



CryoEM workflows in Scipion

EOSC Science Demonstrators and Pilots 13 Sept
Pisa

Pablo Conesa
Biocomputing Unit,
Instruct Image Processing Center, CNB-CSIC
Carlos Oscar S. Sorzano



Slides summary

1. Introduction
2. Basics of Cryo-EM
3. Basic of our software: Scipion
4. EM data deposition and standards
5. Goals
6. Workplan



Introduction



- Pablo Conesa (pconesa@cnb.csic.es)
- Marine biologist + 20 years developing software
- 5 years at EBI + 2 at BCU
- Technical project leader of **Scipion**



- <http://biocomp.cnb.csic.es/>
- 20+ Developing algorithms and software to extract the maximum biological knowledge in 3D-EM and X-ray microscopy
- Lead by Prof. JoseMaria Carazo



- National Center for Biotechnology (CNB)
- Since 1992
- Over 600 people
- Focused on human and animal health, agriculture and environment



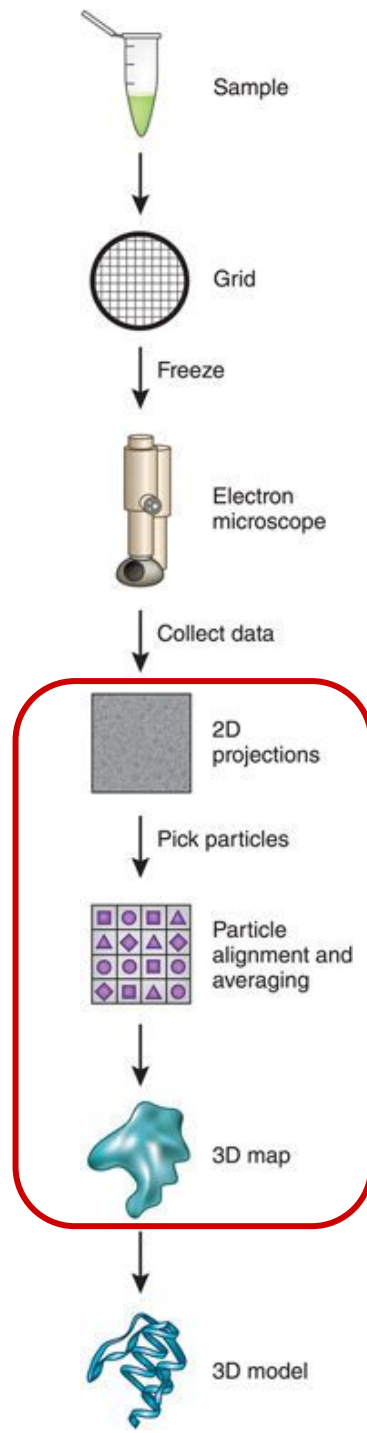
Basics of Cryo-EM

1. Aims at getting the shape of a macromolecule or even a virus.
2. Live view is not possible today, all we can get is a very noisy image/movie.
3. Requires computational approach: SPA or Tomography
4. Video:

<https://www.youtube.com/watch?v=BJKkC0W-6Qk>



Basics of Cryo-EM



Scipion deals with this part

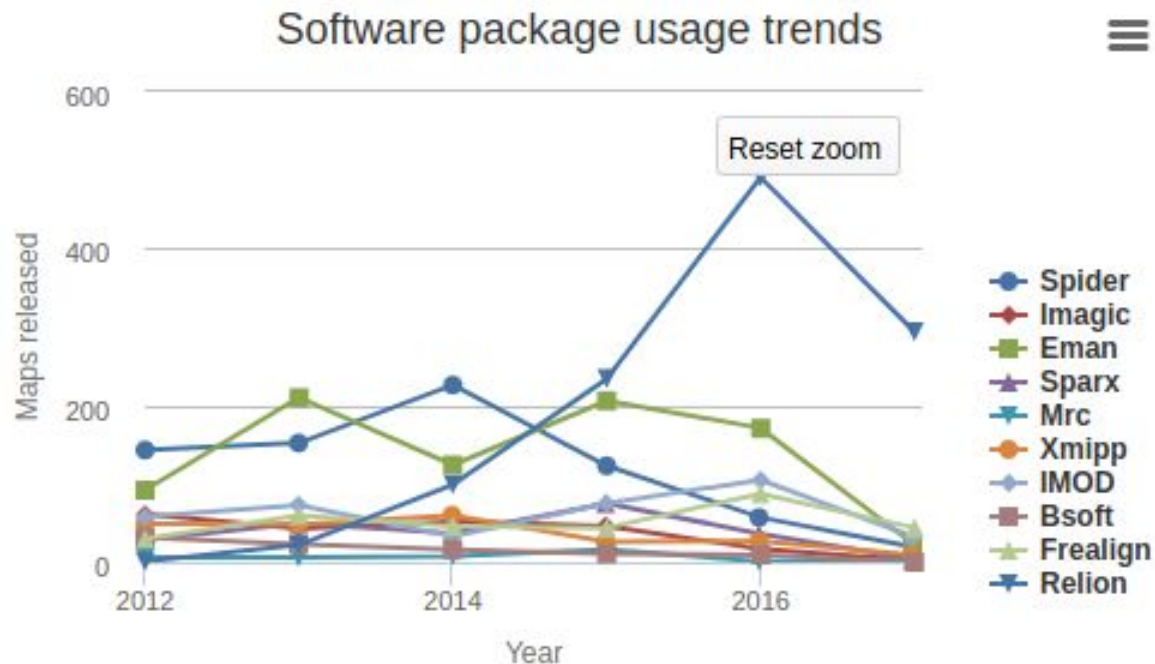
Basics of our software: Scipion

1. Scipion is an image processing framework
2. There are different image processing SW for EM
3. There is only one EM data standard (EMX) not incomplete and not implemented by SW producers
4. Scipion “glues” EM software to explode workflow combinations.
5. Keeps track of all steps of the workflow (Traceability)
6. First steps can be run in “Streaming mode”, useful for facilities.



Integration: Which software is used

EM software reported at EMDB database.

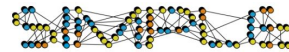
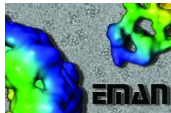
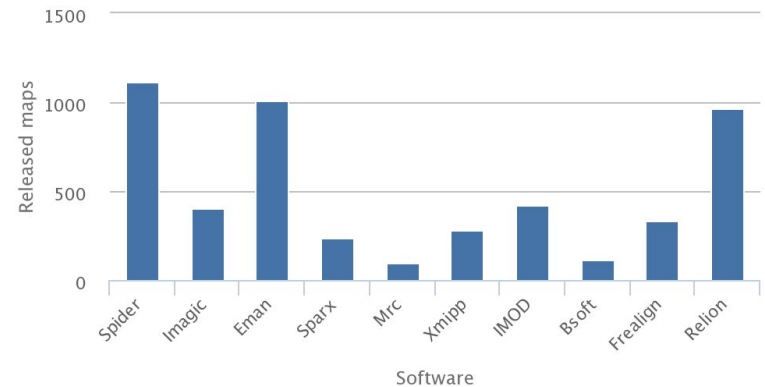


Integration: The EM field needs software integration

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



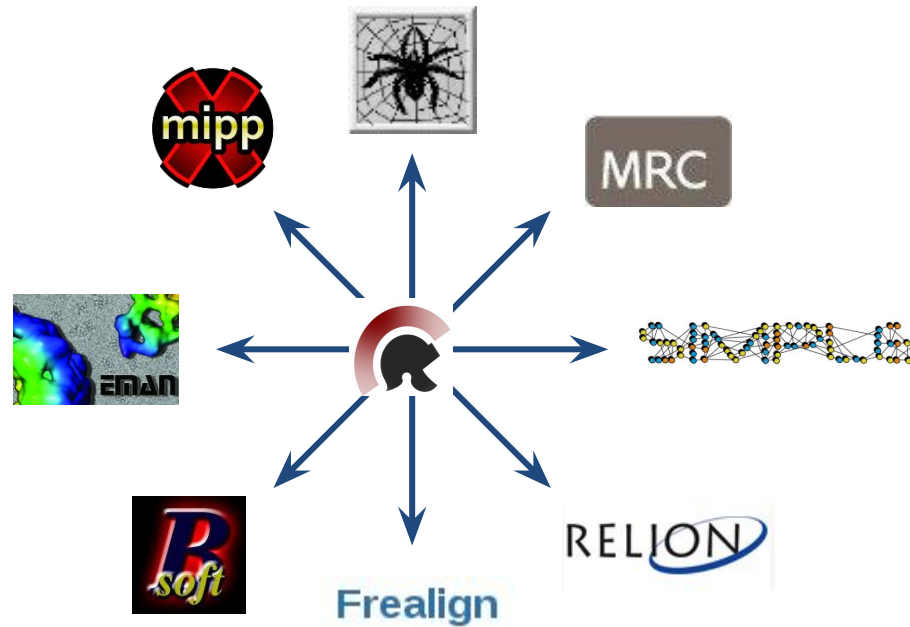
Software package usage distribution



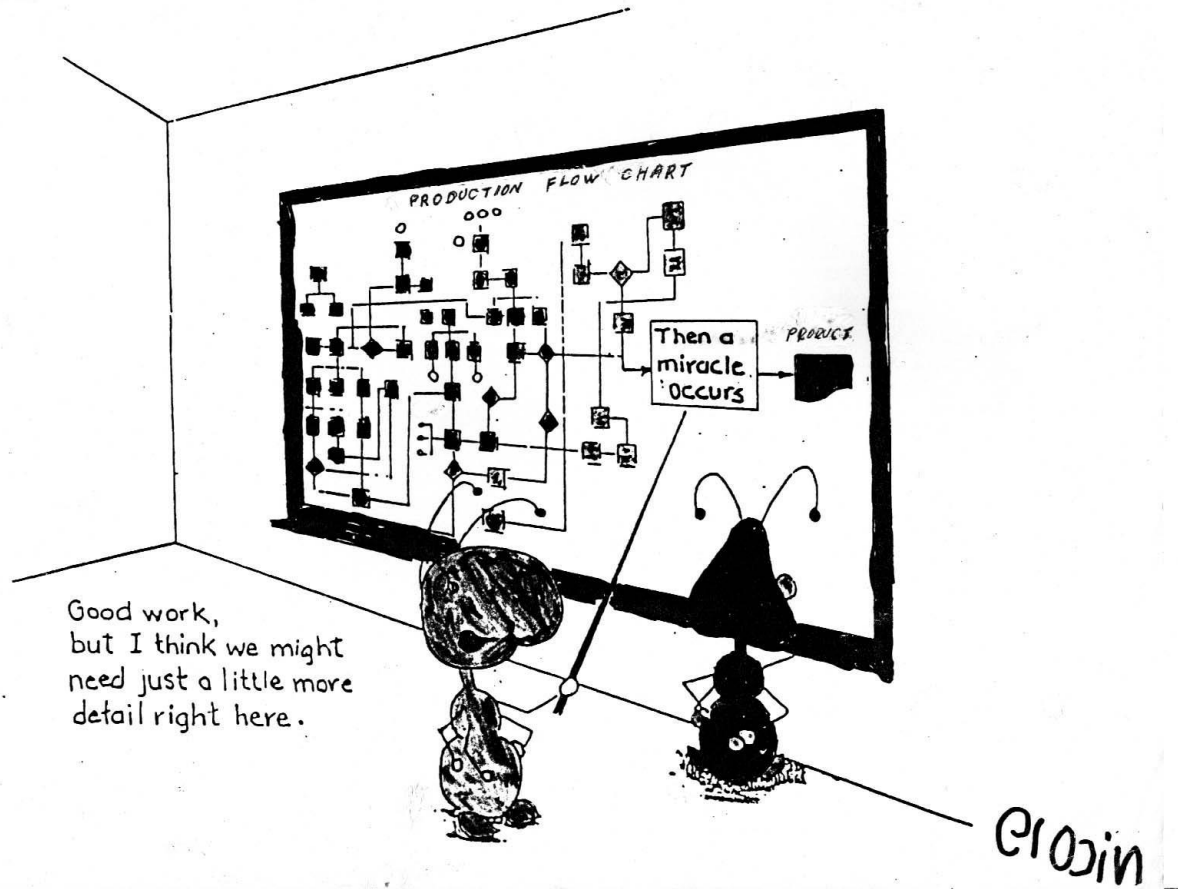
Frealign



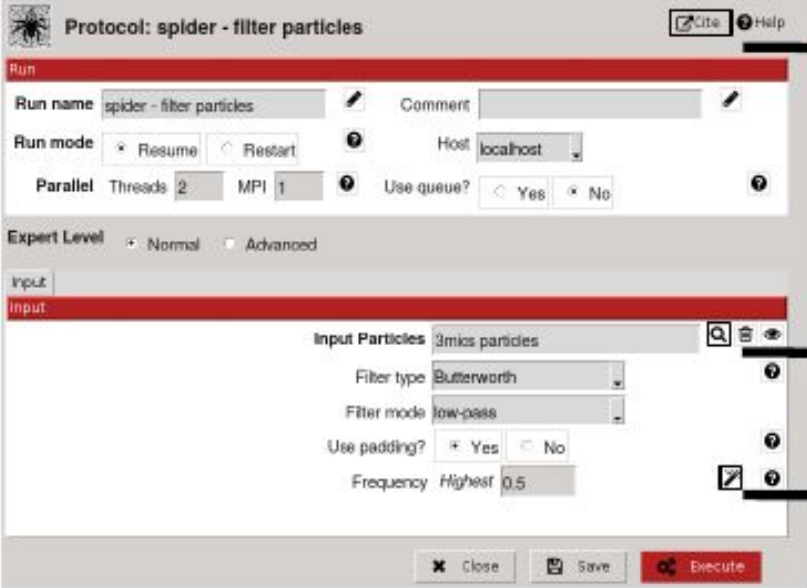
It is better to have a common format



Results should be reproducible, no 'black-boxes'



All parameters are also stored


A  Protocol: spider - filter particles

Run name: spider - filter particles
Run mode: Resume | Restart
Host: localhost
Parallel: Threads 2 | MPI 1
Use queue? Yes | No

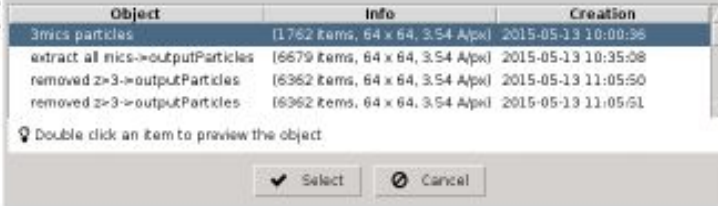
Expert Level: Normal | Advanced

Input Particles: 3mics particles
Filter type: Butterworth
Filter mode: low-pass
Use padding? Yes | No
Frequency: Highest 0.5

Buttons: Close, Save, Execute

B  Package References:
Sheikh, et al. Nature Protocols, 2008
Frank, et al. JSR, 1996

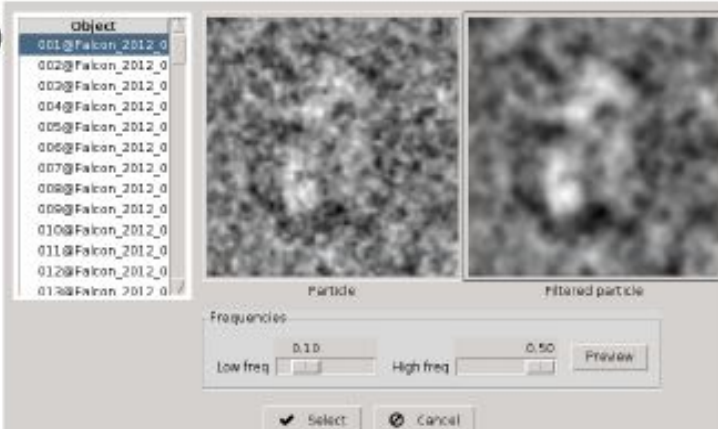
Buttons: OK

C 

Object	Info	Creation
3mics particles	(1762 items, 64 x 64, 3.54 Apx)	2015-05-13 10:00:36
extract all mics->outputParticles	(6679 items, 64 x 64, 3.54 Apx)	2015-05-13 10:35:08
removed z>3->outputParticles	(6362 items, 64 x 64, 3.54 Apx)	2015-05-13 11:05:50
removed z>3->outputParticles	(6362 items, 64 x 64, 3.54 Apx)	2015-05-13 11:05:51

Double click an item to preview the object

Buttons: Select, Cancel

D 

object

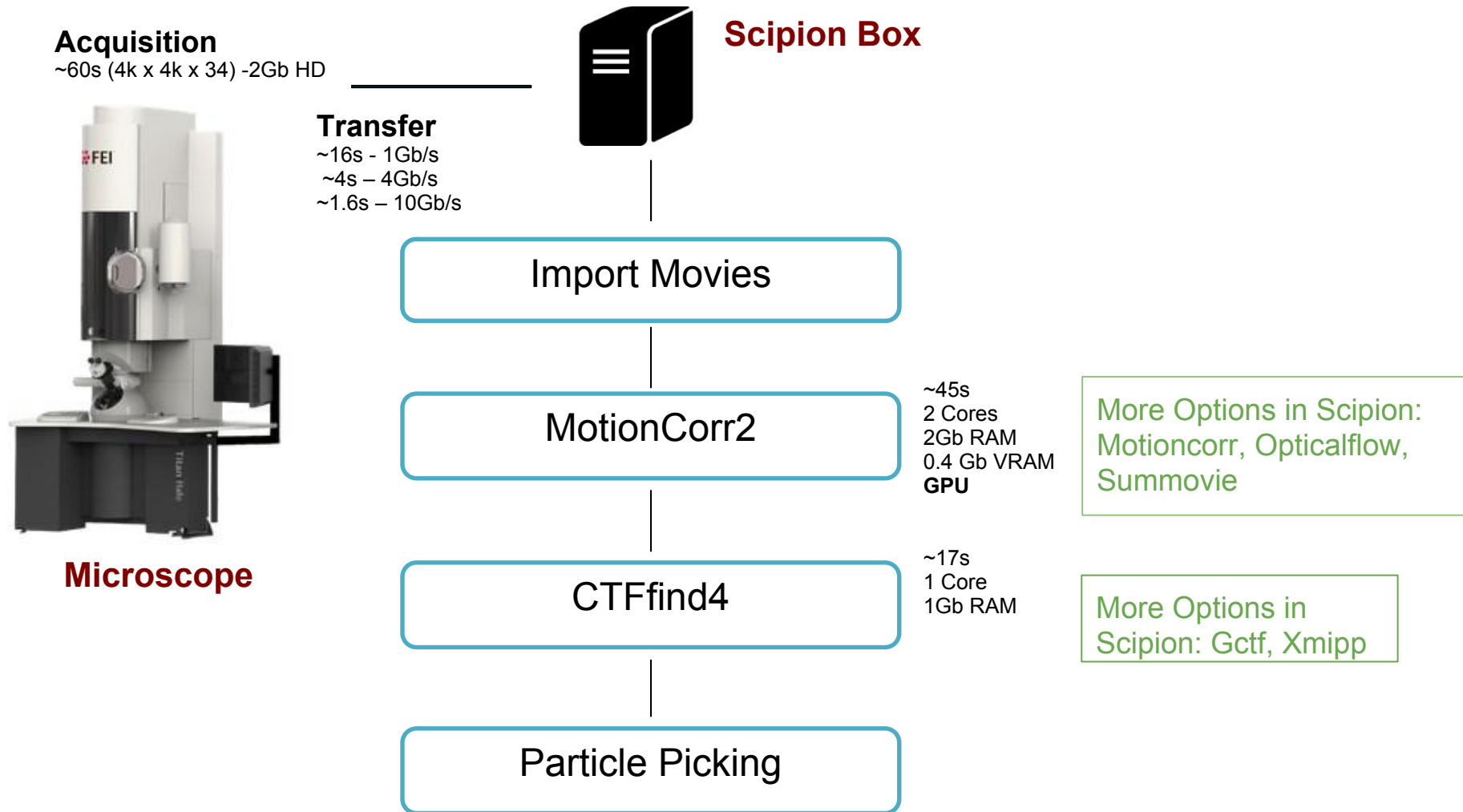
- 001@Falcon_2012_0
- 002@Falcon_2012_0
- 003@Falcon_2012_0
- 004@Falcon_2012_0
- 005@Falcon_2012_0
- 006@Falcon_2012_0
- 007@Falcon_2012_0
- 008@Falcon_2012_0
- 009@Falcon_2012_0
- 010@Falcon_2012_0
- 011@Falcon_2012_0
- 012@Falcon_2012_0
- 013@Falcon_2012_0

Particle | Filtered particle

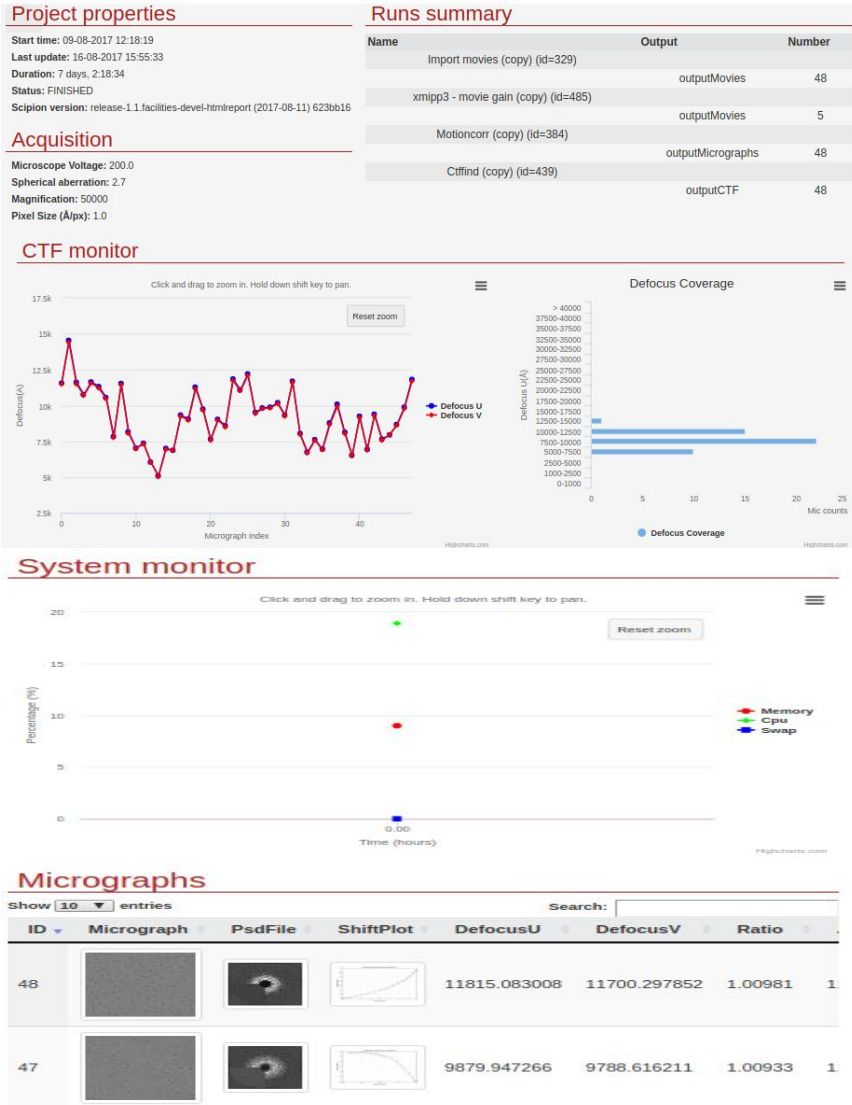
Frequencies: Low freq 0.10 | High freq 0.50 | Preview

Buttons: Select, Cancel

Scipion box: run workflows automatically in streaming



Scipion box: Monitoring

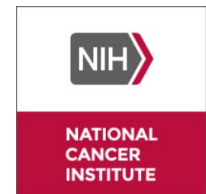


- Generic project info and items count (movies, ctf, micrographs)
- Defocus U and V changes, coverage
- System monitor: Memory, Swap, cpu
- HTML output and alarms



Scipion box: Status

- Just released with v1.1: June 2017
- Next: picking, extraction and initial volume in streaming mode. improve report
- In production at:
 - CNB (here)
 - SciLifeLab
 - eBIC (Diamond Light Source)
- Being evaluated at:
 - FEMR - McGill university - Canada
 - Center for Cancer Research - NIH (Bethesda)
 - Necen - (The netherlands)
 - ESRF (Grenoble)
 - Ceitec (Brno)



Scipion Cloud

Image instances at [Amazon Cloud](#) and European Academic Cloud ([Federated cloud](#)).

The screenshot shows the Amazon EC2 console interface. At the top, there are navigation menus for 'Services' and 'Resource Groups'. On the left, a sidebar contains links for 'EC2 Dashboard', 'Events', 'Tags', 'Reports', 'Limits', and 'INSTANCES'. The main content area features a 'Launch' button and an 'Actions' dropdown. Below this is a search bar for 'Public images' with the search term 'scipion'. A table lists the search results:

<input type="checkbox"/>	Name	AMI Name	AMI ID	Source	Owner	Visibility	Status	Creation Date	Platform	Root Device T	Virtualization
<input checked="" type="checkbox"/>	ScipionCloud-1.0.3		ami-bd663fce	510009172324/...	510009172324	Public	available	December 1, 2016 at 11:58:...	Other Linux	ebs	hvm

The screenshot shows the 'Applications Database' website. The header includes the logo and navigation buttons for 'Home', 'Software Marketplace', and 'Cloud Marketplace'. The breadcrumb trail is 'Home > Software > Science Workflows > Scipion'. Below this, there are tabs for 'Information', 'Publications (0)', 'Software Releases', and 'Comments & Ratings'. The main content area displays the 'Scipion' entry with a rating of five stars (unrated) and a 'follow' button. The entry is categorized under 'Science Workflows' and includes disciplines like 'Health-related biotechnology' and 'Pharmaceutical biotechnology'. It also has tags for 'Electron Microscopy' and 'Spain'. A description at the bottom states: 'Scipion is an image processing framework to obtain 3D models of macromolecular complexes using Electron Microscopy.'

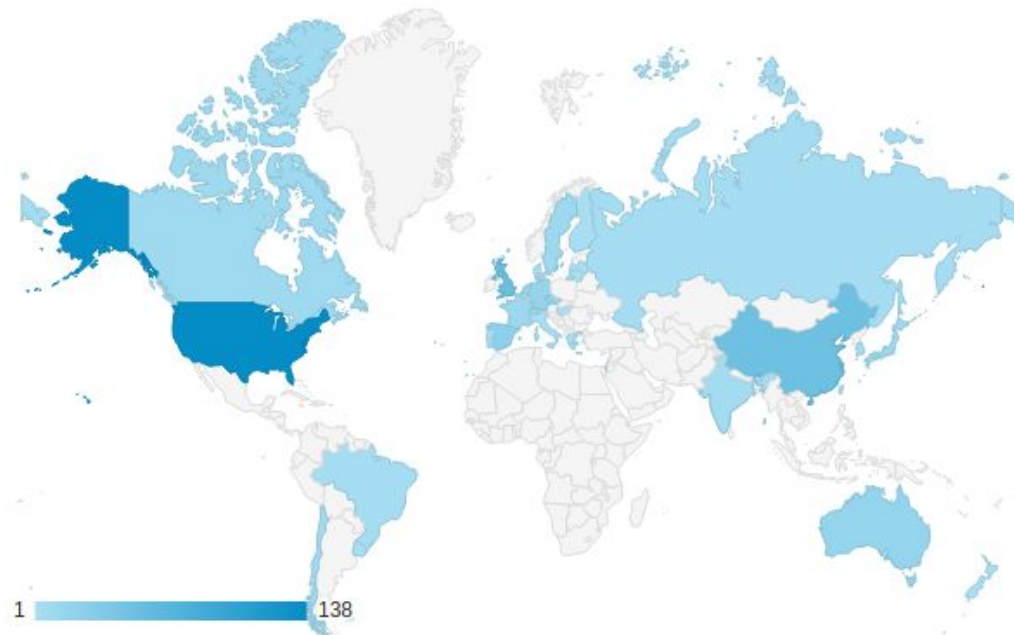


Scipion stats

- Almost at 1300 downloads since February, 2016
- Worldwide used
- Second release, v1.1, out since June.

Downloads by country

Número de visitas a páginas ▾



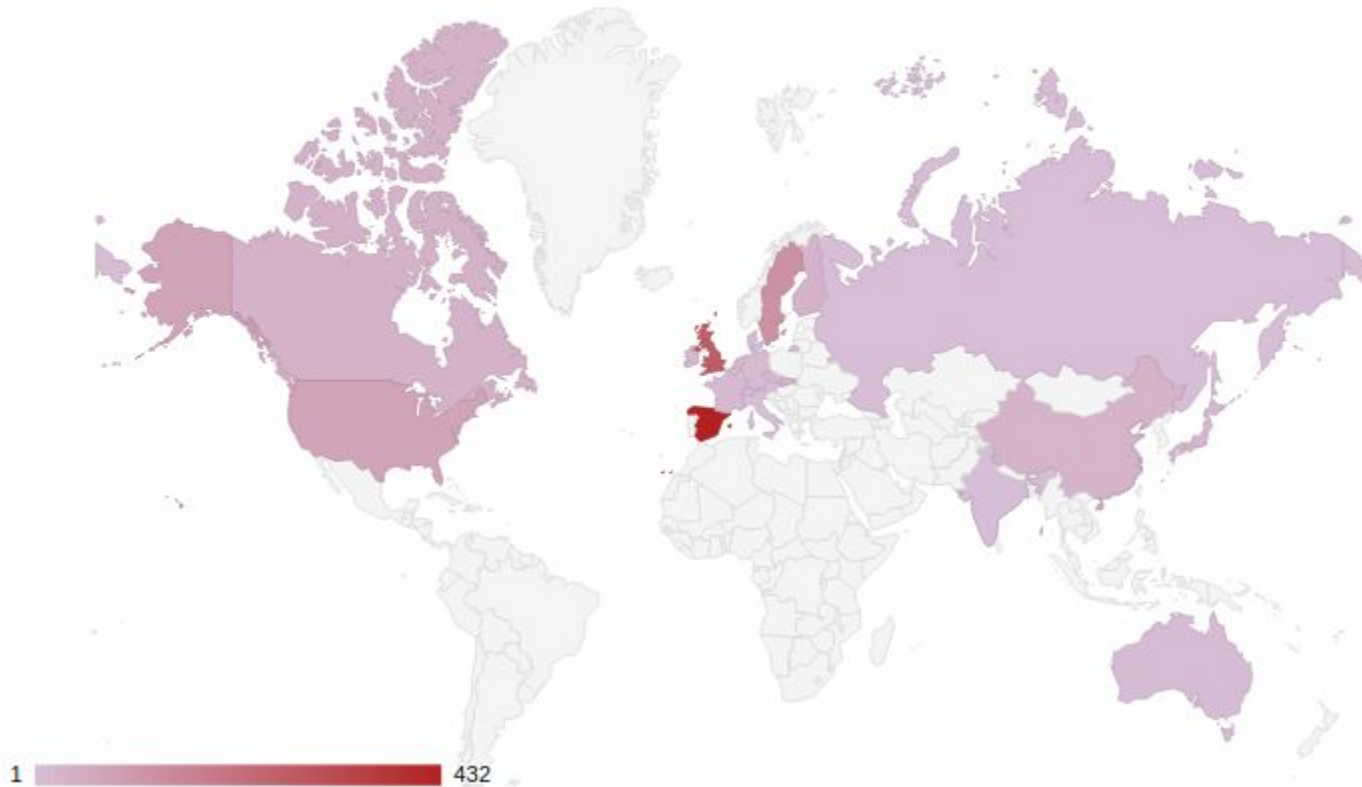
Scipion stats

- Worldwide used

Scipion usage map

Map showing where Scipion it's being used. Number refer to the number of Scipion projects created.

A big, big "thank you" to all of those that have allowed scipion to send usage statistics. You should see you country in the map!



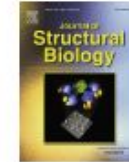
EM data deposition and standards

- Only 1 EM standard: EMX, incomplete, implemented by some SW



Journal of Structural Biology

Volume 194, Issue 2, May 2016, Pages 156-163



The Electron Microscopy eXchange (EMX) initiative

Roberto Marabini ^a ✉, Steven J. Ludtke ^b, Stephen C. Murray ^f, Wah Chiu ^b, Jose M. de la Rosa-Trevin ^c, Ardan Patwardhan ^d, J. Bernard Heymann ^e, Jose M. Carazo ^c

^a Escuela Politecnica Superior, Universidad Autonoma de Madrid, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain

^b National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

^c Biocomputing Unit, National Center for Biotechnology (CSIC), c/Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain

^d European Molecular Biology Laboratory – European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

^e Laboratory of Structural Biology Research, National Institute of Arthritis, Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA

^f Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030 USA

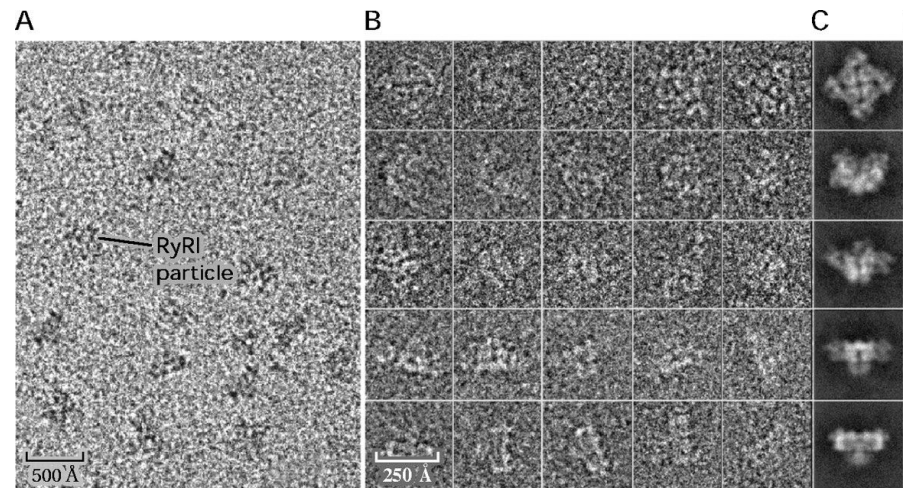
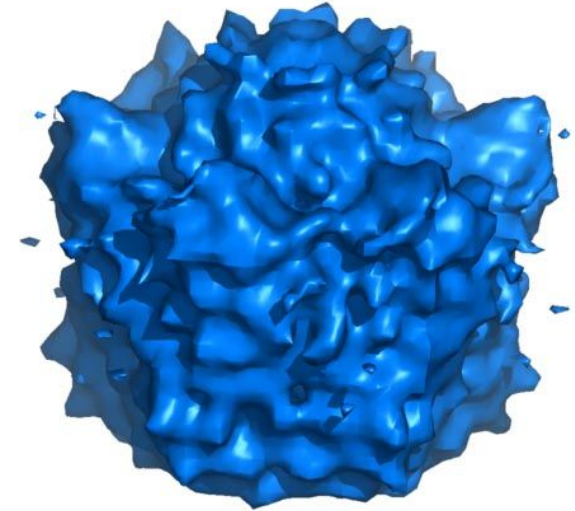
Received 3 September 2015, Revised 2 February 2016, Accepted 5 February 2016, Available online 9 February 2016.



EM data deposition and standards

EMDB: Accepts final outcome of an EM workflow (the volume)

EMPIAR: Accepts RAW and intermediate binary files like movies, micrographs (A), particles (B) or averages (C)



Goals within this project

1. Improve our workflow export file to meet FAIR principles to:
 - a. contain detailed information enabling the reproduction of processing steps
 - b. be accepted to be deposited in cryo-EM databases like EMDB and EMPIAR
2. Easy to browse and analyze over the Web
 - a. Create a widget to visualize our workflow
 - b. Easy to “plug” in deposition databases like EMDB
3. Help facilities to run Scipion processes on the Cloud when required.



Workplan

Work planned for Q1

Jul 2017- Sep 2017

The work will be centred in the analysis of the specific requirements for the implementation. Including:

- Ways to express workflows: Initially our proposal contemplated the use of the so called CWL, but our analysis so far indicates that CWL is mostly oriented to command line interfaces and could result in cumbersome workflow description in our case. This is so since most of our software work at the level of "Functional Units", which handle both sophisticated cryo-EM specific operations as well as many preparatory steps. Currently we consider that these Functional Units are the ones to be specified in the workflow, and that a lightweight data-interchange format such as Json format would be adequate for this task
- Analysis of the precise information to be stored to assure reproducibility and summarization, considering both current preprocessing workflows and new trends extending them.
- Work with EMDB/EMPIAR to assure that no incompatibilities of any kind would exist for the submission

CWL was analyzed and considered unsuitable for our case. Our JSON file will be extended to meet new requirements. Still there is a chance that we can use CWL.



Workplan

Work planned for Q2	<p>Oct 2017- Dec 2017</p> <ul style="list-style-type: none">-Version 1 of the implementation of a Json-based detailed workflow description covering all steps of current preprocessing workflows within Scipion.-Version 1 of the Implementation of a web browser for our workflow description files <p>Note: We would aim to have a first working version of developments until this point for the EOSC meeting 27th-28th of November.</p>
Work planned for Q3	<p>Jan 2018-Mar 2018</p> <ul style="list-style-type: none">-Final version of the implementation of a Json-based detailed workflow description covering all steps of current preprocessing workflows within Scipion.-Final version of the Implementation of a web browser for our workflow description files.-Implementing Scipion capability to import and reproduce workflow description files.
Work planned for Q4	<p>Apr 2018-Jun 2018</p> <ul style="list-style-type: none">-Deployment at Facility sites

