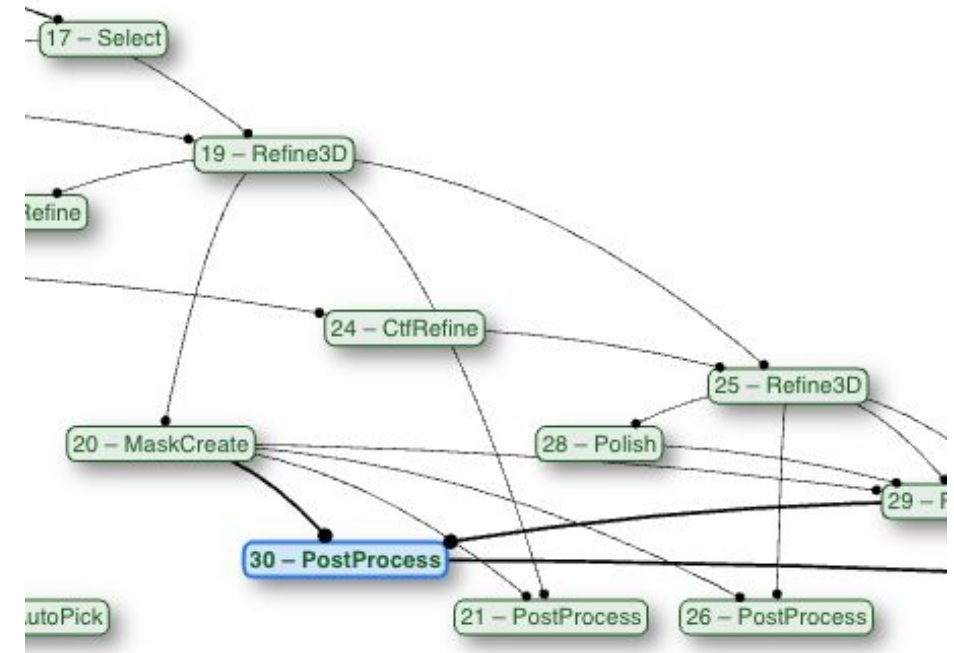
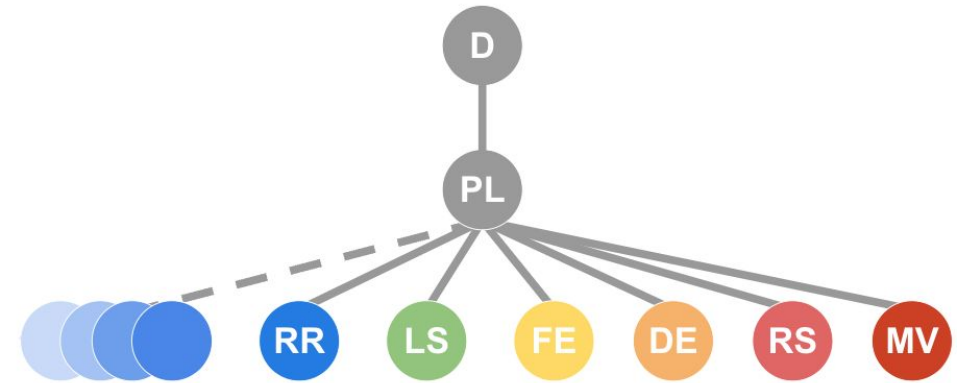


Model  
Validation



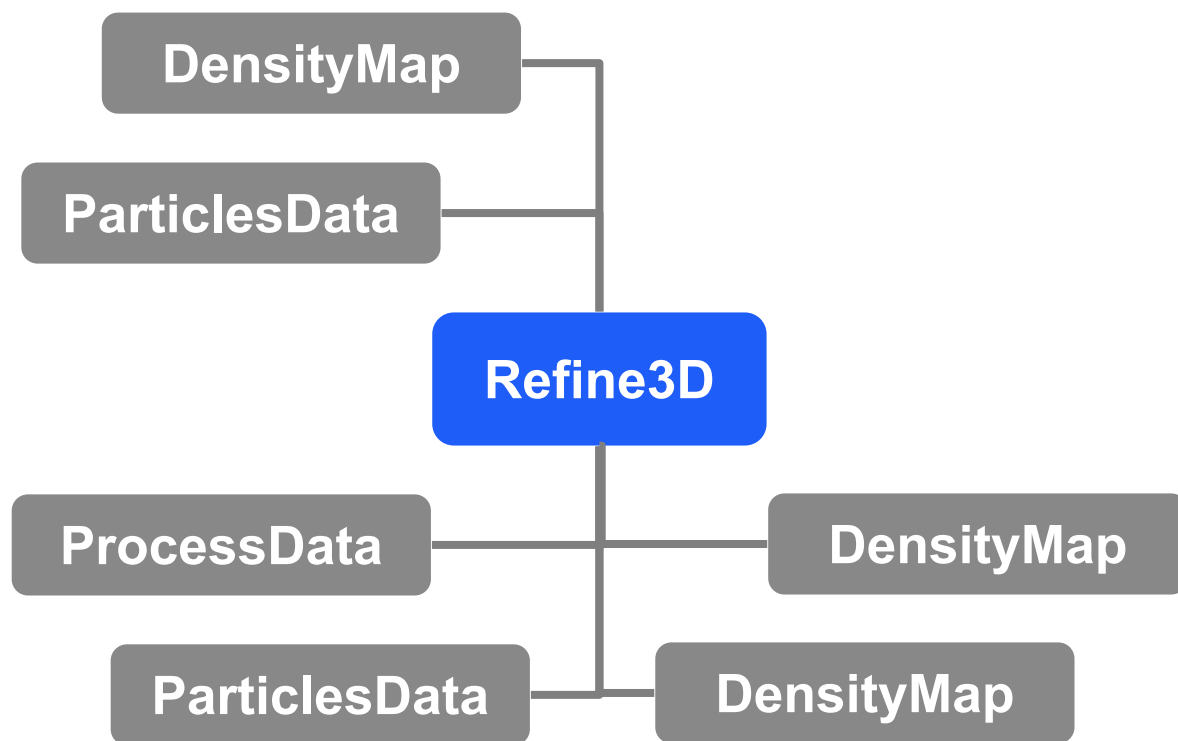
# ccpem-pipeliner

- **Business logic layer decoupled from tasks and UI**
- Python 3
- MPL 2.0 license
- <https://gitlab.com/ccpem/ccpem-pipeliner>
- Directly Acyclic Graph dataflow
  - Metadata tracking
  - Archiving
  - Jobs have input/output nodes
- Plugin architecture
  - 100-500 lines per app
  - Open - supports external apps
- CL / UI / scripting APIs



# Job Nodes

*Each job has input and output nodes & nodes are specific data types*



CCP-EM Doppio

RESULTS LOGS I/O PARAMS Job 25 – Refine3D

Open with: CHIMERA X COOT PDF VIEWER RELION DISPLAY TEXT EDITOR UCSF CHIMERA

Inputs to this job:

- ☐ CtfRefine/job024/particles\_ctf\_refine.star **ParticlesData** relion ctfrefine
- ☐ Refine3D/job019/run\_class001.mrc **DensityMap** relion refine3d

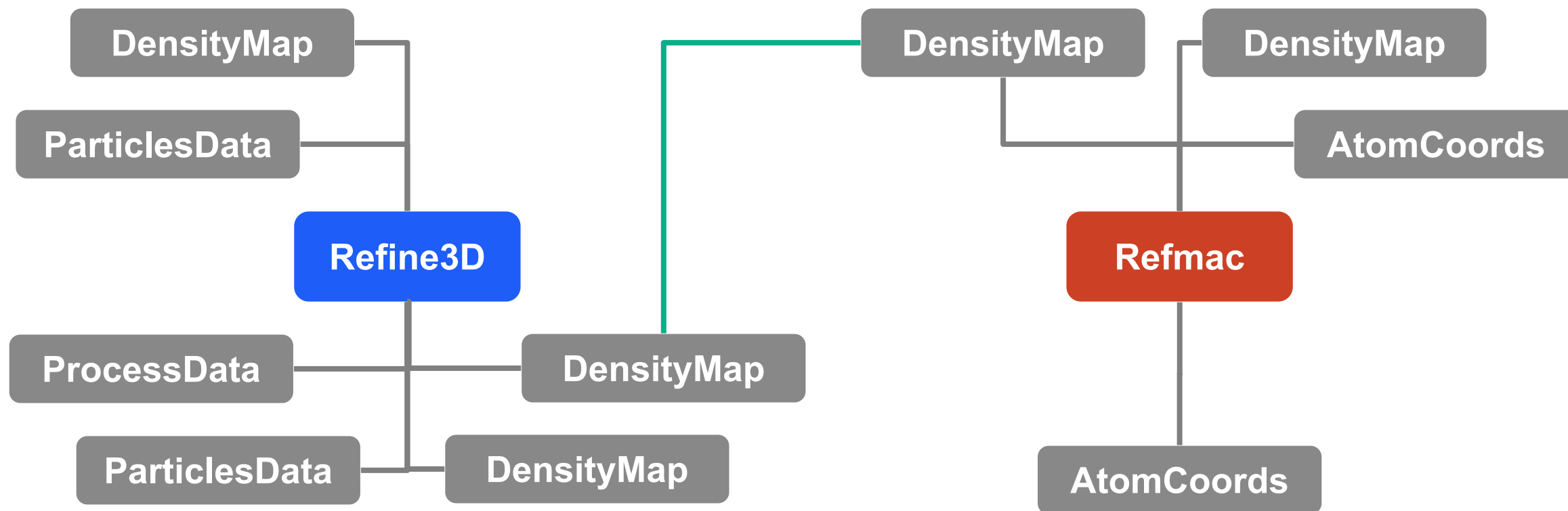
Outputs from this job:

- ☐ Refine3D/job025/run\_data.star **ParticlesData** relion refine3d
- ☐ Refine3D/job025/run\_optimiser.star **ProcessData** relion optimiser refine3d
- ☐ Refine3D/job025/run\_half1\_class001\_unfil.mrc **DensityMap** relion halfmap refine3d
- ☒ Refine3D/job025/run\_class001.mrc **DensityMap** relion refine3d



# Job Nodes

*Each job has input and output nodes & nodes are specific data types*



# Automating complex workflows



- **CLI**

```
#!/user/bin/bash
pipeliner --start_new_project
pipeliner --schedule_job Import_job.star
pipeliner --schedule_job MotionCorr_job.star
pipeliner --schedule_job CtfFind_job.star

pipeliner --run_schedule --name Schedule1 --jobs Import/job001/ job002 CtfFind/new_alias --min_between 15
--nr_repeats 3 --min_wait_before 2 -sec_wait_after 15

pipeliner --metadata_report job003
```

- **Python API**

```
from pipeliner.api.manage_project import PipelinerProject

my_project = PipelinerProject()

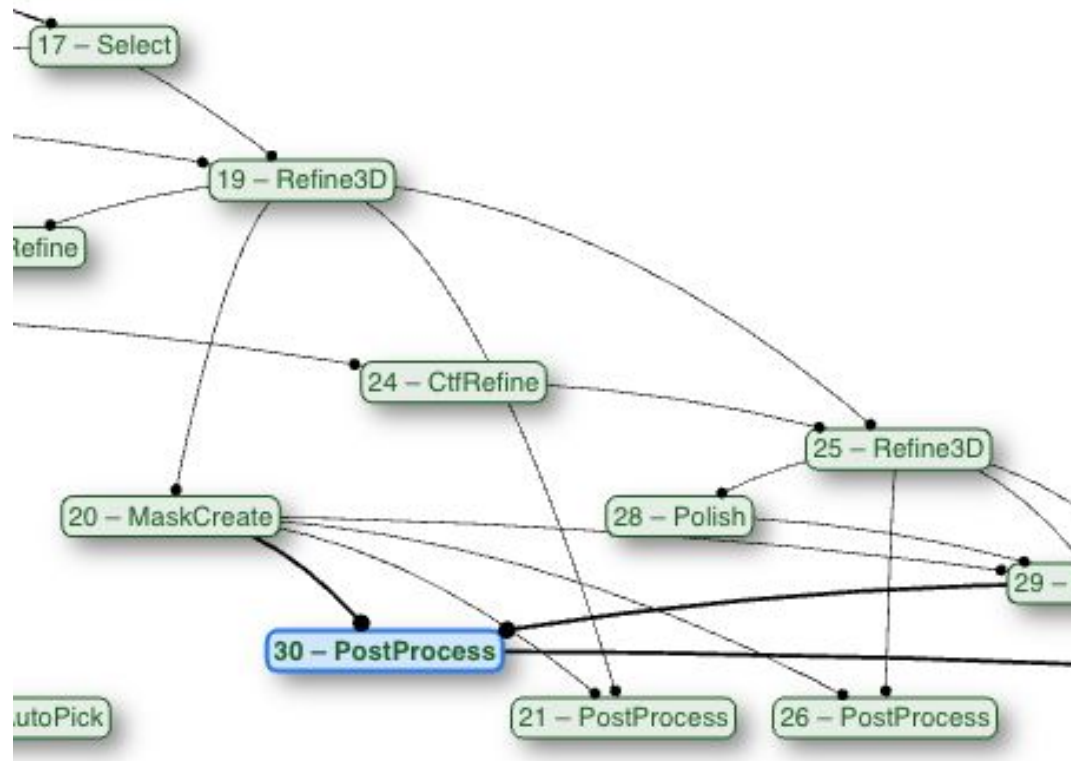
my_project.schedule_job("Import_job.star") # adds Import/job001/ to the pipeline
my_project.schedule_job("MotionCorr_job.star") # adds MotionCorr/job002/ to the pipeline
my_project.schedule_job("CtfFind_job.star") # adds AutoPick/job003/ to the pipeline

my_project.run_schedule(
    name="Schedule1",
    job_ids=["Import/job001/", "MotionCorr/job002/", "CtfFind/job003/"],
    nr_repeat=3,
    minutes_wait=15,
    minutes_wait_before=2,
    seconds_wait_after=15,
)
my_project.get_network_metadata("CtfFind/job003/", "my_metadata.json") # returns metadata
                                                                    # and all upstream jobs
```





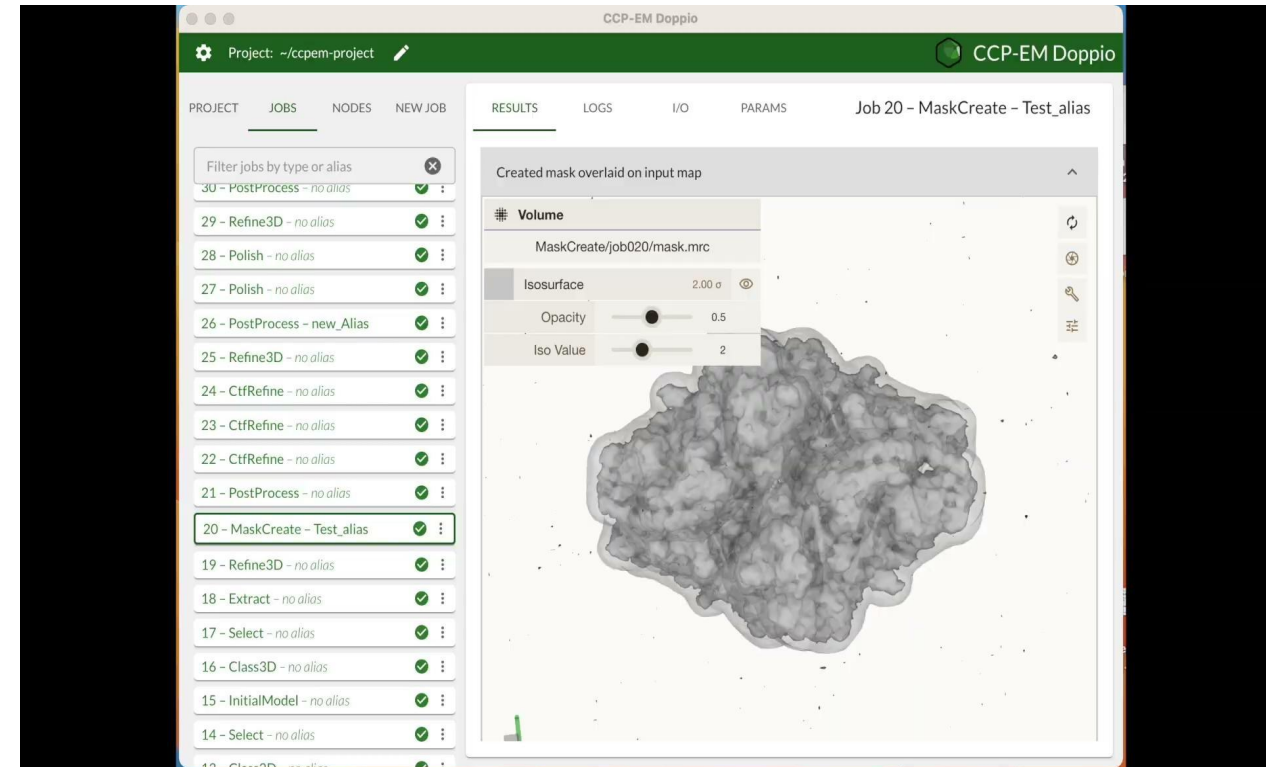
# Metadata Collection and Analysis



- Every job records the complete history of all operations
- Every job defines the metadata it returns
- Generate a metadata report for an entire workflow upstream from any job
- Generate a reference report for whole workflow - ***please cite all***

# Metadata gap

- FAIR 'findable', accessible, interoperable and re-useable
- Metadata required for efficient re-use of stored data, AI
- *Efficient metadata deposition requires automation*
- *CCP-EM Doppio, Scipion, others have metadata tools but need to finish links with facilities and repositories*
- **EMPIAR (2024)**
  - Entries with micrographs: **1679**
  - Entries with no metadata: **1578 (94%)**
  - ~15% can be reprocessed automatically
  - ~40% reprocessed with manual intervention



CCP-EM Doppio

- *Raw data ~GB-TBs | Metadata ~KB-MBs*
- *Cost not in storage but investment to produce metadata gathering, annotation and deposition pipelines*



Matt Iadanza CCP-EM

# Automated deposition

- **EMPIAR deposition**

- Launched from any job with movies, micrographs, or particle images
- Prepares upload directory
- Creates deposition data file
- Complete deposition via empiar-depositor
- Request 'EMPIAR\_depositor script' from EBI

- **EMDB deposition**

- Launch from any job with map or model
- Pre-populated deposition forms
- Implementation and testing completed
- Automated OneDep deposition available soon

- *N.B. some details e.g. sample prep will need to be provided by depositor*

## Metadata reports

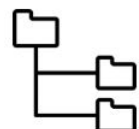
To create a new metadata report, right-click a job in the Jobs panel and select Metadata report from this job.

```
PostProcess/job030/          Contains data from 28 jobs  [trash] [full screen] [share] [up arrow]
{
  3 Items
  "CCPEM pipeline job metadata trace": {
    2 Items
    "Terminal job": "PostProcess/job030/"
    "Number of parent jobs": 27
  }
  "Jobs metadata": {
    20 Items
    "PostProcess/job030/": { ... } 2 Items
    "MaskCreate/job020/": { ... } 2 Items
    "Refine3D/job029/": { ... } 2 Items
    "Refine3D/job019/": { ... } 2 Items
    "Polish/job028/": { ... } 2 Items
    "Refine3D/job025/": { ... } 2 Items
    "Extract/job018/": { ... } 2 Items
    "Class3D/job016/": { ... } 2 Items
    "CtfRefine/job024/": { ... } 2 Items
    "CtfFind/job003/": { ... } 2 Items
    "Select/job017/": { ... } 2 Items
    "Select/job014/": { ... } 2 Items
    "InitialModel/job015/": { ... } 2 Items
    "CtfRefine/job023/": { ... } 2 Items
    "MotionCorr/job002/": { ... } 2 Items
    "Class2D/job013/": { ... } 2 Items
    "CtfRefine/job022/": { ... } 2 Items
    "Import/job001/": { ... } 2 Items
    "Extract/job012/": { ... } 2 Items
    "AutoPick/job011/": { ... } 2 Items
  }
  "Jobs network edges": [
    27 Items
    0: [ ... ] 3 Items
  ]
}
```

# Preserving workflows

## Full Archive

Preserve the project and all associated files



Project directory structure



All input files, except raw data

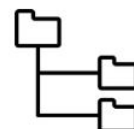


All results files



## Script archive

Preserve and repeat the workflow



Project directory structure



All input files, except raw data



Script to re-run the project





# Doppio demo



CCP-EM Doppio

Project: ~/IUCr\_demo

CCP-EM Doppio

PROJECT JOBS NODES NEW JOB

Filter jobs by type or alias

- 39 - Import - no alias ✓
- 38 - Flip - no alias ✓
- 37 - ReboxRescale - no alias ✓
- 36 - Select - no alias ✓
- 35 - Molrep - no alias ✓
- 34 - Molrep - no alias ✓
- 33 - Fetch - no alias ✓
- 32 - Fetch - no alias ✓
- 31 - LocalRes - no alias ✓
- 30 - PostProcess - no alias ✓**
- 29 - Refine3D - no alias ✓
- 28 - Polish - no alias ✓
- 27 - Polish - no alias ✓
- 26 - PostProcess - no alias ✓
- 25 - Refine3D - no alias ✓
- 24 - CtfRefine - no alias ✓
- 23 - CtfRefine - no alias ✓

RESULTS LOGS I/O PARAMS Job 30 - PostProcess

PostProcessed map info

Resolution: 2.84 Å

Sharpening b-factor: -8.26995

Masked map preview (downsampled)

Volume

PostProcess/job030/postprocess\_masked...

Isosurface 5.45  $\sigma$

Opacity 1

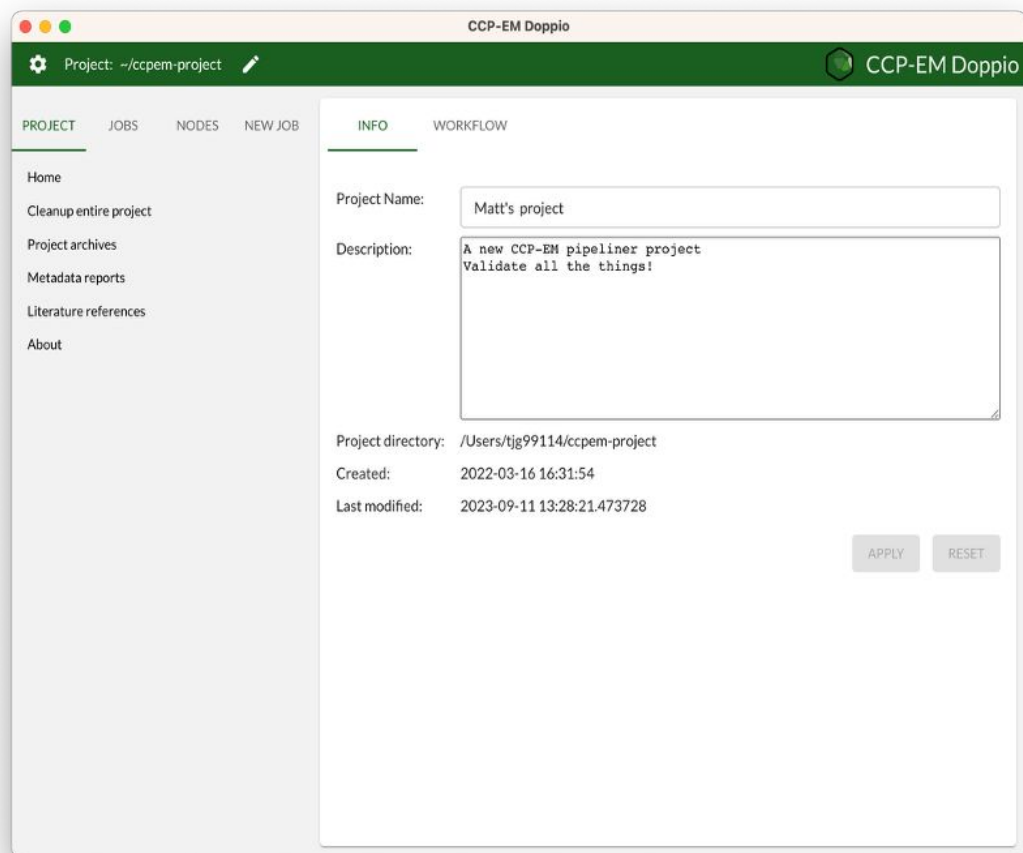
Iso Value 5.45



Science and  
Technology  
Facilities Council

# Doppio modes: app & web

## Doppio app

A screenshot of the CCP-EM Doppio application running as a desktop app. The window has a title bar with standard macOS window controls. The interface includes a top navigation bar with a settings icon, the project path "/ccpem-project", and the application name. A sidebar on the left contains links: Home, Cleanup entire project, Project archives, Metadata reports, Literature references, and About. The main content area is divided into two tabs: "INFO" (selected) and "WORKFLOW". Under the "INFO" tab, there are input fields for "Project Name" (containing "Matt's project") and "Description" (containing "A new CCP-EM pipeliner project" and "Validate all the things!"). Below these are read-only fields for "Project directory", "Created", and "Last modified". At the bottom right of the main area are "APPLY" and "RESET" buttons.

CCP-EM Doppio

Project: /ccpem-project

PROJECT JOBS NODES NEW JOB

Home

Cleanup entire project

Project archives

Metadata reports

Literature references

About

INFO WORKFLOW

Project Name: Matt's project

Description: A new CCP-EM pipeliner project  
Validate all the things!

Project directory: /Users/tjg99114/ccpem-project

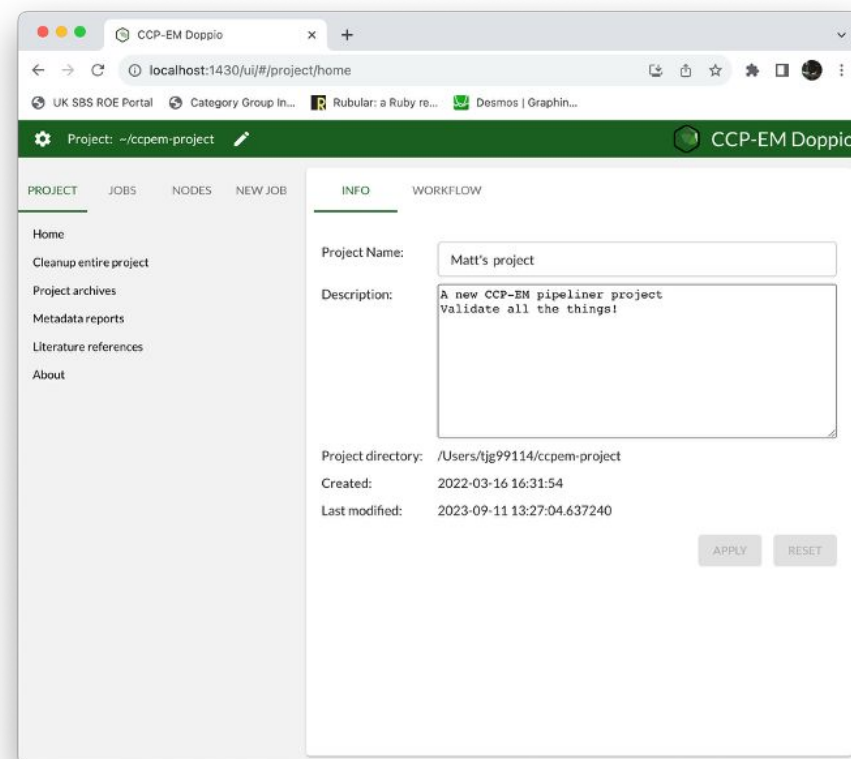
Created: 2022-03-16 16:31:54

Last modified: 2023-09-11 13:28:21.473728

APPLY RESET

## doppio-web

```
$ ./doppio-web
INFO: Started server process [2142]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://127.0.0.1:1430
```

A screenshot of the CCP-EM Doppio web interface accessed via a browser. The browser window shows the URL "localhost:1430/ui/#/project/home". The interface is identical to the desktop app version, with a top navigation bar, a sidebar, and a main content area with "INFO" and "WORKFLOW" tabs. The "INFO" tab is active, showing the same project details as the desktop app.

CCP-EM Doppio

Project: /ccpem-project

PROJECT JOBS NODES NEW JOB

Home

Cleanup entire project

Project archives

Metadata reports

Literature references

About

INFO WORKFLOW

Project Name: Matt's project

Description: A new CCP-EM pipeliner project  
Validate all the things!

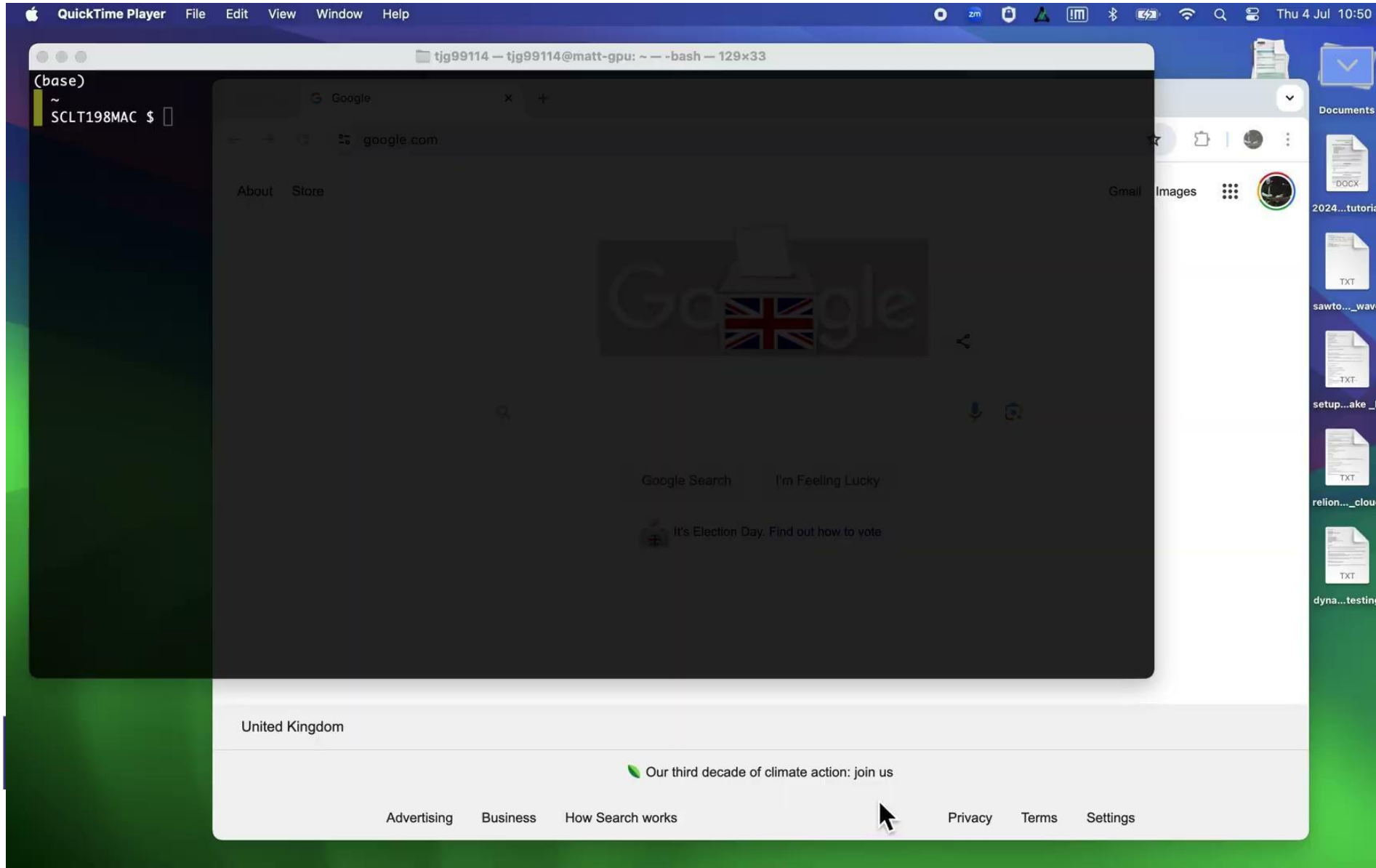
Project directory: /Users/tjg99114/ccpem-project

Created: 2022-03-16 16:31:54

Last modified: 2023-09-11 13:27:04.637240

APPLY RESET

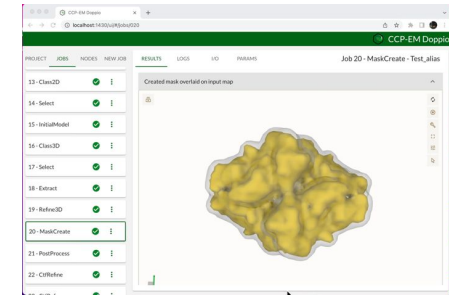
# Doppio modes: web & remote



# Doppio modes



Local

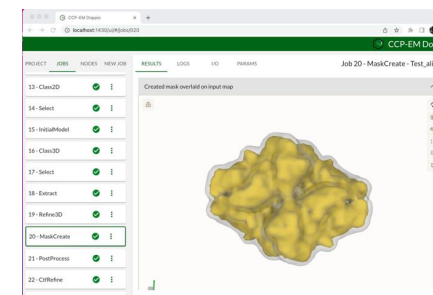




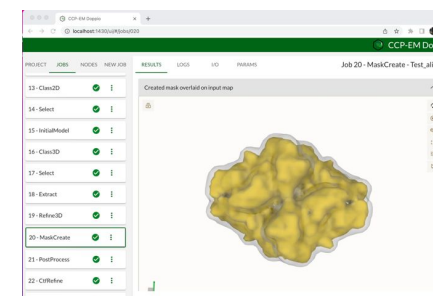
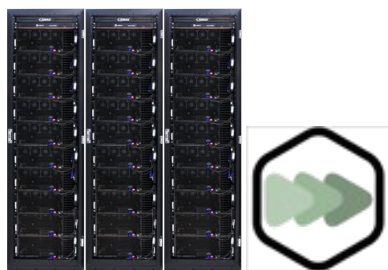
# Doppio modes



Local



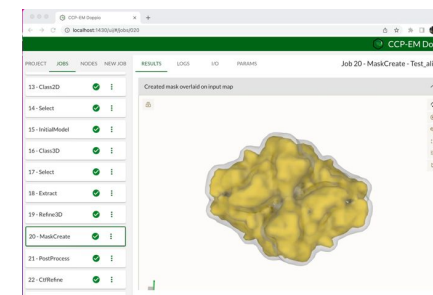
Remote



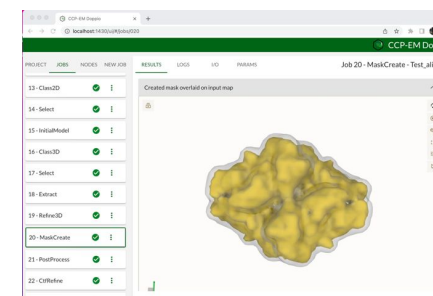
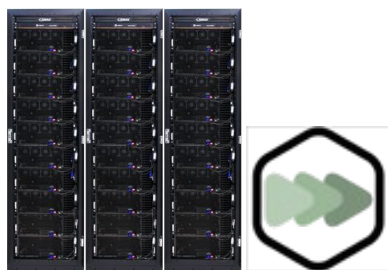
# Doppio modes



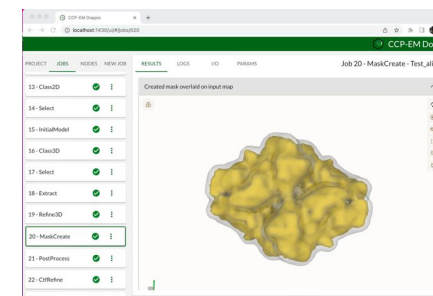
Local



Remote



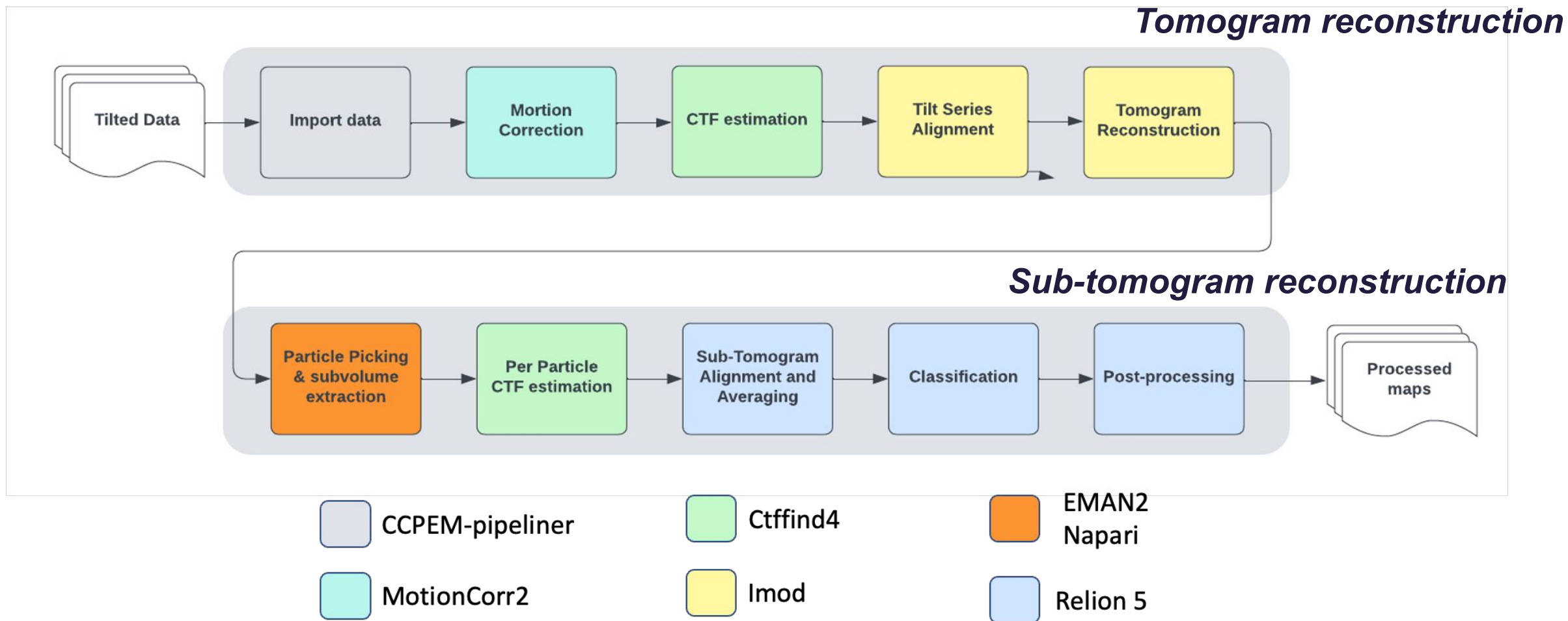
STFC Cloud



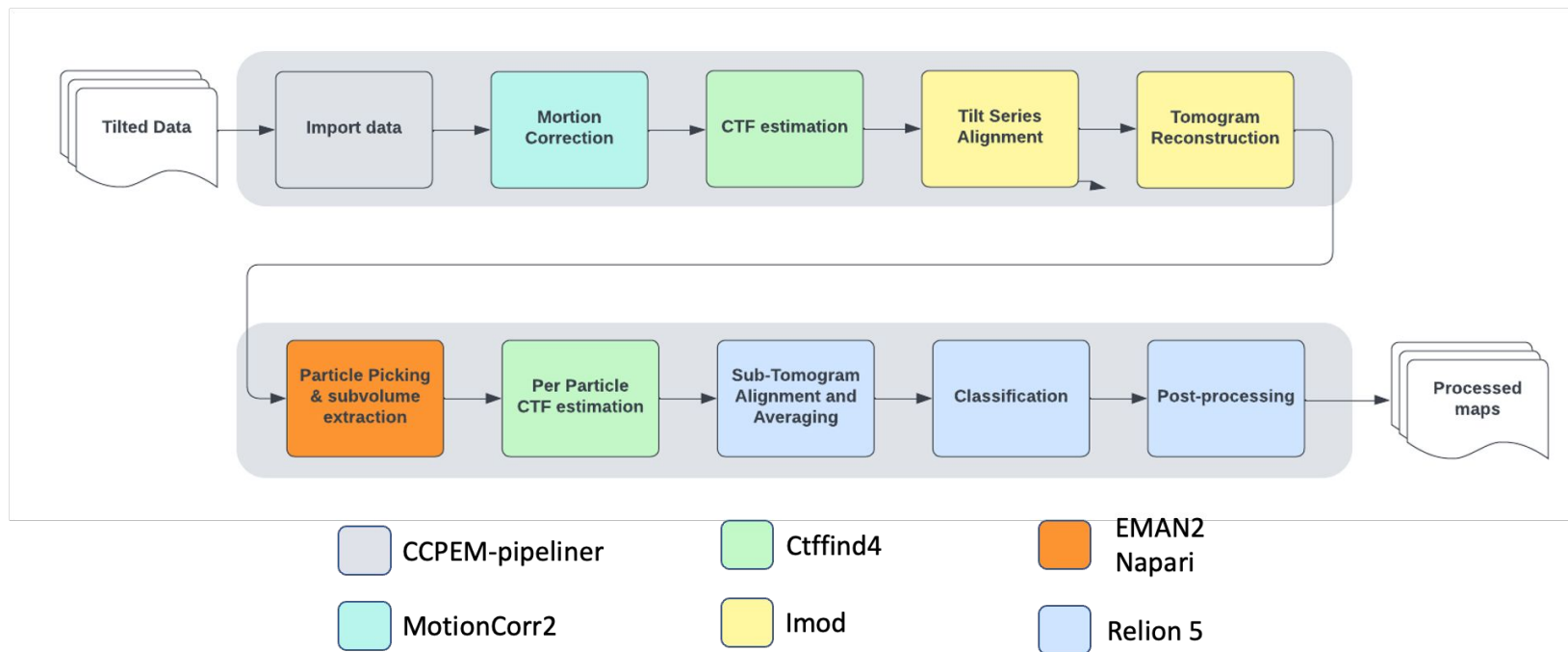
Science and  
Technology  
Facilities Council



# Doppio: tomography and STA



# Doppio: subtomogram averaging



**Relion 5 workflow**

**Metadata flows through all stages** (centralized handling)

**Remote-friendly:** Port forwarding (not Napari, some IMOD visualisation)

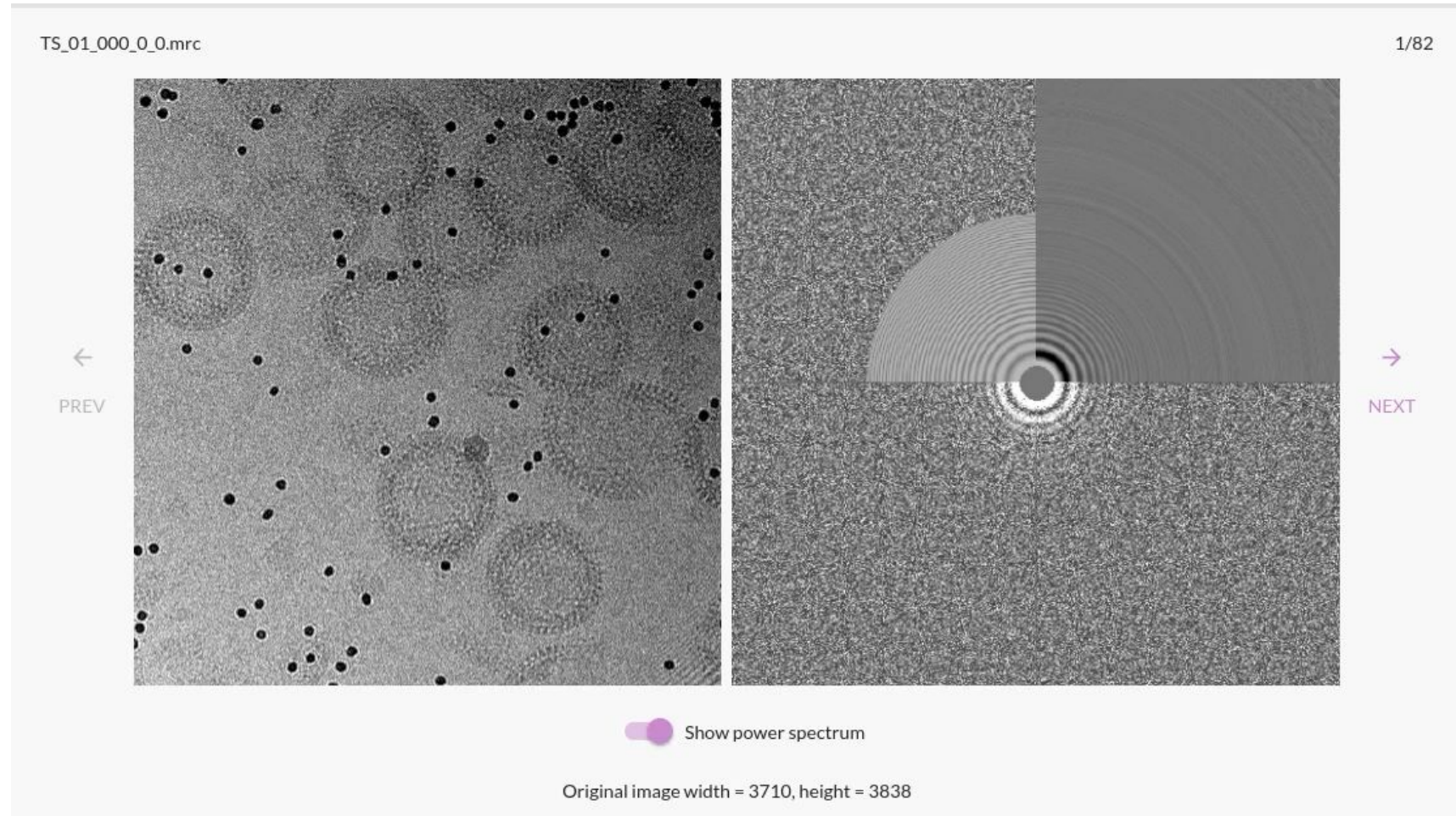


# Doppio: tomography and STA

## Intermediate visualization matters

## CTF Estimation

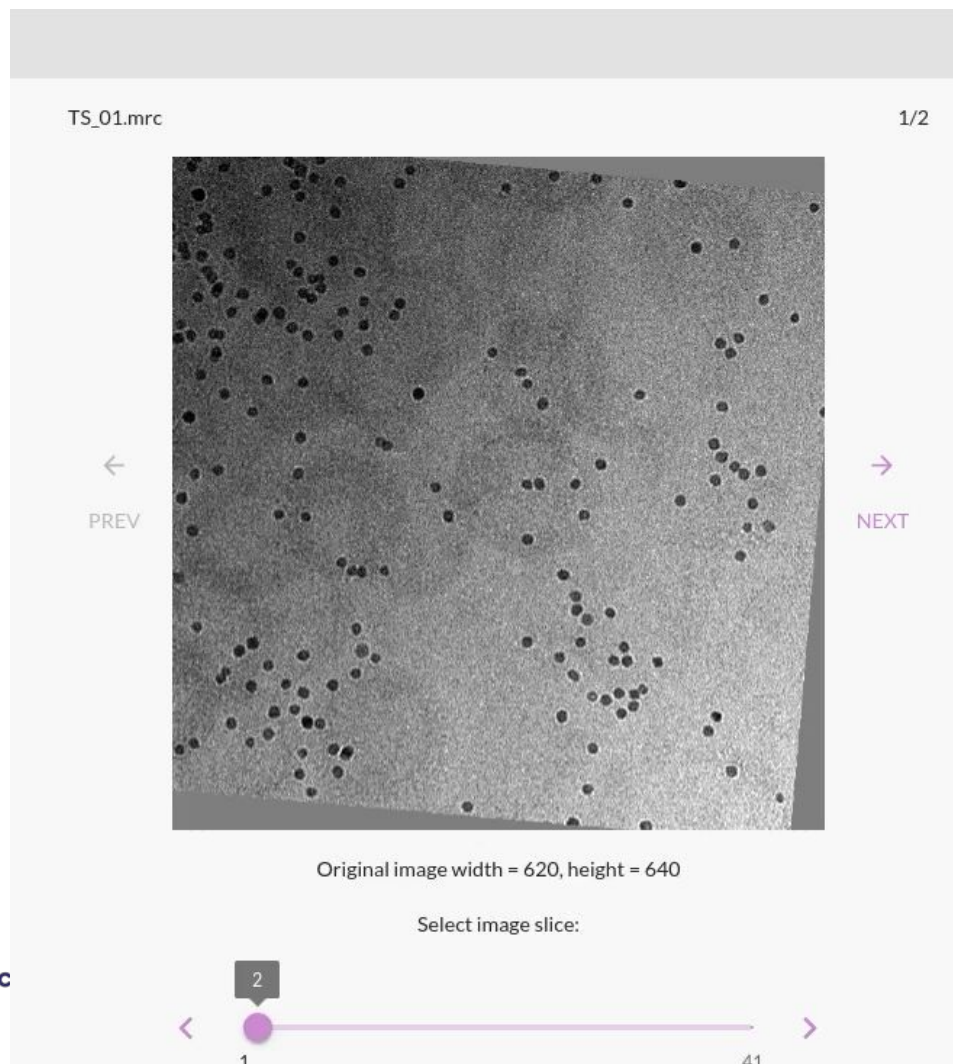
### Verify contrast transfer fit



# Doppio: tomography and STA

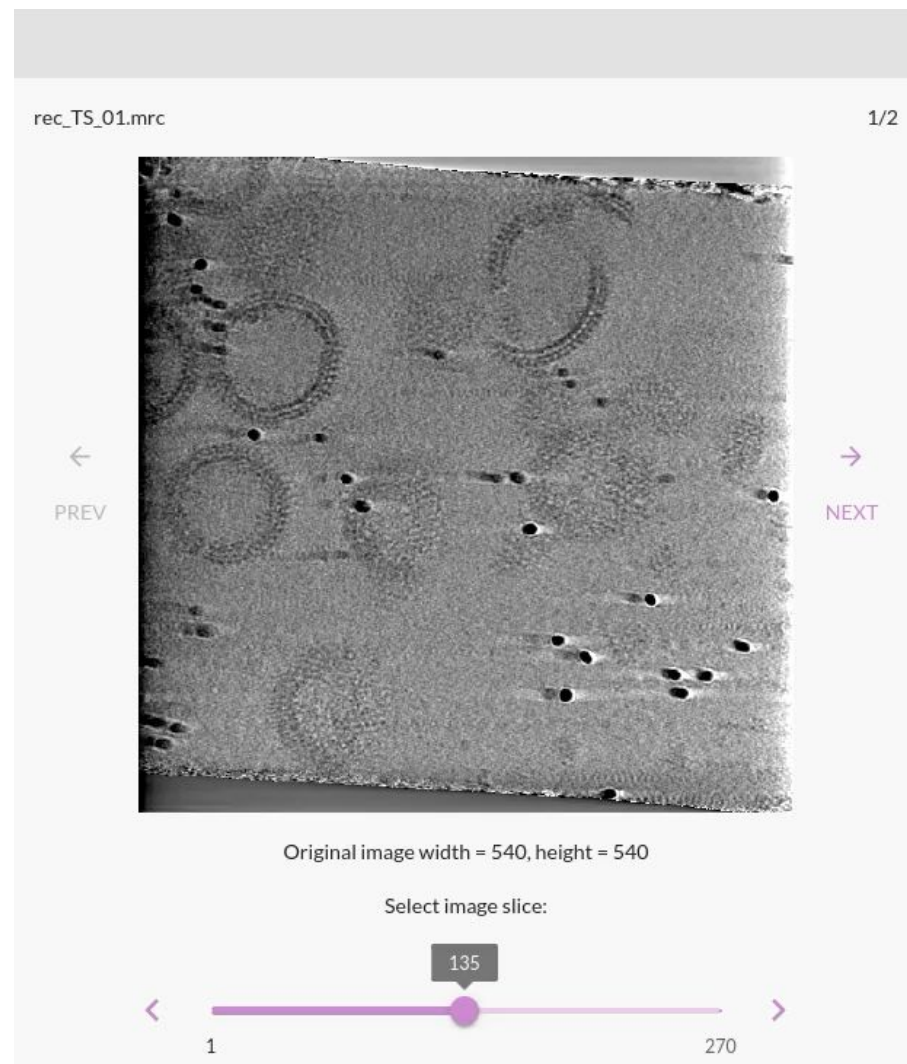
## Alignment

*Inspect shifts, angles, and quality of tilt series alignment*



## Tomogram Inspection

*Evaluate reconstruction quality and look for artifacts*



# Doppio: heterogeneous reconstruction

Many heterogeneous reconstruction algorithms have now been developed

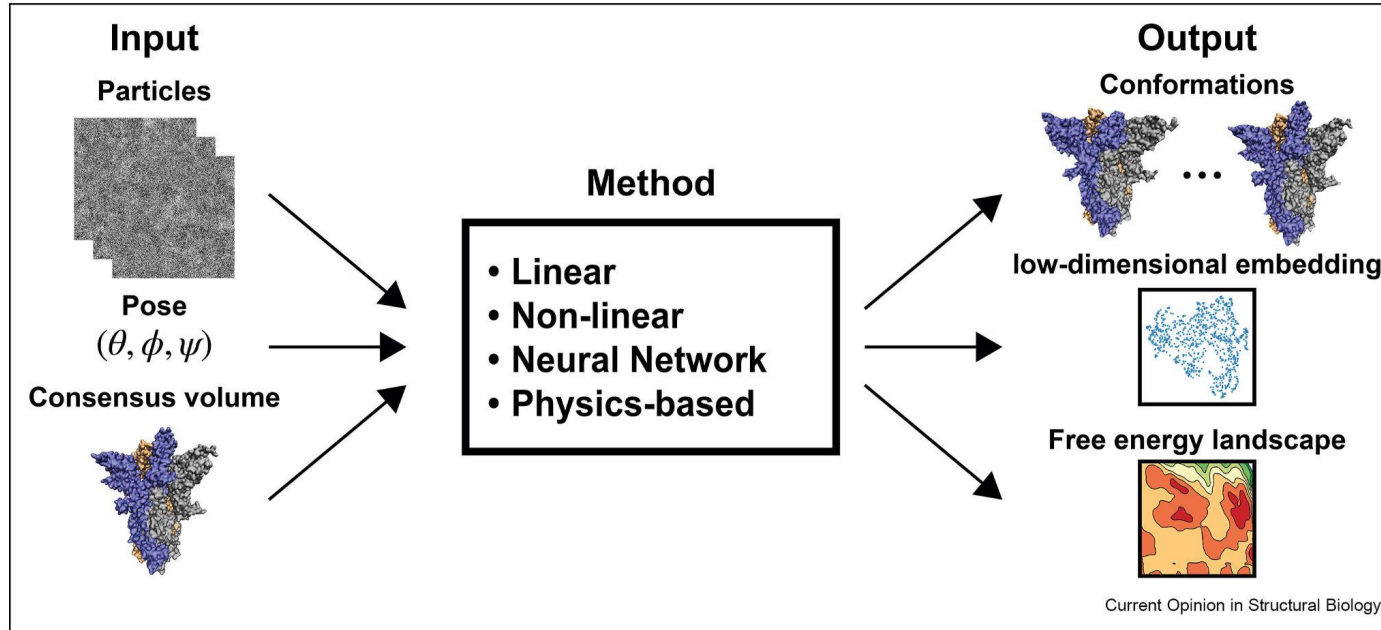


Fig 2, Tang et al., Current Opinion in Structural Biology, 2023

Benchmark datasets and common metrics are useful:

- To evaluate the range of heterogeneity which HRAs are capable of faithfully interpreting and reconstructing
- Each method would otherwise require benchmarking “by eye” from an expert



Conformational ensembles by cryoEM:

Where are we? (my personal view)

José-María Carazo  
CNB-CSIC, Madrid  
Instruct-ES, INB-ELIXIR

Cryo electron microscopy and conformational ensembles - Jose-Maria Carazo (CSIC)

Symposium talks:

Amit Singer ('25)  
Jose-Maria Carazo ('25)  
Sjors Scheres ('24)

RECOVAR: CRYO-EM HETEROGENEITY ANALYSIS USING REGULARIZED COVARIANCE ESTIMATION AND KERNEL REGRESSION

[Gilles & Singer, PNAS, 2025]

Amit Singer  
Princeton University  
Department of Mathematics and PACM  
CCPEM Symposium - 24/04/2025

Github: [ma-gilles/recover](https://github.com/ma-gilles/recover)

RECOVAR - Amit Singer (Princeton)

CCP-EM 150... Subscribed

0

Share

DynaMight:

"modelling structural **Dynamics** that **Might** improve your map"

RELION 5 - Sjors Scheres (MRC-LMB) | CCP-EM Spring Symposium 2024

# HRA jobs in Doppio

CryoDRGN (Ellen Zhong - Princeton):

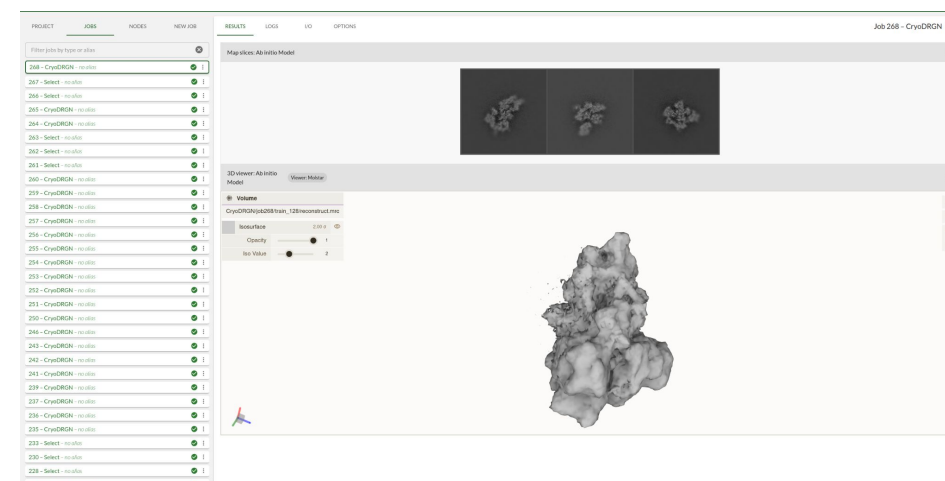
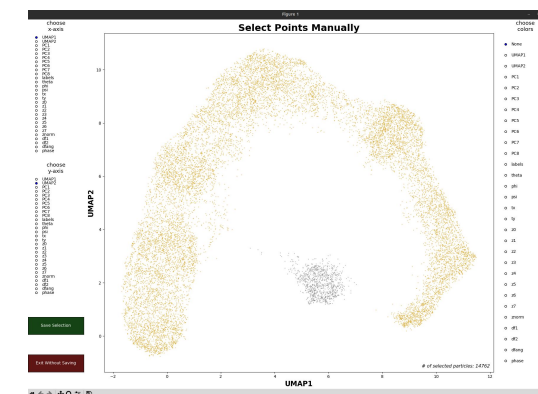
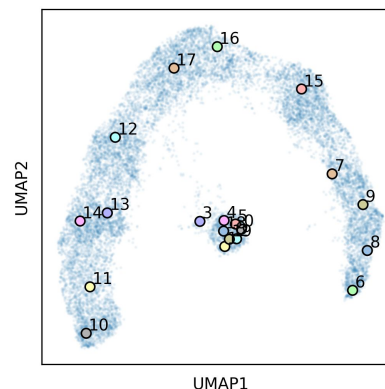
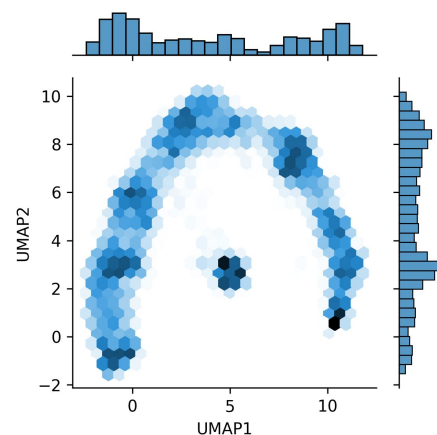
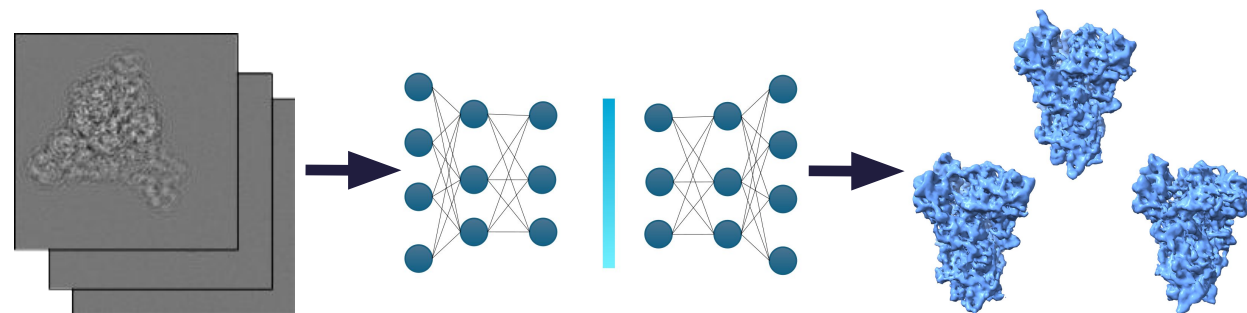
- CryoDRGN 1 – heterogeneous reconstruction from refined particles
- CryoDRGN 2 – *ab initio* homogeneous and heterogeneous reconstruction
- Interactive subset selection
- Neural network based

DynaMight (Johannes Schwab, Sjors Scheres - MRC-LMB)

- Neural network based
- Built-in error estimation

Adding RECOVER (Amit Singer - Princeton) is WIP:

- Performs consistently well across CryoBench benchmark datasets
- Linear approach



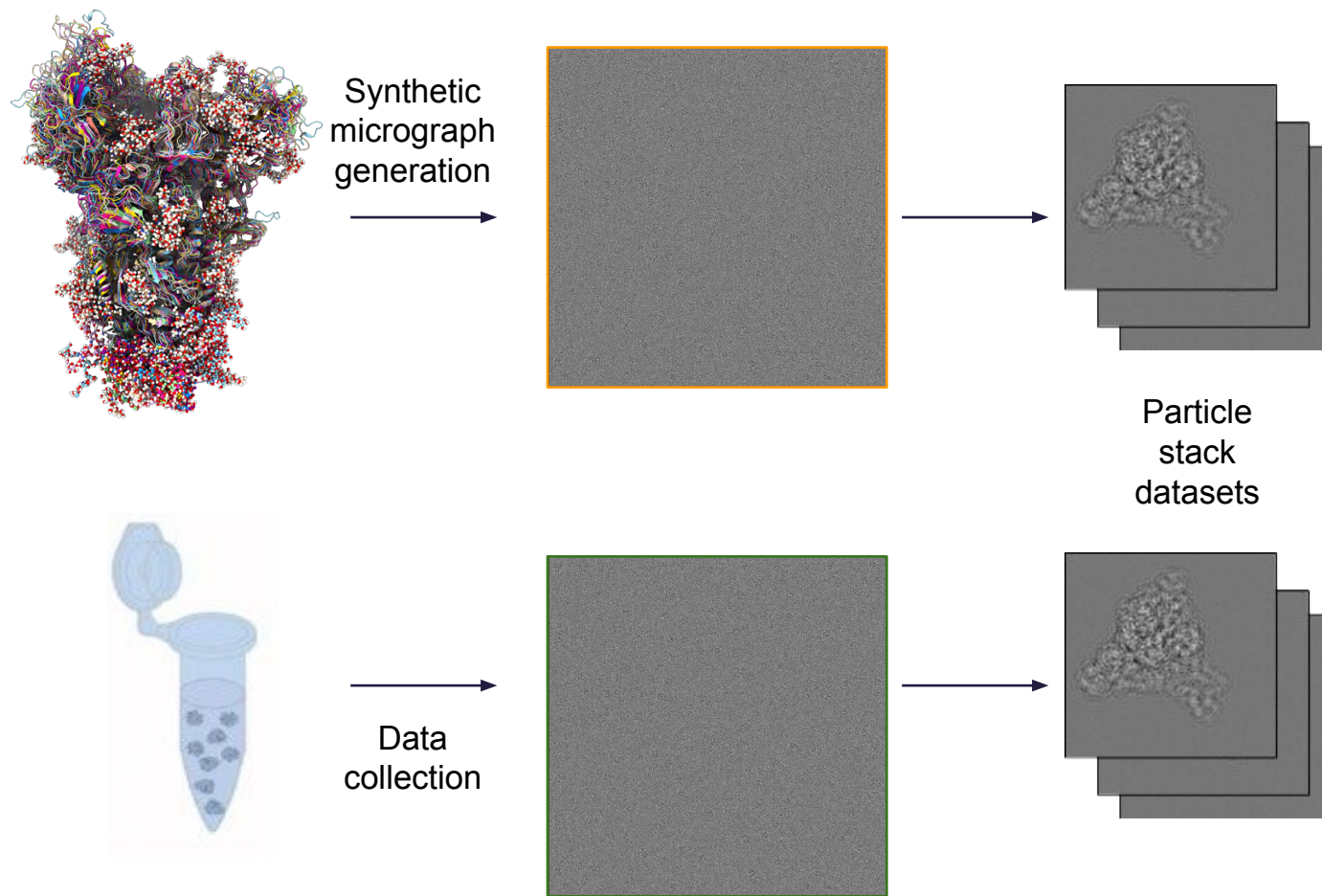


# Heterogeneity Challenge

We are going to run a CASP-inspired  
**heterogeneity community challenge**

- Assess how faithfully heterogeneity present in cryoEM images is recovered
- Facilitate the development and evaluation of **heterogeneous reconstruction algorithms (HRAs)**
- Drive the development of metrics

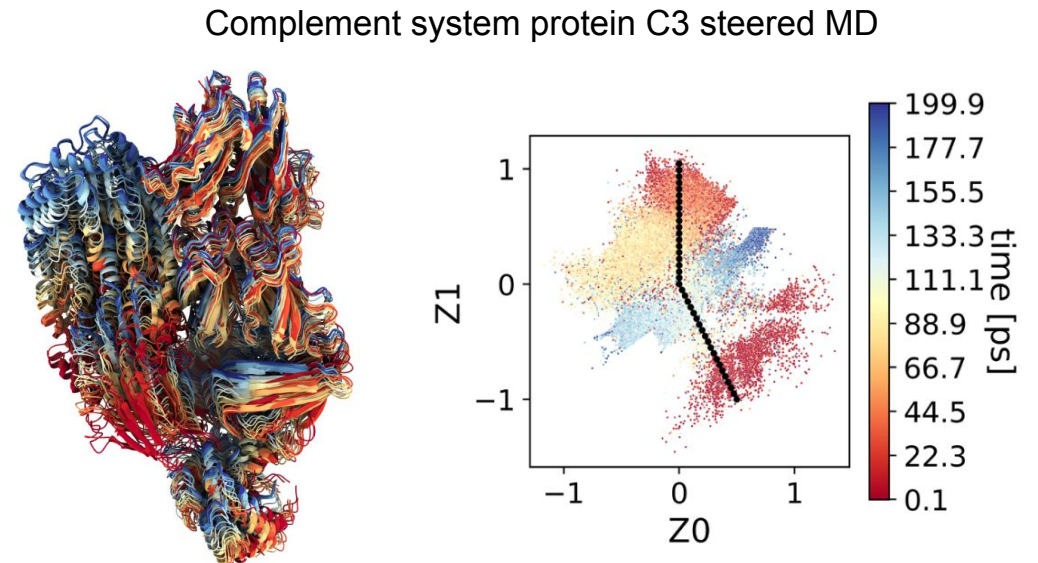
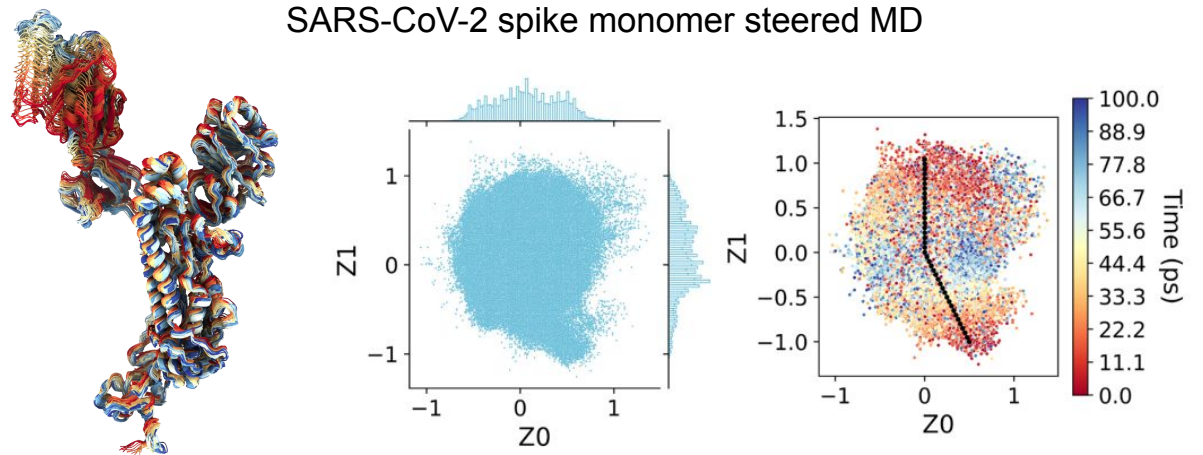
**We will release a set of experimental and synthetic ground-truth datasets**



# Recovering Heterogeneity Present in Cryo-EM Datasets

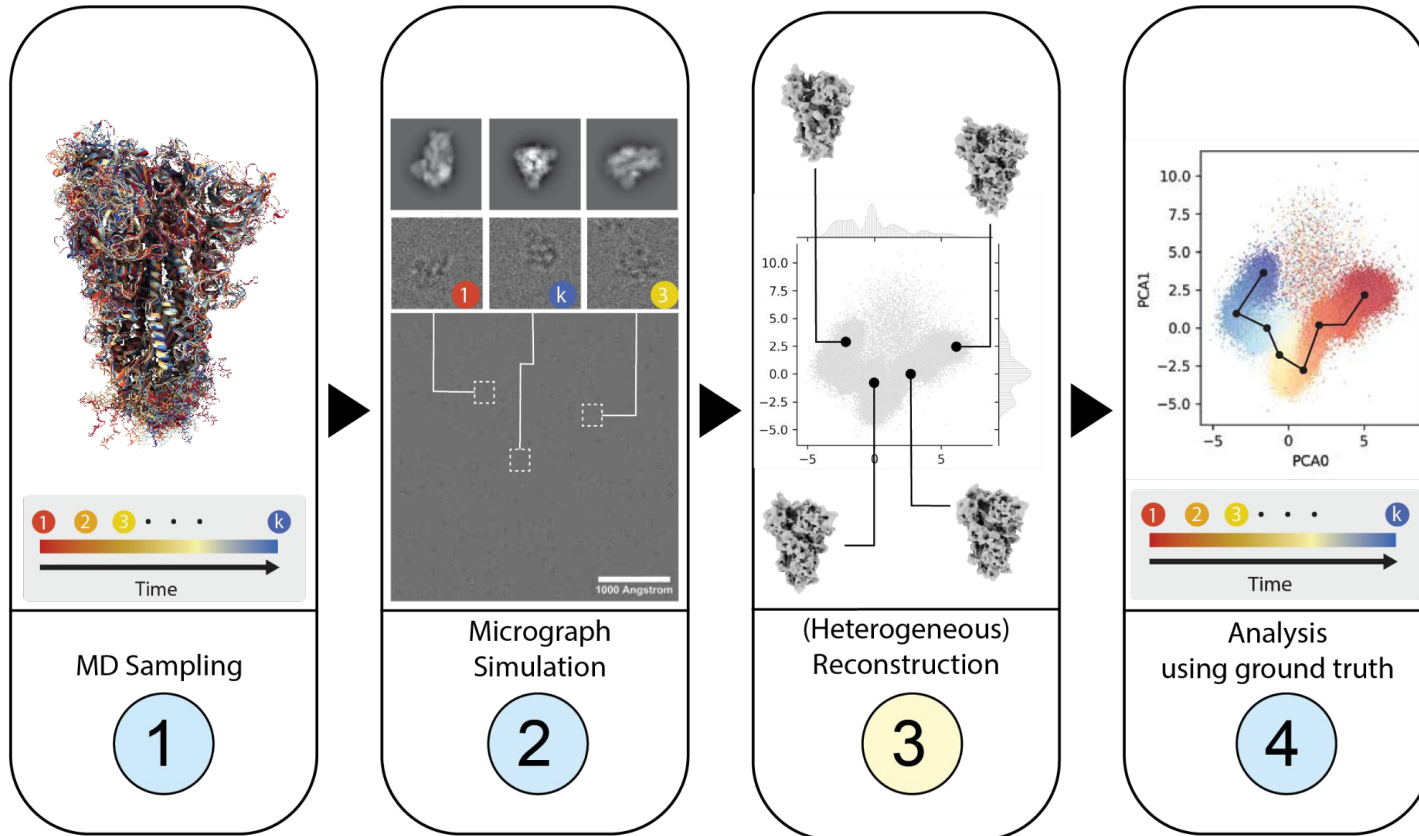
Interpreting the results from HRAs can be challenging:

- Need to be cautious when interpreting latent space representations/embeddings
- Non-linearity can obscure complex heterogeneity and validation should be performed to avoid hallucinations
- Reproducibility can be poor
- Difficult to define metrics and create benchmark datasets due to the wide variety of approaches and outputs
- See Amit Singer's and Jose-Maria Carazo's presentations from CCP-EM Symposium



# Creating Benchmarks

Roodmus software toolkit - CCP-EM & TU Delft



- MD simulation for 'ground truth' dynamics
- Generate physically realistic SPA data from MD trajectories
- Highlighted need for more investigation into impact of data quality and hyperparameters



Maarten Joosten



Arjen Jakobi



Joel Greer



Tom Burnley



James Parkhurst



# Inaugural Flatiron Community Challenge

Launched in June 2023

- Entrants asked to create 80 density maps describing PC of motion and relative populations
- Anonymised submissions and analysis pipeline available at: <https://osf.io/8h6fz/>

Highlighted the difficulty of:

- Developing metrics to benchmark performance
- Validating heterogeneity in experimental datasets

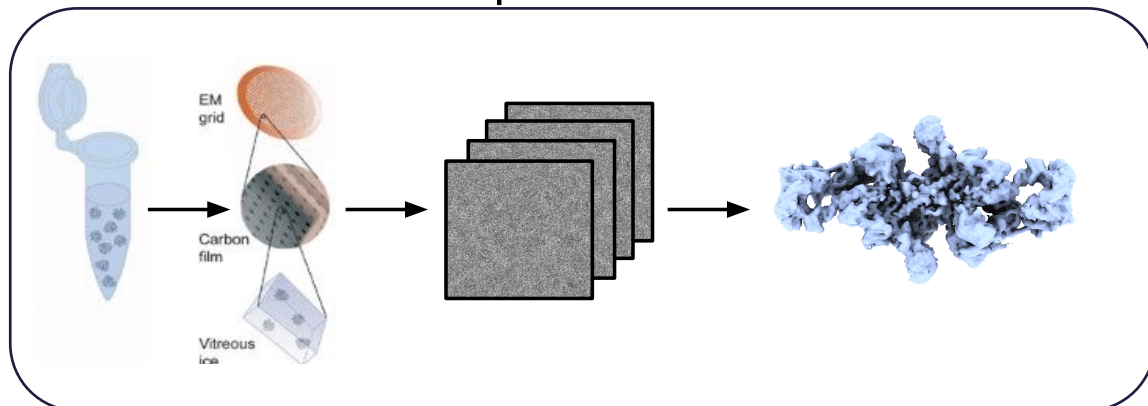


Pilar Cossio

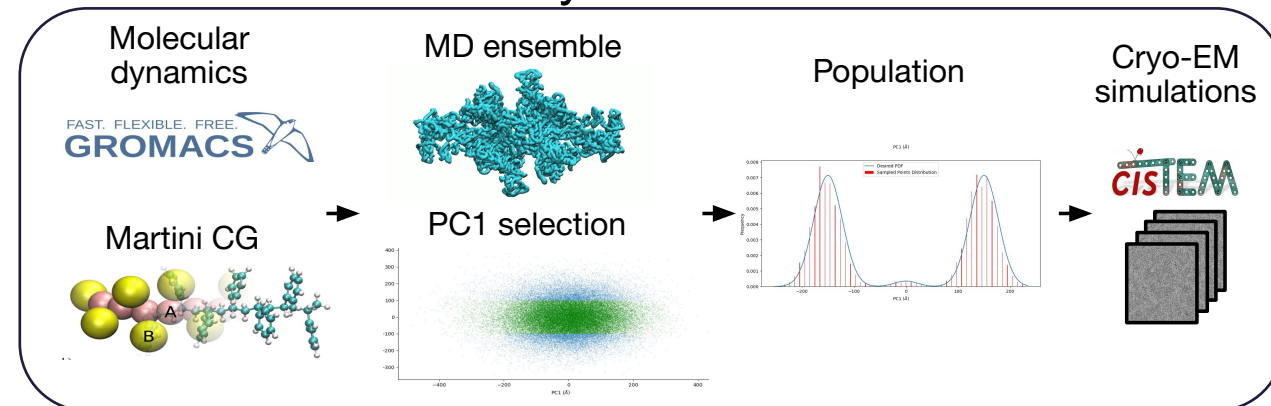


Sonya Hanson

## Experimental



## Synthetic



# Heterogeneity Community Challenge

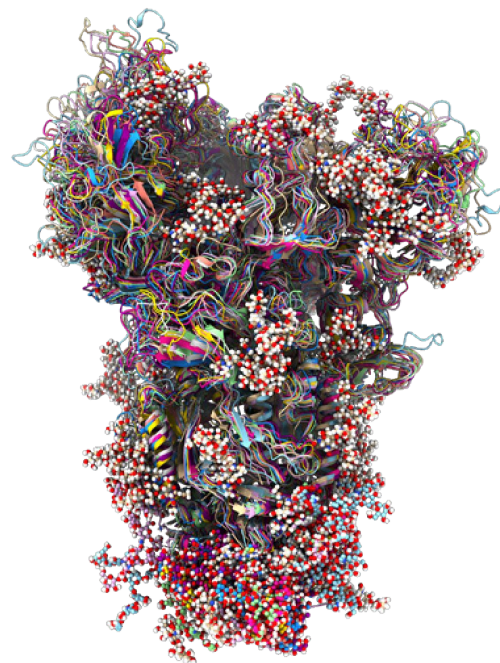
We want to attach developers of HRAs or a users at different levels of experience

Complex mixtures

Non-uniform pose distribution

Microstates from MD simulations

Ligand binding



Ellen Zhong



Ryan Feathers



Pilar Cossio



Sonya Hanson

Launch at the [Flatiron Institute CryoEM Summer Workshop](#) (24th & 25th July 2025)

Review findings at a workshop in the UK the following year



Tom Burnley



Joel Greer



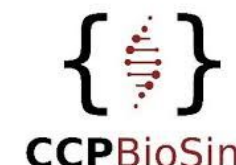
Science and  
Technology  
Facilities Council



CCP-EM



FLATIRON  
INSTITUTE



Engineering and  
Physical Sciences  
Research Council

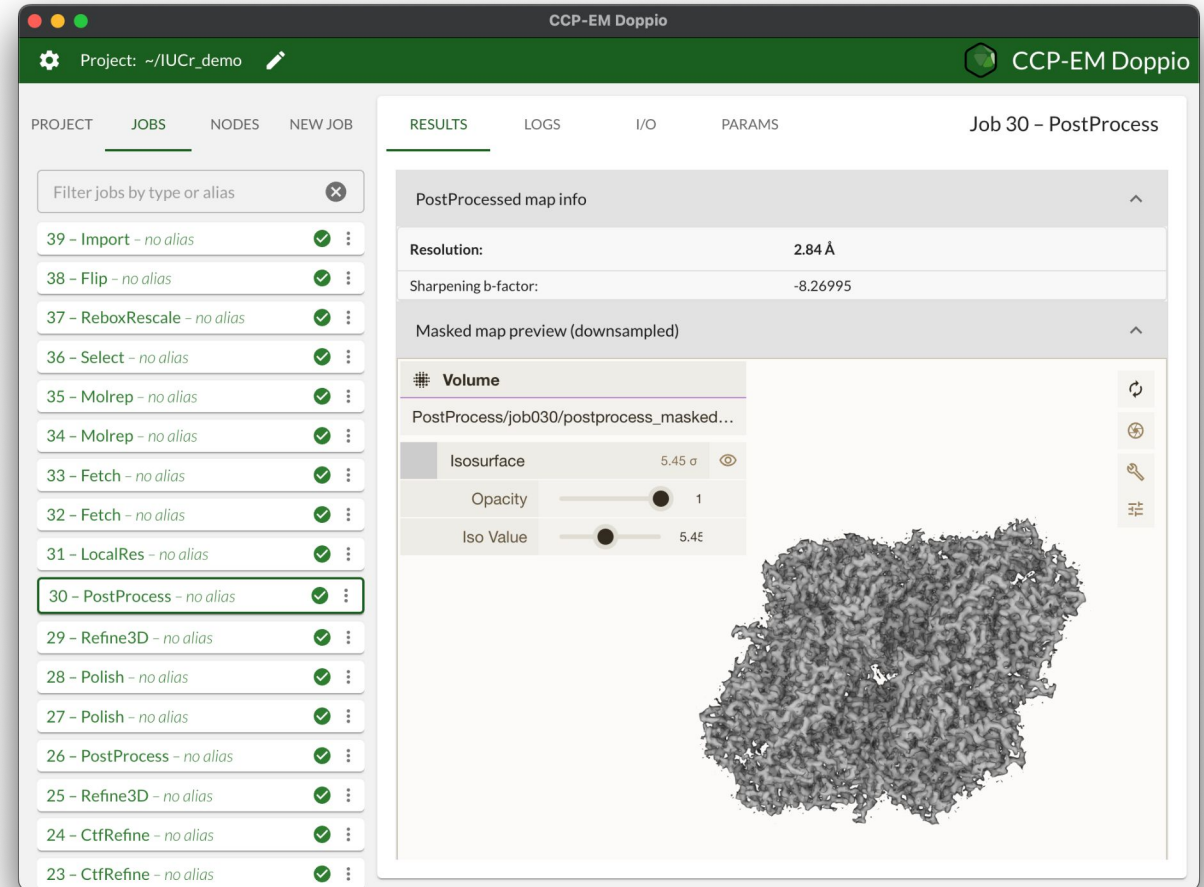


Research Complex  
at Harwell



# CCP-EM Doppio 1.3 Release

- **1.3 - April 2025:**
  - Linux and Mac
  - Requires CCP-EM 1.x
  - Optionally CCP4 9.x & Relion 5.x/4.x
- **Workflows for:**
  - Single particle analysis
  - Subtomogram averaging
  - Heterogeneous reconstruction
  - Map optimisation
  - Atomic model building and validation



CCP-EM Core Team

CCP4 Core Team

STFC Business & Innovation

CCP-EM Commercial License Holders

CCP-EM Collaborators

CCP-EM Users

MRC Core funding

Website & Downloads: [www.ccpem.ac.uk](http://www.ccpem.ac.uk)

Mailing list: [www.jiscmail.ac.uk/ccpem](http://www.jiscmail.ac.uk/ccpem)

Bluesky: [ccpem.bsky.social](https://bsky.social/ccpem)

Email: [ccpem@stfc.ac.uk](mailto:ccpem@stfc.ac.uk)



# Practical Tutors



Tom  
Burnley



Colin  
Palmer



Rob  
Nicholls



Arjen  
Jakobi

