Testing the use of OMERO as a bioimage informatics platform for HIV viral particle tracking

Caterina Strambio De Castillia
Jeremy Luban
Mario Valle (ETH-CSCS)
What is our problem?

People living with HIV (2007)

25 years after the identification of the virus, we still know remarkably little about the cell biology of HIV-1!
What is our approach?

• **The question:** How does the HIV-1 virus gains access to the cellular genome to ensure proviral DNA integration?

• **The obstacle:** the reductionist approach has reached its limit

• **The answer:** a systematic quantitative strategy requires a multidisciplinary team
  – Virology
  – Cell Biology
  – Bio-Image informatics = integration of OME + analysis algorithms + integrated image analysis workflows

33 mil people living with HIV
Who are we?

Caterina Strambio
UniGe
Cell biology
bioimaging

Jeremy Luban
UniGe
HIV-1
Virology

Mario Valle
ETH-CSCS
Visualization
data analysis and
management

...and the team is growing!
The question: HIV-1 viral entry

- HIV-1 enters the cell and moves towards the nucleus
- Key virus-host interactions influence viral trajectory within the cell
- Biochemical and cell-fractionation studies have reached their limit
- Real-time tracking of viral particles and mathematical analysis of movement can provide key insights into the mechanism of infection
Huge challenge: What do we need to understand HIV-1 viral movement?

1. Infect primary human cells under strict bio-safety regimens
2. Visualize HIV-1 particles as they move within the cell with high spatial and temporal resolution
3. Keep track and manage multiple large datasets
4. Mathematical analysis of viral trajectories
5. Statistical analysis on multiple trajectories
6. Integrate with other kind of knowledge

Essential to develop reproducible scientific workflows
The beginning of the answer: real-time 3D video microscopy of HIV-1

HIV-1 moves along the cytoskeleton

Mean Square Displacement analysis of trajectories
The workflow for real-time HIV tracking

- SPT: identification of viral trajectories in 3D image stacks over limited time
- Semi automated, supervised trajectory segmentation
- Manual statistical analysis of numerical trajectory data


Stand alone software is being developed...

...but for systematic use it has to be integrated
Can OMERO provides the glue?

Diagram:
- Digital Image Acquisition System
- Raw Data
- Metadata
- Quantitative Analysis
- Data Management, Tagging, Querying
- Visualisation
- Processed Data
- Metadata

2009 OME User’s Meeting, Paris / May, 19th-20th 2009
What do we need from OMERO?

✓ Image data management
✓ Image data visualization
✓ Metadata (both acquisition and custom) integration

➡ Image analysis:
  - Modular image analysis
  - Keeping track of analysis outputs
  - Managing complex workflows: Reproducibility across users, time and space

➡ Facilitate integration of different data types and information source: semantic web
What we have done so far:
OMERO testing on two fronts

- **End user**
  - Provide feedback on various aspects of OMERO usability and relevance
    - Data model
    - Data integrity and security issues
    - Metadata editability
    - OMERO.editor

- **Developer**
  - Test feasibility of a local OMERO analysis paradigm
  - Build a small custom C++ client that would serve as the interface between images stored in OMERO and the stand-alone SPT software
OMERO.analysis: local client model
OMERO.cpp: building a small custom client

• **Goals**
  – Understand Client API writing a small C++ client
  – Why C++?
    • Because the SPT algorithm is in C++
    • Because I know it.
    • Because ICE make languages interchangeable.

• **Results**
  – Limited success…
  – Connect, list user data, list user projects, list datasets for one project, list images for one dataset, access the image data, access the pixels
OMERO.cpp: building a small custom client

• On the positive side
  – It works, and compiles too!
  – New API simplifies common tasks

• On the negative side
  – Documentation is not available or at least not usable (not only about C++ API in general, but also about individual functions)
  – Javadoc pages are not up to date (OME vs OMERO not well documented in Javadoc…)
  – Code examples: would be useful to be able to share and re-use code the community develops.
  – Searchable list-serve archives (along the same lines) and/or community wiki would reduce workload from OME team
What we need to allow specialized analysis modules full integration in the image database?
- integrating analysis modules in the server
- analysis modules repository for sharing
- quantitative results stored in the database
- repository for sharing and reusing workflows

OMERO. analysis: modular analysis paradigm
What architecture for analysis will be supported?

Model 1
- Analysis engine GUI (parallel to rendering engine)
- OMERO.analysis (parallel to OMERO.insight)
- OMERO.script (workflow repository)

Model 2
- Analysis results in a parallel db (linked OMERO)

OMERO.server powered by ZeroC
- LAN Users (.insight & .importer)
- Data Processing/Scripting (C, C++, Python, Matlab)
- WAN Users (.web)

OMERO Beta 4.0.0 (released March 09)
- Search Index
- OMERO.ts image repositories
- Firewall

Integrated in OMERO
- Analysis results
OMERO.scripts and .grid: what is their future?
Scientific workflows for bio-image informatics: what do we need...

- Will OMERO support distributed modular analysis architectures?
- OMERO vs BISQUE?
- OMERO.workflow integrating OMERO with software such as TAVERNA to develop standardized bioimage informatics workflows
Thank you!

Caterina Strambio De Castillia
Jeremy Luban
Mario Valle (ETH-CSCS)