Science and Technology Facilities Council

Introduction to CCP-EM

Software and support for the cryoEM community

Tom Burnley

July 2025

Padova, Italy

CCP-EM

What is CCP-EM?

CCP-EM

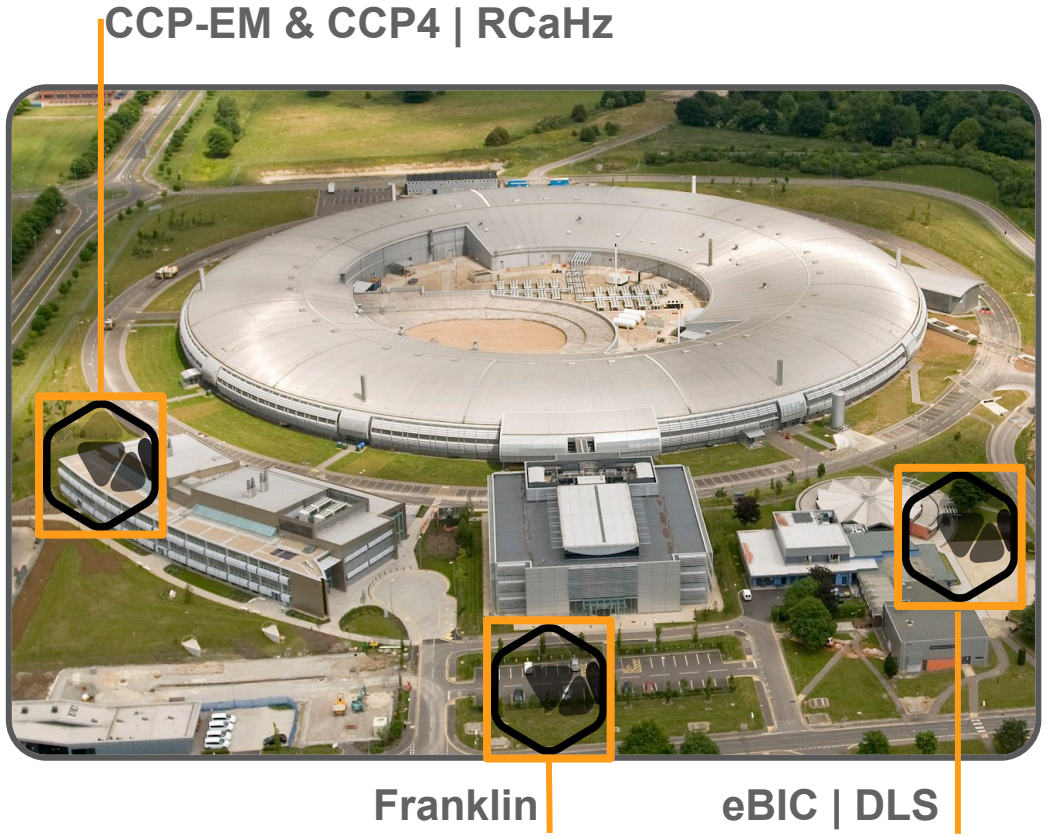
Collaborative Computational Project for Electron cryo-Microscopy

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



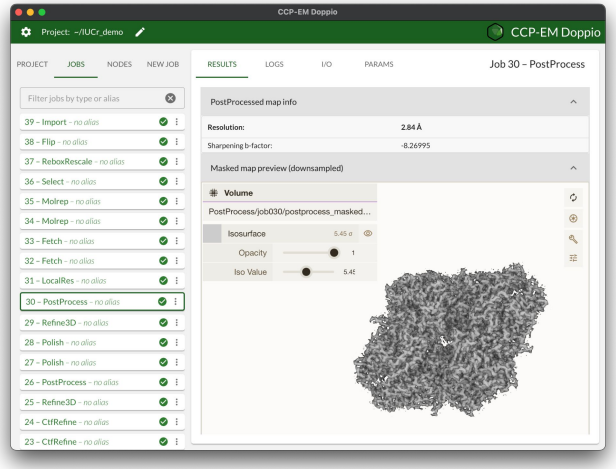
Science and Technology Facilities Council

Medical Research Council

Research Complex at Harwell

CCP-EM

Collaborative Computational Project



Science and Technology Facilities Council

Tom Burnley

George Coldstream

Joel Greer

Lauren Giles

Matt Iadanza

Agnel Joseph

Sony Malhotra

Jola Mirecka

Colin Palmer

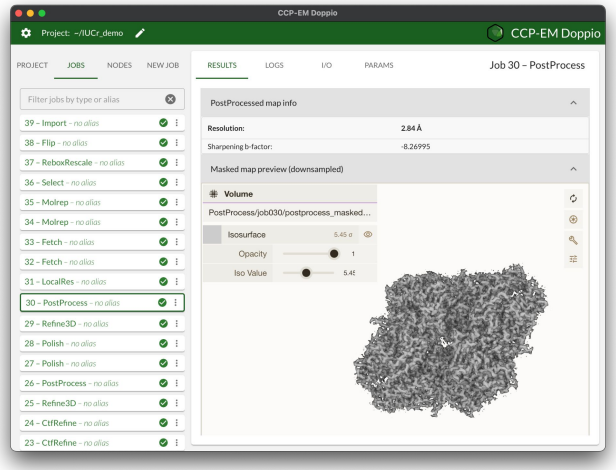
Rangana Warshamanage

Martyn Winn

Nick Whyatt

CCP-EM

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

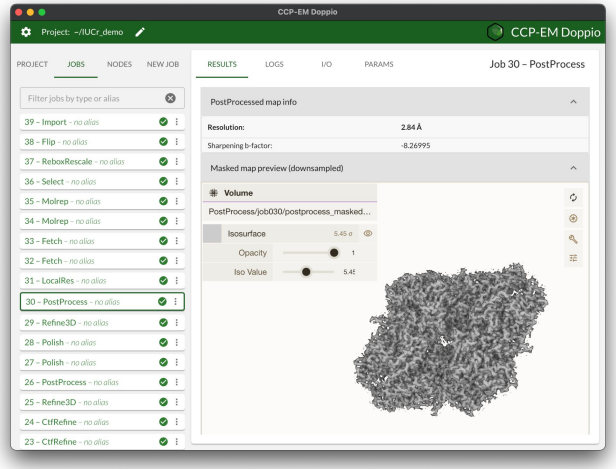
UK PIs, provide community oversight for the project and helps to set priorities

Science and Technology Facilities Council

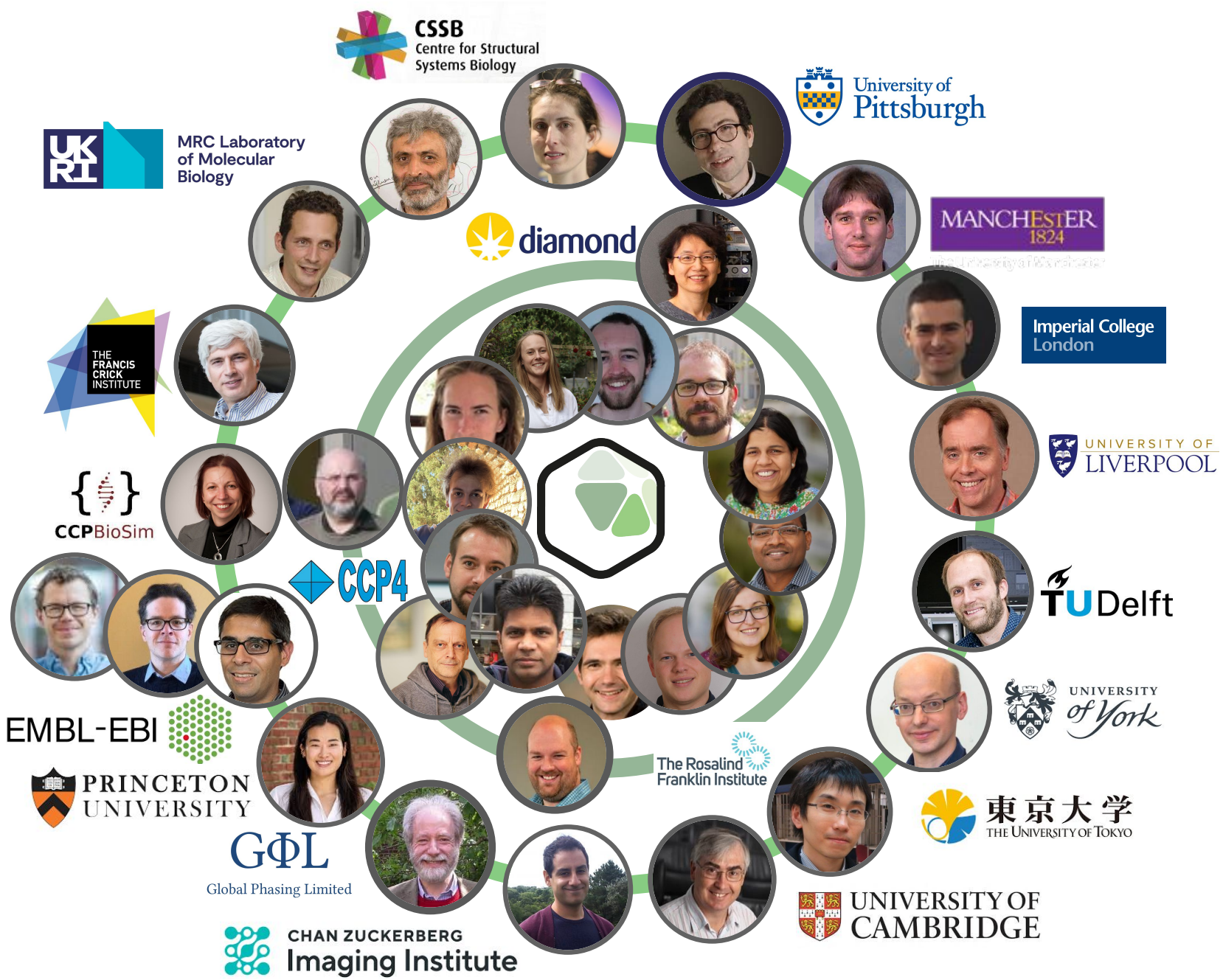
www.ccpem.ac.uk/about/working-groups

CCP-EM

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
- **11th Anniversary May 2025 Nottingham / Zoom**
  - >360 in-person, >800 virtual
  - Free virtual attendance
- **12th 22-24 April 2026, Nottingham, UK**
  - Bursaries available

Science and Technology Facilities Council

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)

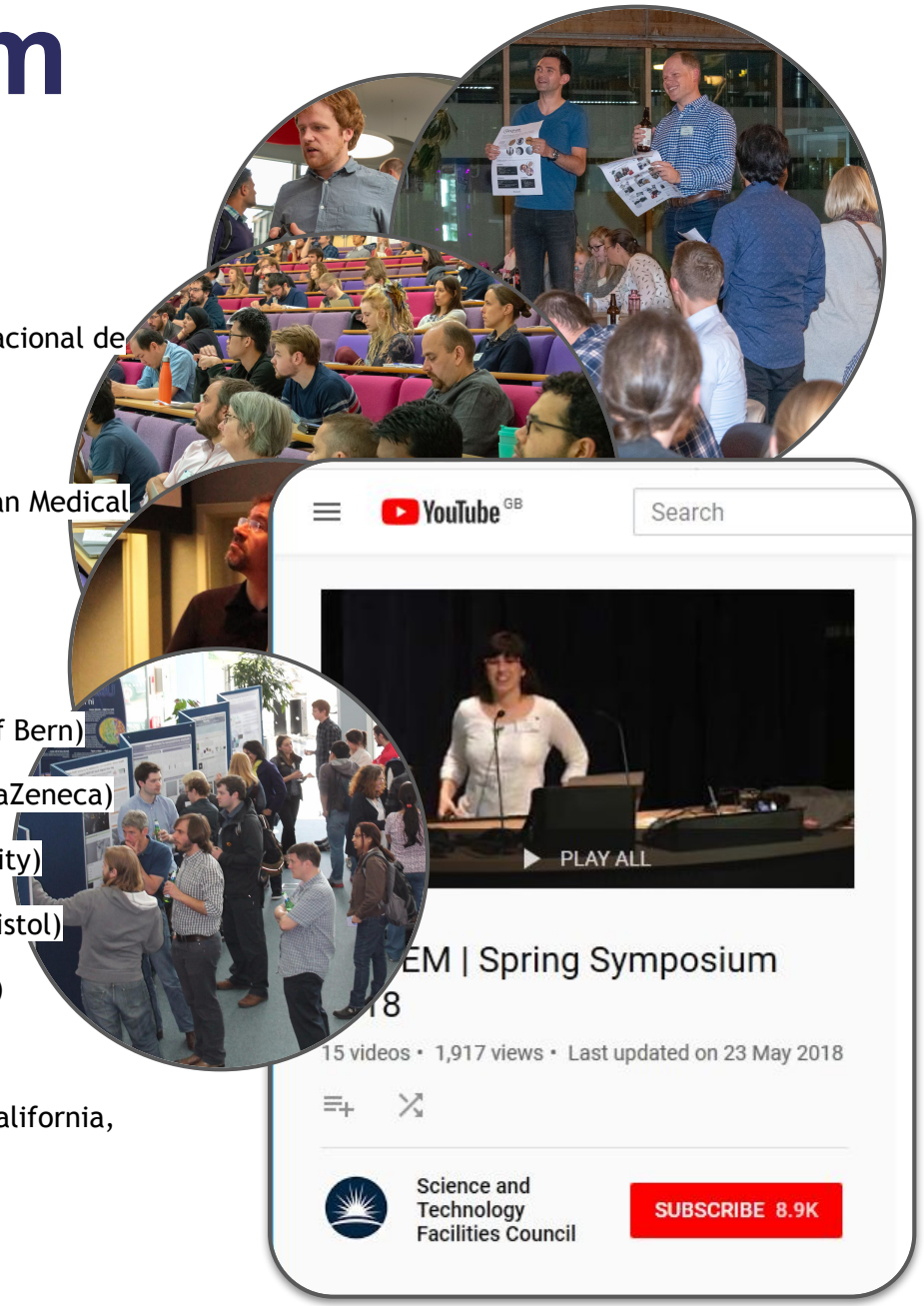
Thom Sharp (University of Bristol)

Alexander Shtyrov (MRC-LMB)

Amit Singer (Princeton)

Jessie Zhang (University of California, Berkeley)

Kai Zhang (Yale)



CCP-EM Software Suite

CCP-EM software suite

30+ tasks in a common Python framework

Raw micrographs to validated atomic model

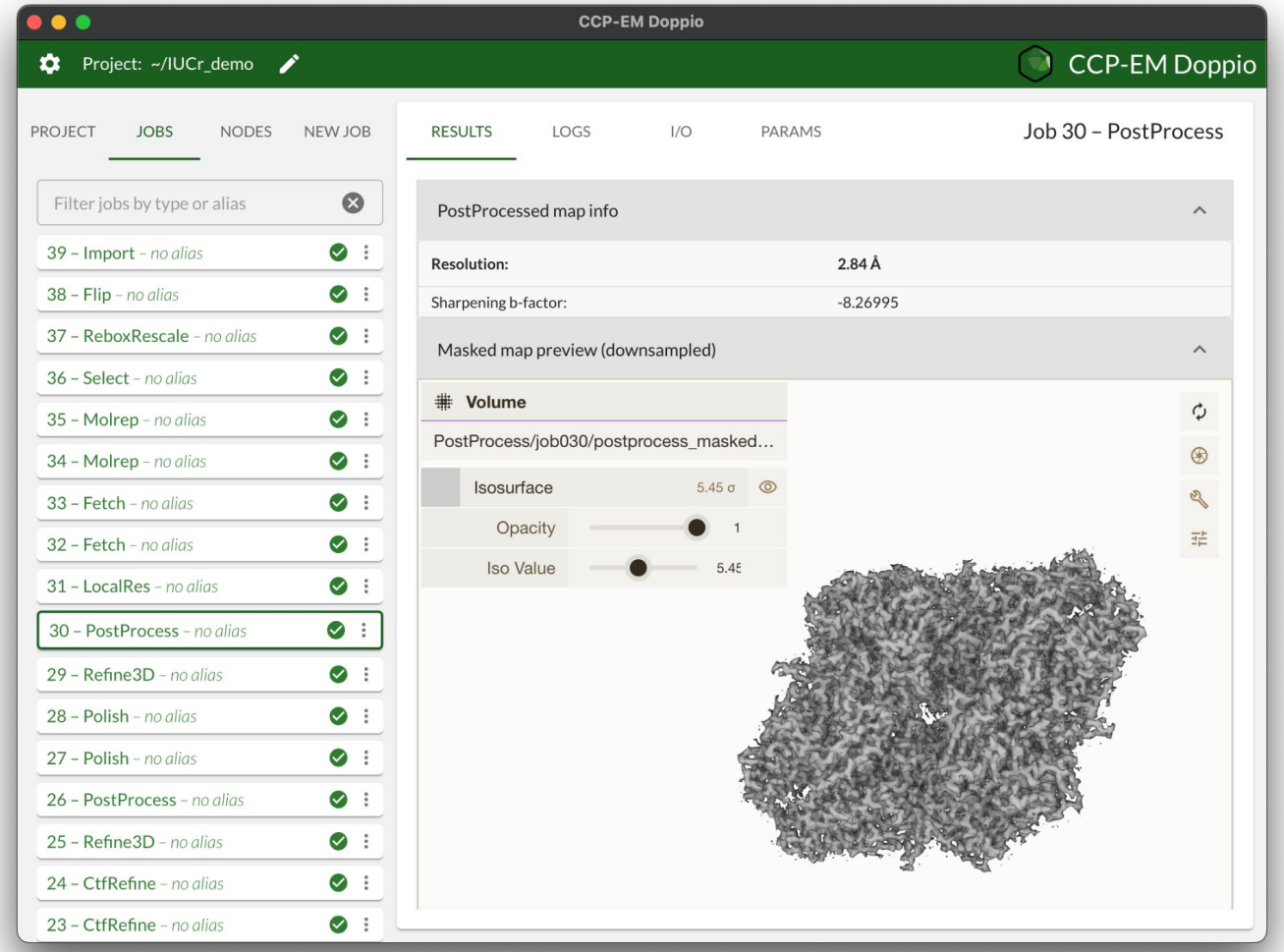
Download from ccpem.ac.uk

Linux & Mac

Free for academic use, fee for commercial

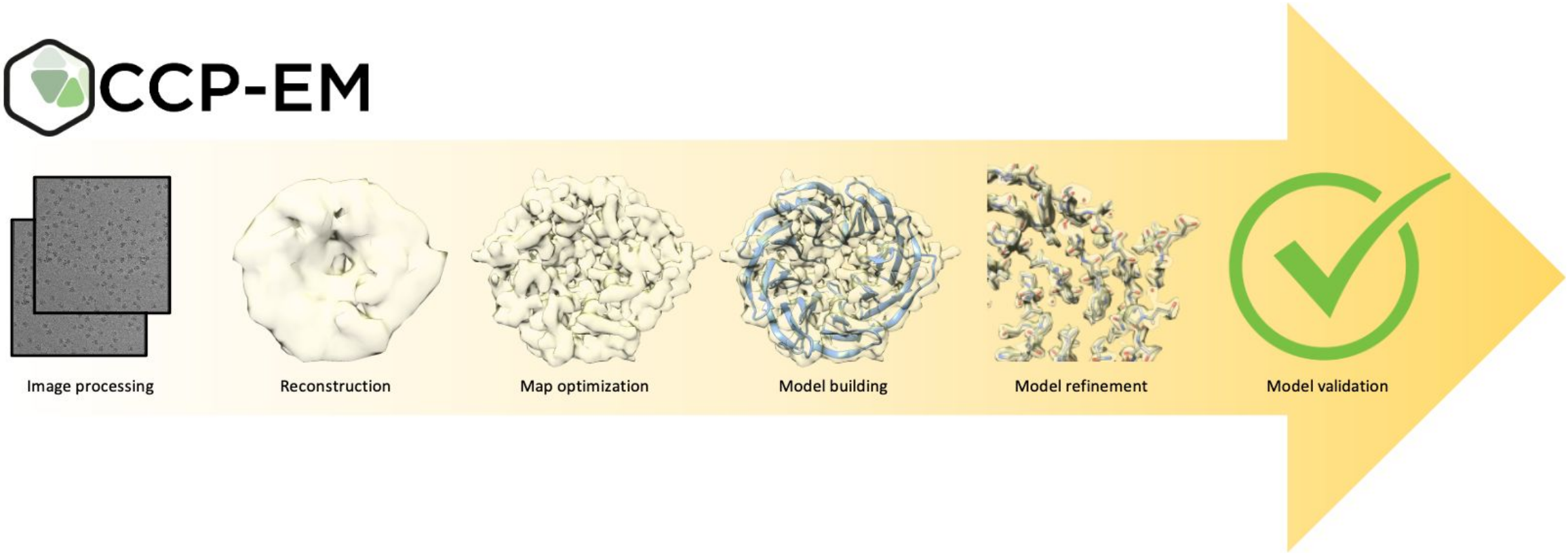
Get in touch:

ccpem@stfc.ac.uk



Science and Technology Facilities Council

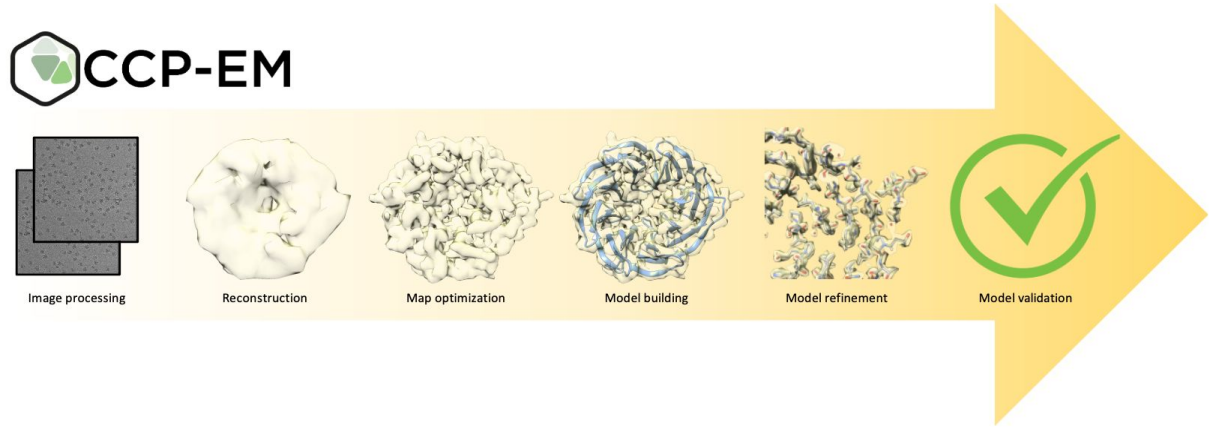
CCP-EM workflow



Science and Technology Facilities Council

SPA & atomic modelling workflow

Single Particle Analysis	Map Optimisation	Model Building	Model Docking	Automated Refinement	Interactive Refinement	Map/Model Validation
Relion	LocScale	ModelCraft	EM_ placement	Refmac Servalvat	Coot	CCP-EM Validation
Topaz	Lafter	Model Angelo	MOLREP	AceDRG	Moorhen	Find/Check MySeq
CrYOLO	Confidence Map	Fetch AF2		TEMPy -REFF	Chimera	CryoEF
Icebreaker						



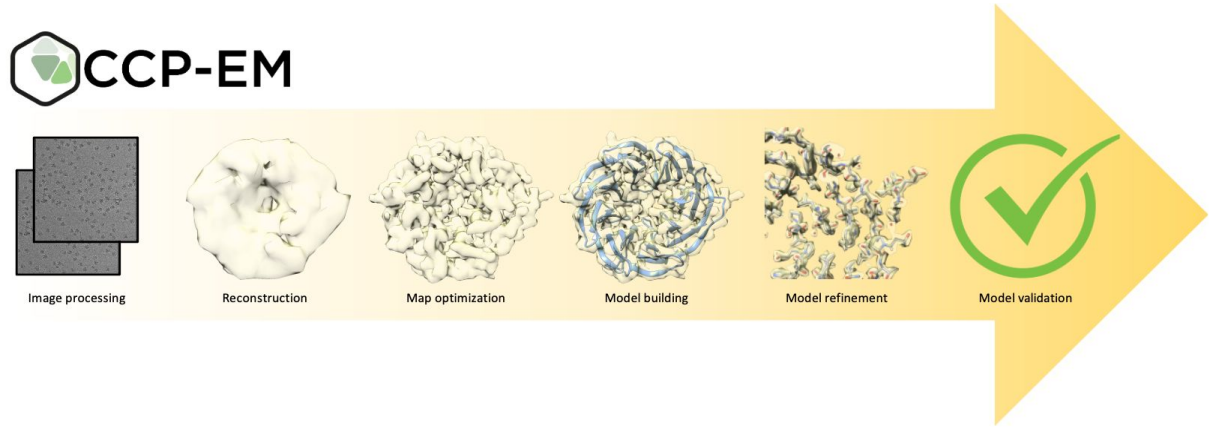
Model & Map Utilities

e.g. Map trimming

Science and Technology Facilities Council

SPA & atomic modelling workflow

Single Particle Analysis	Map Optimisation	Model Building	Model Docking	Automated Refinement	Interactive Refinement	Map/Model Validation
Relion	LocScale	ModelCraft	EM_ placement	Refmac Servalvat	Coot	CCP-EM Validation
Topaz	Lafter	Model Angelo	MOLREP	AceDRG	Moorhen	Find/Check MySeq
CrYOLO	Confidence Map	Fetch AF2		TEMPy -REFF	Chimera	CryoEF
Icebreaker						

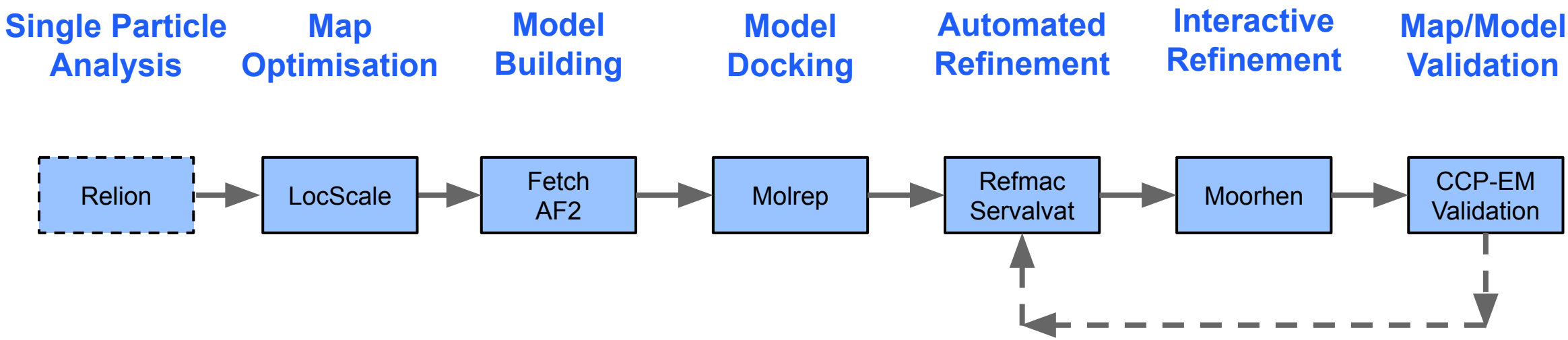


Model & Map Utilities

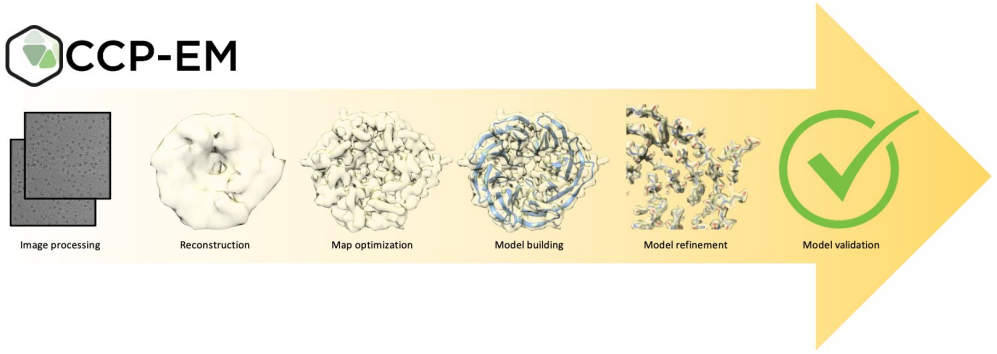
e.g. Map trimming

Science and Technology Facilities Council

Tutorial workflow



xMultiple human in the loop iterations



CCP-EM version 1.x

CCP-EM

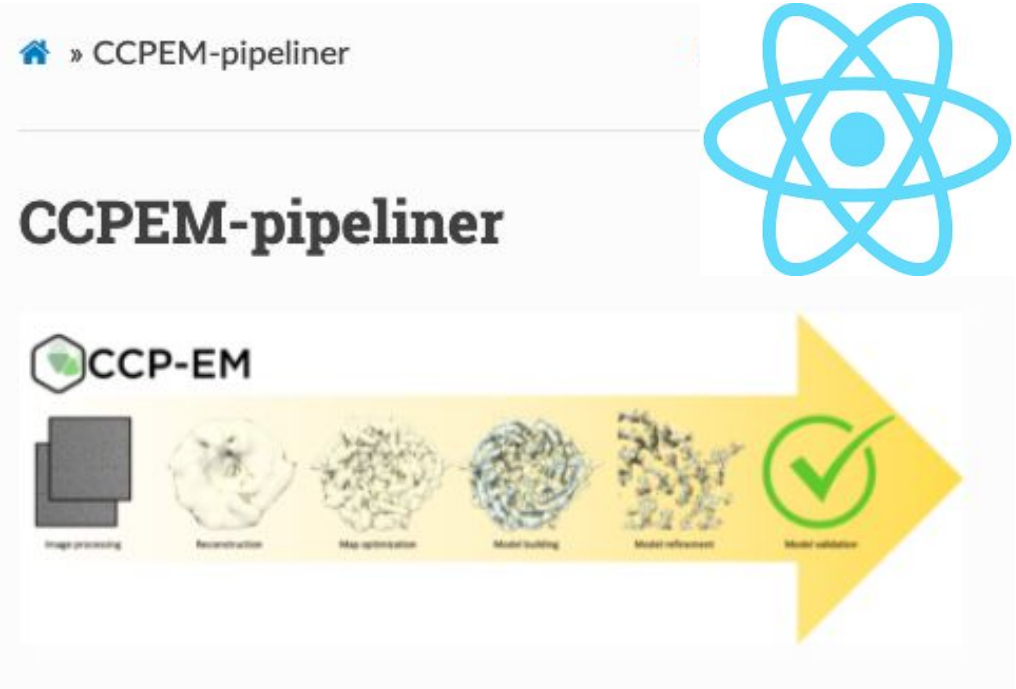
- Map post processing
- Atomic modelling
- PyQt5 GUI
- No data model
- Basic project management

RELION

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

Science and Technology Facilities Council

CCP-EM version 2.x



Images to structures

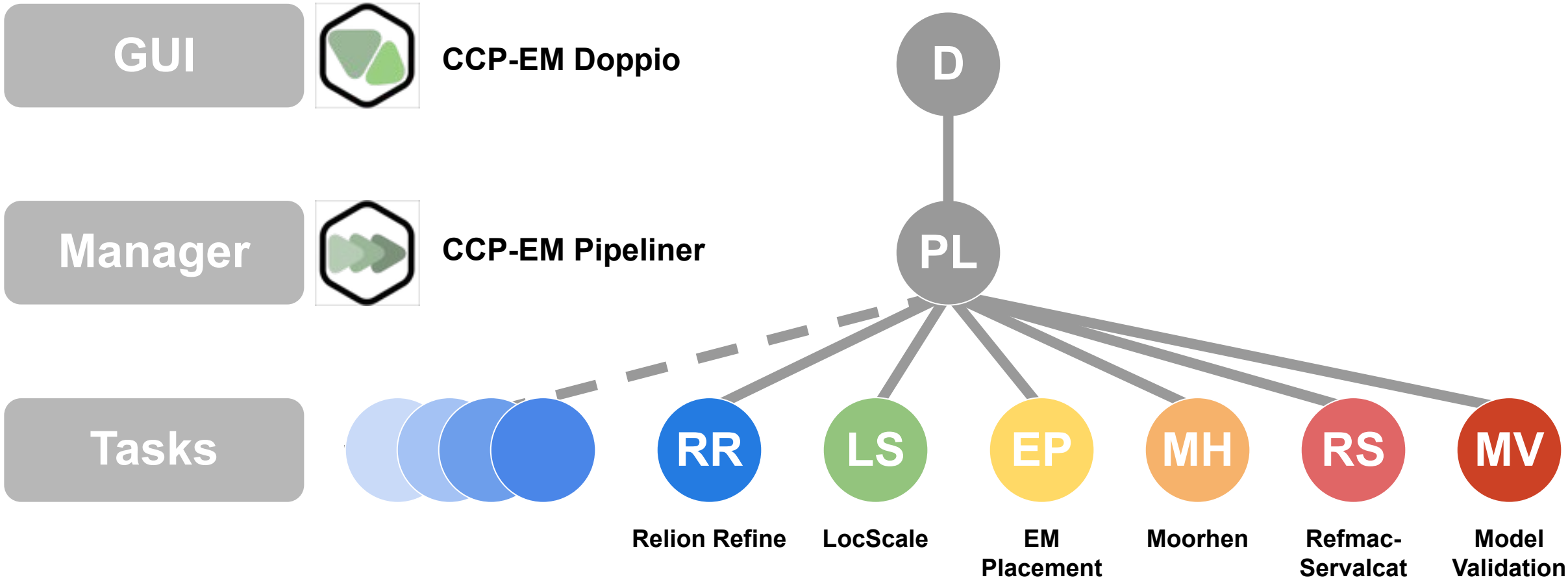
- Python pipeliner python API
- JS react GUI

- Updated data model
- Project management

- Single particle
- Atomic modelling
- Tomography (coming soon)

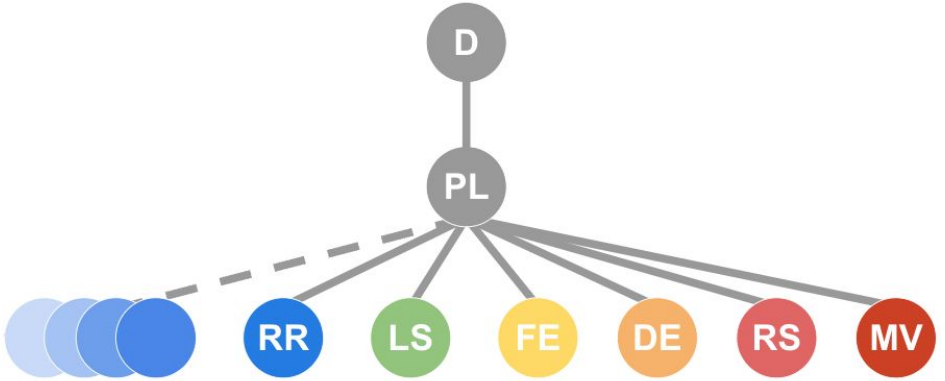
Science and Technology Facilities Council

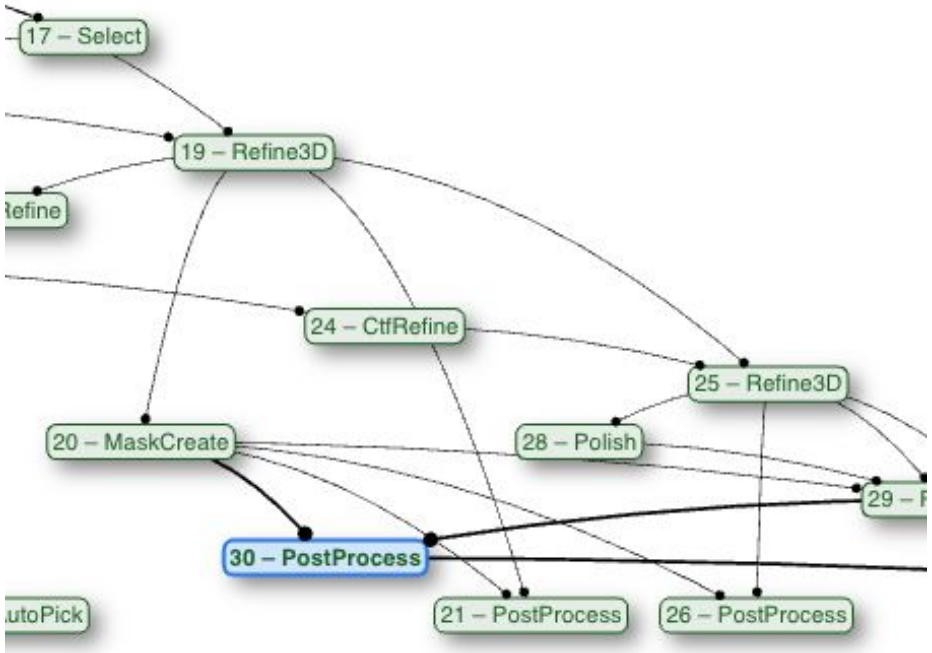
CCP-EM software plan



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



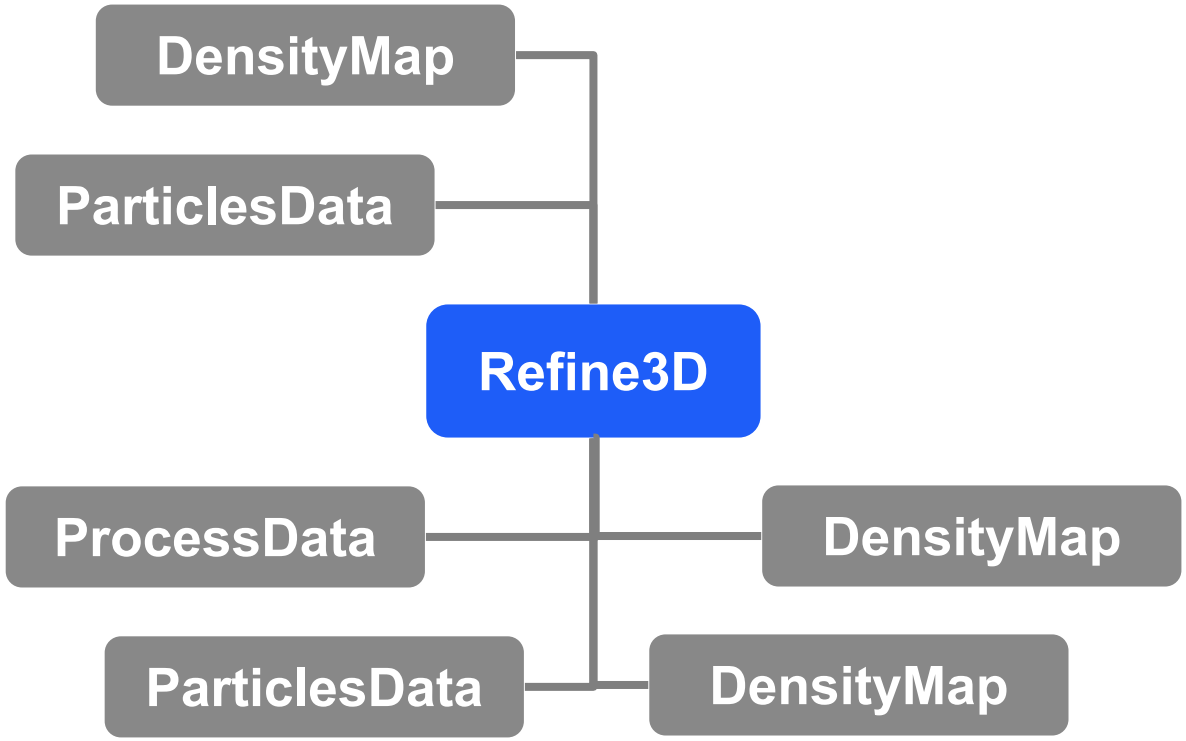


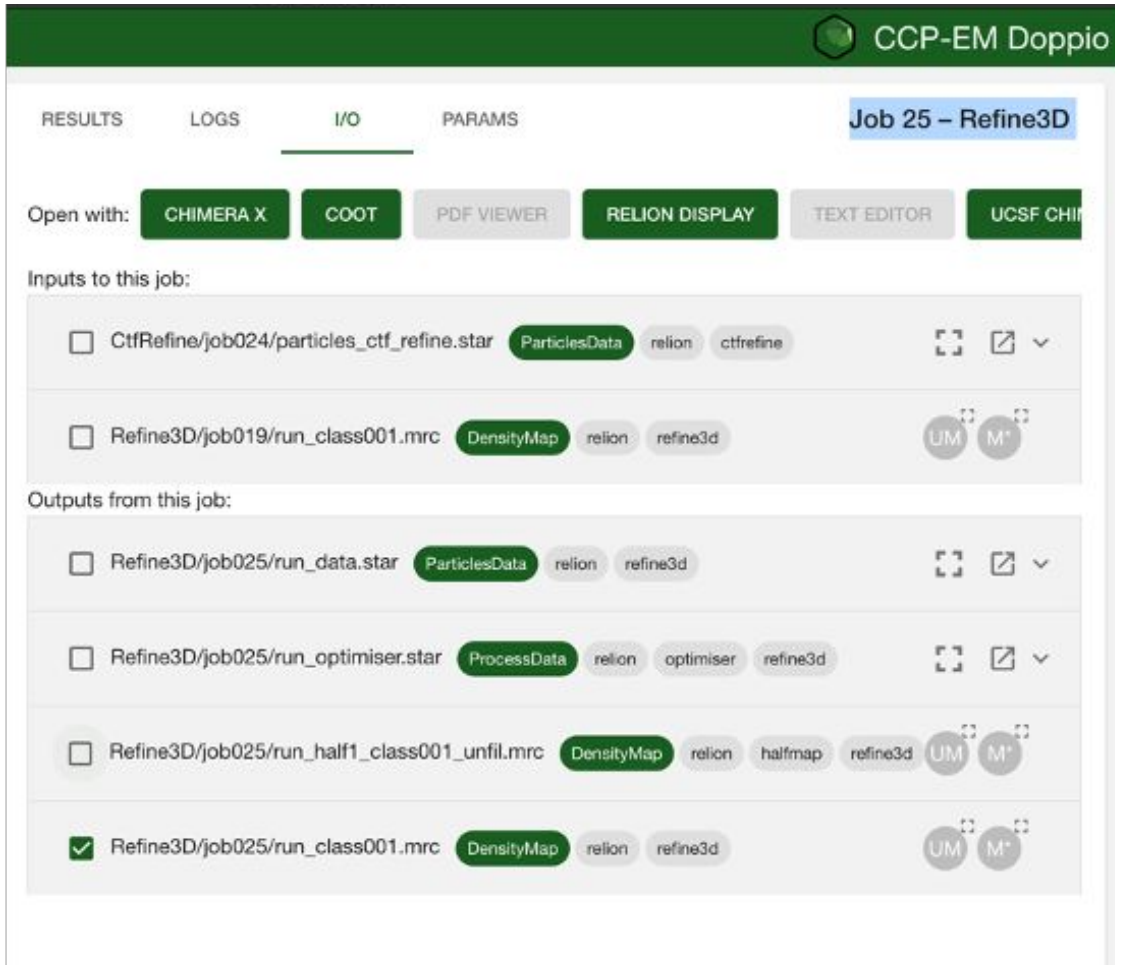
Science and Technology Facilities Council



Job Nodes

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



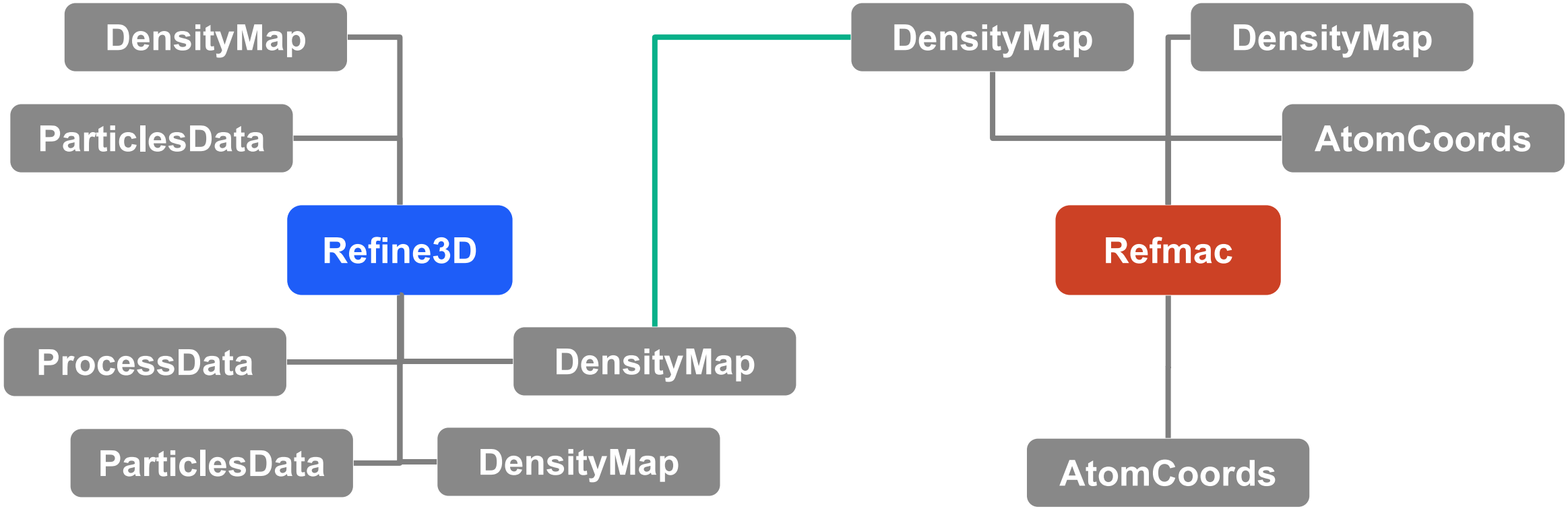


Science and Technology Facilities Council



Job Nodes

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



Science and Technology Facilities Council



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
```

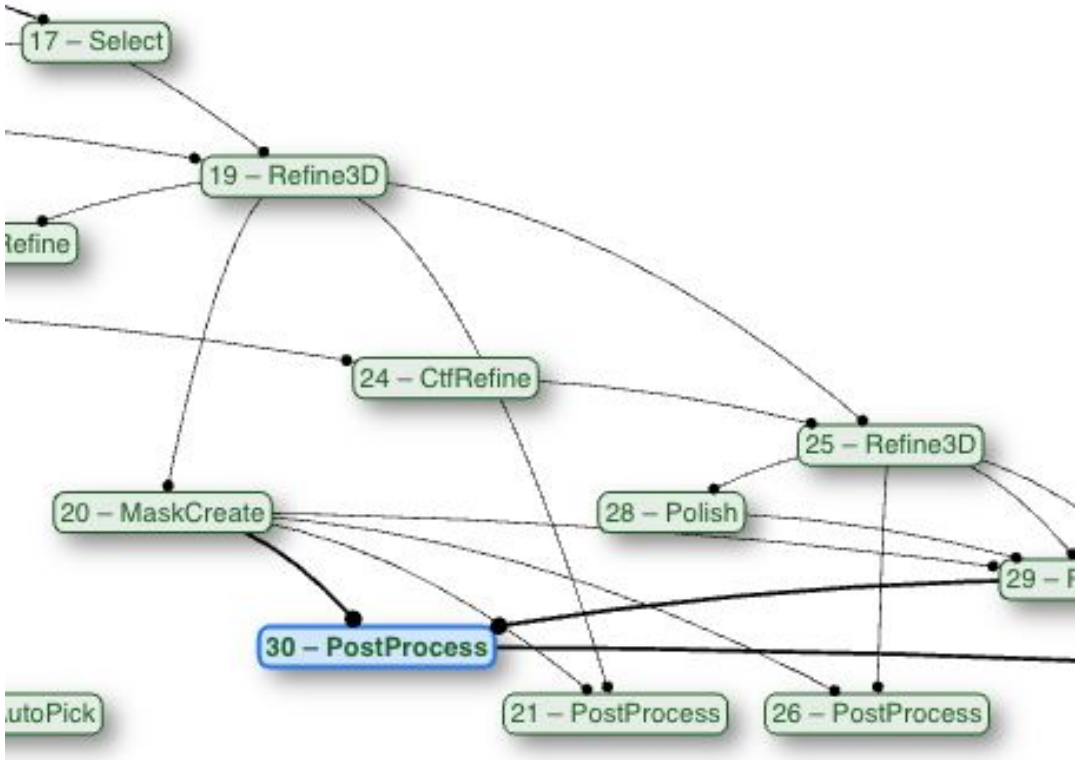
- ***Python API***

```
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
```

Science and Technology Facilities Council

diamond

Metadata Collection and Analysis



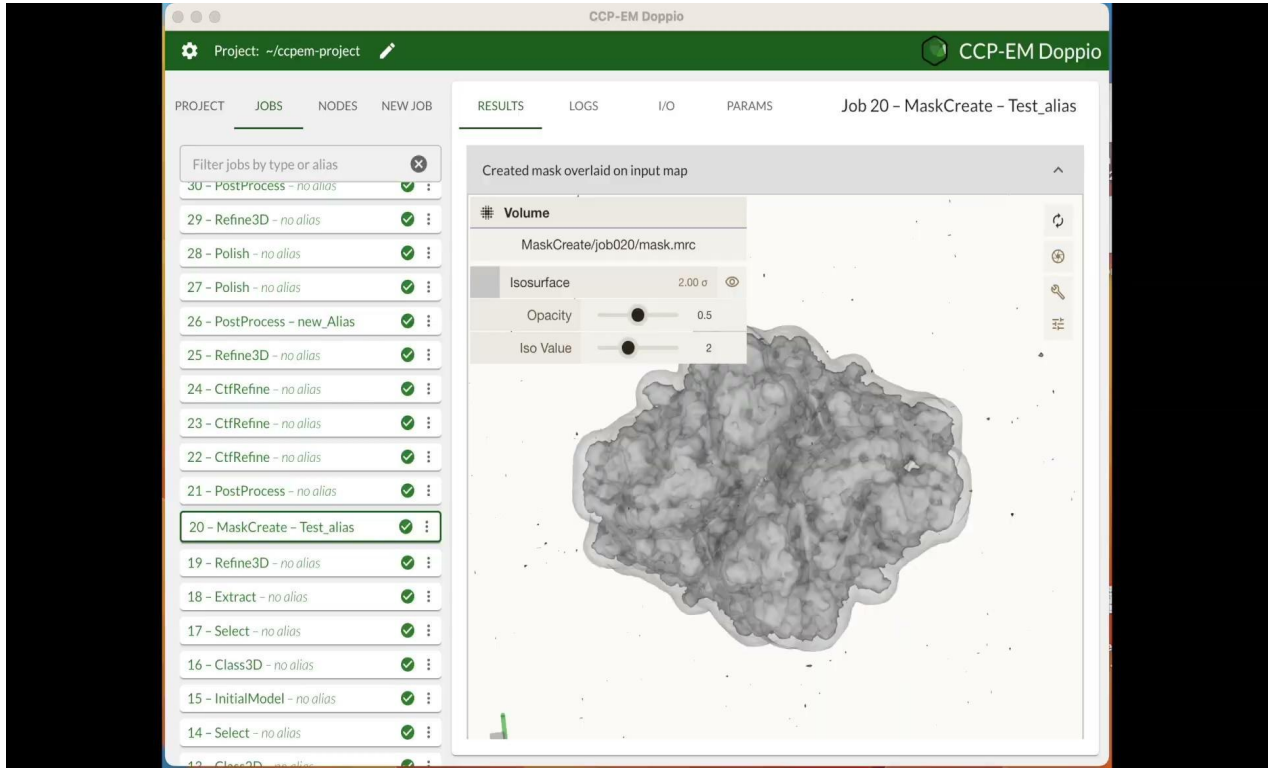
- Every job records the complete history of all operations
- Every job defines the metadata it returns
- Generate a metadata report for an entire workflow upstream from any job
- Generate a reference report for whole workflow - ***please cite all***

Science and Technology Facilities Council

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

Science and Technology Facilities Council

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

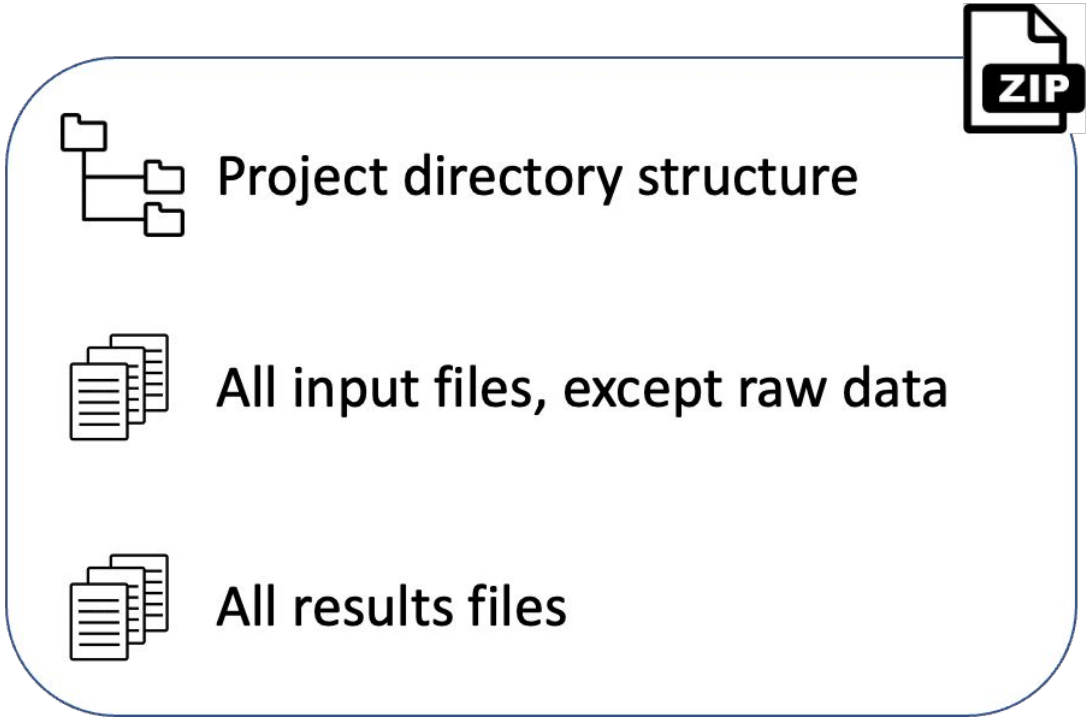
- *N.B. some details e.g. sample prep will need to be provided by depositor*

Science and Technology Facilities Council

Preserving workflows

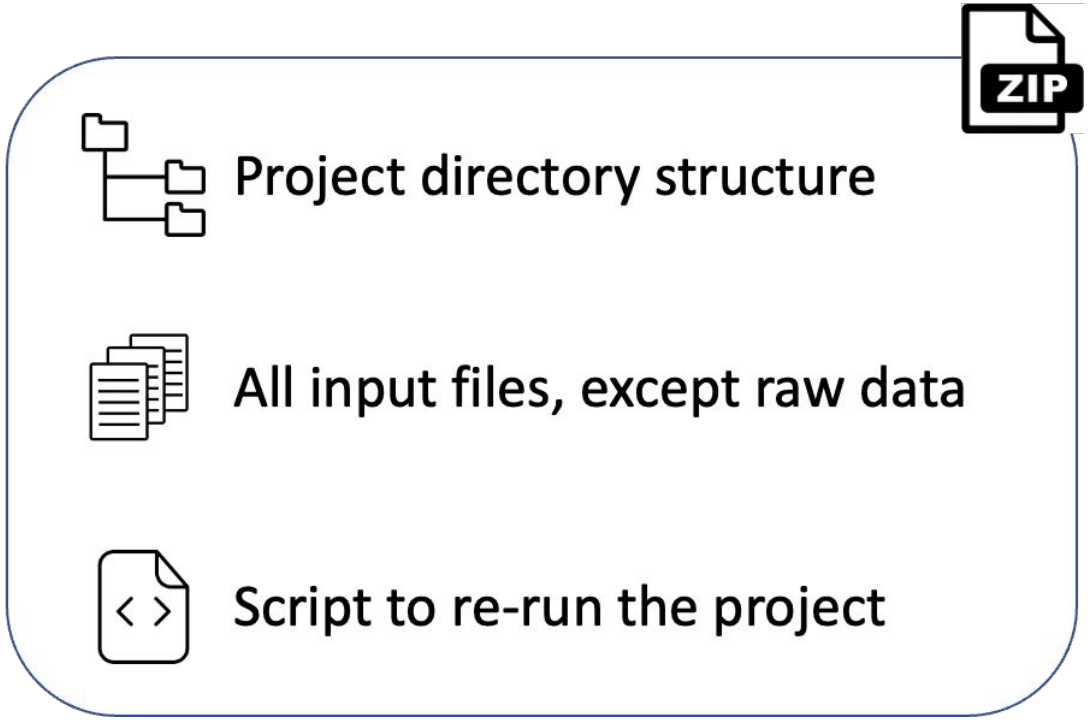
Full Archive

Preserve the project and all associated files



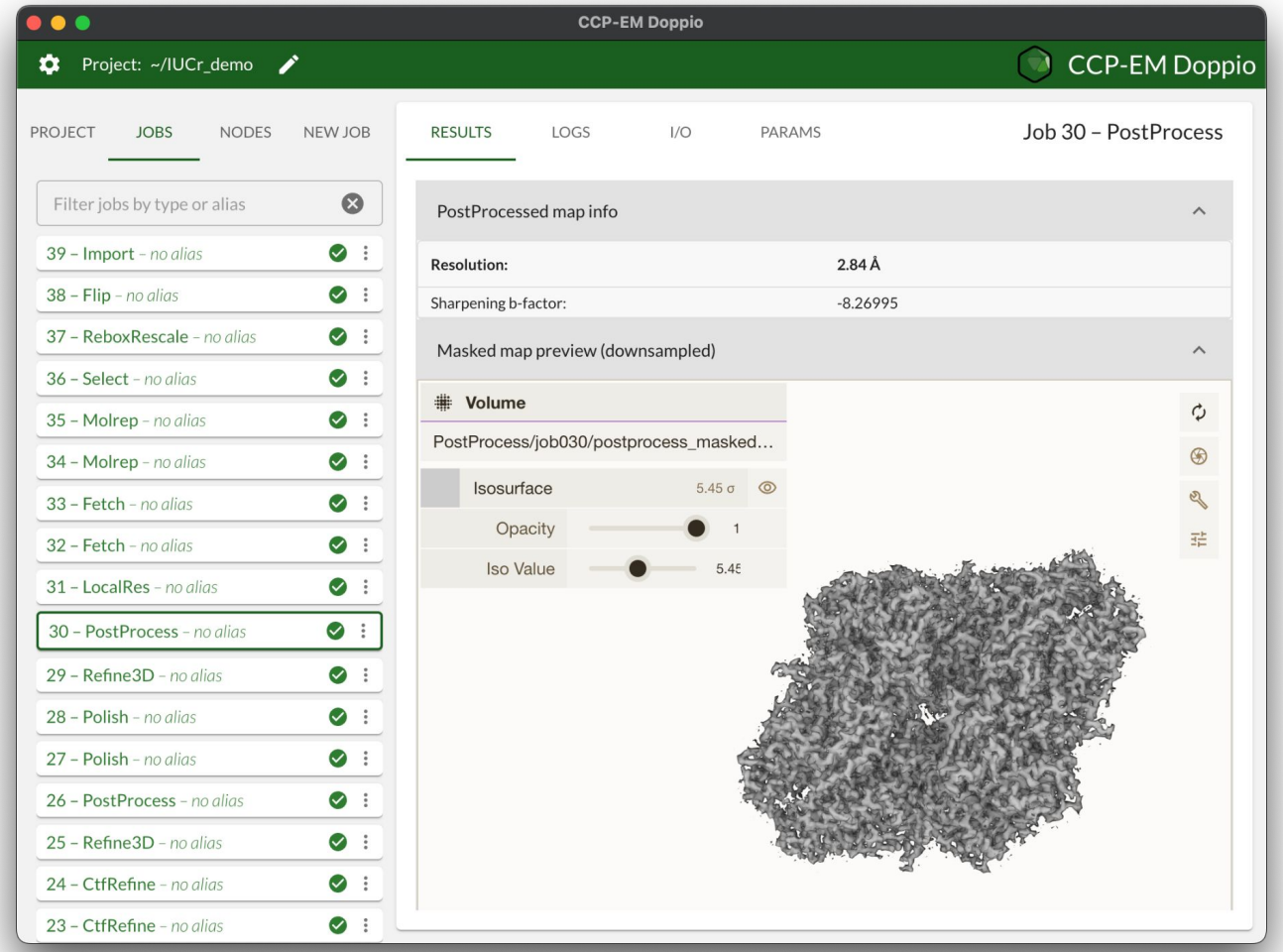
Script archive

Preserve and repeat the workflow



Science and Technology Facilities Council

Doppio demo

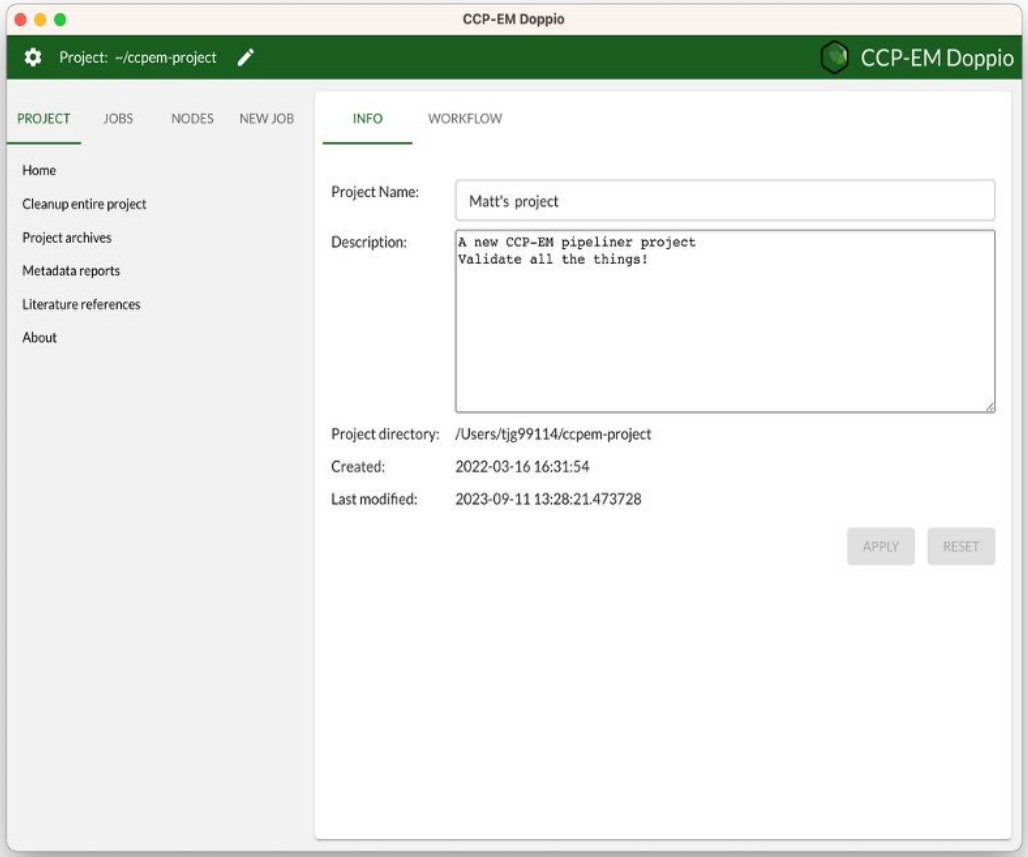


Science and Technology Facilities Council

Doppio modes: app & web

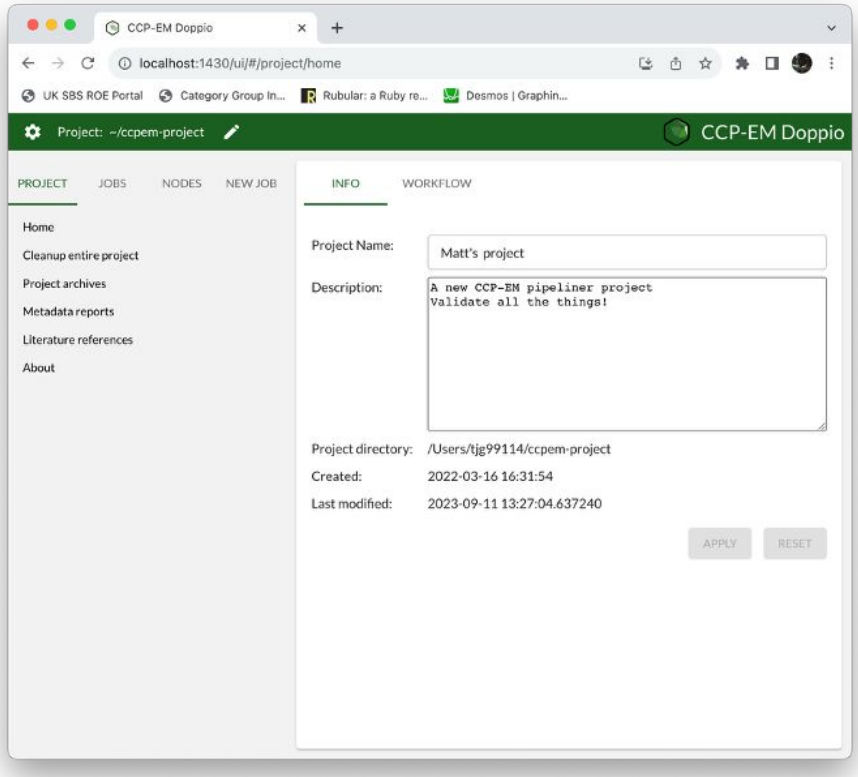
Doppio app

CCP-EM Doppio

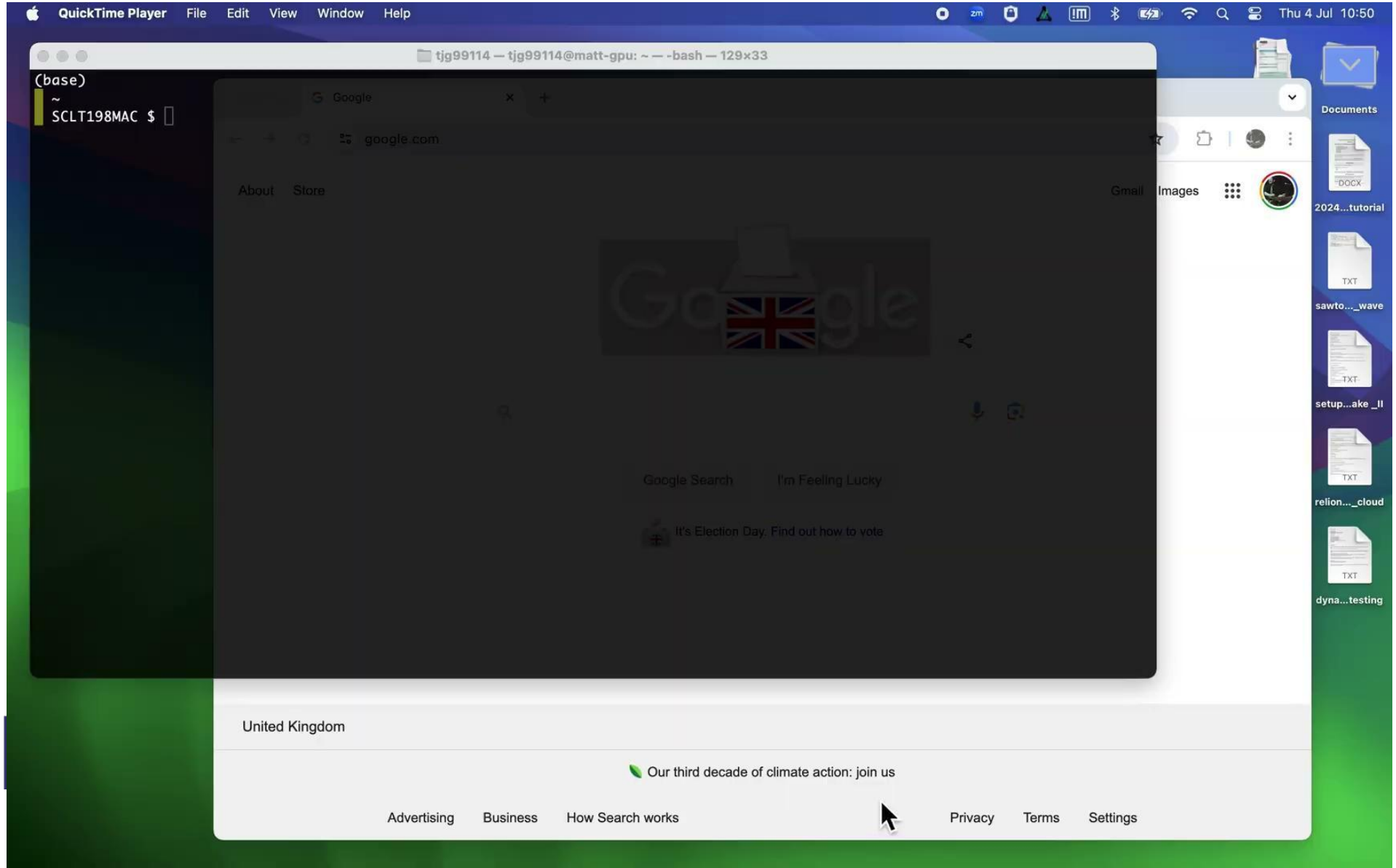


doppio-web

```
$ ./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
```



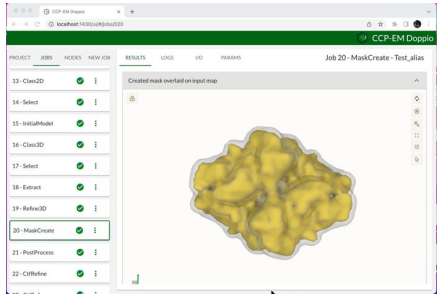
Doppio modes: web & remote



Doppio modes

Lo

Local

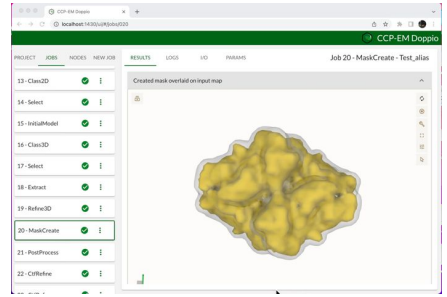


Science and Technology Facilities Council

Doppio modes

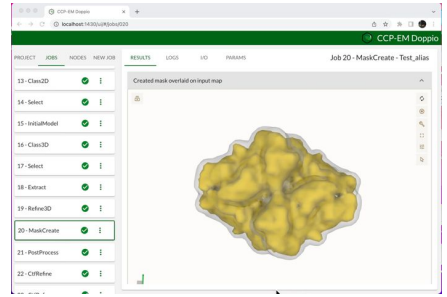
Lo

Local



Rm

Remote

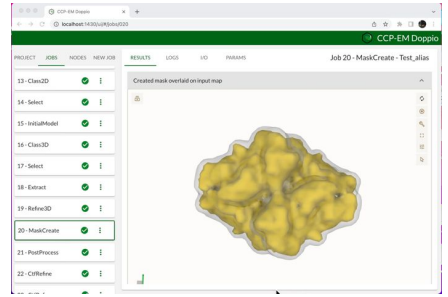


Science and Technology Facilities Council

Doppio modes

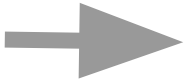
Lo

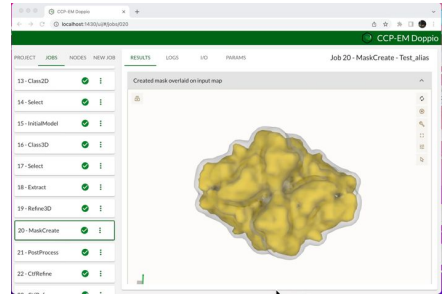
Local



Rm

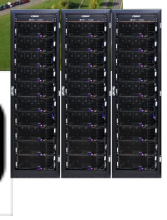
Remote

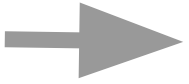


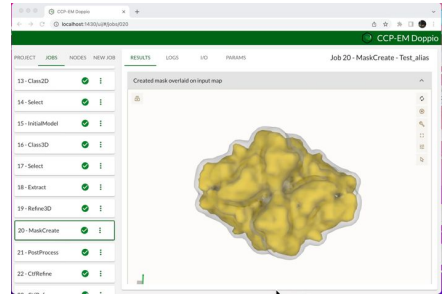


SV

STFC Cloud

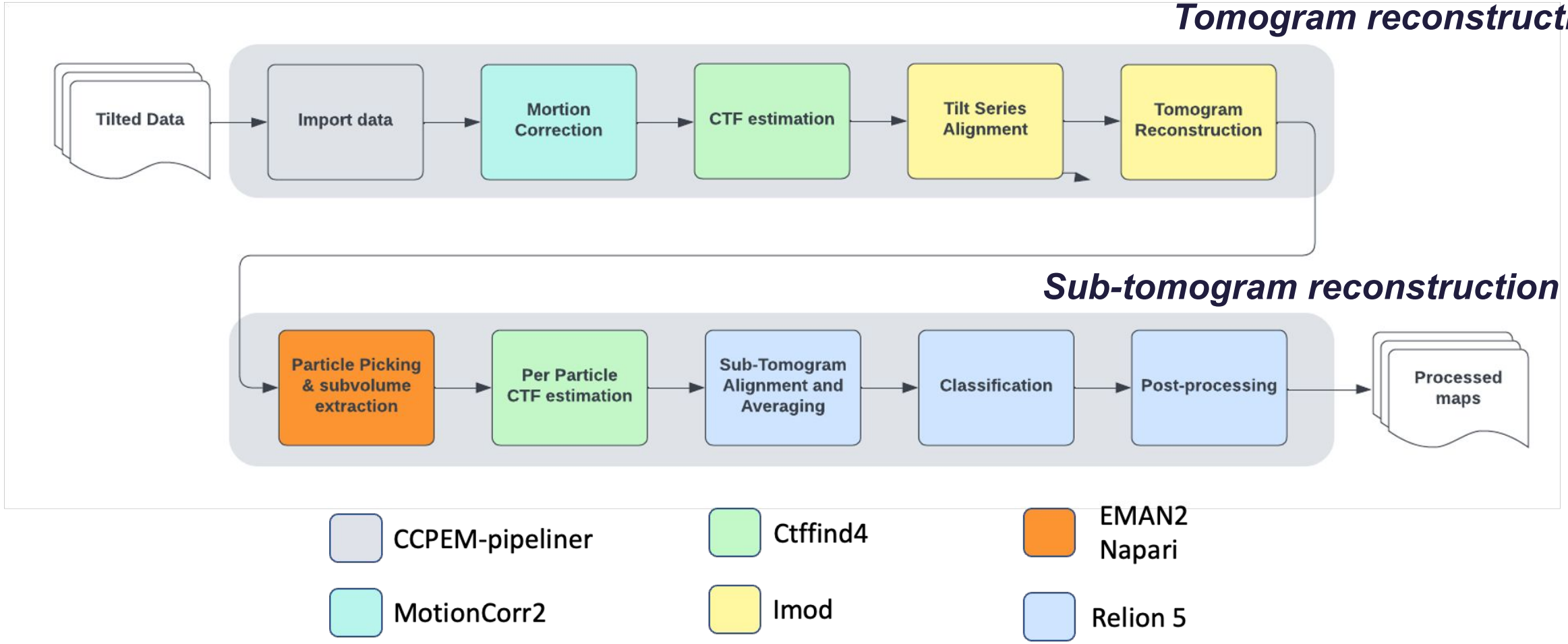






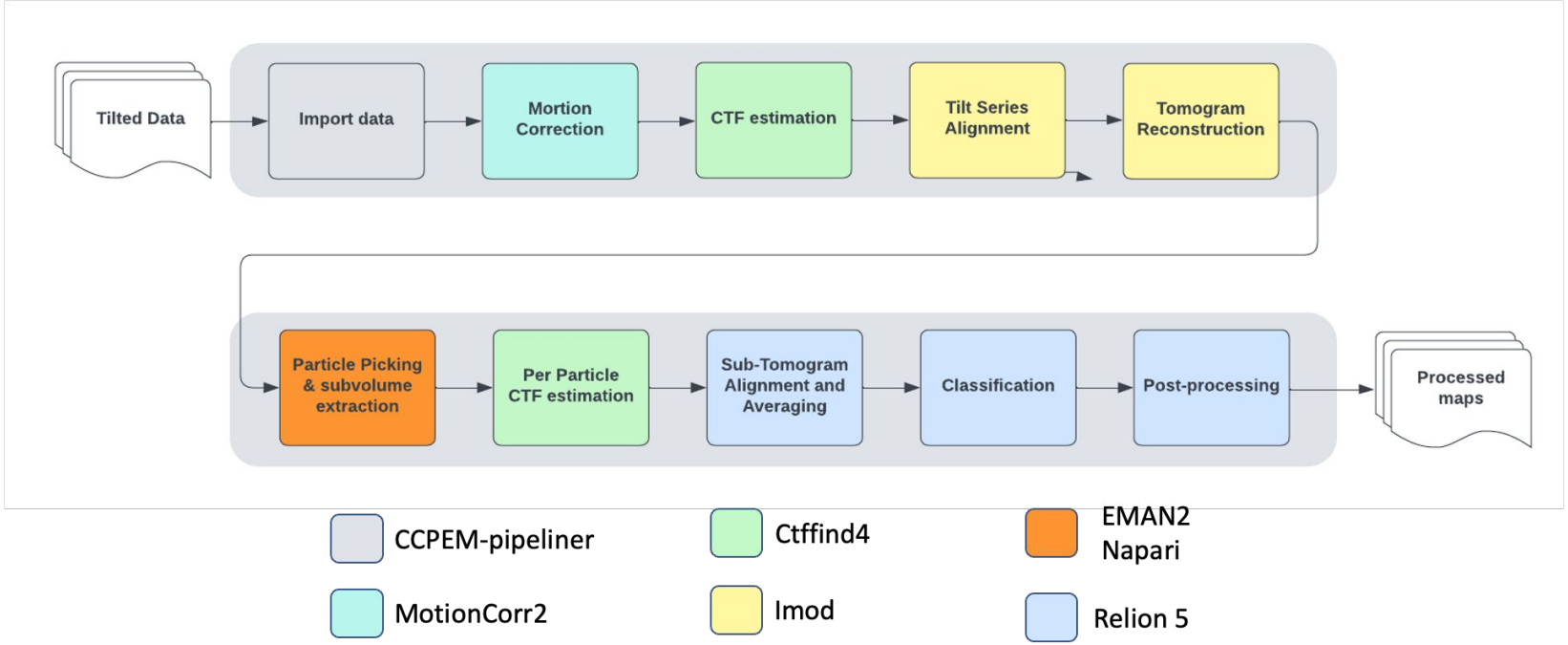
Science and Technology Facilities Council

Doppio: tomography and STA



Science and Technology Facilities Council

Doppio: subtomogram averaging



Relion 5 workflow

Metadata flows through all stages (centralized handling)

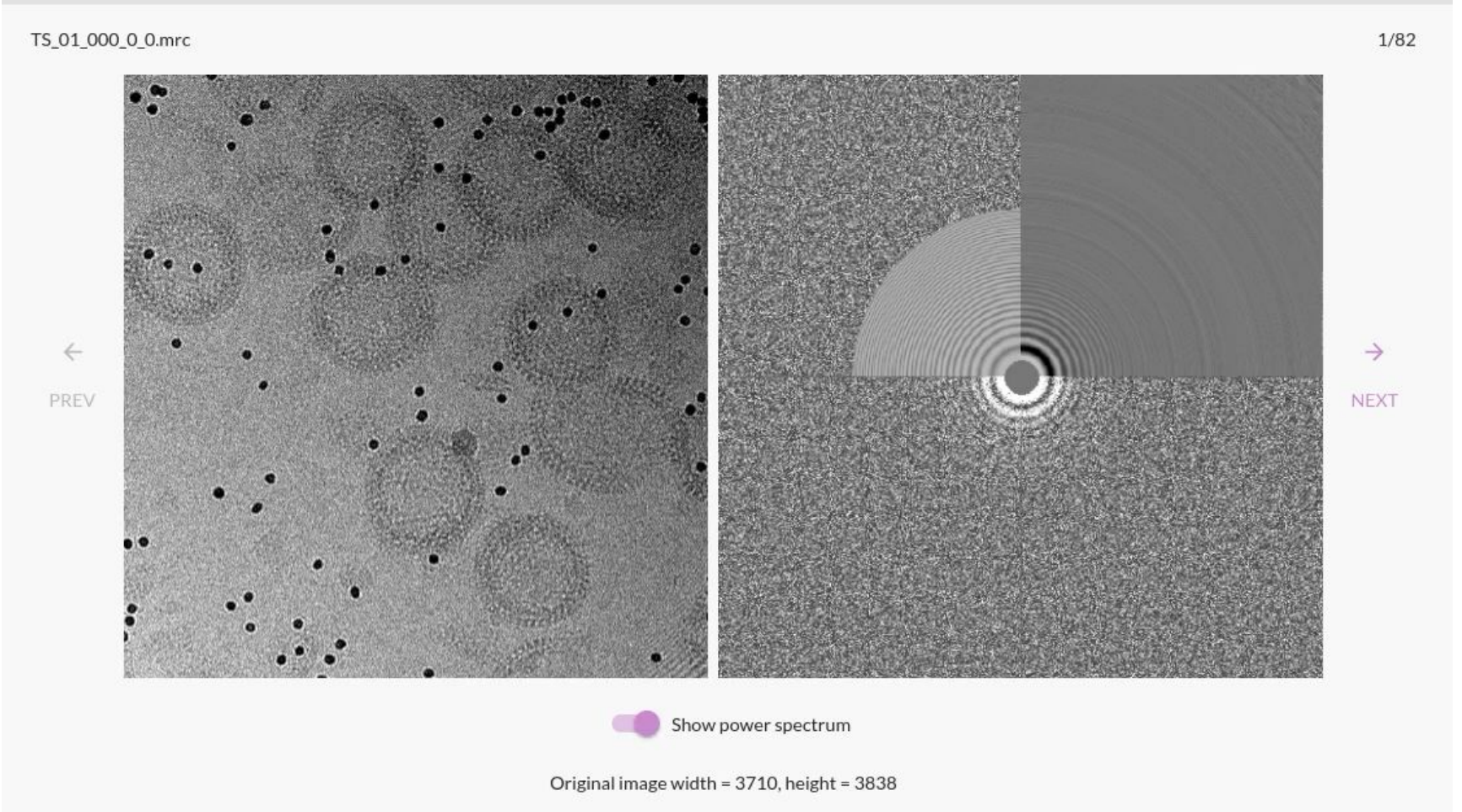
Remote-friendly: Port forwarding (not Napari, some IMOD visualisation)

Science and Technology Facilities Council

Doppio: tomography and STA

Intermediate visualization matters

CTF Estimation
Verify contrast transfer fit

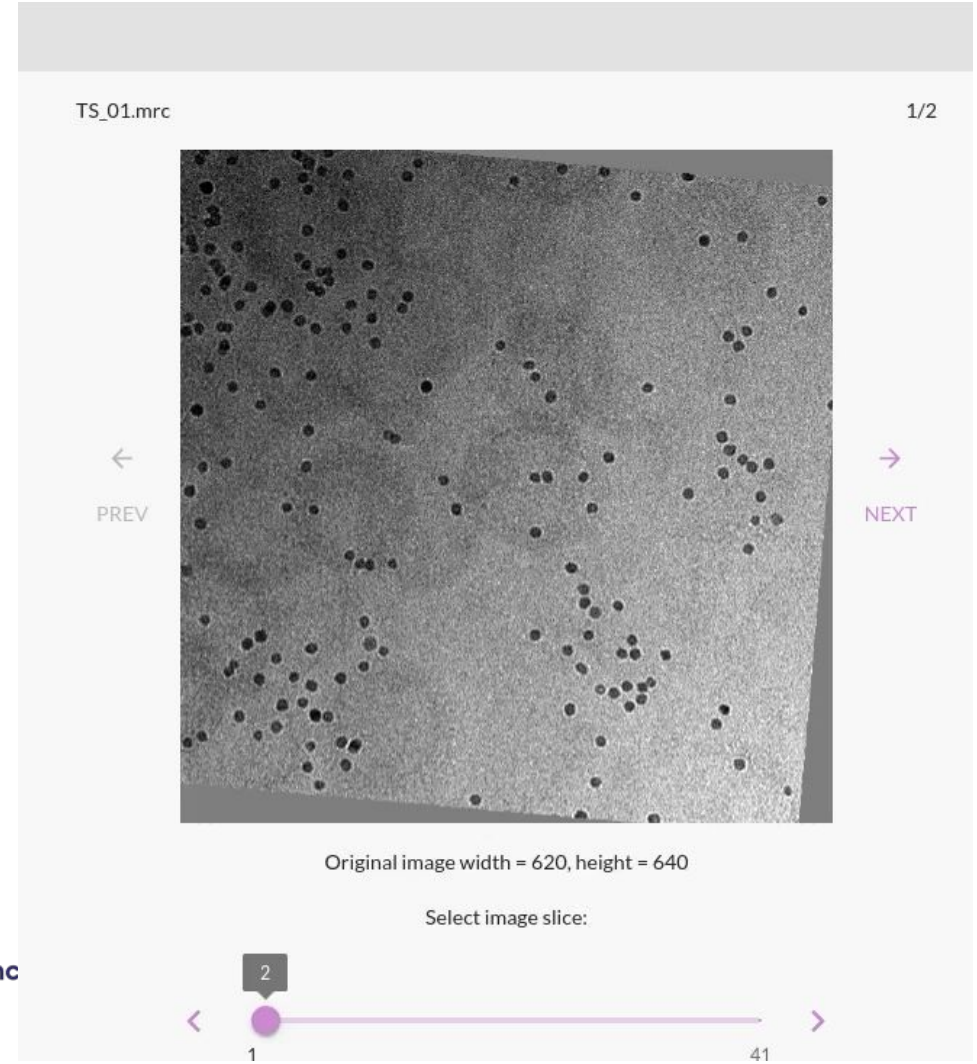


Science and Technology Facilities Council

Doppio: tomography and STA

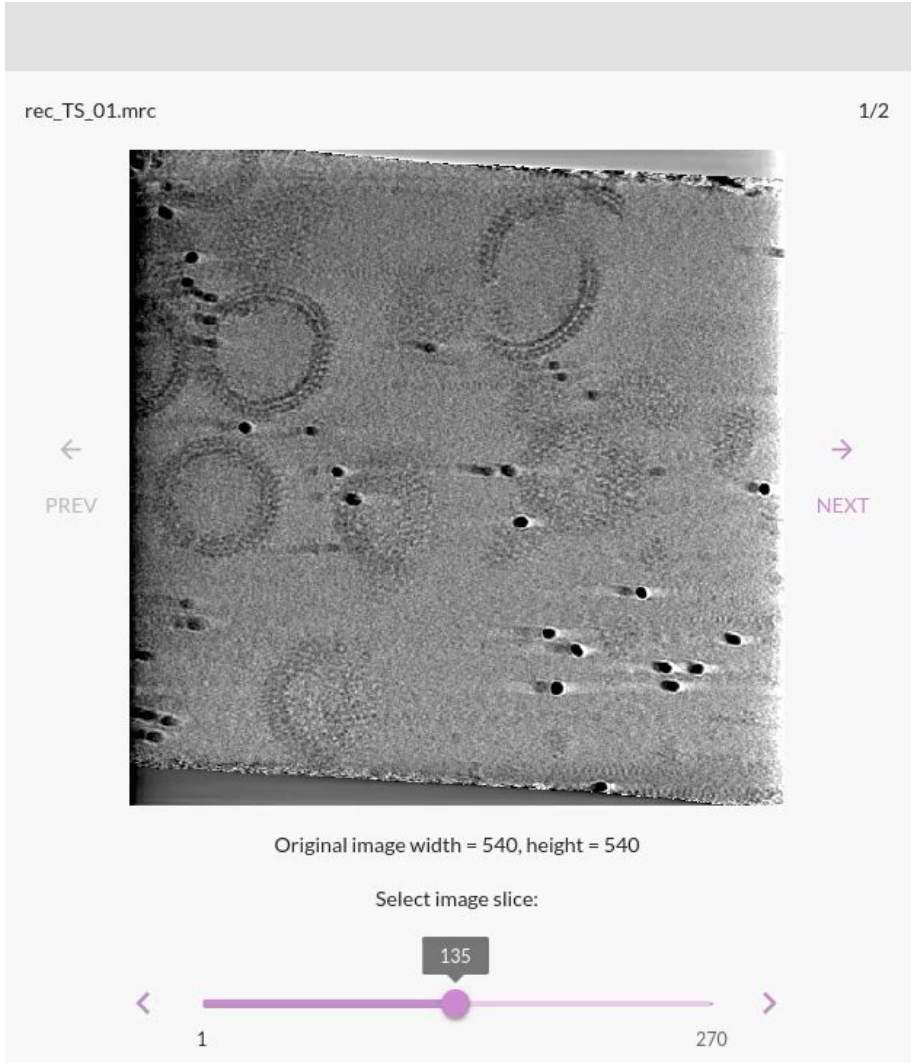
Alignment

Inspect shifts, angles, and quality of tilt series alignment



Tomogram Inspection

Evaluate reconstruction quality and look for artifacts



Science and Technology Facilities Council

Doppio: heterogenous reconstruction

Many heterogeneous reconstruction algorithms have now been developed

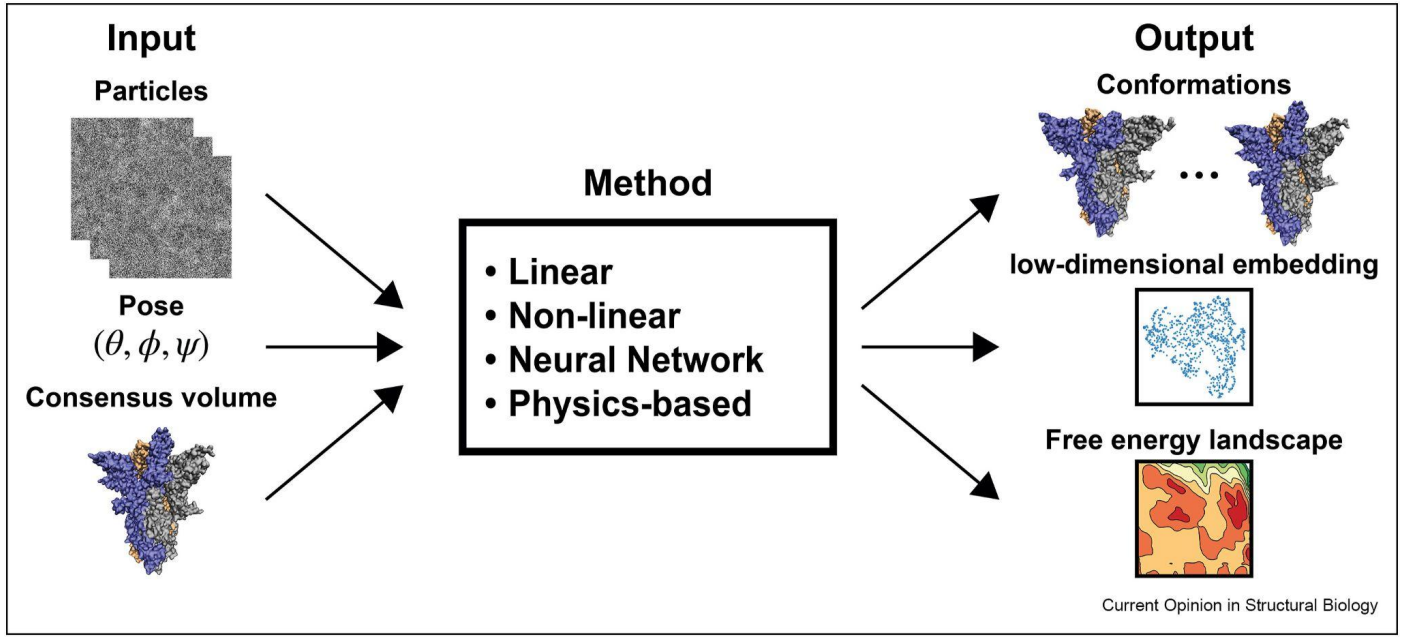
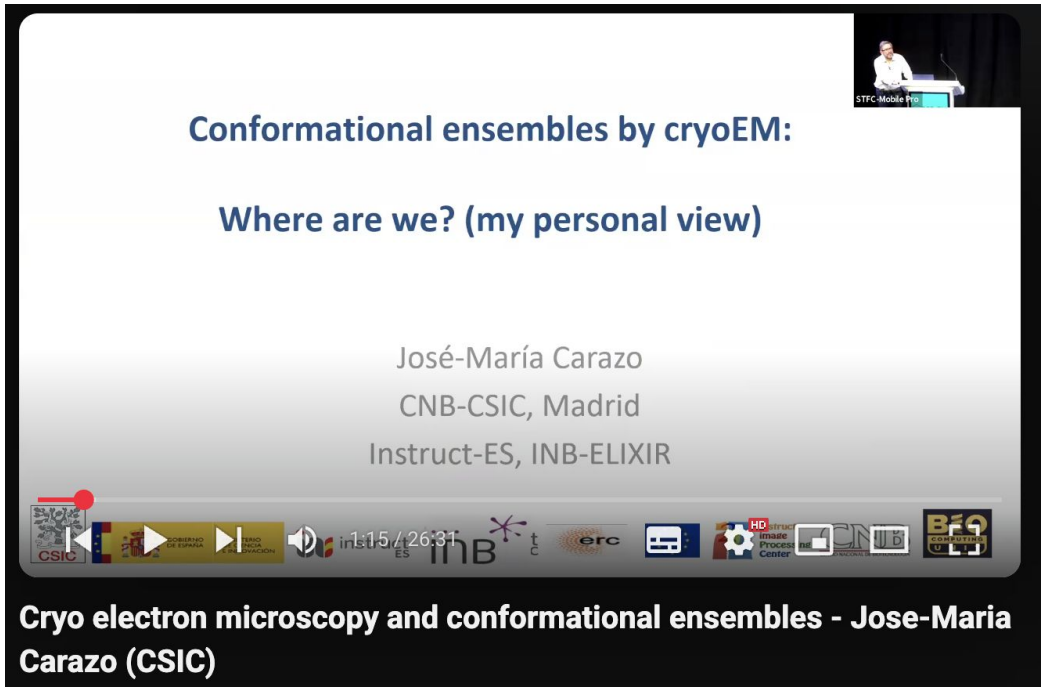


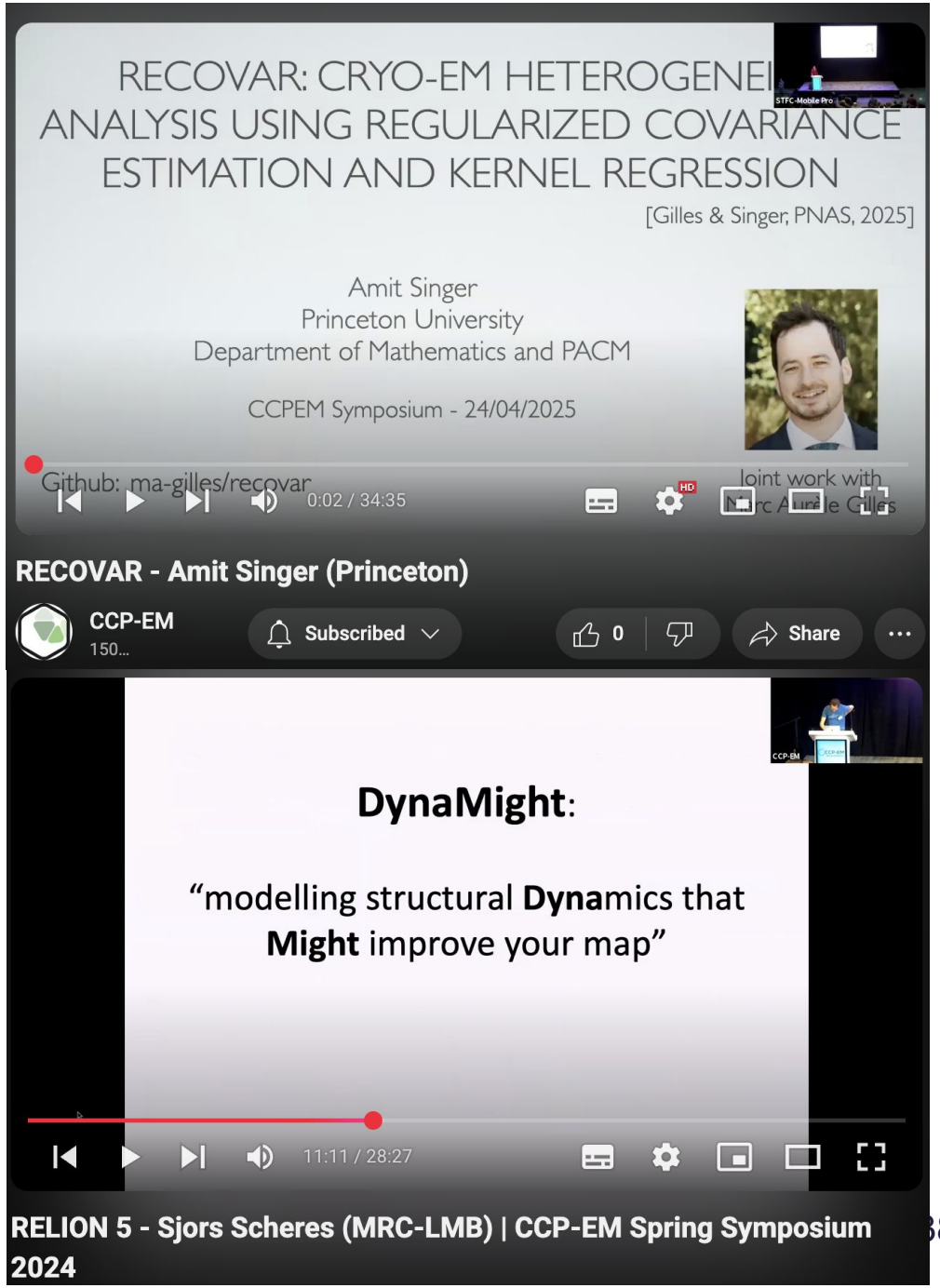
Fig 2, Tang et al., Current Opinion in Structural Biology, 2023

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



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



Symposium talks:

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

HRA jobs in Doppio

CryoDRGN (Ellen Zhong - Princeton):

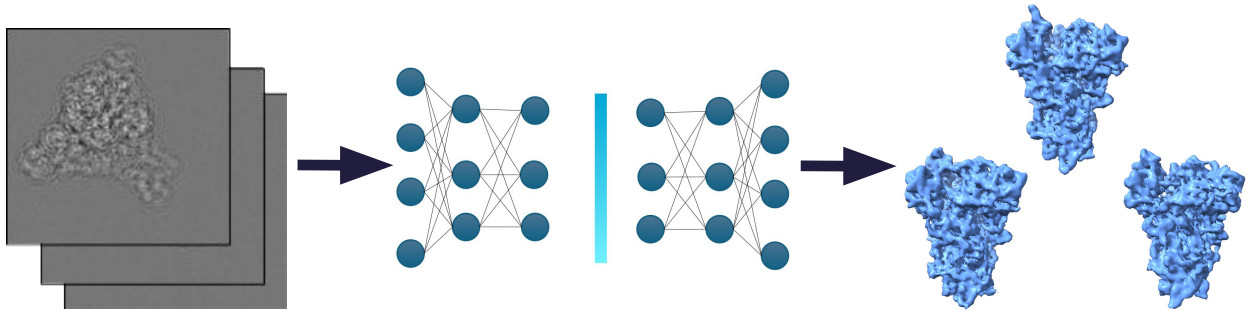
- CryoDRGN 1 – heterogeneous reconstruction from refined particles
- CryoDRGN 2 – *ab initio* homogeneous and heterogeneous reconstruction
- Interactive subset selection
- Neural network based

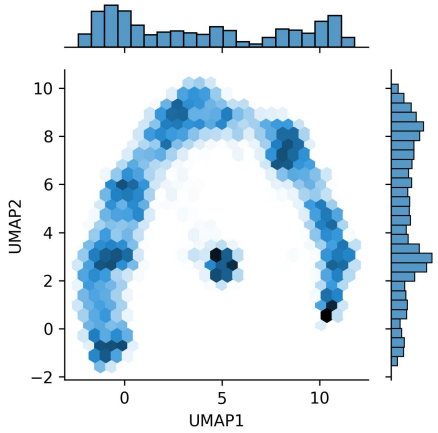
DynaMight (Johannes Schwab, Sjors Scheres - MRC-LMB)

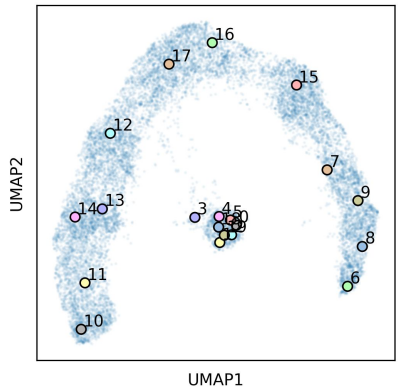
- Neural network based
- Built-in error estimation

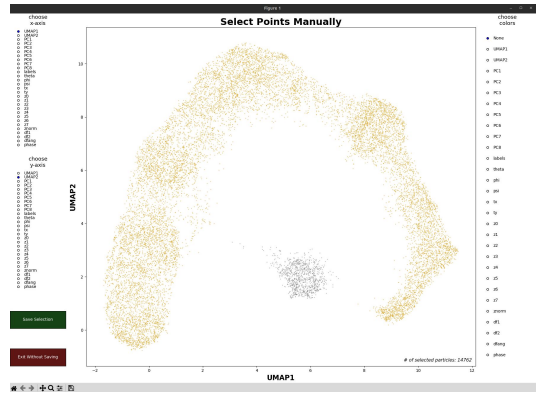
Adding RECOVAR (Amit Singer - Princeton) is WIP:

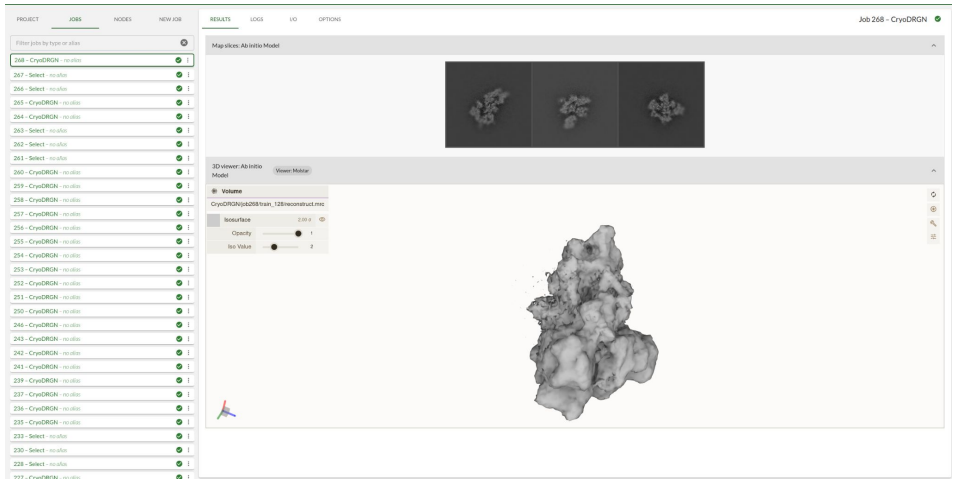
- Performs consistently well across CryoBench benchmark datasets
- Linear approach











Science and Technology Facilities Council

CCP-EM

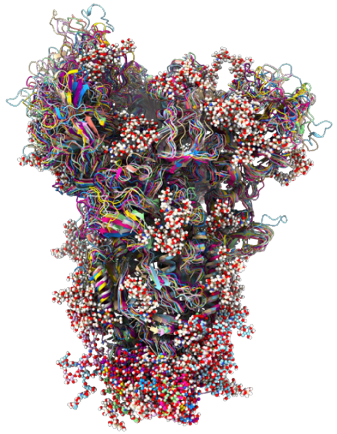
CoSeC

Heterogeneity Challenge

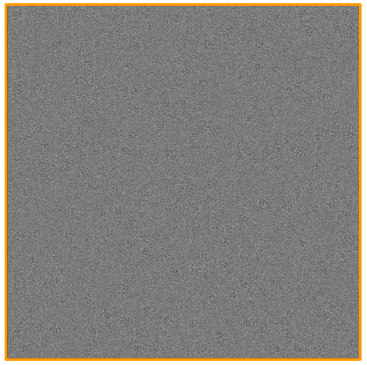
We are going to run a CASP-inspired **heterogeneity community challenge**

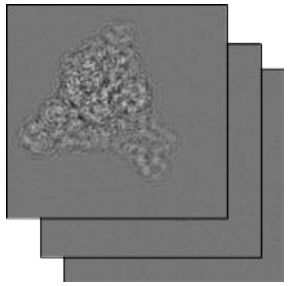
- Assess how faithfully heterogeneity present in cryoEM images is recovered
- Facilitate the development and evaluation of **heterogeneous reconstruction algorithms (HRAs)**
- Drive the development of metrics

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



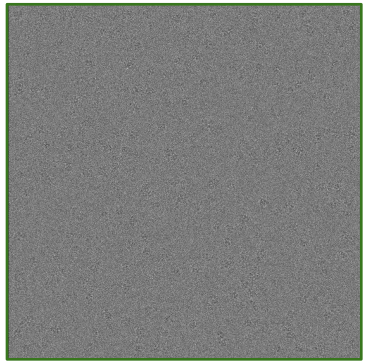
Synthetic micrograph generation

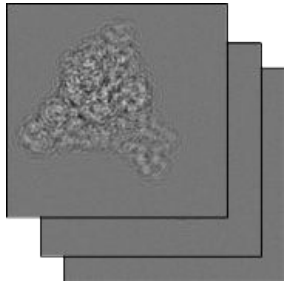




Particle stack datasets

Data collection



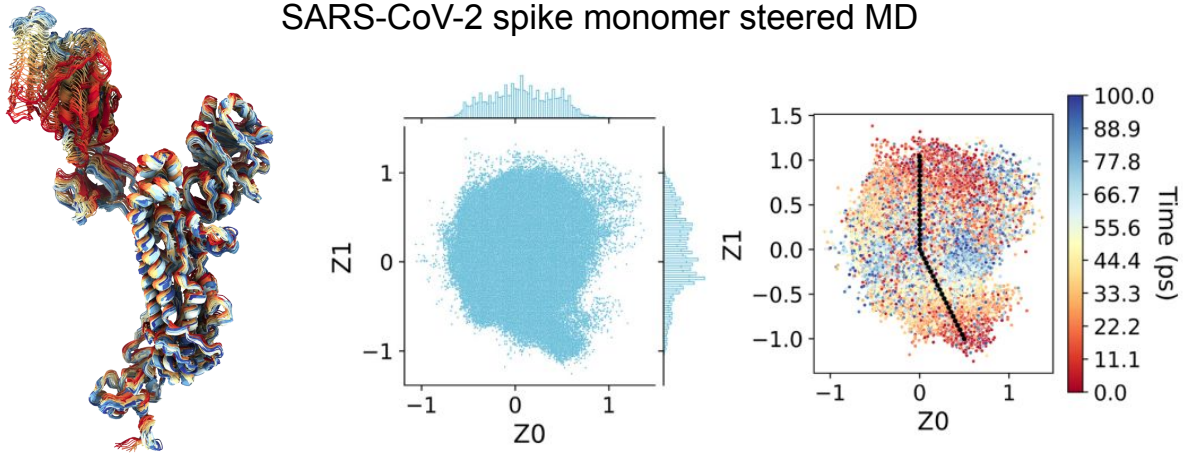


Recovering Heterogeneity Present in Cryo-EM Datasets

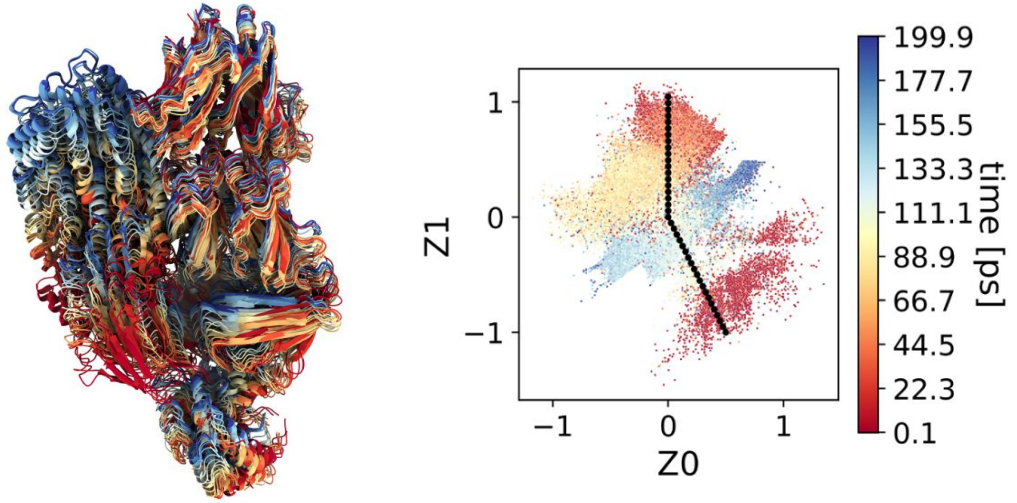
Interpreting the results from HRAs can be challenging:

- 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

SARS-CoV-2 spike monomer steered MD



Complement system protein C3 steered MD



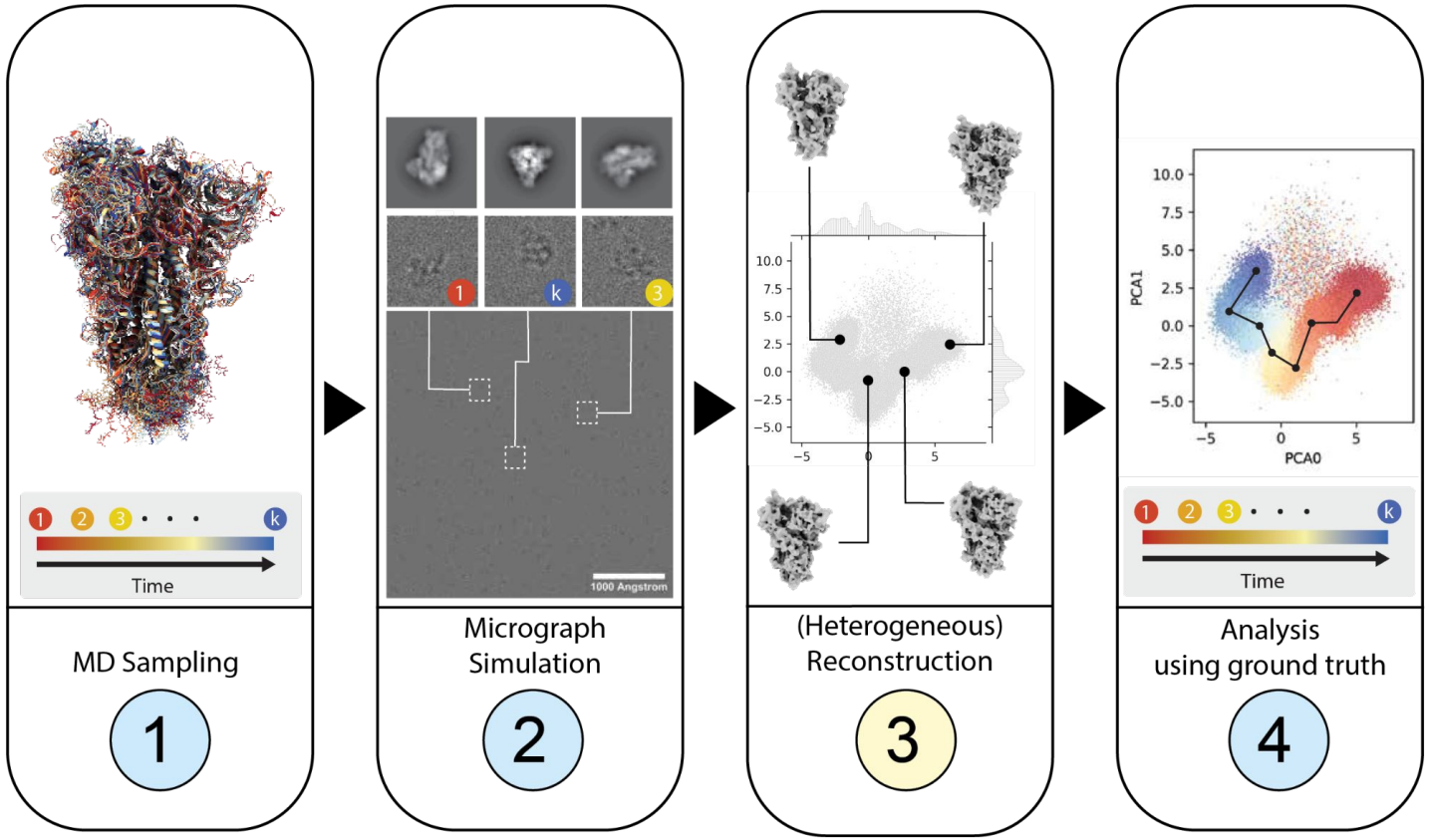
Science and Technology Facilities Council

CCP-EM

CoSeC

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

Science and Technology Facilities Council

CCP-EM CoSeC TU Delft

IUCrJ, 11:6, 2024

Inaugural Flatiron Community Challenge


Pilar Cossio


Sonya Hanson

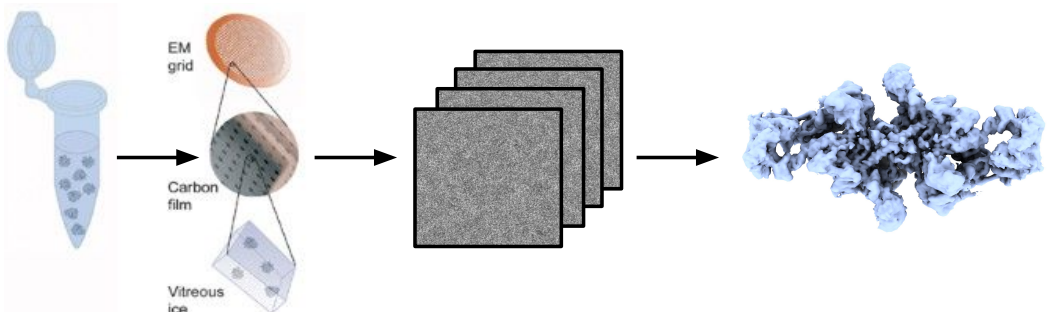
Launched in June 2023

- 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/

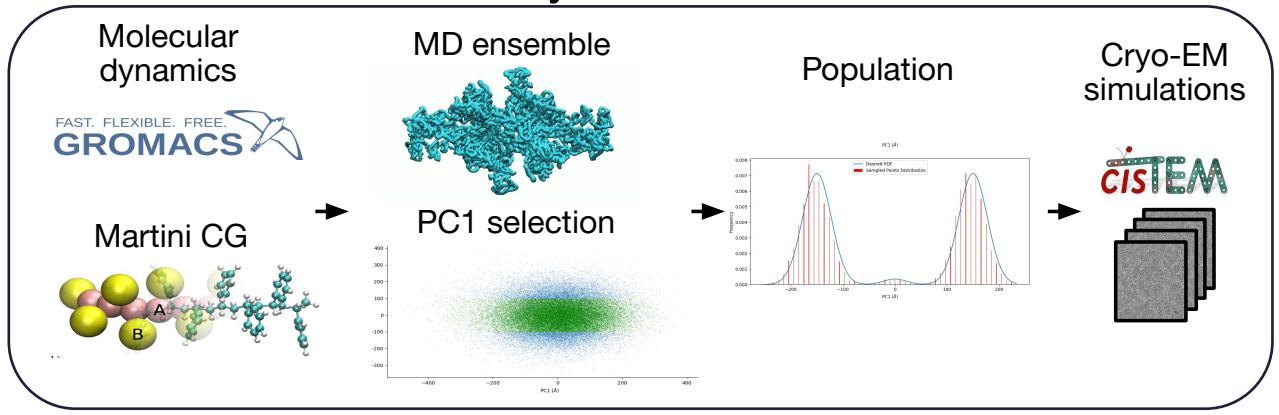
Highlighted the difficulty of:

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

Experimental



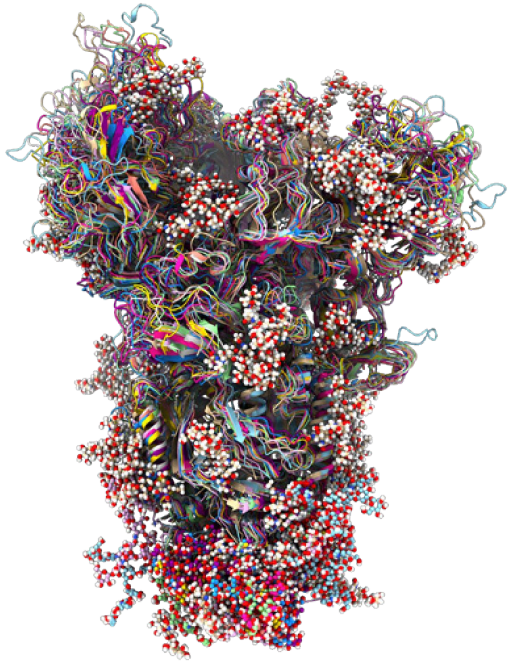
Synthetic



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

Engineering and Physical Sciences Research Council

CCP-EM

Research Complex at Harwell

FLATIRON INSTITUTE

CoSeC

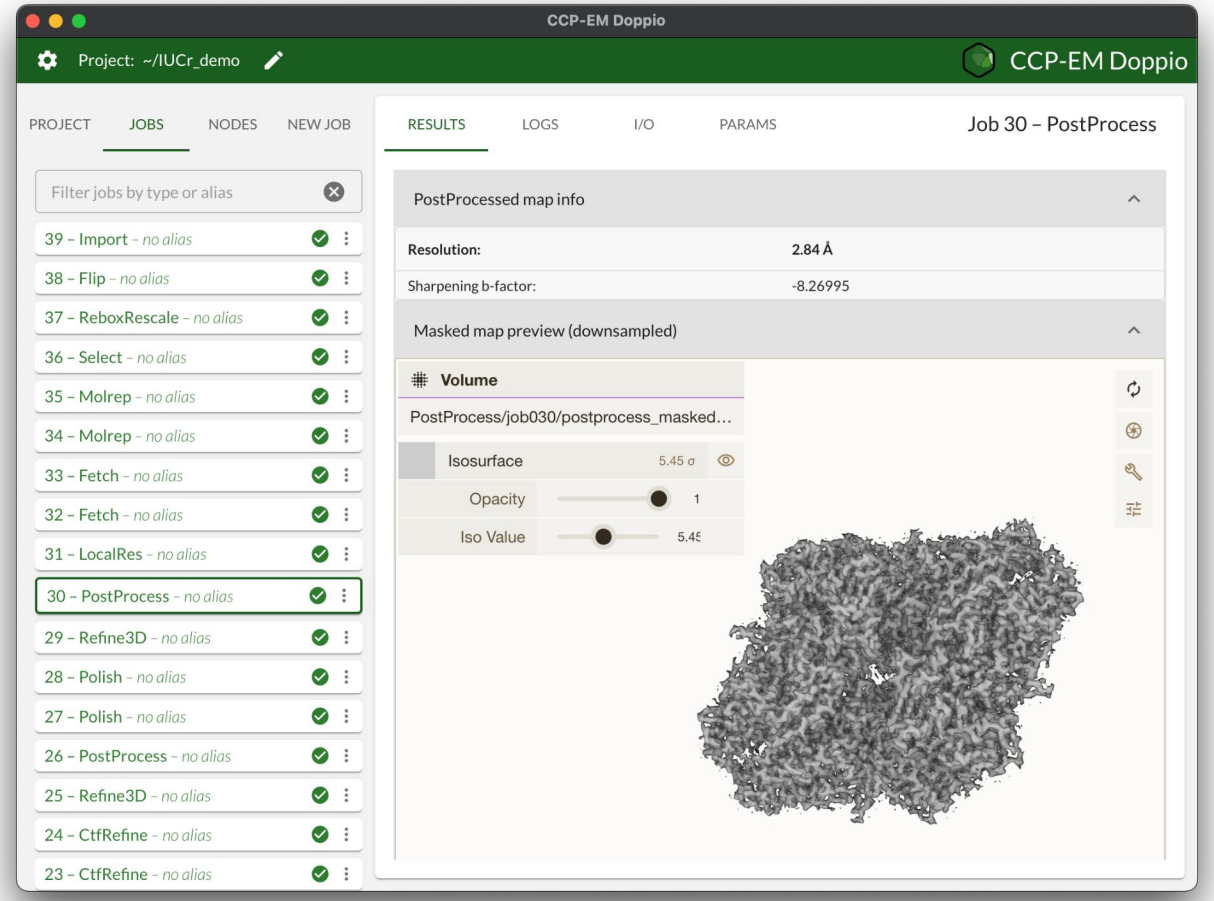
HECBioSim

CCPBioSim

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



Science and Technology Facilities Council

www.ccpem.ac.uk

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

Mailing list: www.jiscmail.ac.uk/ccpem

Bluesky: ccpem.bsky.social

Email: ccpem@stfc.ac.uk

Science and Technology Facilities Council
Medical Research Council
Research Complex at Harwell
wellcome
University of Pittsburgh
Department for Science, Innovation & Technology
GΦL Global Phasing Limited
ada lovelace centre
CoSeC

Practical Tutors

Tom Burnley

Colin Palmer

Rob Nicholls

Arjen Jakobi

Science and Technology Facilities Council

MRC Laboratory of Molecular Biology