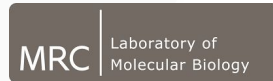


Scipion - possibilities, traceability, reproducibility and reporting.

2nd Instruct workshop on best practices - Leiden 2018



Summary

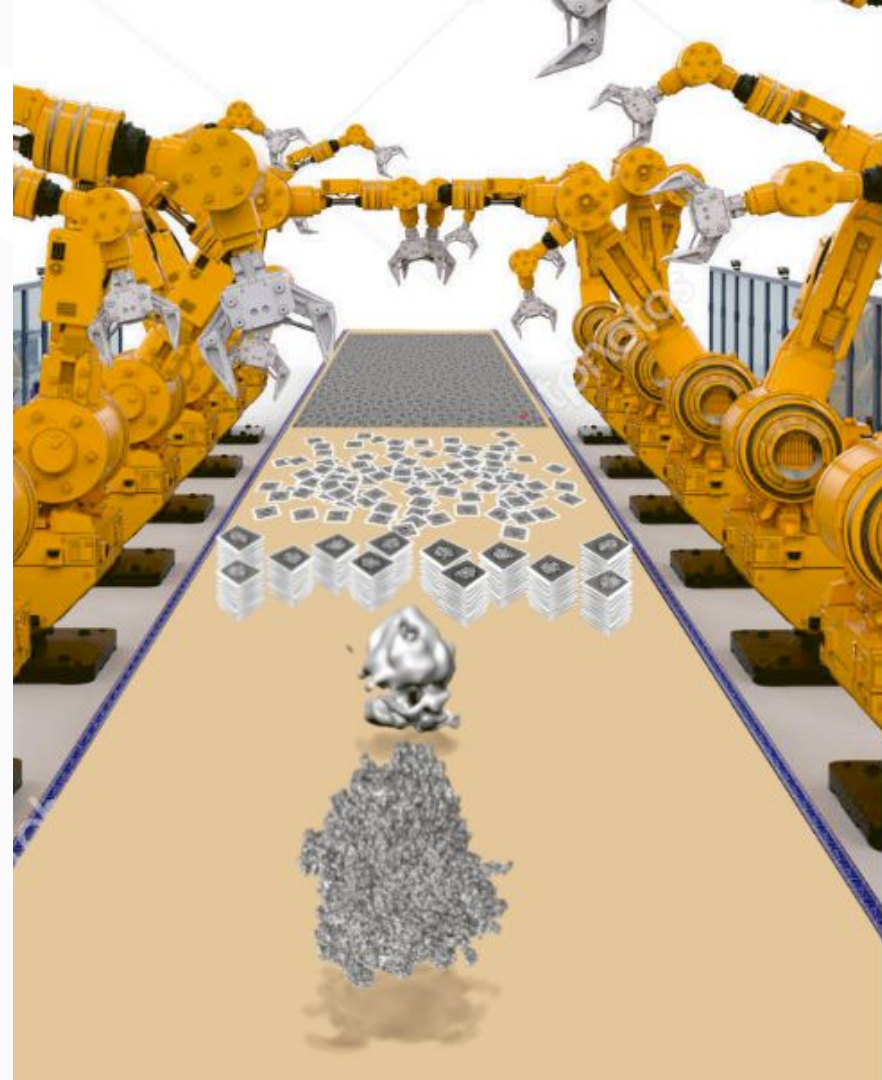
- Scipion intro
- Automation and streaming
- Integration at CNB
- Current status
- Features summary

Why Scipion?

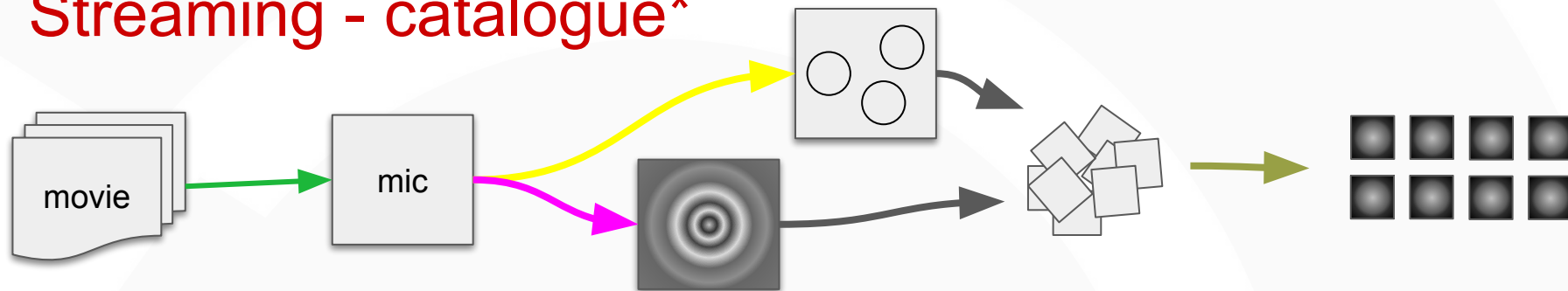
- Execute complete workflows in an automated manner.
- Do image processing on-the-fly (streaming)
- Customize/create/combine your own workflow at wish.
- Combine common EM software methods without programming.
- Have future state-of-the-art methods available with one click
- Know what have you done or repeat what has worked before.

Workflow, automation, streaming

- Automation: SW “orchestration”
- Streaming: Processing while acquiring
- To feedback the acquisition
- To feedback the user
- Increase **ADDED VALUE** of your service



Streaming - catalogue*



Movie alignment	Ctf estimation	Picking	Extraction	2D classification
Motioncor2 Motioncor Xmipp optical flow Xmipp crosscor... Unbur Summovie	ctffind4 gCTF Xmipp ctf ctffind	Xmipp auto Relion auto Eman boxer Eman sparx Appion dogpicker GAutomatch gemPicker CrYOLO (coming)	Xmipp Relion	Xmipp gL2D

*21, Some of them finished but not released yet.

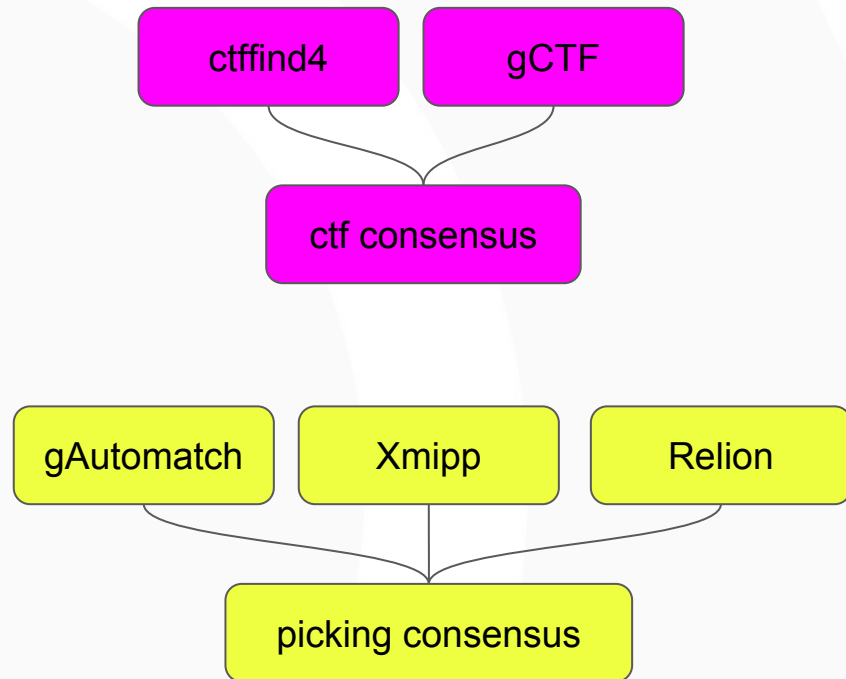
Streaming - catalogue tools

- streaming tools

- CTF Consensus, filtering
- Picking consensus, filtering

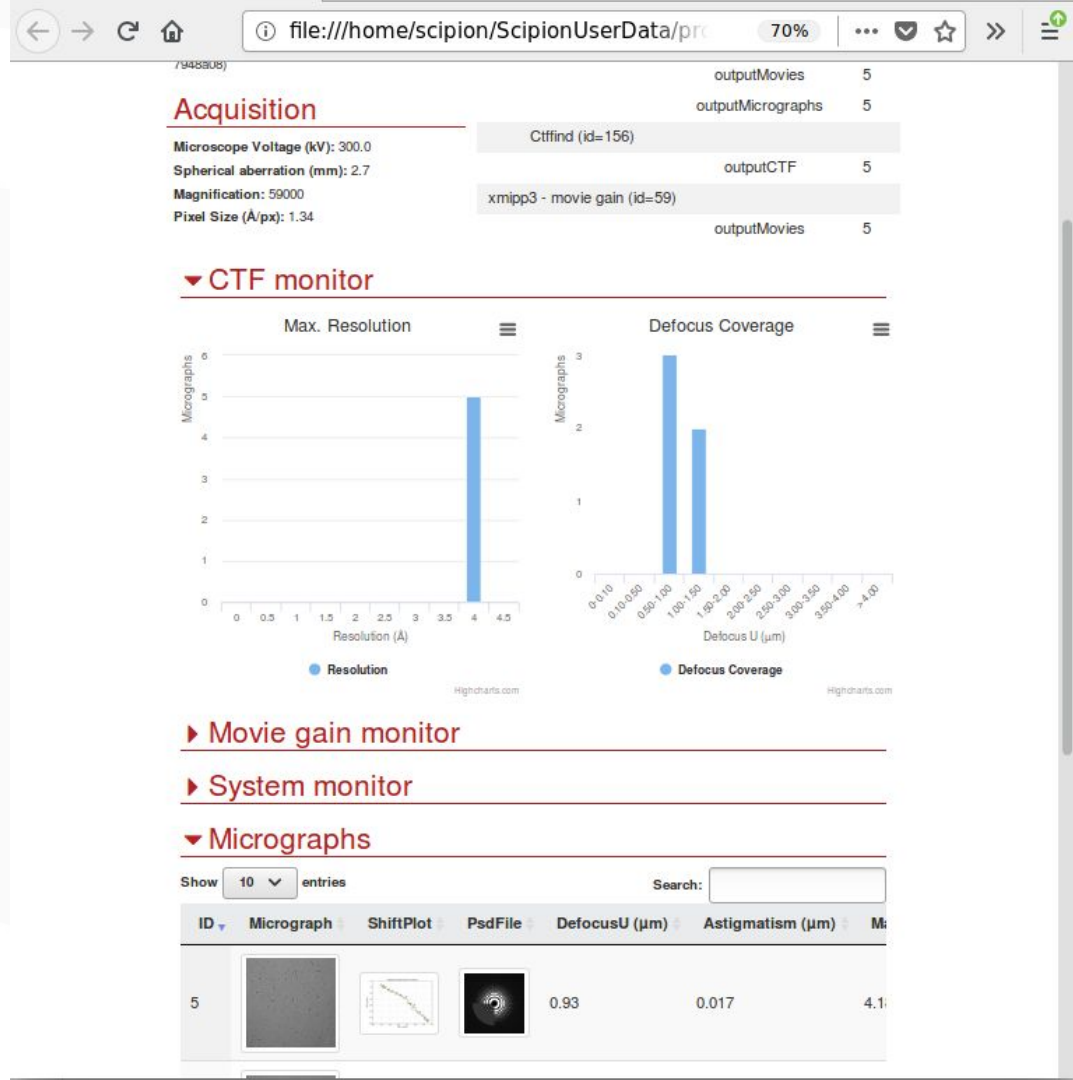
- monitoring - 📏 + 🕒

- System: GPU, memory, network, I/O
- CTF: defocus, astigmatism, phaseShift
- Gain

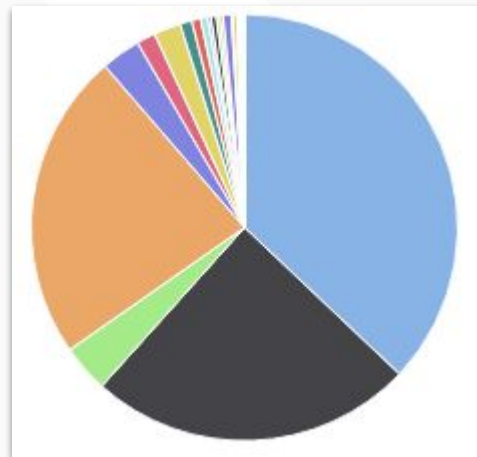
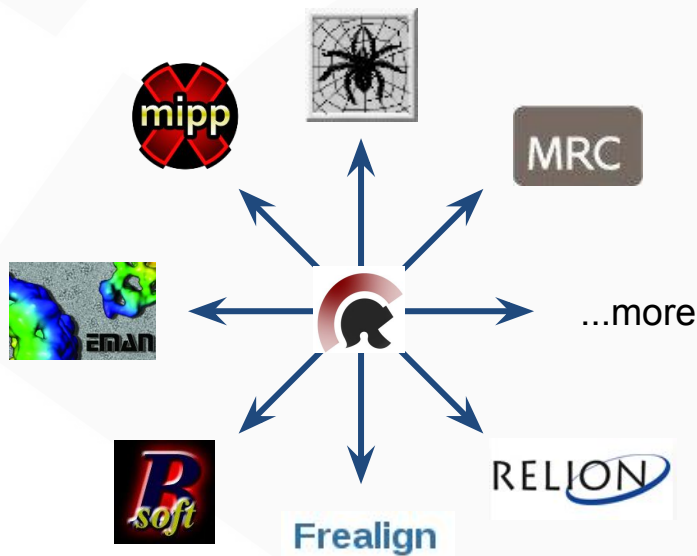


Acquisition report

- Ready to publish on the web
 - Basic acquisition info
 - CTF charts
 - Monitors output
 - Summary mic table
- “rsyncable”



Combine EM software



- Scipion
- gCTF
- Resmap
- Coot
- ESRF-European synchrotron
- Bramford lab

- Relion
- Eman
- Grigorieff lab
- Not classified
- cryomethods

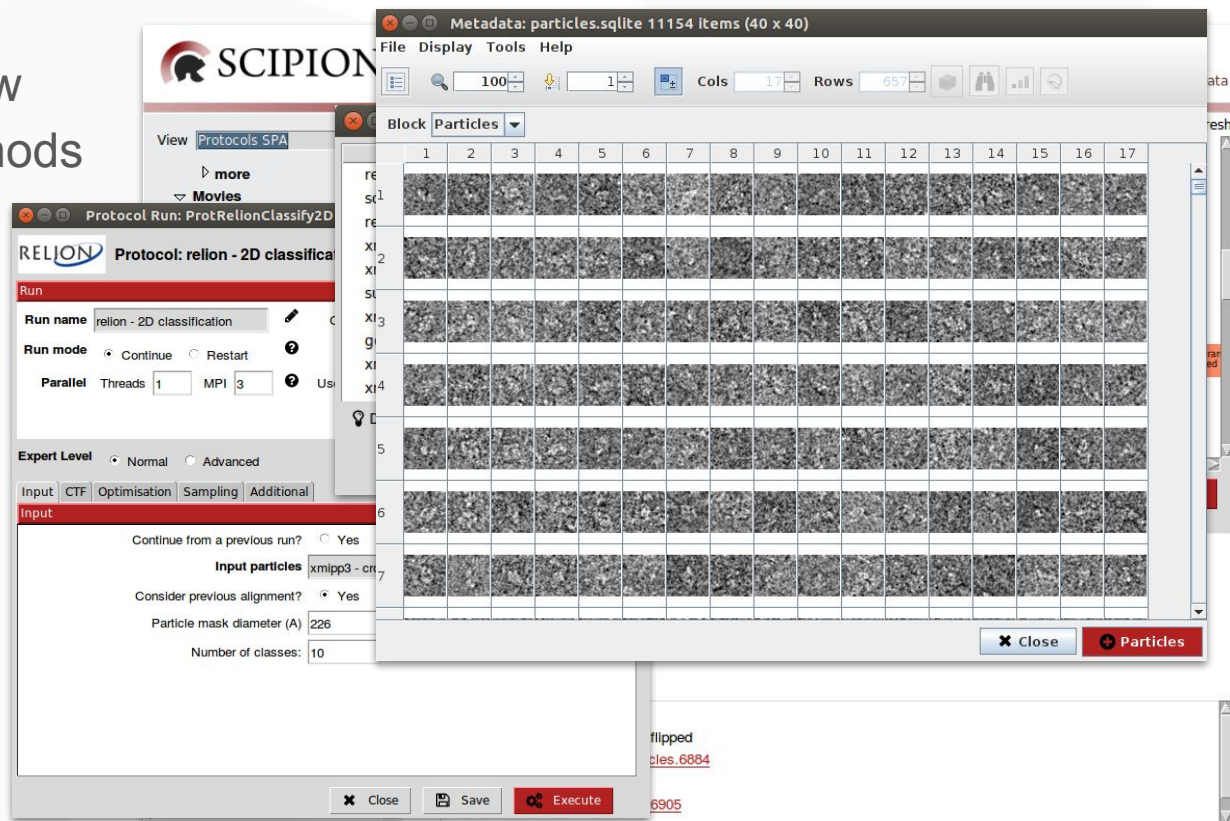
- CTFfind
- Gautomatch
- DLS-Diamond Light Source
- Emx
- Phenix

- Xmipp
- Sphire
- Chimera
- Igbmc

- Motioncor/2
- Opic
- Spider
- Appion
- BSOFT

Easy to use

- Visual project overview
- Easy to combine methods
- Wizards
- Help per parameter
- Easy data inspection
- Link to articles
- Bibtext citation export










Traceability and reproducibility

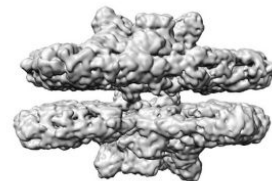
- Everything is recorded
- Logs are carefully stored
- Everything can be exported or imported
 - Json format
 - easy to view on a browser
- Submit to EMPIAR
 - The workflow (metadata)
 - The data (movies/mics)

EMPIAR-10198

Paired C2S2M PSII-LHCII supercomplexes from thylakoid membranes of *Pisum sativum*

Publication: Pea PSII-LHCII supercomplexes form pairs by making connections across the stromal gap
Albanese P , Melero R , Engel BD , Grinzato A , Berto P, Manfredi M, Chiodoni A, Vargas J, Sorzano COS, Marengo E, Saracco G, Zanotti G , Carazo JM , Pagliano C 
Sci Rep 7 (2017)
PMID: 28855679
DOI: [10.1038/s41598-017-10700-8](https://doi.org/10.1038/s41598-017-10700-8)

Contains:
 micrographs



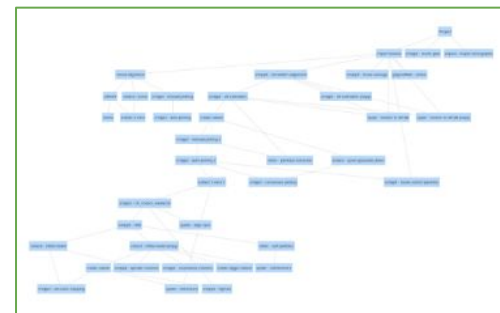
Related EMDB entry: [EMD-3825](#)
Deposited: 2018-06-25
Released: 2018-10-05
Last modified: 2018-10-05
Dataset size: 433.1 GB
Dataset DOI: [10.6019/EMPIAR-10198](https://doi.org/10.6019/EMPIAR-10198)

+ Image set

Browse All Files

 Download

- 10198 433.1 GB
- data 433.1 GB
- C2S2M_PSII_LHCII 433.1 GB
- 2.outputMicrographs 433.1 GB
 - deposition.json 5.7 KB
 - workflow.json 28.7 KB
 - 10198.xml 4.8 KB



Integration

- Scipion plays very well with scripts and any python app
- Designed as an API from the beginning
- Already integrated with:
 - ISPyB: eBIC, ESRF
 - EMAdmin: CNB
 - SciLifeLab
- Ready for clusters/queues

The screenshot displays the ISPyB @ ESRF interface. At the top, a blue header reads "ISPyB @ ESRF: Usage on CryoEM". Below this, the "Data Collections" section is visible, with navigation links for "Visit Stats", "Users", and "Parcels". The main "Processing" section is active, showing "Auto Processing" for "Movie: 1". It includes a "Motion Correction" panel with two micrograph images and a drift plot. The drift plot shows a yellow line representing drift over time, with axes ranging from -20 to 20. Below the motion correction panel is the "CTF Correction" panel, which displays various parameters for the movie.

Motion Correction Parameters:

Movie Number: 1	First Frame: 1
Last Frame: 0	Dose Per Frame: e ⁻¹ /Å ²
Dose Weight: ?	Total Motion: 65.5Å
Average Motion / Frame: 3.45Å	Patches Used: x
Comment: aligned	

CTF Correction Parameters:

Movie Number: 1	Boxsize: xµm
Resolution: - Å	Defocus: - Å
Defocus Step Size: Å	Astigmatism: 1.01Å
Astigmatism Angle: 158.2°	Estimated Resolution: 3.10Å
Estimated Defocus: 2855Å	Amplitude Contrast:
CC Value:	
Comment:	

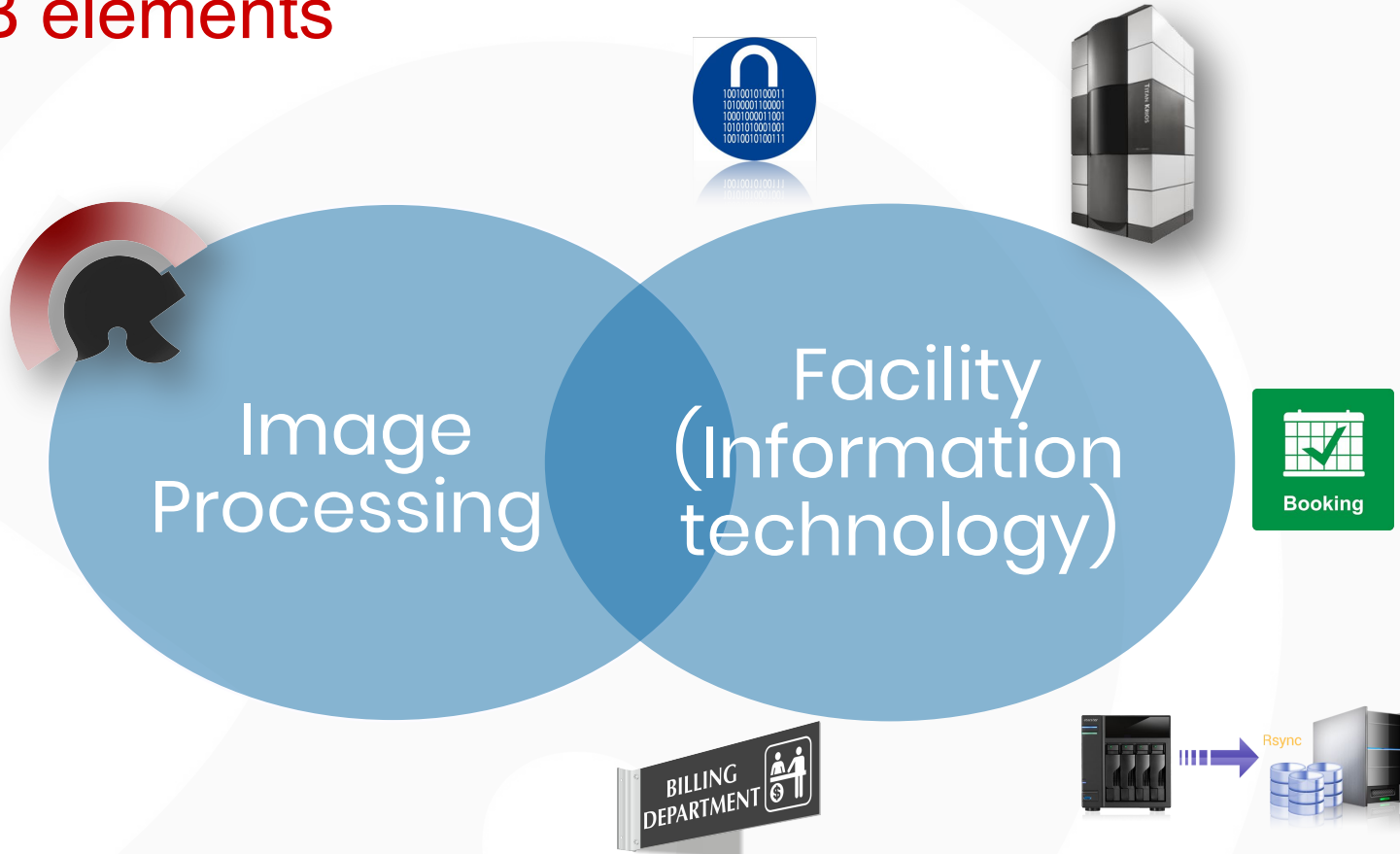
Additional parameters shown in the interface include "Motion Correction: 4017" and "CTF: 4017" with green checkmarks. The interface also features "Micrograph" and "Q Log file" buttons in the motion correction section, and "Output" and "Q Log file" buttons in the CTF correction section.

- Have output from Motion Correction & CTF Correction
- Motion Correction: Drift plots, some correction parameters
- CTF: Some parameters, FFT theoretical vs. real

Scipion at CNB



CNB elements



EM Admin

Users management

Directory structure aware

Acquisition setup

Scipion report aware

Facility statistics

Workflow selection

Launch Scipion

<https://github.com/rmarabini/webservices/tree/master/EMAdmin>

M Inbc x Facil x 18 Goo x Hon x Inst: x Scip x Scip x bR Real x GitH x SF Pró x

← → ↻ 127.0.0.1:8000

☰ CNB/CIB-CSIC Electron Microscopy Facility Project Report Statistics About Help

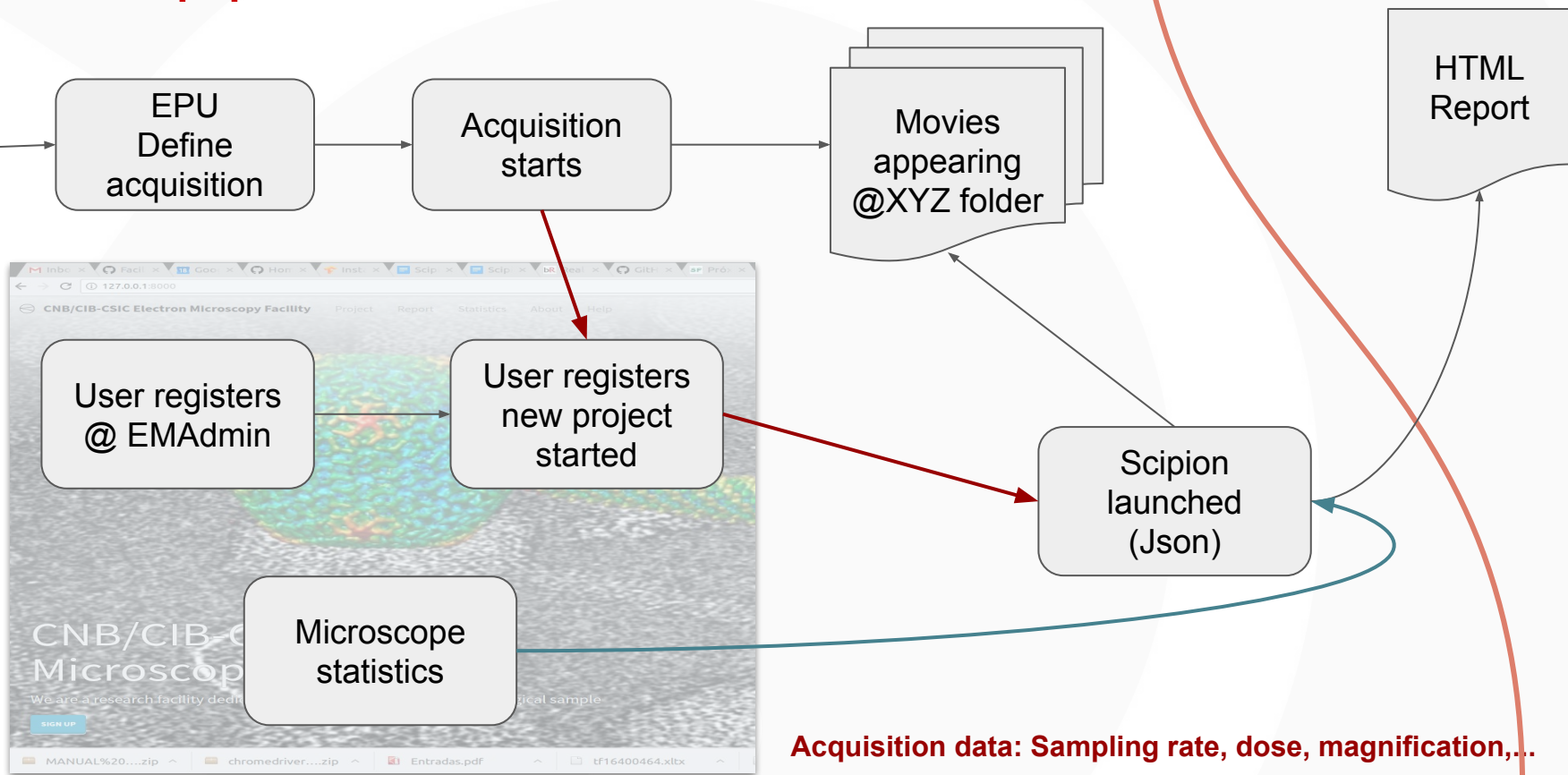
CNB/CIB-CSIC Electron Microscopy Facility

We are a research facility dedicated to ultrastructure analysis of biological sample

[SIGN UP](#)

MANUAL%20....zip ^ chromedriver....zip ^ Entradas.pdf ^ tf16400464.xltx ^

CNB pipeline



Current status facilities



eBIC at Diamond Light Source

European Synchrotron R.F.

National Cancer Institute - NIH

Science For Life Laboratory

National Institute of Environmental Health Sciences -NIH

University of Virginia Health System
University Hospital

University of Umeå

National Center for Biotechnology
CNB-CSIC

University of Copenhagen

CEITEC - Central European Institute of Technology

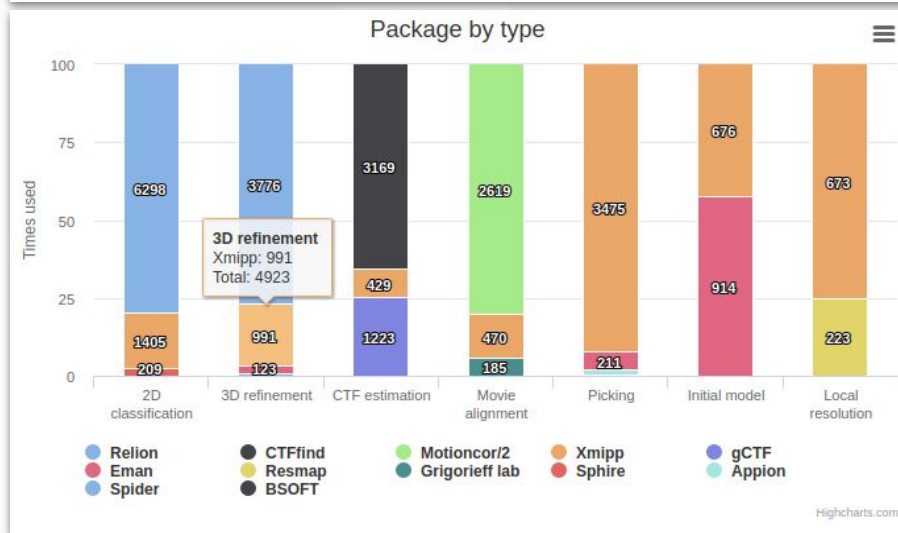
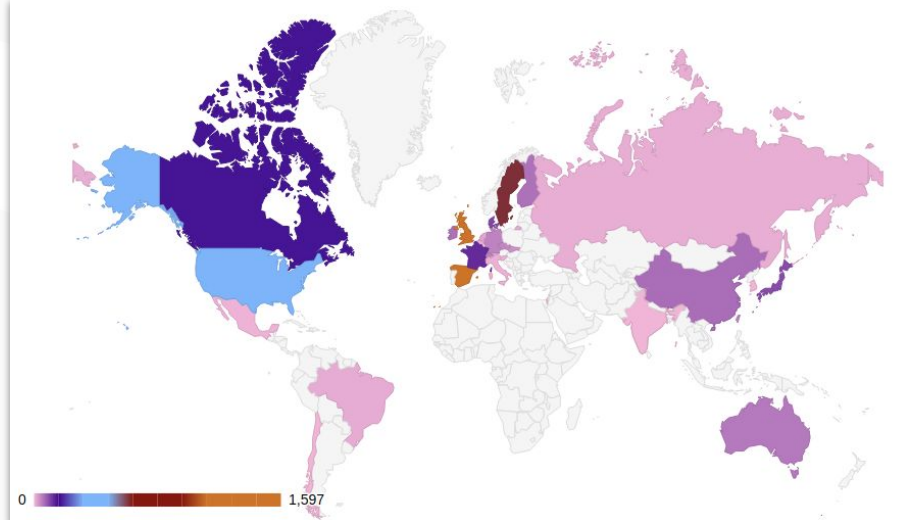
McGill University

NYU School of Medicine

Aarhus University

Current status - Users

- Data sent after user's consent.
- 3500+ projects, subset.
- Developers blacklisted
- Data sent clearly documented
- Important feedback:
 - Geographic use
 - What is used



Features summary

- Streaming processing
- High traceability & reproducibility, full workflow.
- Full API access (programmable) --> automation.
- Connectivity with LIMS.
- Integrates EM packages smoothly, possibilities.
- Extensible...new best methods in the field will appear.
- GPUs, clusters/queues aware.
- Citation aware.
- Workflow customization without coding.
- Resumable
- Strong support: Email, private slack channel for the facilities.

Scipion team



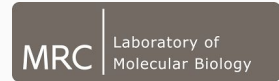
- José María Carazo
- Carlos Oscar Sorzano
- Roberto Marabini
- Pablo Conesa
- Yaiza Rancel
- Laura del Cano
- David Maluenda
- Yunior Fonseca
- Péter horváth
- Jose Luis Vilas
- Marta Martinez
- Amaya Jimenez
- Javier Mota
- Roberto Melero
- Ruben Sanchez
- CNB Facility
- CNB betatesters



- Jose Miguel de la Rosa
- Cryo-EM facility
- Marta Carroni
- Alexey Amunts
- Erik Lindahl



- Josué Gómez
- Javier Vargas



- Grigory Sharov
- Joaquín Otón

Others:

- Juha Huiskonen
- Vahid Abrishami
- Mohsen Kazemi
-

Acknowledgements



Gregory Sharov

For almost everything, making Scipion better, adding new packages and bugfixing.



SciLifeLab

For its contribution to the development of Scipion and having Scipion as it's processing framework in the facility.



Xmipp team

For adding new methods and strongly collaborating with scipion and testing the latest Scipion core developments, sometimes suffering. :-)



ESRF

For early adopting Scipion in their EM processing pipeline and enable connectivity to their LIMS system: ISPyB.



eBic -DLS

For early adopting Scipion in its automated pipeline and improving Scipion with useful feedback.



CORBEL

European Union and Horizon 2020 through grant (INFRADEV-1-2014-1, Proposal: 654248)



EGI-Engage

This work used the EGI Infrastructure and is co-funded by the EGI-Engage project (Horizon 2020) under Grant number 654142.



CNB Cryo EM Facility

For beta testing our latest developments of Scipion and feedback.



Scipion team

For obvious reasons...



Javier Vargas Balbuena

For his contribution to Scipion methods



CNB Biocomputing unit

From its logistic support, administrative effort to the financial support of Scipion.



Spanish Ministry of Economy and Competitiveness

through Grants BIO2013-44647-R, BIO2016-76400-R(AEI/FEDER, UE)



Comunidad Autónoma de Madrid

through Grant: B2017/BMD-3817



Instituto de Salud Carlos III

through grants PT13/0001/0009, PT17/0009/0010



iNEXT

European Union (EU) and Horizon 2020 through grant iNEXT (INFRAIA-1-2014-2015, Proposal: 653706)



Elixir - EXCELERATE

European Union (EU) and Horizon 2020 through grant (INFRADEV-3-2015, Proposal: 676559)



West-life

European Union (EU) and Horizon 2020 through grant EINFRA-2015-1, Proposal: 675858



Beta testers

For testing Scipion in prerelease stages and giving feedback to make better and stable releases.



Jose Miguel de la Rosa Trevin

For strongly contributing to scipion in all aspects, design, management, training.



Juha Huiskonen

For making scipion better integrating localrec into scipion.



Roberto Melero

For his continuous feedback and use of Scipion and sharing his huge Cryo EM image processing experience with us.



Instruct

For its continuous support and use of Instruct resources, a Landmark ESFRI project.



Josue Gomez-Blanco

For his contribution to Scipion core and packages



Christopher A. Lavender

For early adopting Scipion at NIH-NIEHS and dockerizing scipion.



Mario J. Borgnia

For early adopting Scipion at NIH-NIEHS and dockerizing scipion.



NVIDIA Corporation

For the donation of the Quadro GP100 GPU used for testing GPU methods.