

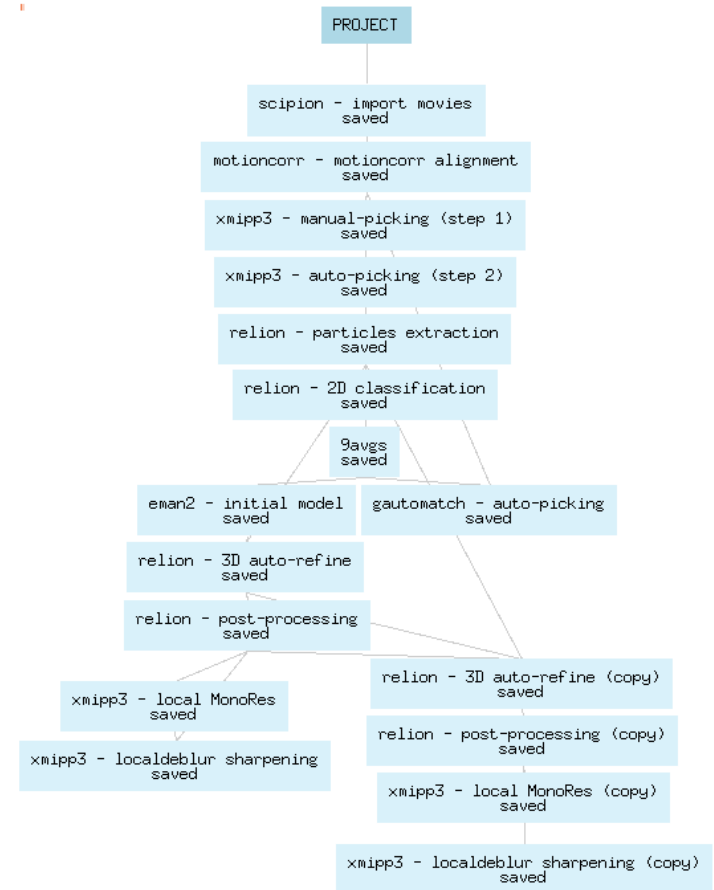
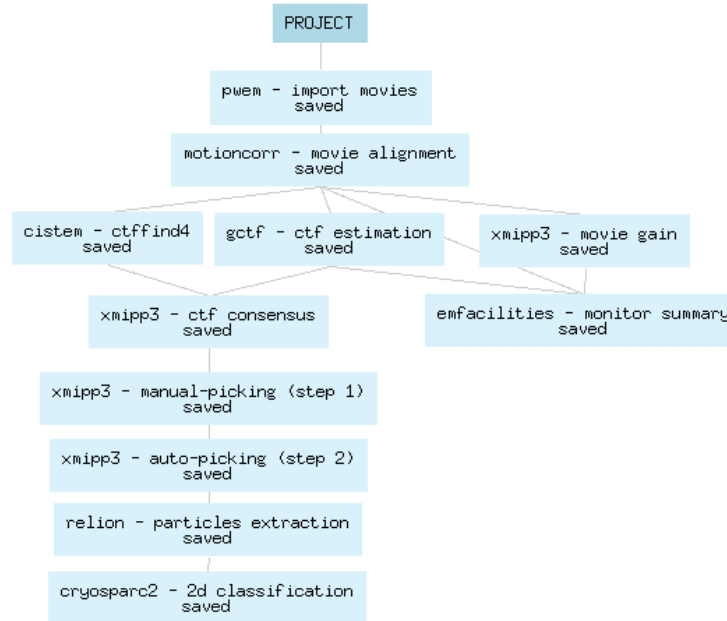
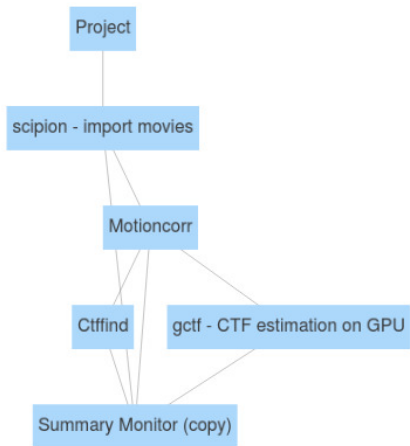


Software developments at CNB-CSIC

Biocomputing Unit
Natl. Center of Biotechnology (CSIC)
Madrid



Scipion: a workflow manager



Large community involvement

Repositories	165
Code	50K
Commits	13K
Issues	1K
Discussions	0
Packages	4
Marketplace	1
Topics	1
Wikis	184
Users	48

48 developers

Languages	
Python	112
Dockerfile	10
C++	5
JavaScript	5
Puppet	3
Shell	3

165 repository results Sort: Best match

- I2PC/scipion** Public archive

Scipion is an image processing framework to obtain 3D models of macromolecular complexes using Electron Microscopy (3...

python c-plus-plus workflow structural-biology image-processing reproducible-science usability electron-microscopy cryo-em single-particle-analysis

☆ 71 Python Updated on Jun 4, 2020
- scipion-em/scipion-pyworkflow**

Underlying pyworkflow module for the Scipion framework

core

☆ 4 Python GPL-3.0 license Updated 8 days ago
- I2PC/xmipp**

Xmipp is a suite of image processing programs, primarily aimed at single-particle 3D electron microscopy.

cryo-em cryoem scipion

☆ 15 C++ GPL-3.0 license Updated yesterday 1 issue needs help
- I2PC/scipion-em-xmipp**

Plugin to use Xmipp programs within the Scipion framework

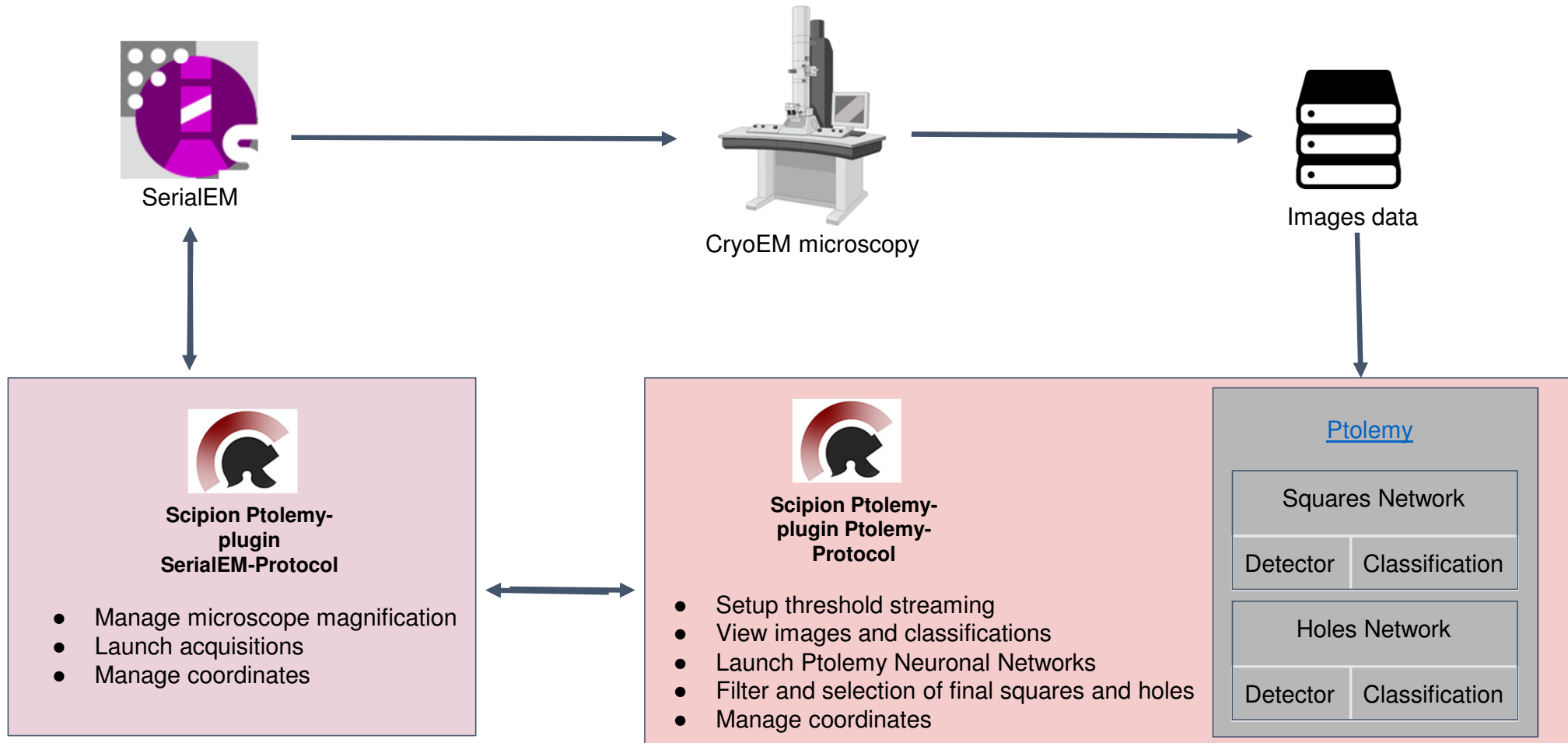
☆ 5 Python GPL-3.0 license Updated yesterday



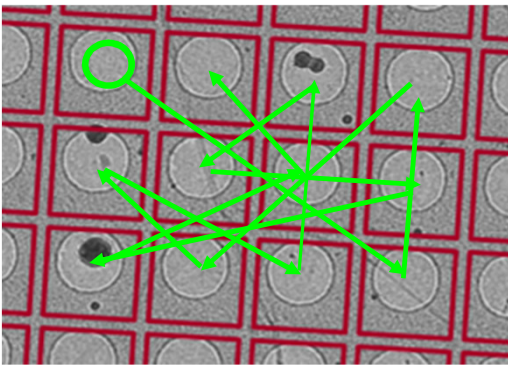
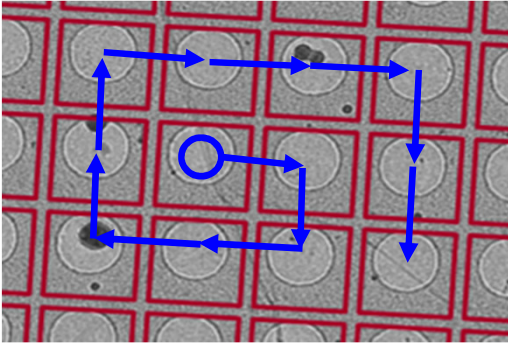
Contents

1. Automatic acquisition
2. JAFIS and minimal hysteresis
3. Stream processing and monitoring
4. Automated sample evaluation
5. Automation in facilities workflows
6. Distributed processing
7. Giving the data back to the user
8. Remote access
9. User machines

Grid squares and holes detection and classification



JAFIS and minimal hysteresis



Each beam shift **accumulate a hysteresis in the microscope lens**, causing aberrations (AFIS). Stopping the acquisition to calibrate the lens is the current solution.

JAFIS software, available for SerialEM, bring all the parameters involve in the aberrations issue for each hole (h).

$h_{i,j,k,l}^{0,0}$	$h_{i,j,k,l}^{0,1}$...	$h_{i,j,k,l}^{0,m}$
$h_{i,j,k,l}^{1,0}$	$h_{i,j,k,l}^{1,1}$...	$h_{i,j,k,l}^{1,m}$
...
$h_{i,j,k,l}^{n,0}$	$h_{i,j,k,l}^{n,1}$...	$h_{i,j,k,l}^{n,m}$

Find a path that minimize the number of changes of sign in those parameters, is carried out by a **genetic algorithm**. Results:

There are 479M possible paths for 12 holes

- The genetic algorithm found a path with only **4 sign changes**:
[6, 11, 9, 3, 5, 2, 8, 0, 10, 4, 1, 7]
- Following a default spiral path would result in **9 sign changes**:
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

Stream processing and monitoring



Micro 1 - 200 kV FEI TALOS Arctica



Micro 2 - 300 kV JEOL CryoARM

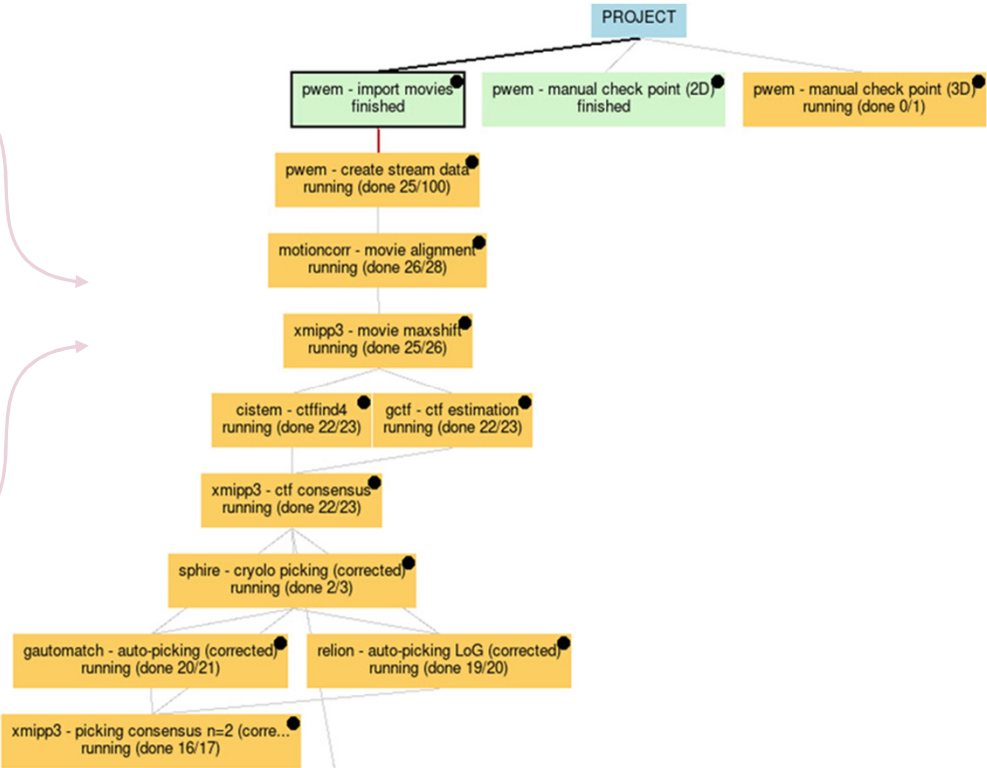


Processing server 1

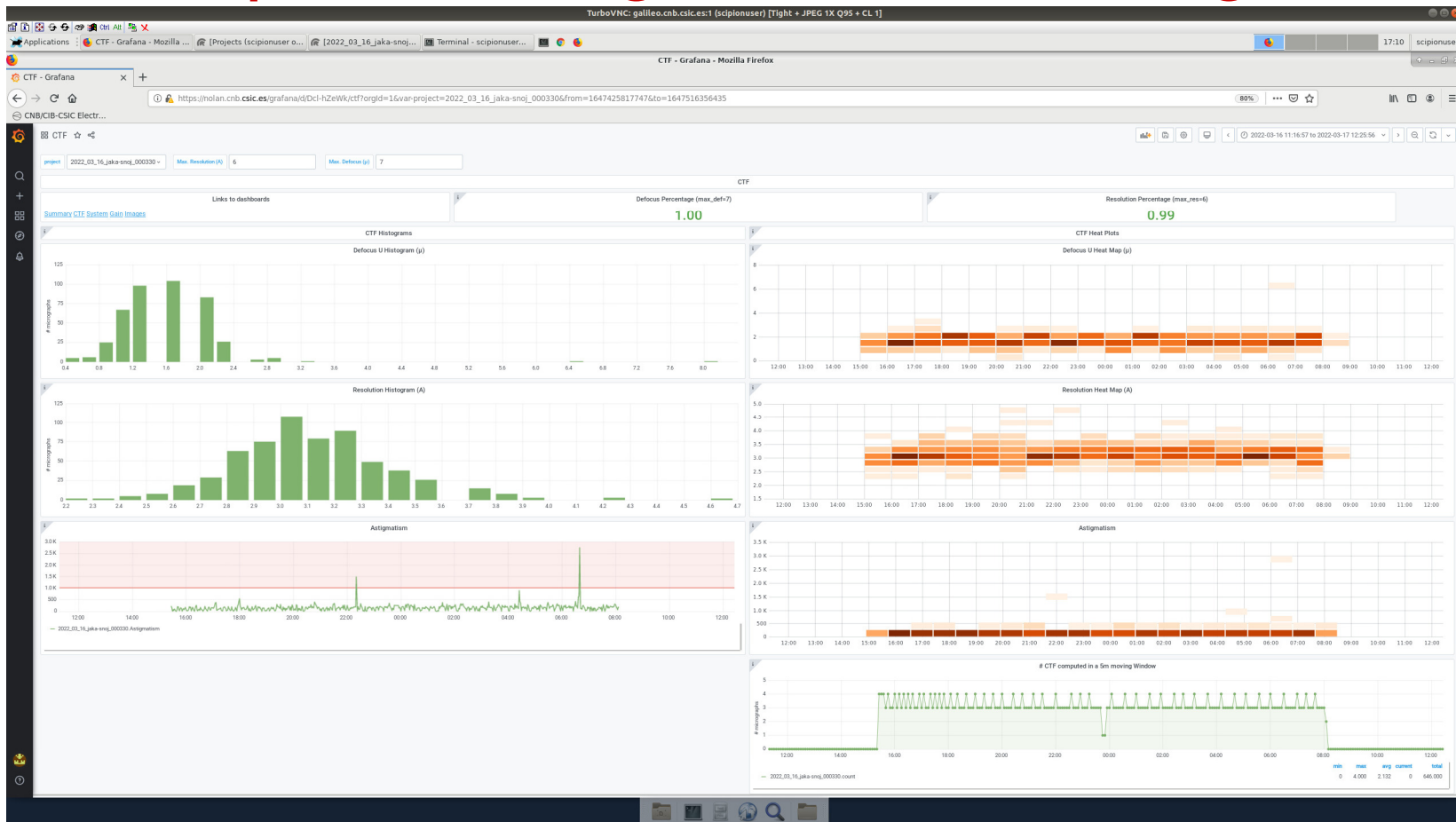


Processing server n

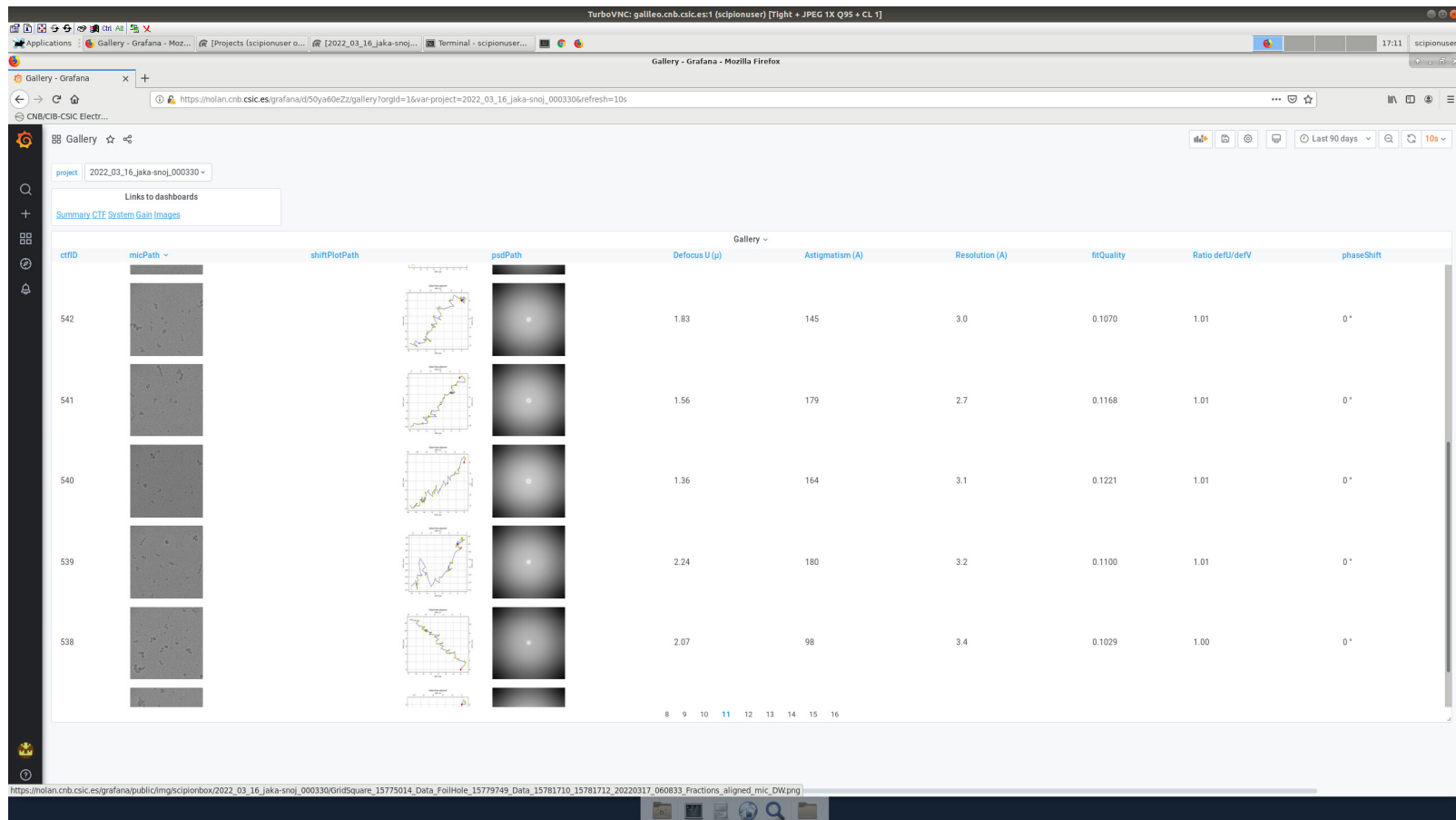
Scipion On-the-fly processing



Stream processing and monitoring

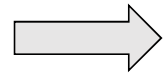


Stream processing and monitoring



Automated sample evaluation

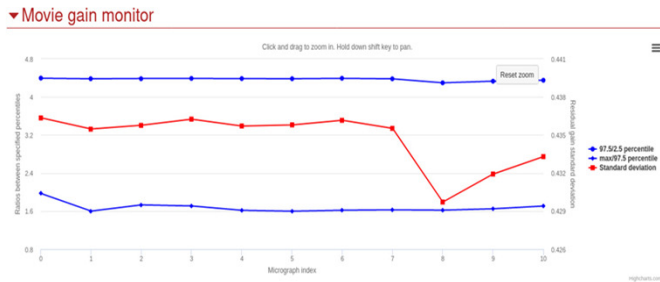
Objectives



- 1) Constant feedback on the **quality of acquisition**
- 2) **Automated intelligent decisions** to filter bad quality images

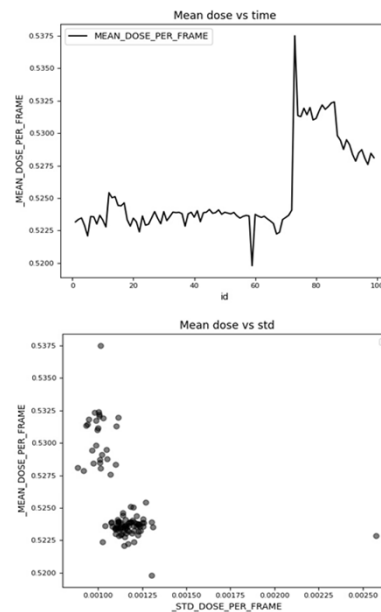
1) **Quality of acquisition:** protocols to monitor the acquisition by extracting quality measurements such as gain, dose analysis or beam tilt analysis.

Movie gain protocol

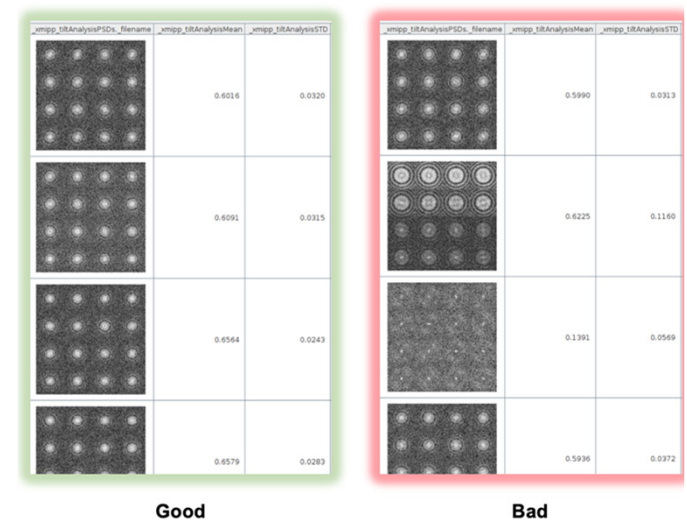


These protocols can be added to your workflows for an extra level of quality checks.

Poisson count protocol



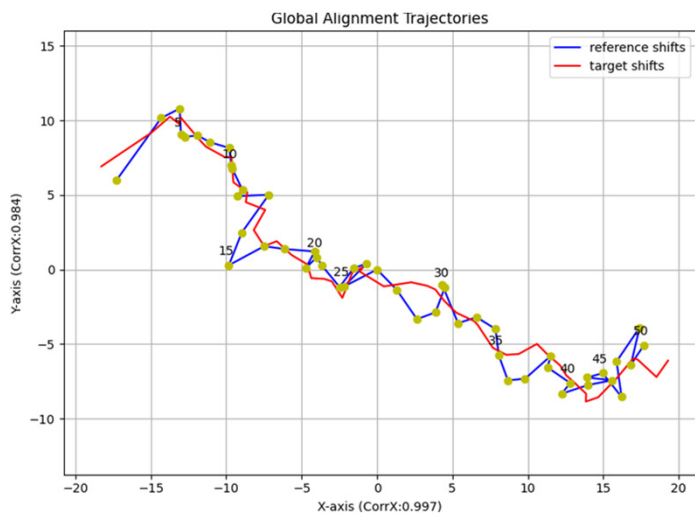
Tilt analysis protocol



Automated sample evaluation

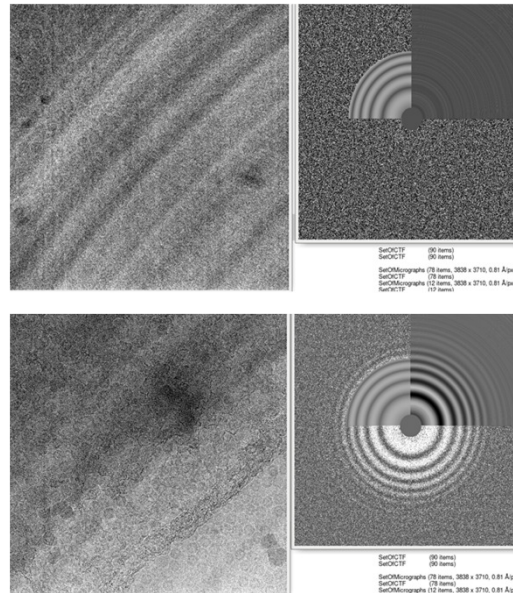
2) Automated intelligent decisions: Consensus protocols combine estimations where at least two different algorithms agree from the same input data. Helps to make more robust estimations and discard bad quality results.

Alignment consensus



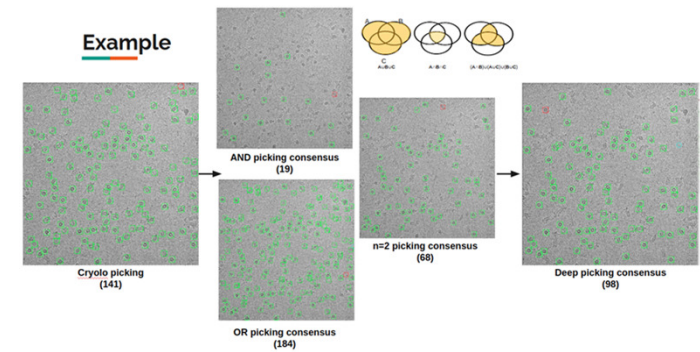
Shifts trajectory correlation between two global alignments

CTF consensus



Discarded micrographs

Particle picking consensus

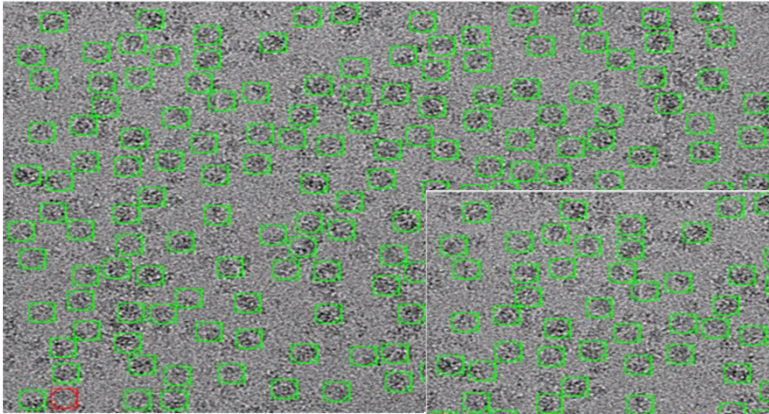


Picking consensus options

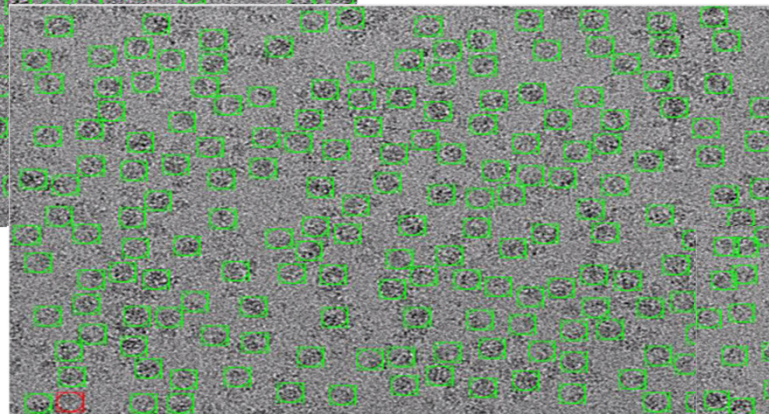
These protocols can be added to your workflows for more robustness and to act as quality filters.

Automation in facilities workflows

Topaz: 7486 particles

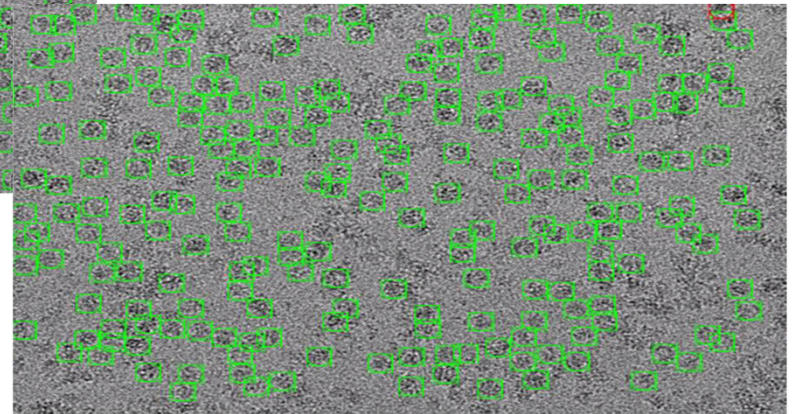


Relion: 2477 particles



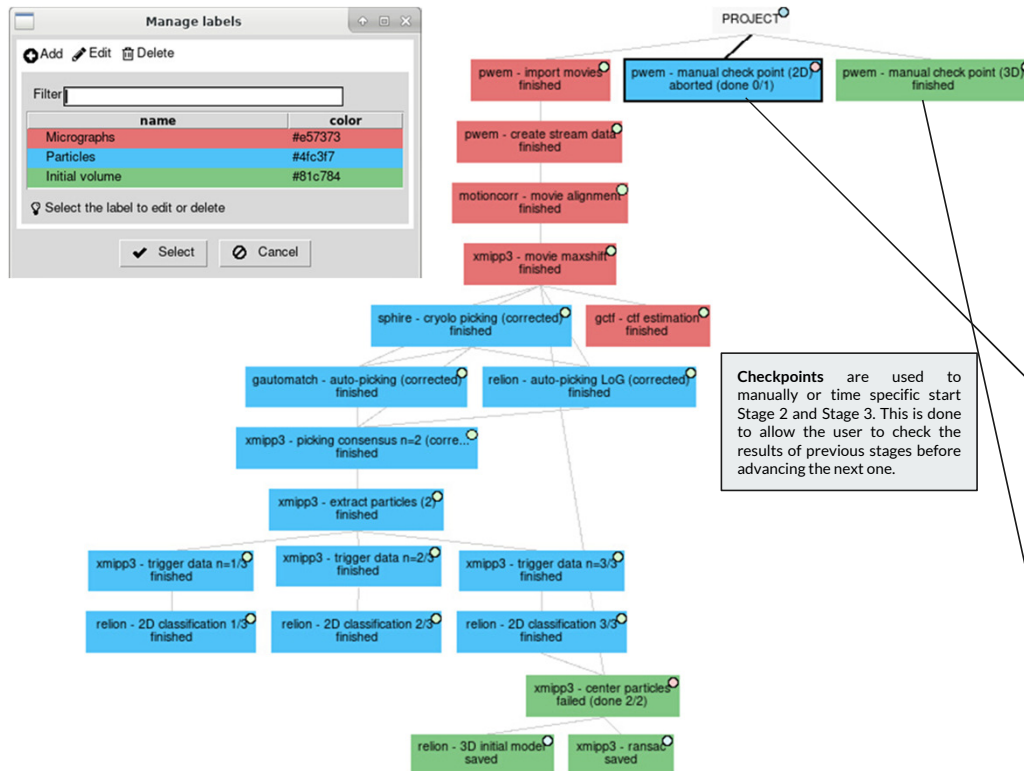
+Deep Consensus
+Micrograph Cleaner

Gautomatch: 3311 particles



Automation in facilities workflows

On-the-fly processing (streaming) with reduced or non-human interactions workflows:



Stage 1:

1. From movies to micrographs

- Alignment options (motioncorr, flexAlign, reion, etc.)

2. CTF estimation

- Different options (gctf, cistem, xmipp, etc)
- CTF consensus recommended

Stage 2:

1. Manual Checkpoint 1

- Manually or time approved starts the picking

2. Particle picking strategy

- Different picking strategies based in normal picking and consensus, using 2D references or training new models.

3. Trigger data

- 2D Classification with eg: n=1,000, n=5,000, n=10,000 number of particles

Stage 3:

1. Manual Checkpoint 2

- Manually or time approved starts centering particles for the initial volume

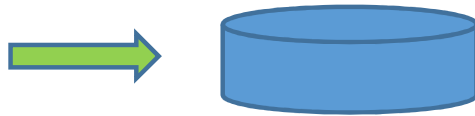
2. Initial volume estimation

- Initial volume (eman2, cryosparc, reion, xmipp3, etc)

Distributed processing



Micro 2 - 300 kV JEOL
CryoARM



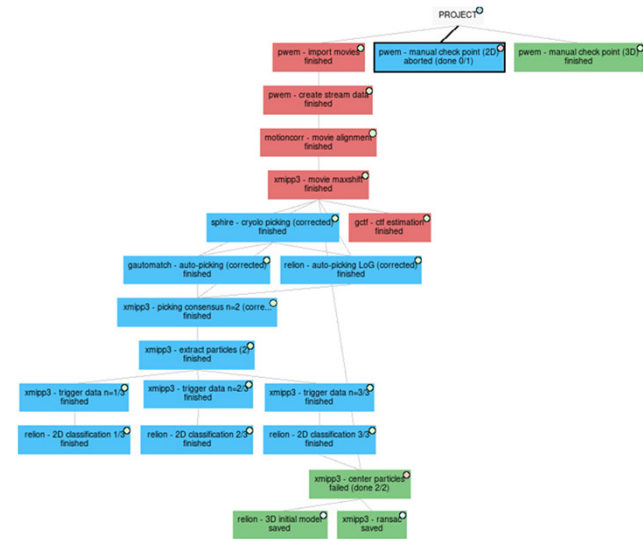
Processing
server 1



Processing
server 2

Movie alignment

All other steps

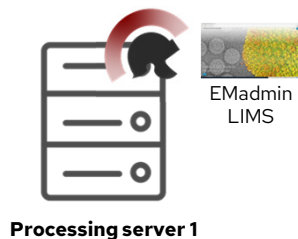


+GPU Queues

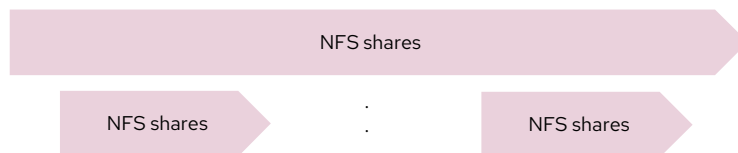
Giving the data back to the user



Micro 1 - 200 kV FEI
TALOS Arctica



Processing server 1



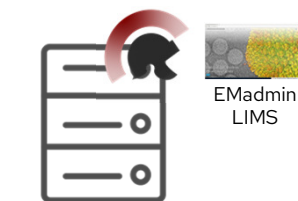
Storage server



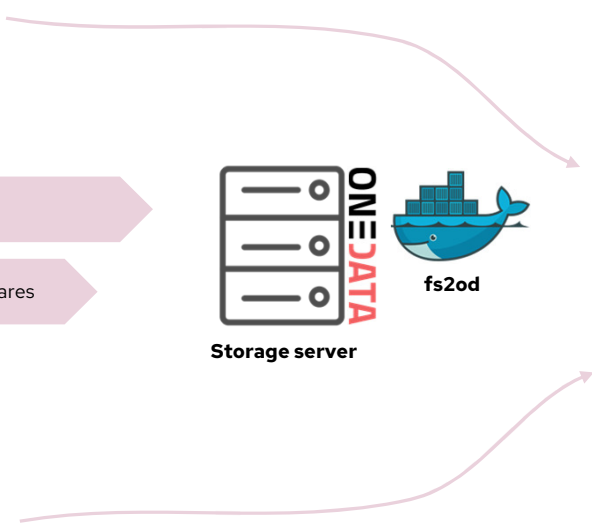
CryoEM Workflow
Viewer



Micro 2 - 300 kV JEOL
CryoARM



Processing server n



Giving the data back to the user

xmipp3 - movie gain

xmipp3 - ctf consensus GCTF+CTFFind4



-6819.outputMicrographsDiscarded:

summary:
29/29 CTF processed (28 accepted and 1 discarded). *General Criteria*: - _Defocus_ Range: 4000 - 40000 - _Astigmatism_ Threshold: 1000 - _Resolution_ Threshold: 4 (1 discarded)
Input: 6631.outputCTF (from cistem - ctffind4) - SetOfCTF (29 items)
Input: 6692.outputCTF (from gctf - ctf estimation) - SetOfCTF (29 items)
Output: 6819.outputCTF - SetOfCTF (28 items)
Output: 6819.outputMicrographs - SetOfMicrographs (28 items, 3710 x 3838, 0.49 Å/px)
Output: 6819.outputCTFDiscarded - SetOfCTF (1 item)
Output: 6819.outputMicrographsDiscarded - SetOfMicrographs (1 item, 3710 x 3838, 0.49 Å/px)

log: Runs/014156_CryoEMWorkflowViewerDepositor/extra/images_representation/6819_XmippProtCTFConsensus.log
label: Micrographs
plugin: xmipp3

reion - auto-picking LoG

eman2 - sparx gaussian picker

sphere - cryolo picking

xmipp3 - auto-picking (step 2)

Remote access: VPN+VNC

The screenshot shows a remote VNC session to a SCIPION workstation. The window title is "TurboVNC: galileo.cnb.csic.es:1 (scipionuser) [Tight + JPEG 1X Q95 + CL 1]". The SCIPION interface displays a project workflow diagram for "2022_03_16_jaka-snoj_000330". The workflow includes steps like "scipion - import movies", "relion - motion correction", "xmipp3 - ctf estimation", "relion - particles extraction", "cryosparc2 - 2d classification", and "cryosparc2 - 3D non-uniform refinement". A summary section at the bottom shows input and output volumes.

```
graph TD
    PROJECT[PROJECT] --> S1[scipion - import movies finished]
    S1 --> S2[relion - motion correction finished]
    S2 --> S3[xmipp3 - ctf estimation finished]
    S2 --> S4[sphire - cryolo picking finished]
    S3 --> S5[xmipp3 - ctf consensus finished]
    S4 --> S6[relion - particles extraction copy 2 finished]
    S5 --> S7[cryosparc2 - 2d classification copy 3 finished]
    S6 --> S7
    S7 --> S8[451786 finished]
    S7 --> S9[388avgps finished]
    S8 --> S10[cryosparc2 - initial model finished]
    S9 --> S10
    S10 --> S11[77063 vol5 finished]
    S10 --> S12[vol5 finished]
    S11 --> S13[cryosparc2 - 3D non-uniform refinement finished]
    S11 --> S14[cryosparc2 - 2d classification copy 4 finished]
    S11 --> S15[cryosparc2 - 2d classification copy 5 finished]
    S11 --> S16[cryosparc2 - 2d classification copy 6 finished]
    S11 --> S17[cryosparc2 - 2d classification copy 7 finished]
    S11 --> S18[cryosparc2 - 2d classification copy 8 finished]
    S13 --> S19[xmipp3 - deepEMhancer finished]
```

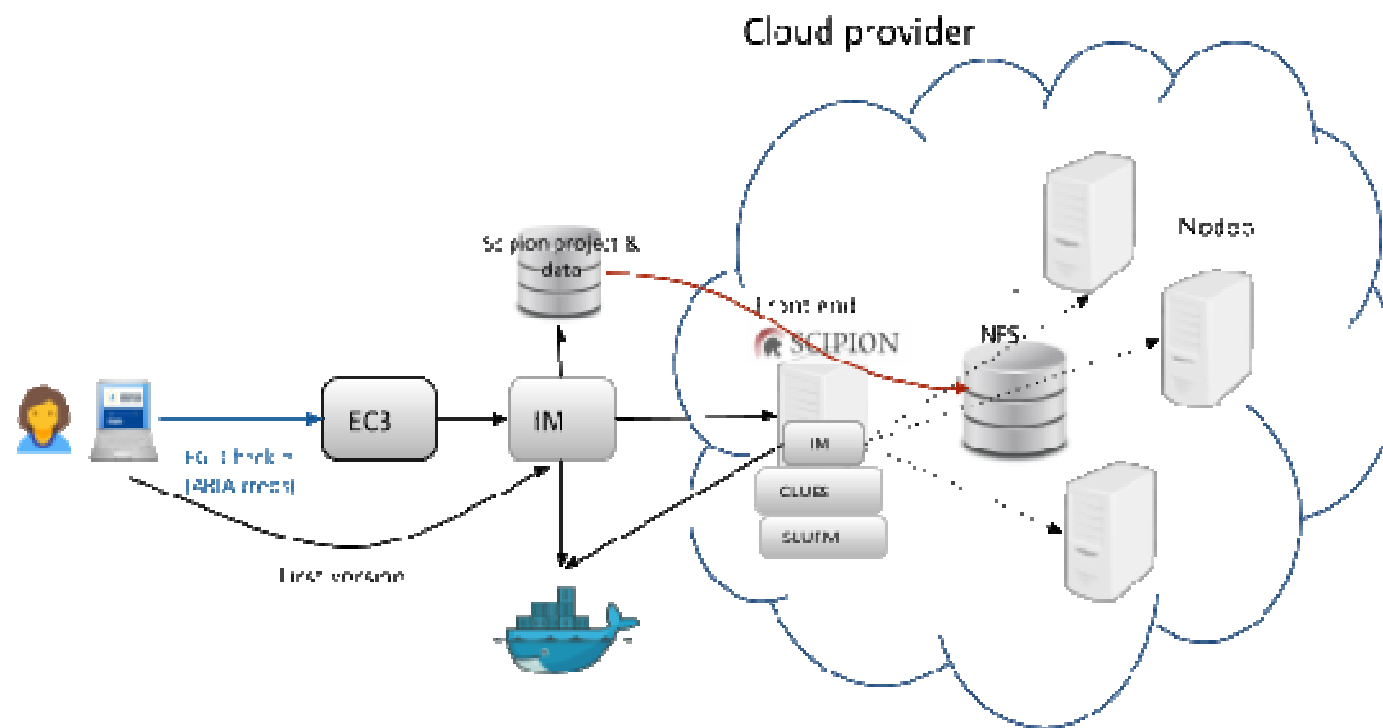
Summary | Methods | Output Log

Input	Output
inputVolume (from cryosparc2 - 3D non-uniform refinement -> outputVolume [outputVolume])	Volume (160 x 160 x 160, 1.71 Å/pix)
output	Volume (160 x 160 x 160, 1.71 Å/pix)
xmipp3 - deepEMhancer -> Volume	Volume (160 x 160 x 160, 1.71 Å/pix)

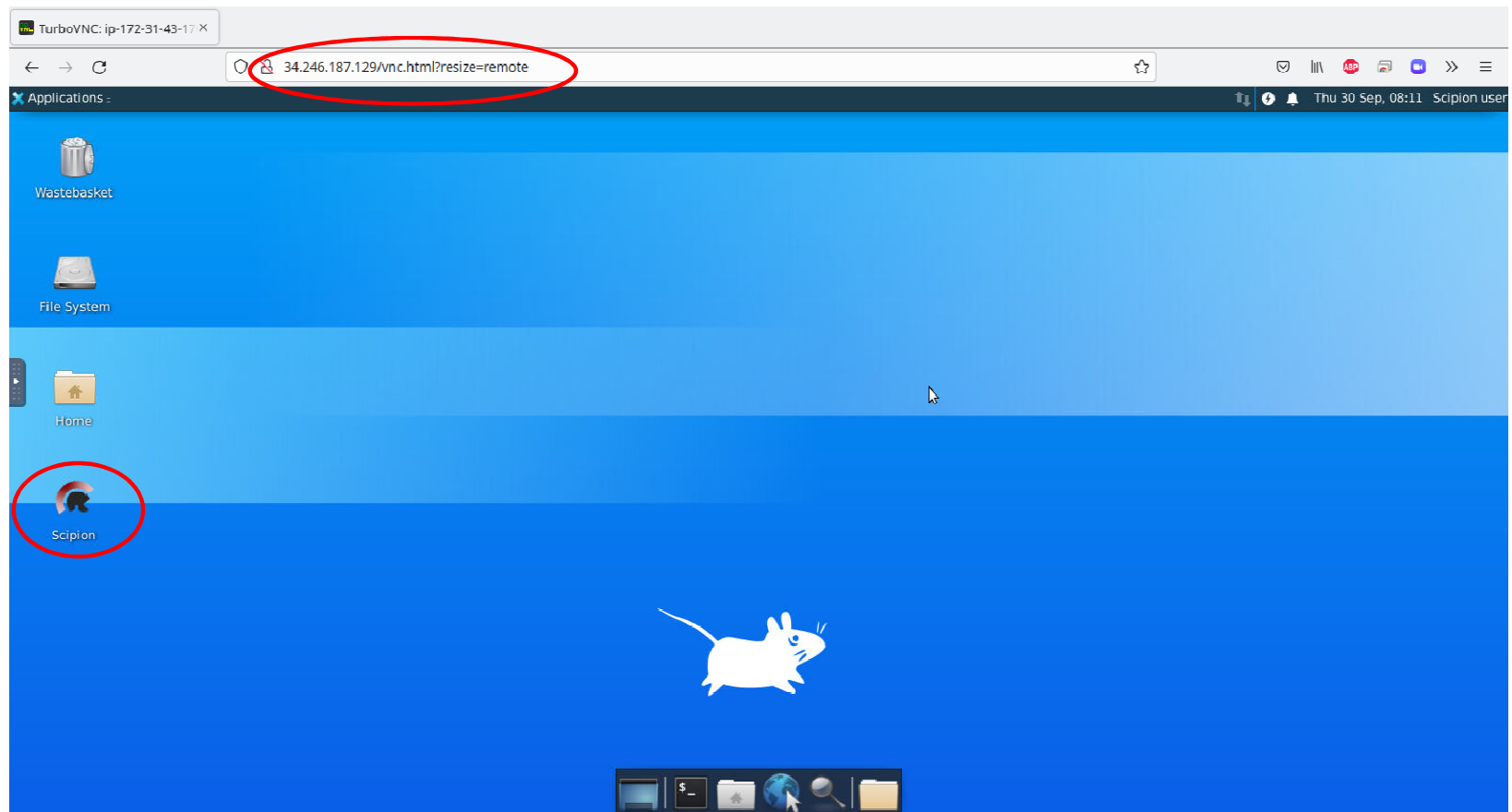
SUMMARY

Input: raw data map
Normalization: auto

User machines



User machines



Contents

1. Automatic acquisition
2. JAFIS and minimal hysteresis
3. Stream processing and monitoring
4. Automated sample evaluation
5. Automation in facilities workflows
6. Distributed processing
7. Giving the data back to the user
8. Remote access
9. User machines

Thanks

