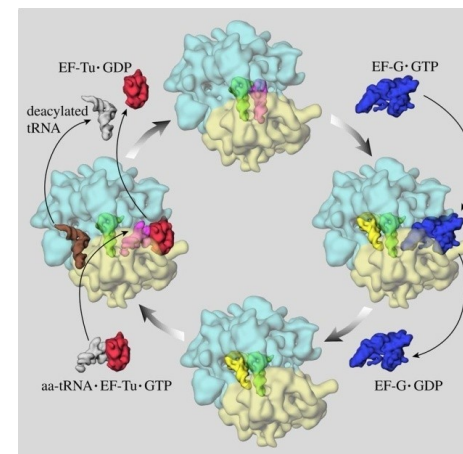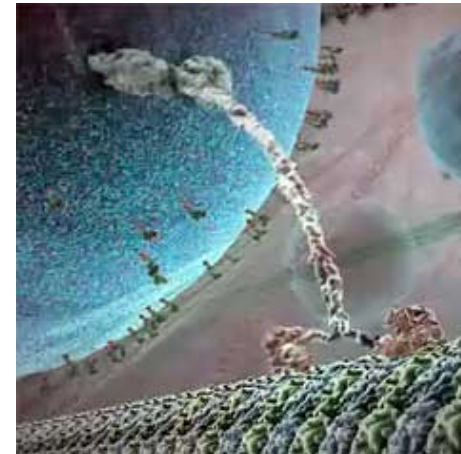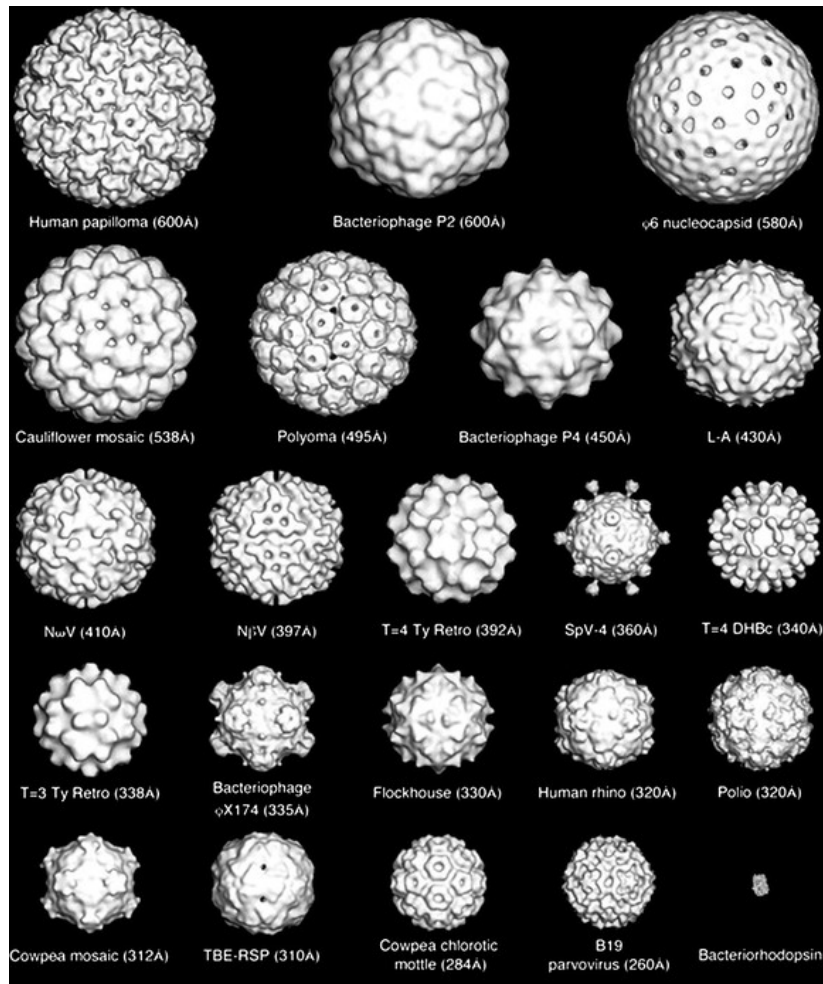# Facing Computing and Data Challenges in Electron Microscopy:
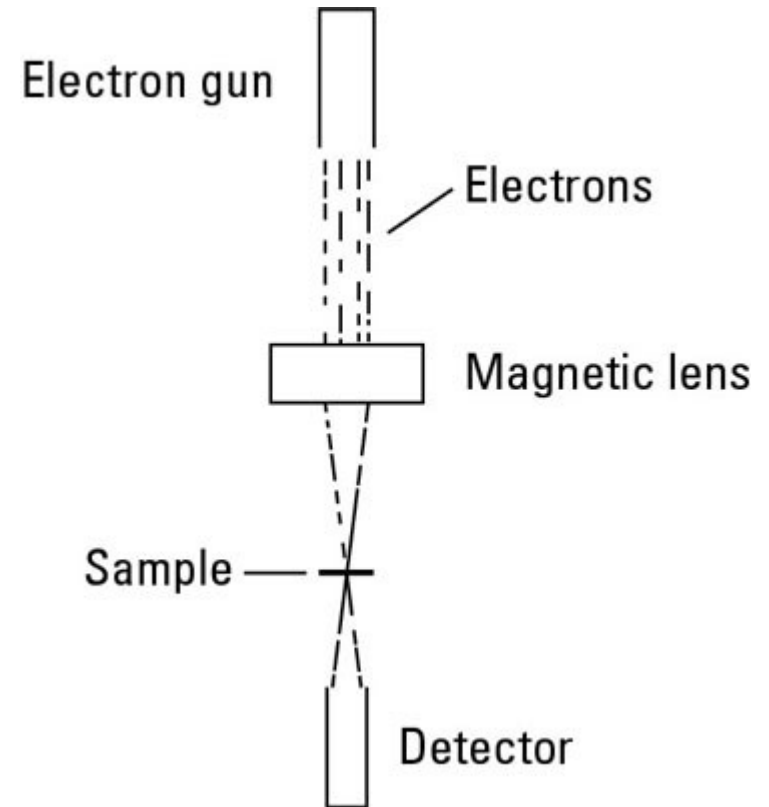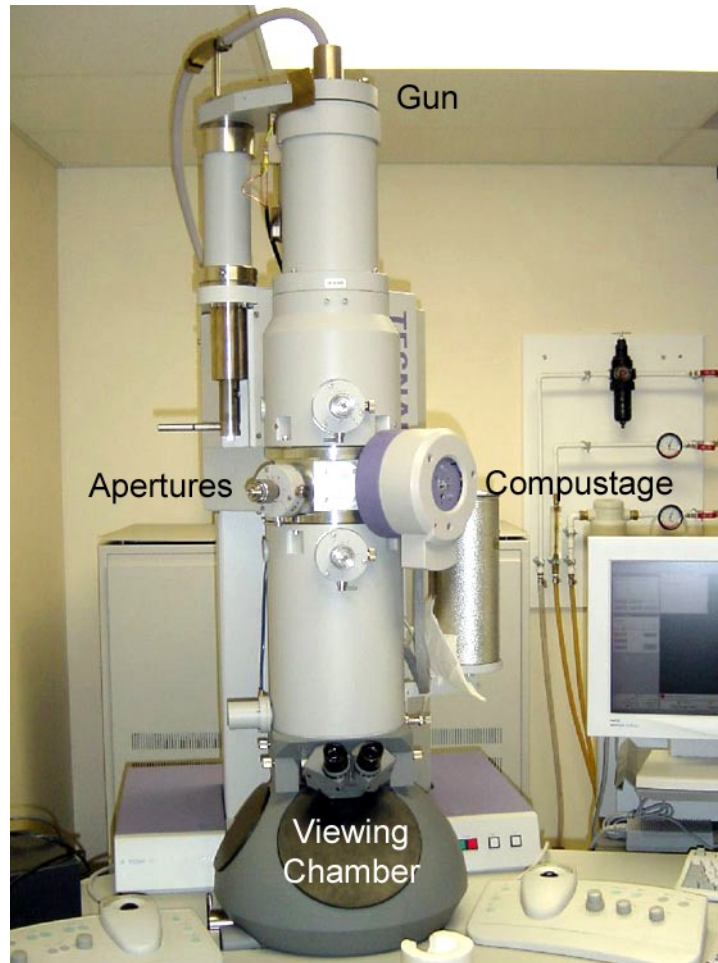
The **SCIPION** Software Framework

Biocomputing Unit,
Instruct Image Processing Center, CNB-CSIC
J.M. de la Rosa Trevín

Life is based on Molecular Machines
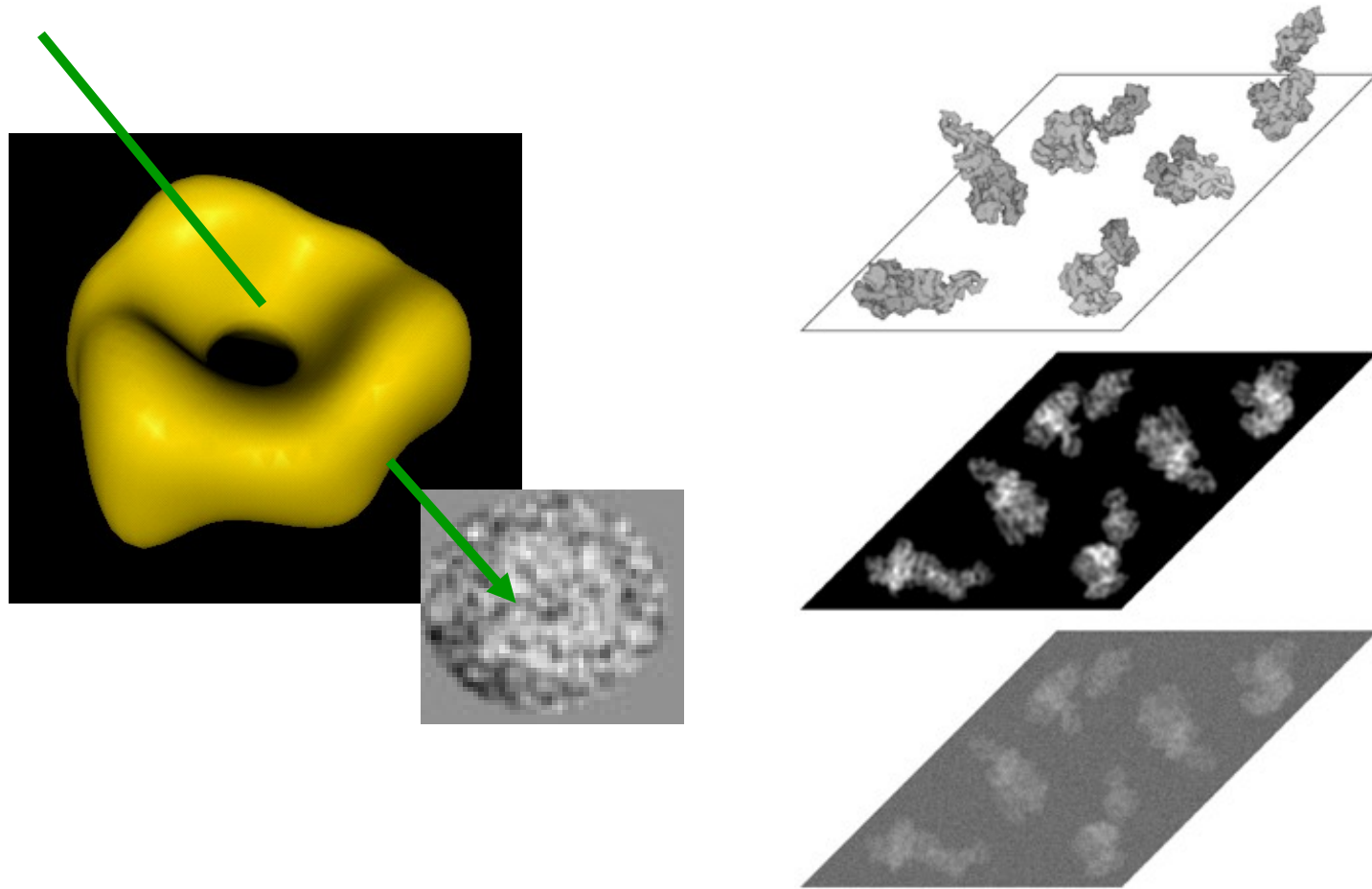
Our microscope use electrons to form images
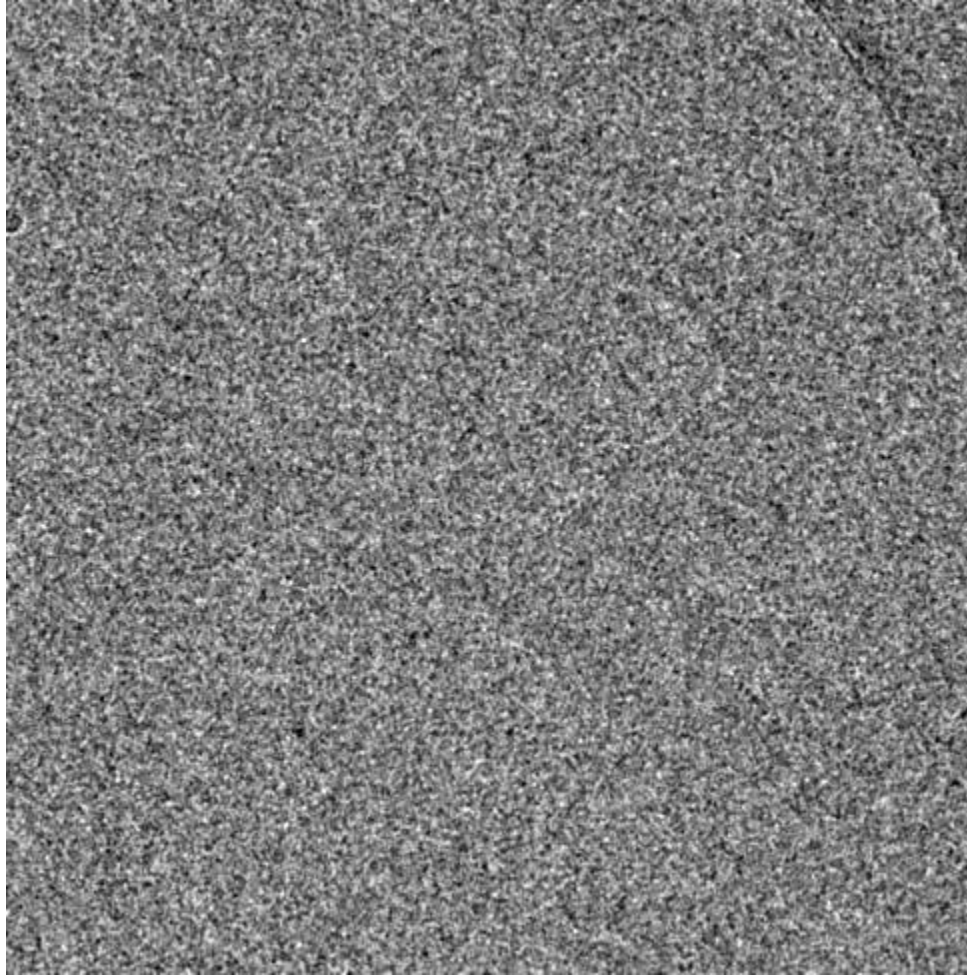
Microscopy can be fancy (and expensive!)

We use a low electron dose

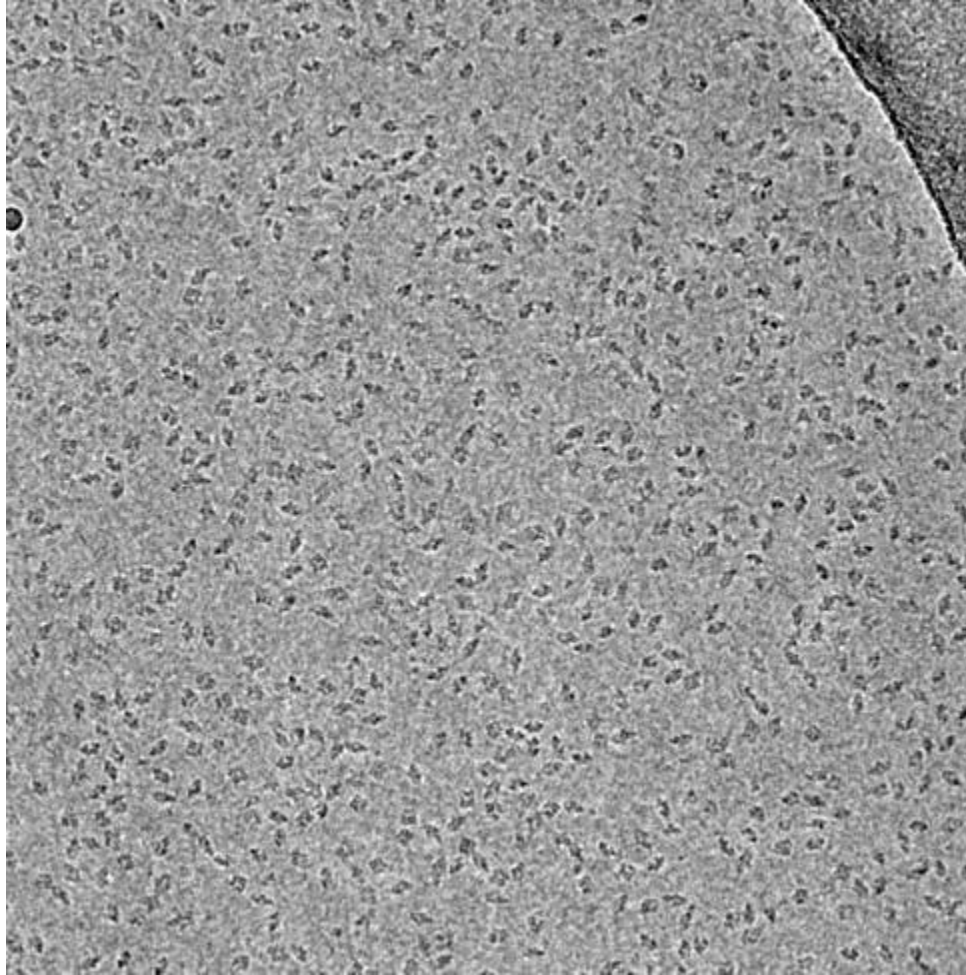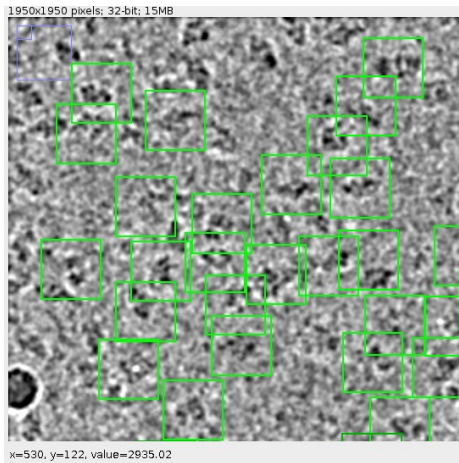# EM-images biological are extremely noisy

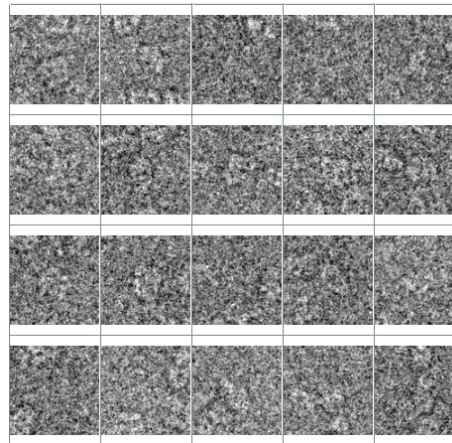# We need to 'average' to extract some signal

A long path from 2D images to 3D volumes
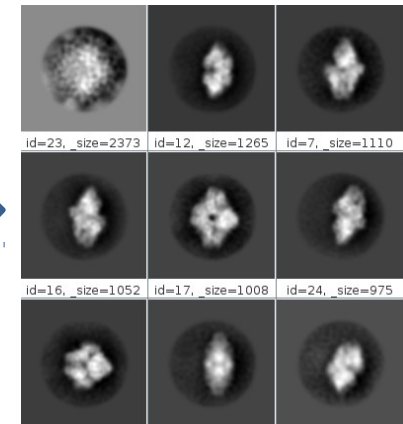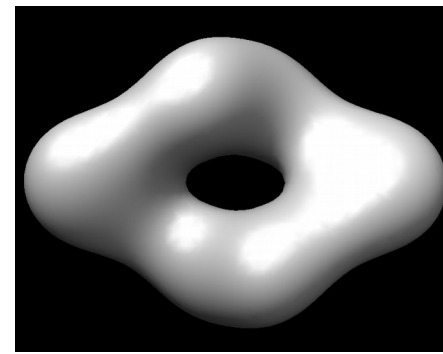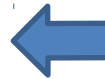
Particles selection

Extraction

2D classification

Final 3D map
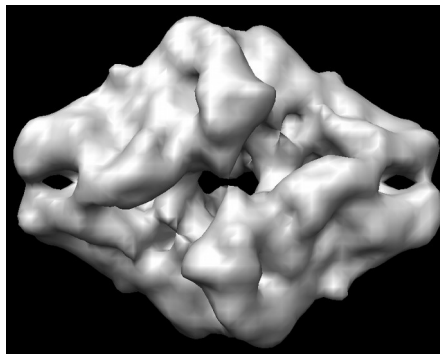
Initial volume

Software Challenges:
Interoperability, Reproducibility and Automation

The EM field needs software integration

Using different EM software packages is now like the tower of Babel
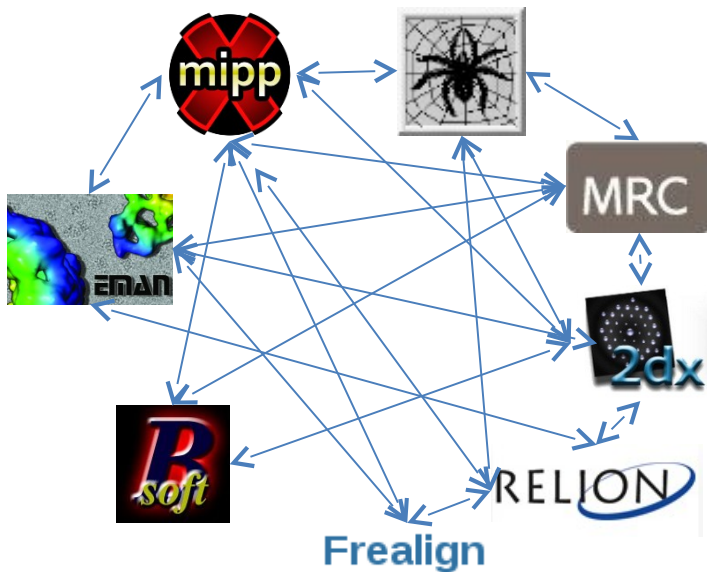
Scipion Principles

1. Integrate EM software packages to be used in the same project.

2. Full project traceability, improving reproducibility.

3. Execute complete workflows in an automated manner.

4. Easy to install and use.

5. Easy to extend with new protocols.

We created a model of the EM domain
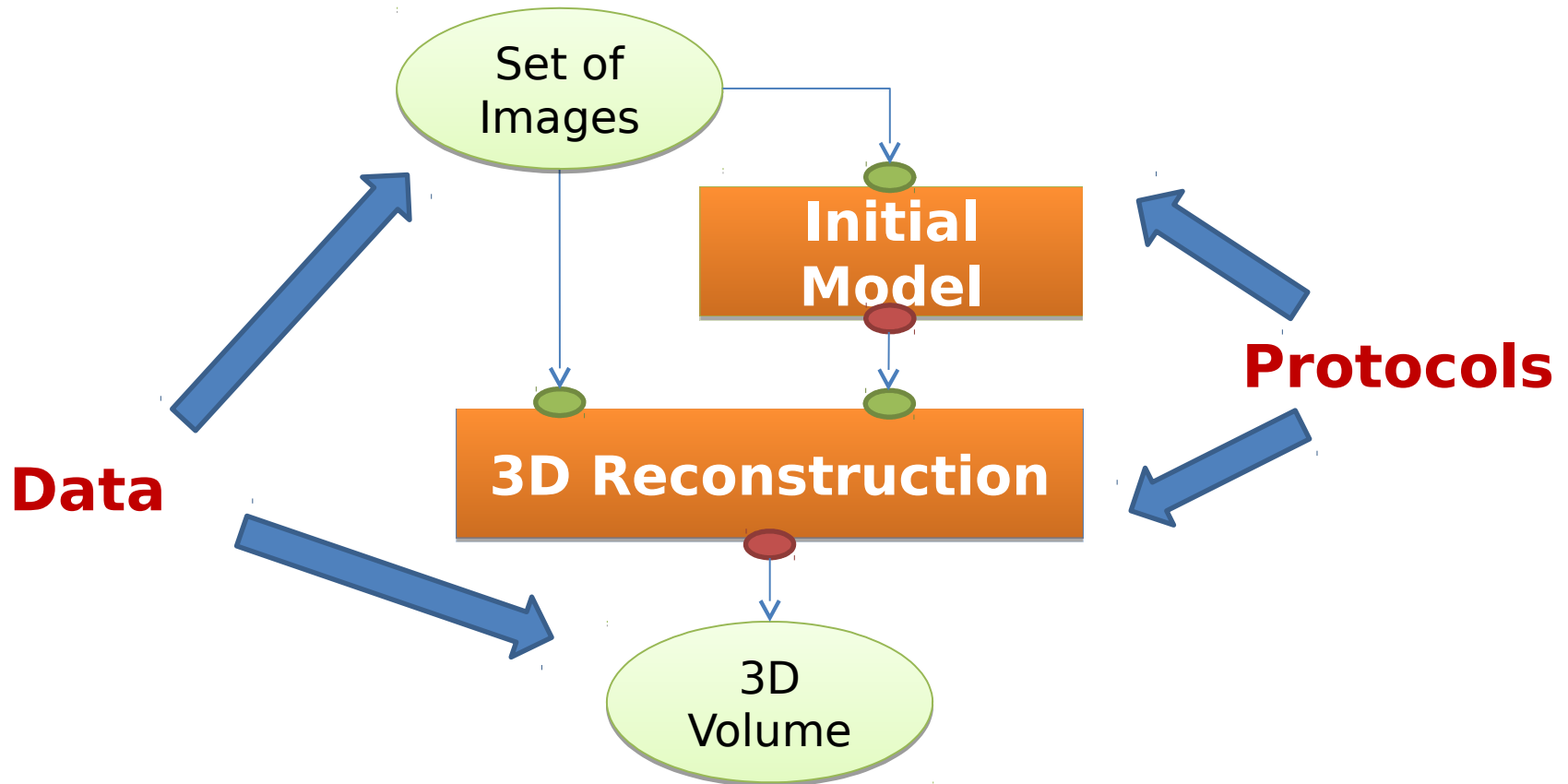
Conversion functions bridge between packages
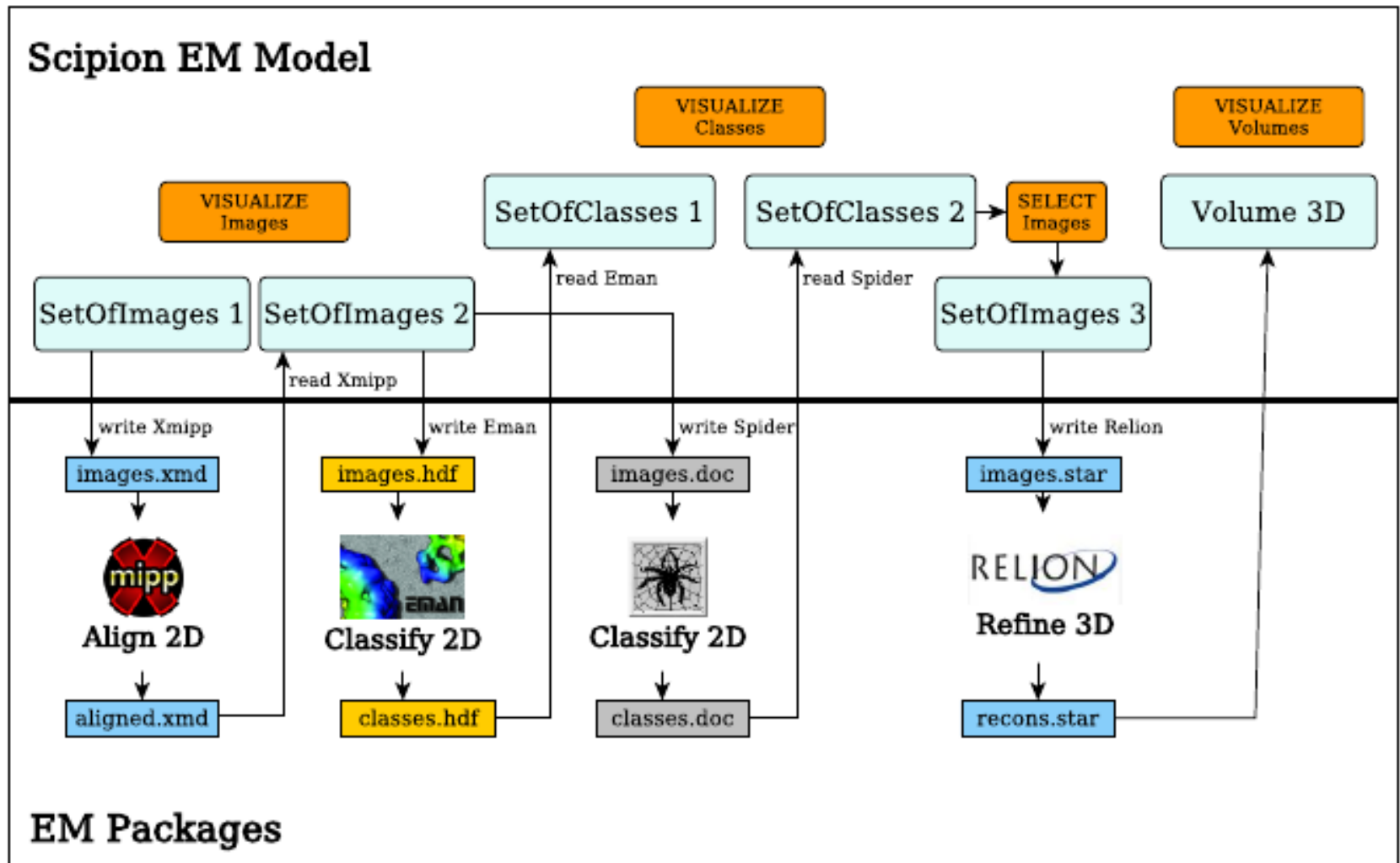
We track all the steps performed in a project
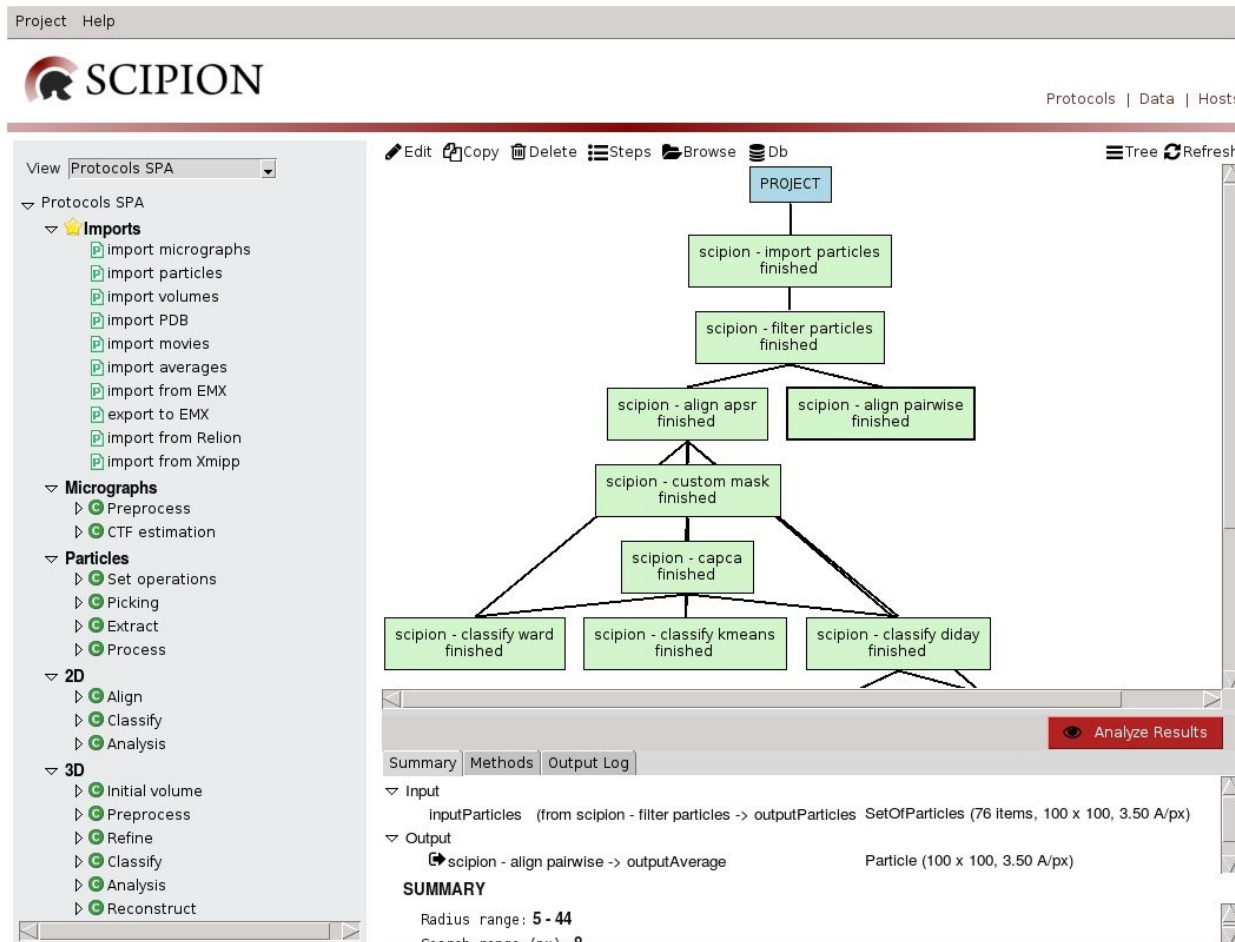
All parameters are also stored

Data Challenges: Accessibility, Transparency

Avoid data duplication as much as possible

2D classification: Only store alignment parameters
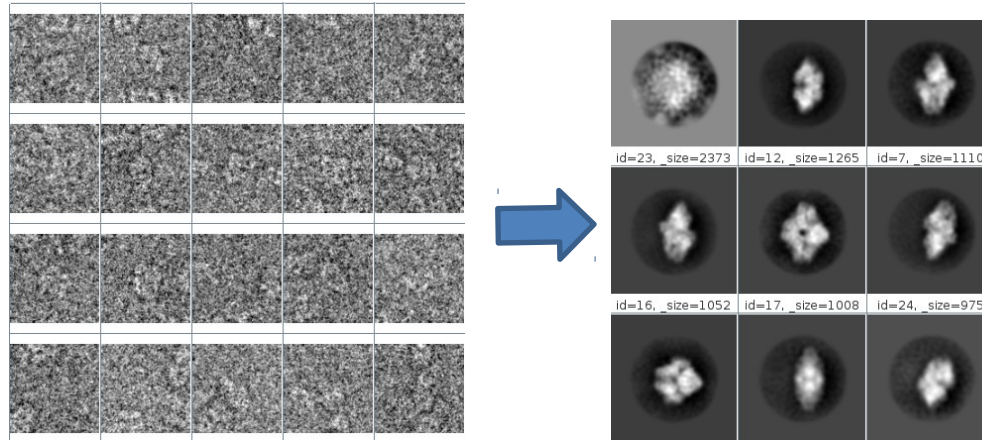
New 3D objects stored per iteration
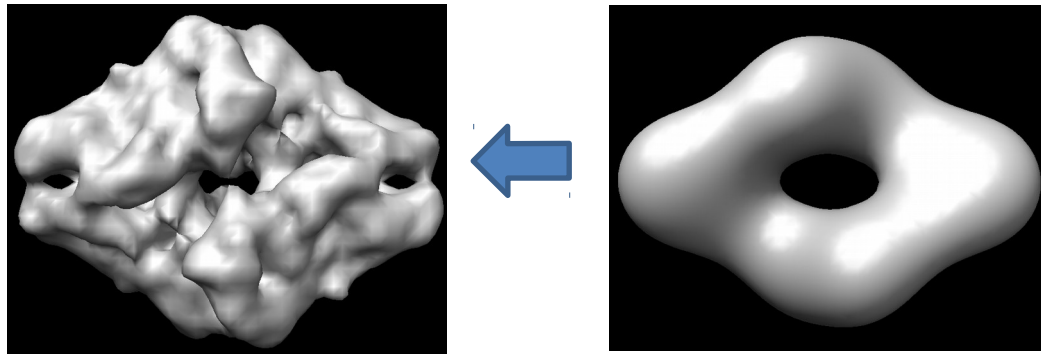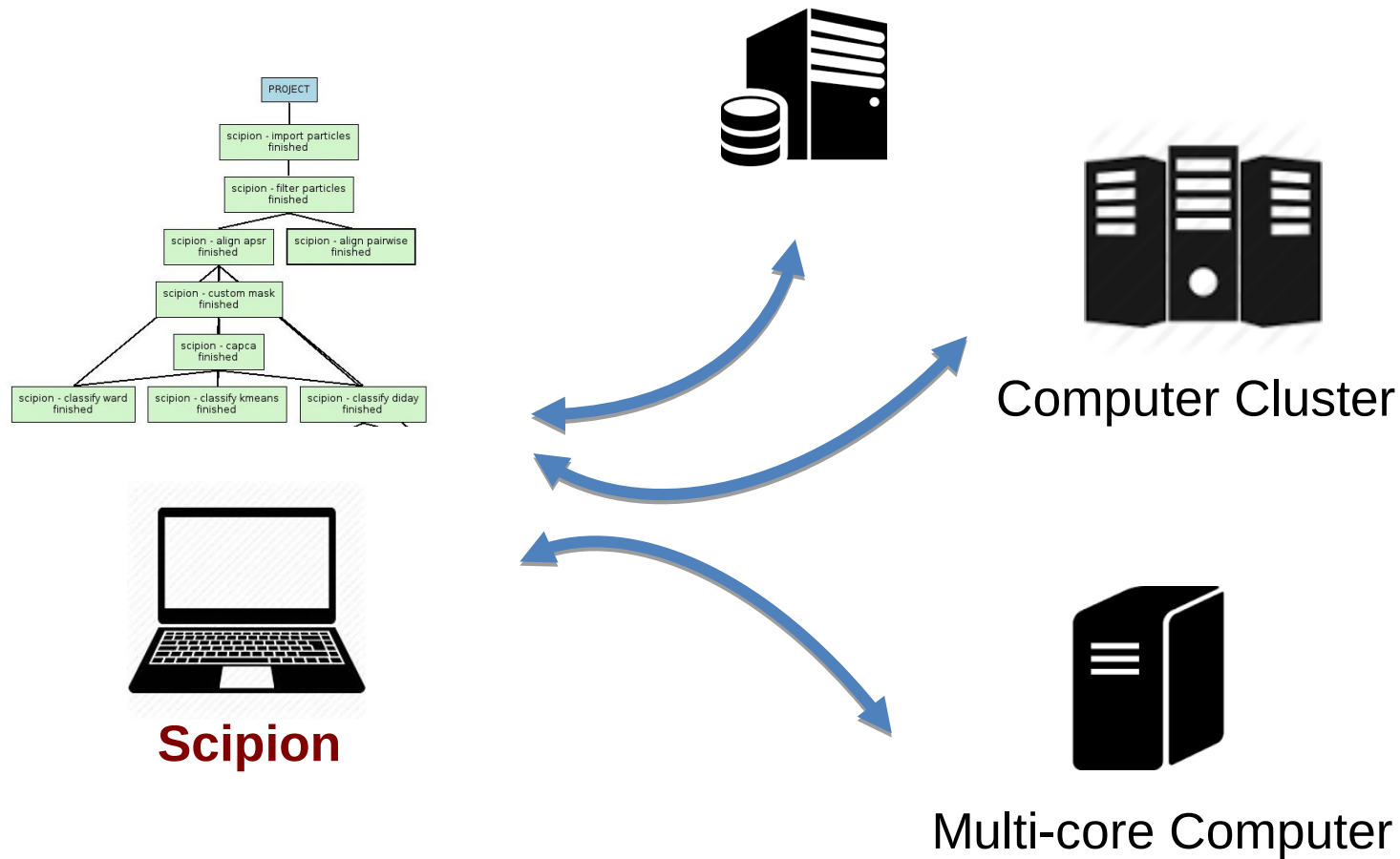
Distribute the processing workflow, dealing with heterogeneous computing requirements

Scipion

Computer Cluster

Multi-core Computer

Distribute the processing workflow, dealing with heterogeneous computing requirements

Computer Cluster

Scipion

Multi-core Computer
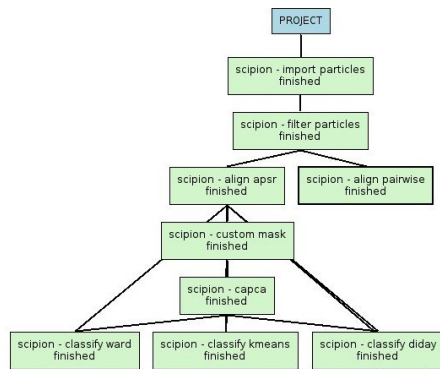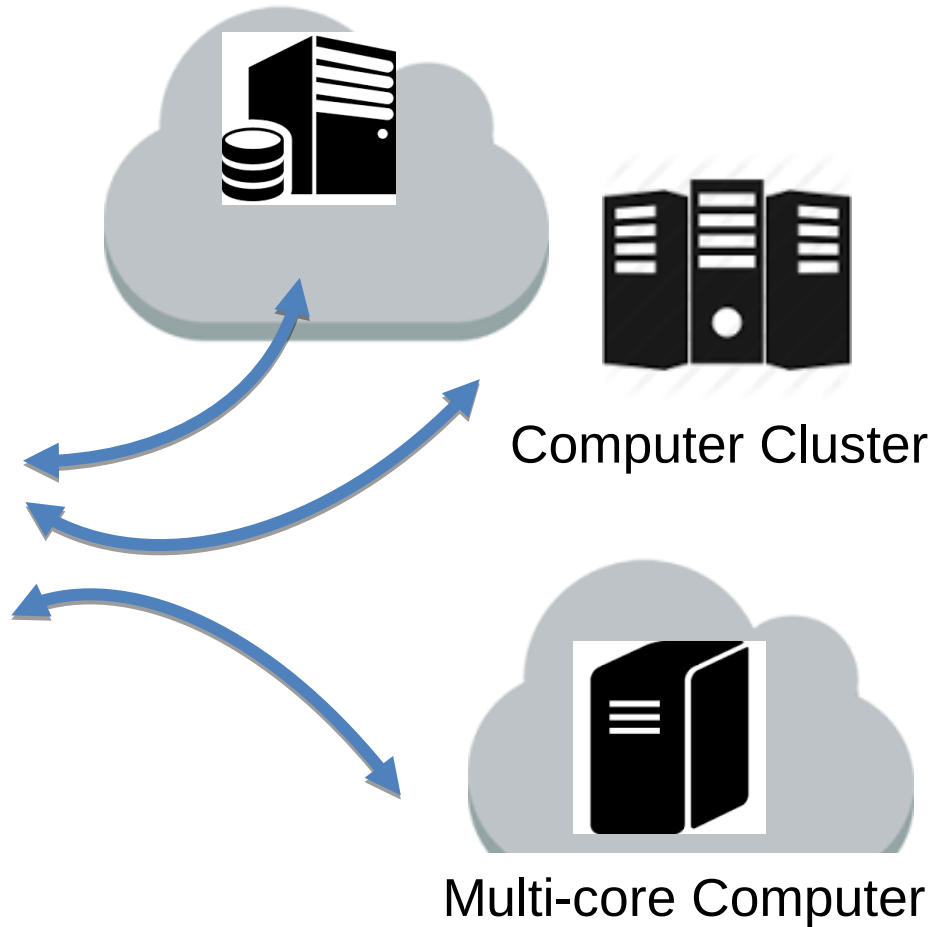
Most of individual jobs involves large size input/output datasets

Import Movies	**~2 Tb**
↓	
Movie Alignment	**~90 Gb (4)**
↓	
CTF Estimation	**~90 Gb (3)**
↓	
Particle Picking	**~20 Mb (3)**

Extract particles	**~22 Gb (2)**
↓	
2D Classification	**~3 Gb (2)**
↓	
3D Classification	**~7 Gb (4)**
↓	
3D Refinement	**~18 Gb (6)**
↓	
Post-process	**~5 Gb (5)**
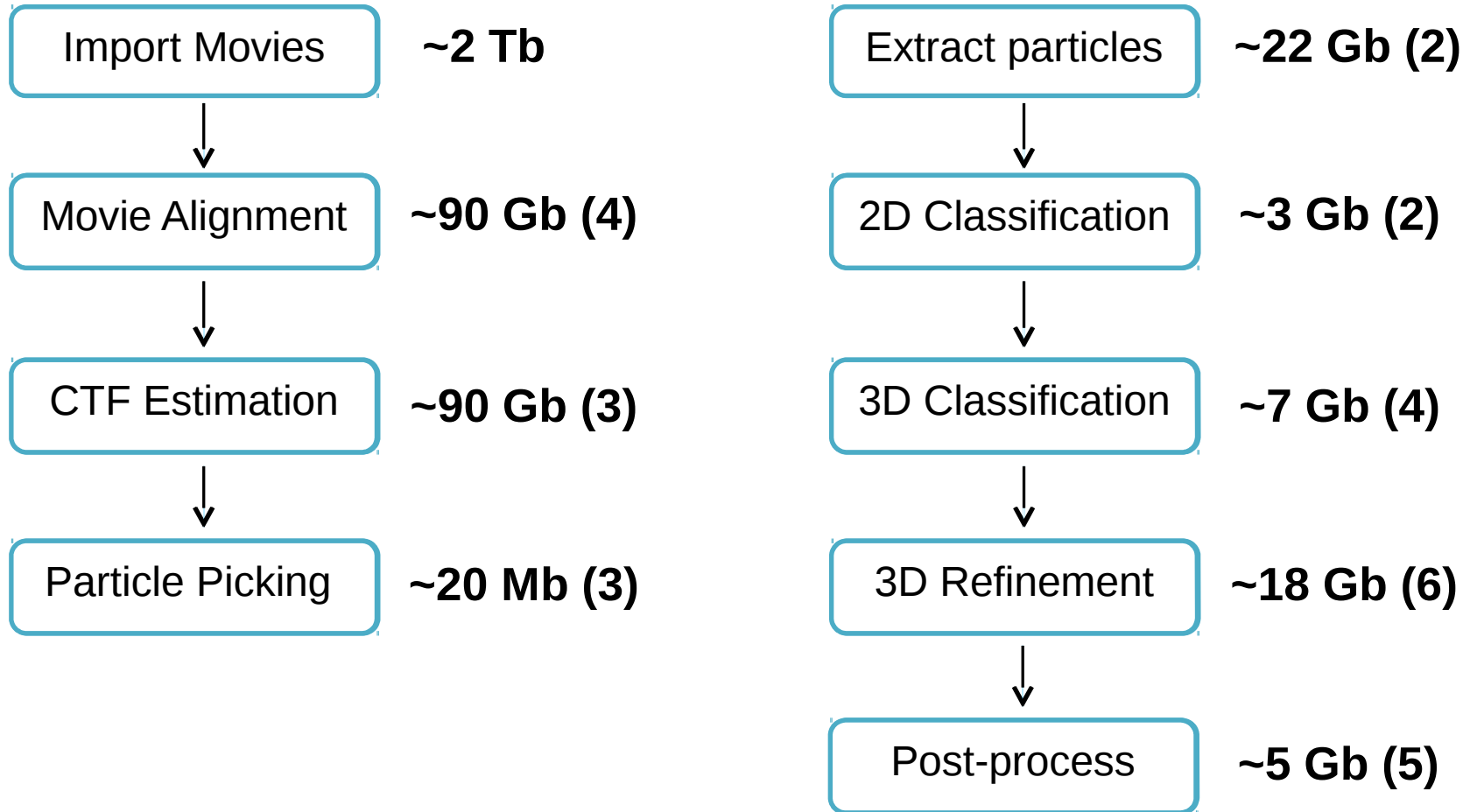
Preprocess data close to the Source

Scipion Box

Microscope

Import Movies

Movie Alignment

CTF Estimation

Particle Picking ???
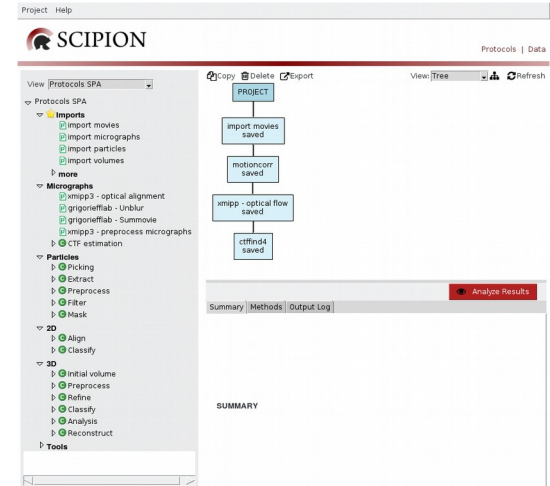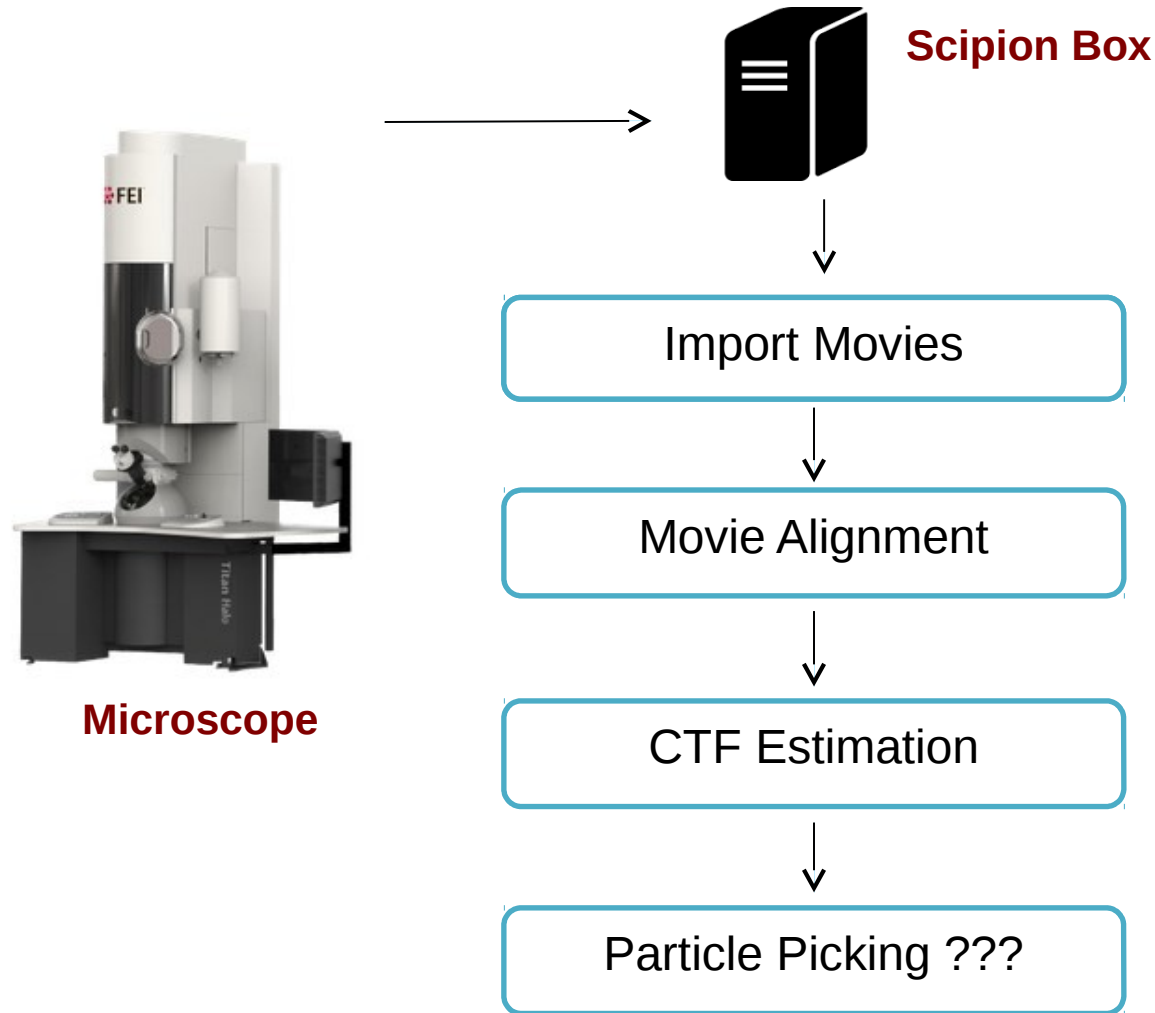
Run workflows automatically in streaming

Acquisition
~60s (4k x 4k x 34) -2Gb HD

Scipion Box

Transfer
~16s - 1Gb/s
~4s – 4Gb/s
~1.6 – 10Gb/s

Microscope

Import Movies

Movie Alignment

MotionCorr
~14s – (22 frames)
2 Cores
2Gb RAM
0.4 Gb VRAM
GPU

Optical Flow
~660s – (22 frames)
1 Core
2Gb RAM
12 Cores < 60s

CTFFIND4
~17s
1 Core
1Gb RAM
gCTF?

Particle Picking ???
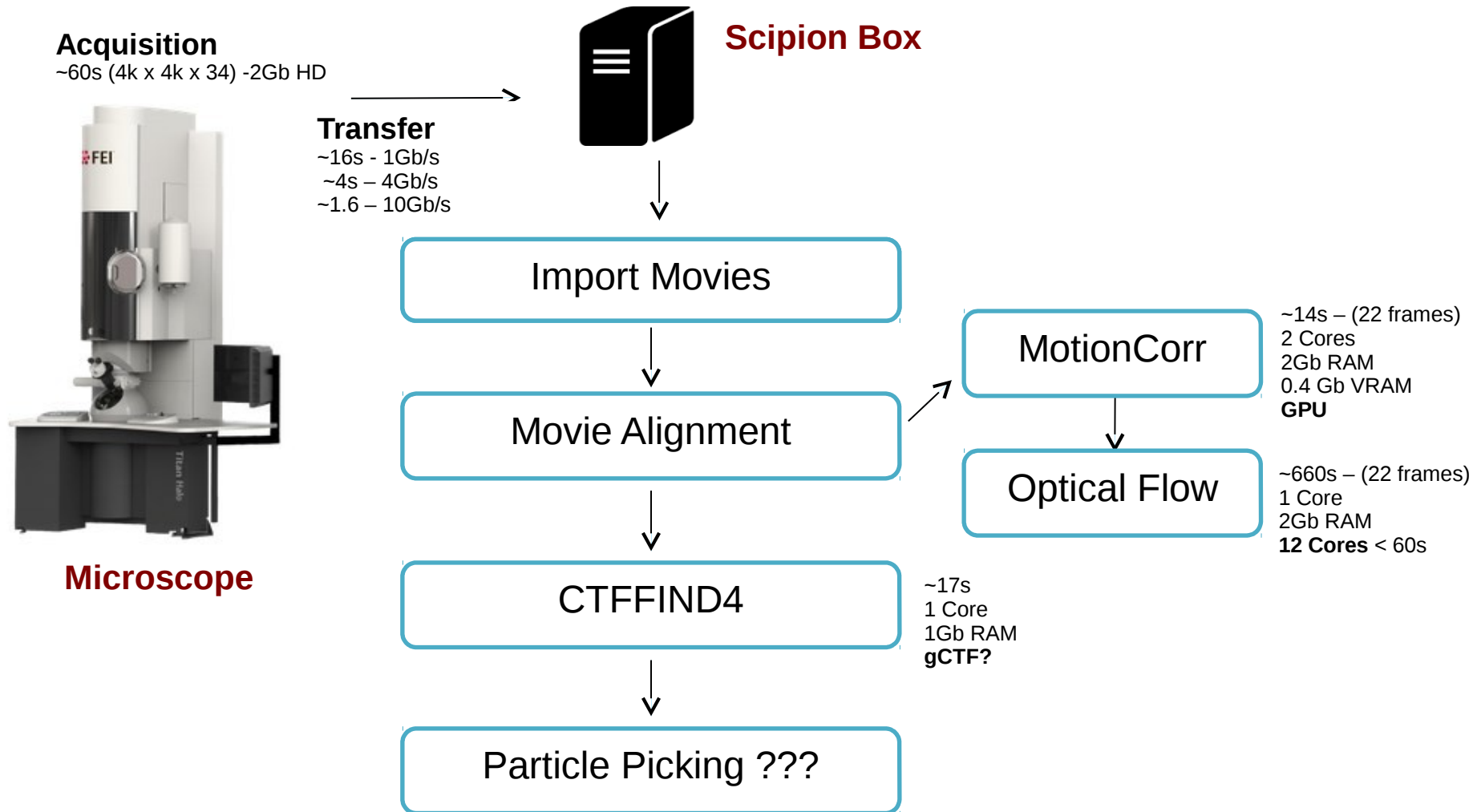
Scipion is open-source and freely available
http://scipion.cnb.csic.es

Instruct Centres > Centre for Image Processing

Instruct Image Processing Center (I2PC)

Instruct Flagship Platform ☆

Personalized support for Image Processing project 🜂

The Instruct Image Processing Center provide personalized support to Instruct projects requiring specialized processing of either electron or X-ray microscopy images. Currently our main offer is focused in the field of single particles analysis using Xmipp and Scipion softwares. Xmipp is a comprehensive suite of image processing algorithms with a strong emphasis in the analysis of single particles, although it is extending towards electron and X-ray tomography. Scipion, in turn, provides a multi-package workflow-oriented platform bridging several software suites, like EMAN, Spider, Relion, Bsoft, Frealign and ctffind. More information can be found at: http://xmipp.cnb.csic.es and http://scipion.cnb.csic.es/docs

Instruct platform information ➜

Conclusions (EM?)

1. We need integrative software frameworks to combine different algorithms

2. Reproducibility should be a first priority task to move science forward.

3. Computing requirements are very heterogeneous (hardware and software).

4. Data Management is a very important issue.

5. We need to join efforts v(user, developers and infrastructures) through collaborative environments.

www.structuralbiology.eu
Follow us on twitter
@instructhub