

OMEGA: viral particle tracking work- and dataflows for images stored in OMERO

Caterina Strambio De Castillia, Ph.D., et al.



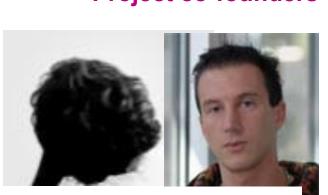
Talk outline

- Who are we:
- Vision (long term!):
 - Developing bio-image informatics as a tool to enable a systems biology approach to the understanding the retroviral life cycle
- Short term goal:
 - Develop integrated tools to analyze viral movement in life cells
- Approach:
 - Develop a prototype OMEGA software tool to deliver usable viral particle tracking, movement analysis and data-exploration capabilities to end-users at the bench side
- Progress report: what we have done so far
- Next steps: where do we want to go from here

The OMEGA team



Project co-founders



software engineering

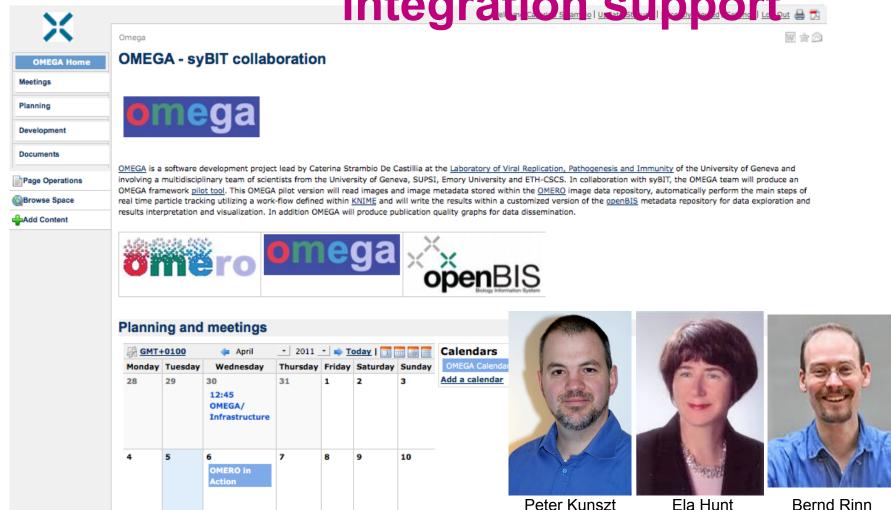




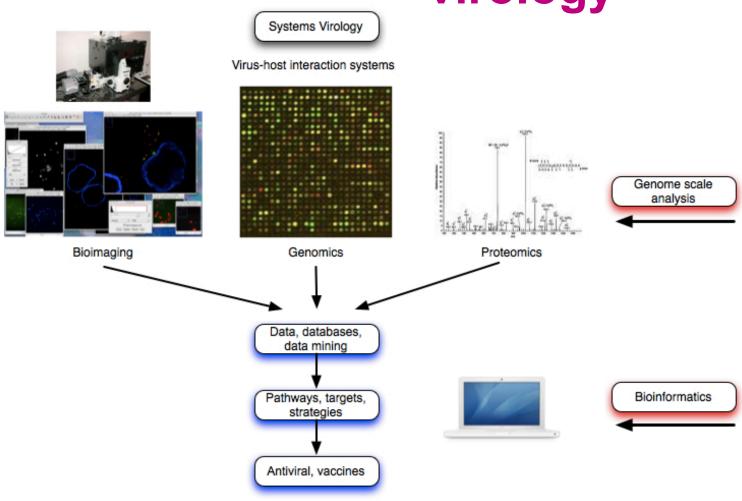
algorithms optimization and integration



syBIT: seed funding for the OMEGA project and integration support

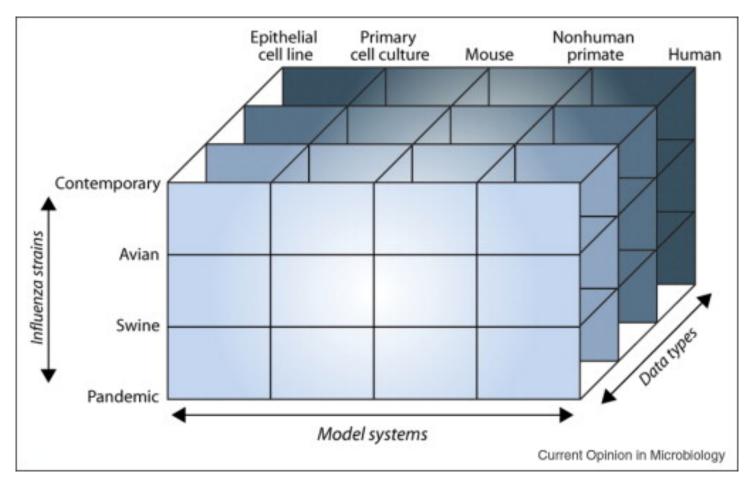


The vision: imaging as an integral part of systems virology



Adapted from: Katze et al. Nature Rev Imm (2002) vol. 2 (9) pp. 675-87

Systems virology: data could be represented using an hypercube

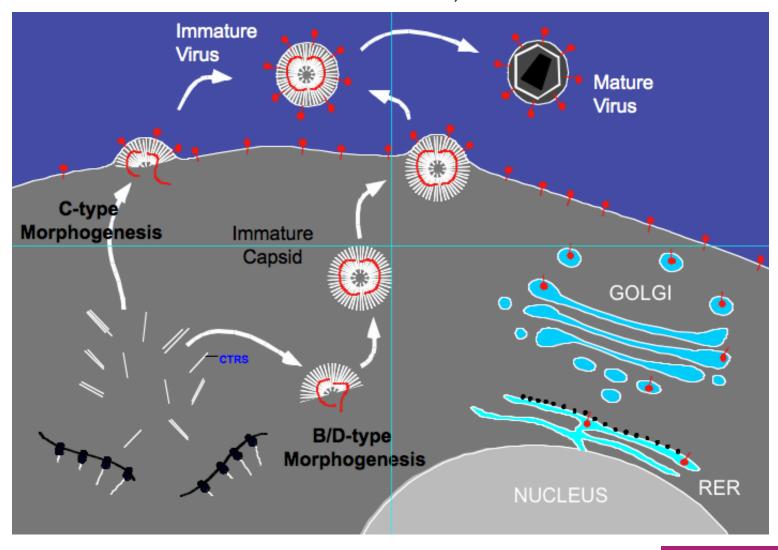


Peng et al. Curr Opin Microbiol (2009) vol. 12 (4) pp. 432-8

The "short" term goal: capturing the motion of viral particles within the infected cell

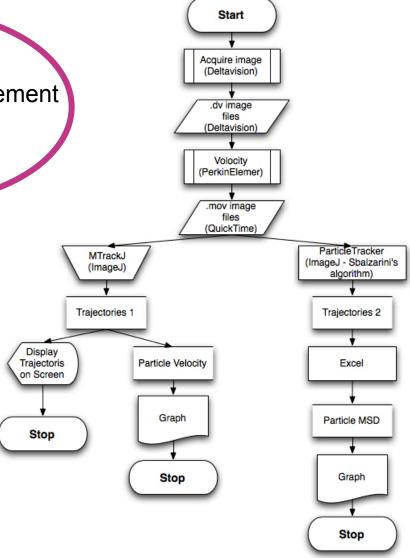
Emory Vaccine Center: tracking retroviral assembly

Jasmine Clark, Lara Pereira and Eric Hunter

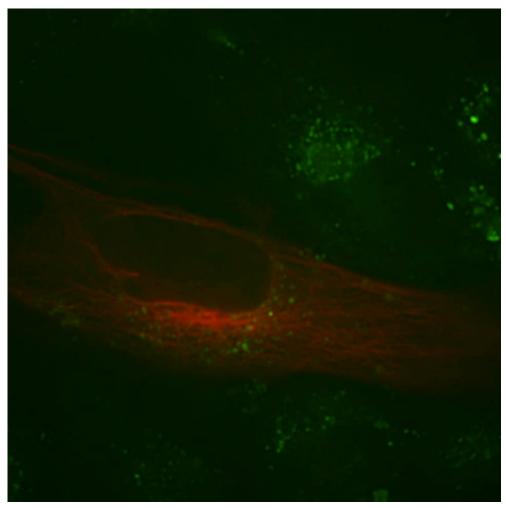


Current situation: software-assisted manual tracking and analysis of trajectories

Manual analysis
No workflow management
No data relation
No data standards



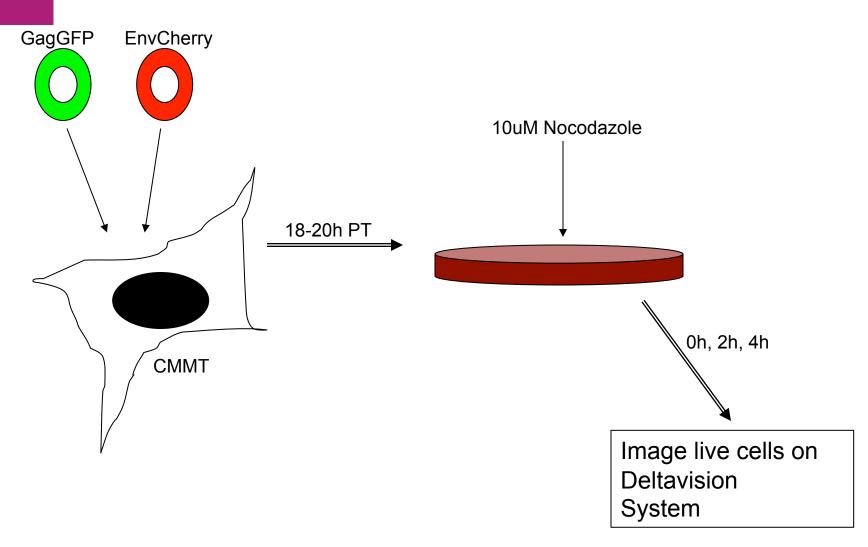
Emory Vaccine Center: retroviral movement from the towards the cell periphery



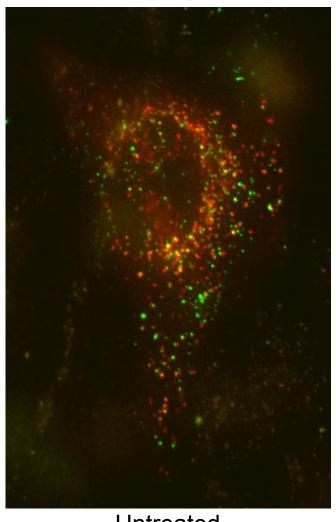
Gag-GFP mCherry-tubulin

Pereira et al. submitted

Emory Vaccine Center: experimental setup



Emory Vaccine Center: live cell imaging

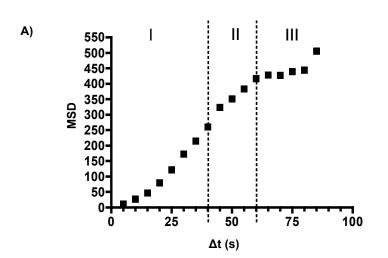


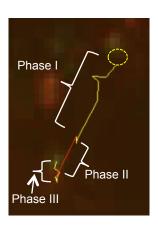
Untreated

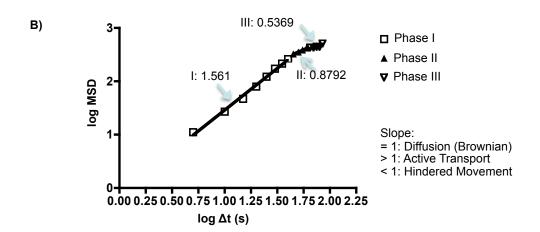
Pereira et al. submitted

2h Nocodazole

Emory Vaccine Center: trajectory analysis

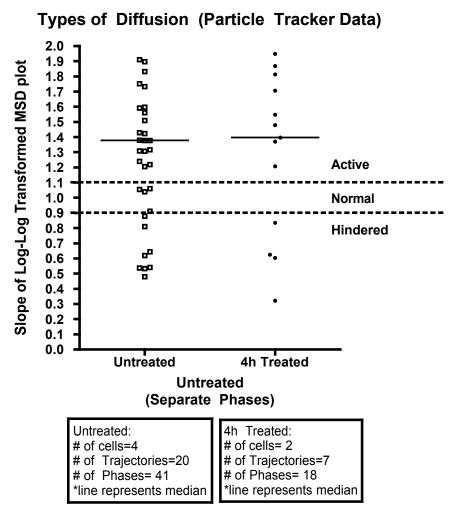






Pereira et al. submitted

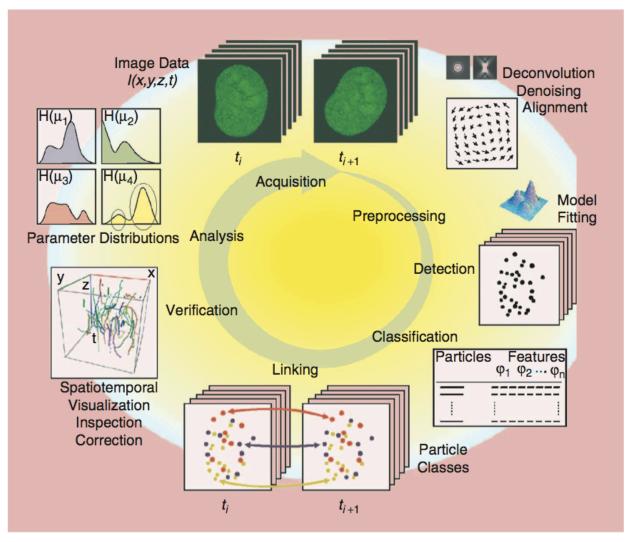
Emory Vaccine Center: trajectory analysis



Pereira et al. submitted

Problem definition: the workflow from infected cells to a mathematical model of movement

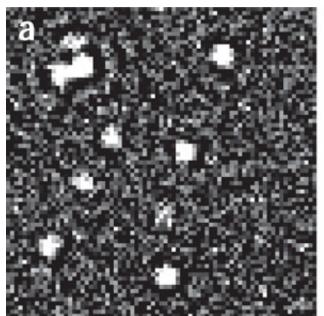
What is single particle tracking?

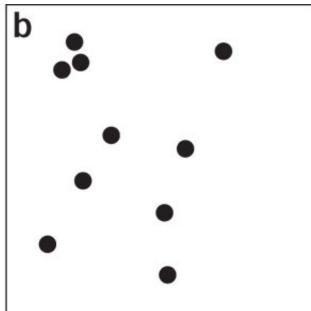


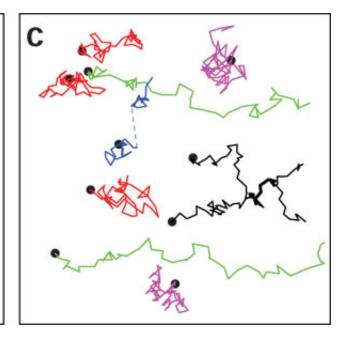
Meijering et al. IEEE signal processing magazine (2006) 23: 46

Single Particle Tracking

Connecting the dots



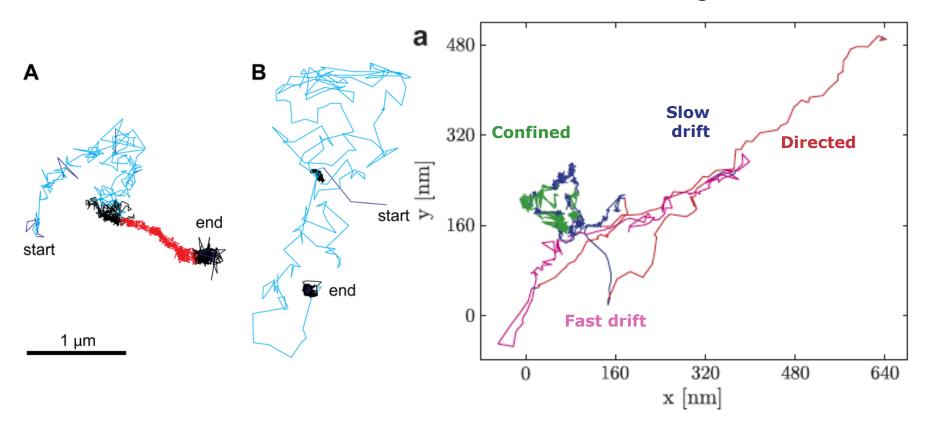




Jacqman and Danuser, 2009

Trajectory segmentation:

Identifying trajectory segments using machine learning

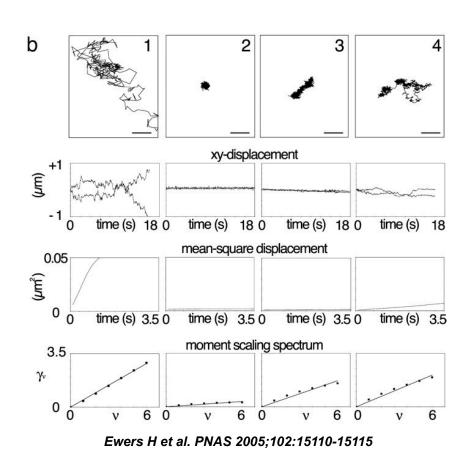


Burckhardt and Greber. PLoS Pathog (2009) 5:e1000621

Helmuth et al. J Struct Biol (2007)159: 347-358

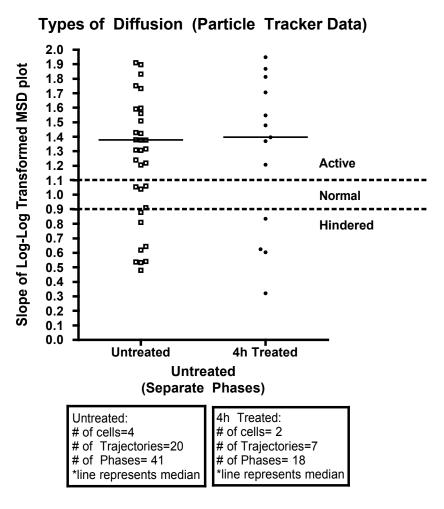
Mathematical analysis of trajectories or trajectory segments:

Computing numerical parameters capable of discriminating between types of movement



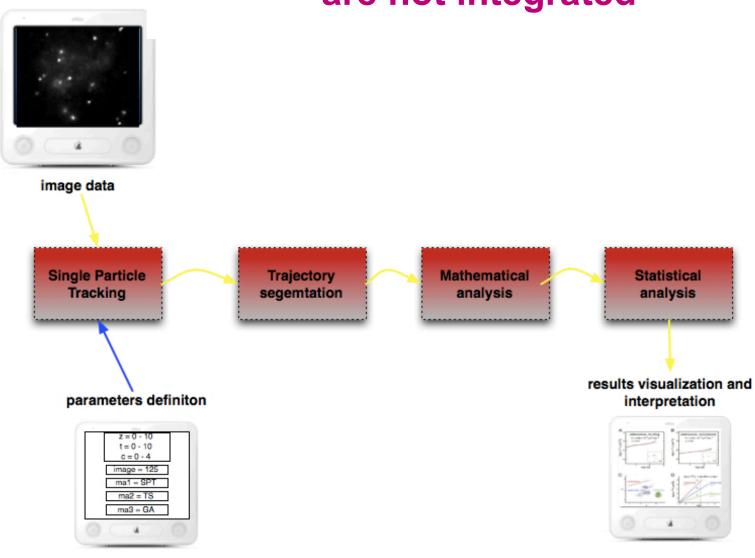
Statistical analysis of results:

Comparing results across datasets



Pereira et al. submitted

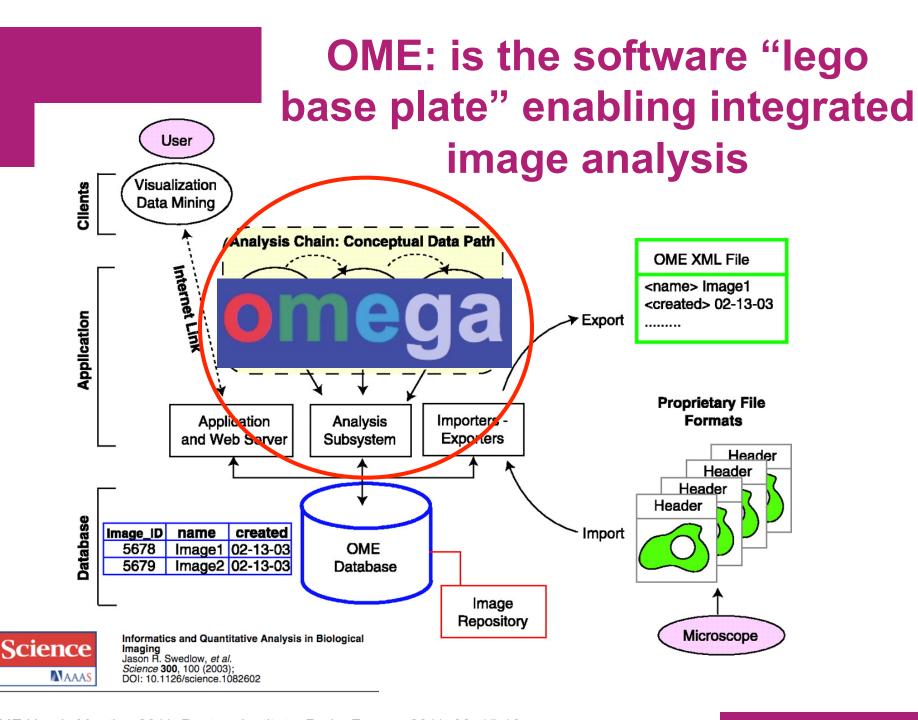
Current situation: individual steps are not integrated



One solution: integrated computational image analysis

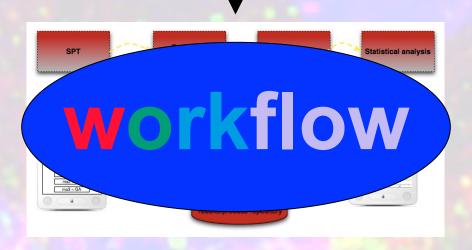


Open Microscopy Environment inte Grated Analysis

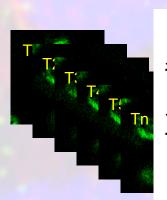


omega

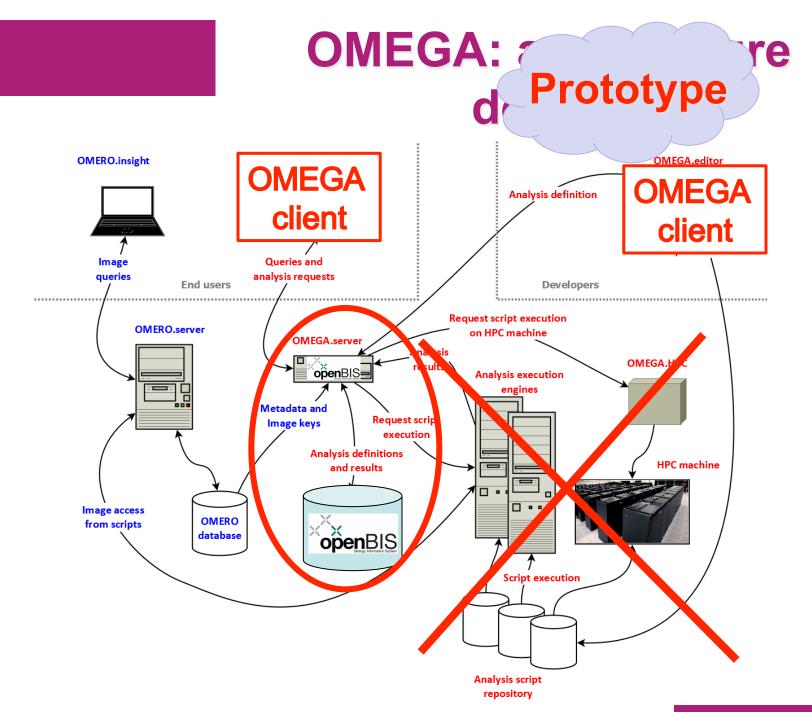
Integration of the analysis process



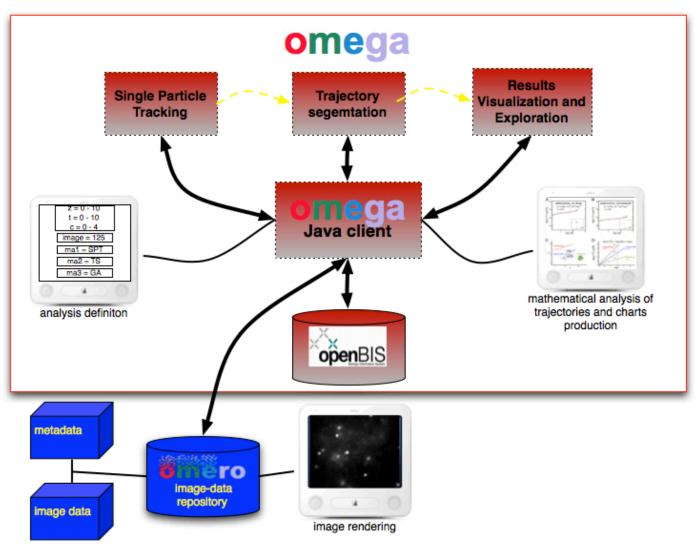
omega Integration of the data pipeline







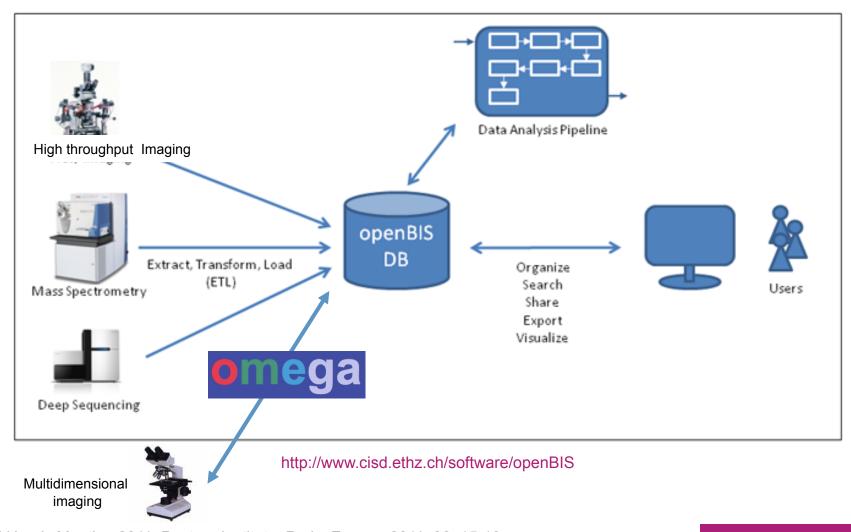
Goals of the OMEGA prototype





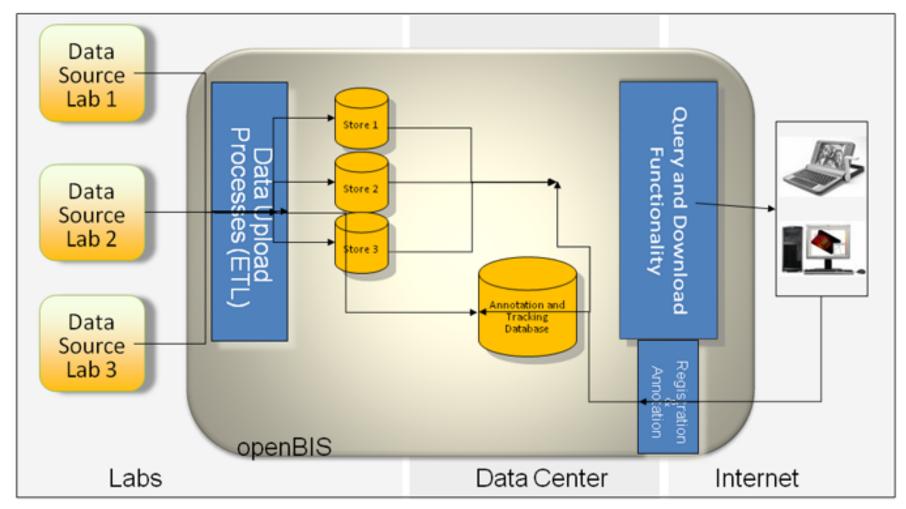


What is openBIS? a generic repository and management system for biological data





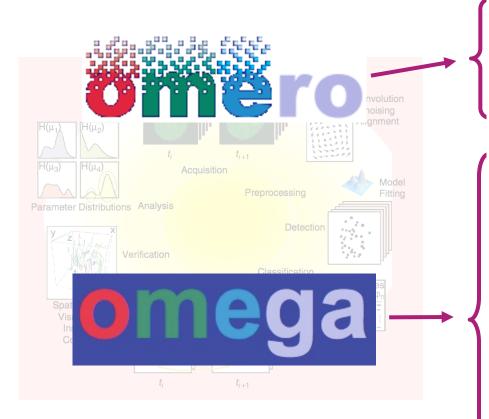
openBIS: Support for distributed storage of analysis results



http://www.cisd.ethz.ch/software/openBIS

The concrete plan

Integrated work- and dataflows for viral particle tracking



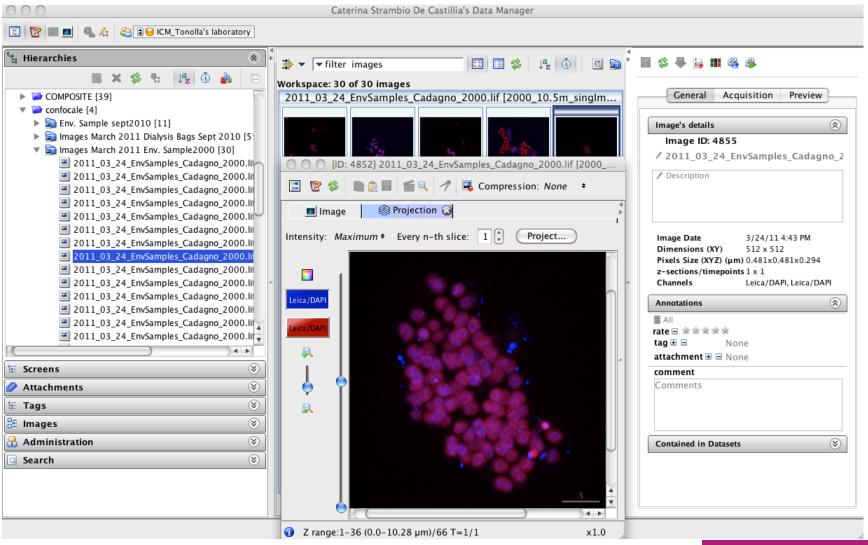
- Image pixel data/metadata management
- Image processing
- Work- and data-flow management
- Single Particle Tracking
- Trajectory segmentation
- Trajectory segments classification
- Analysis results visualization and exploration
- Statistical analysis of analysis results

OMERO installations: the starting point

- ☑ Microbiolgy Institute of Canton Ticino: image data management
- ☑ University of Applied Sciences of Southern Switzerland (SUPSI): benchmarking installation for software development
- □ Emory Vaccine Center: image data management

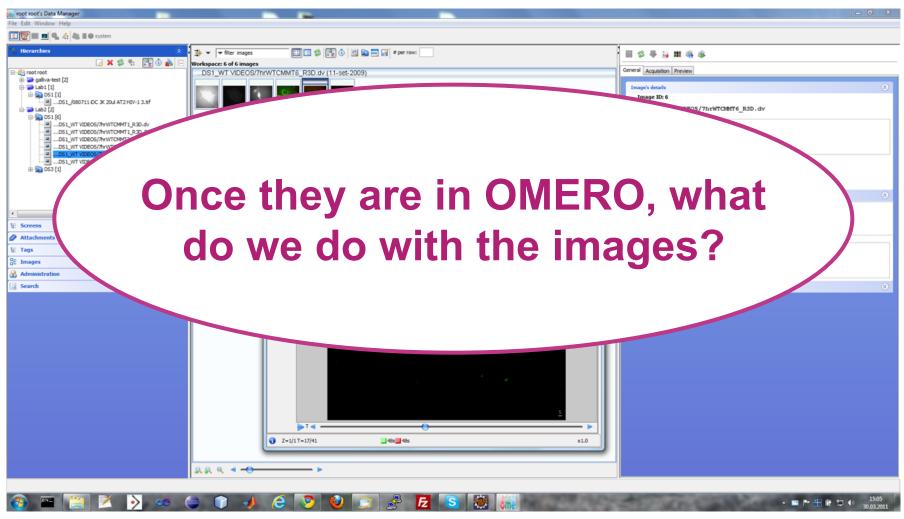
OMERO (1): environmental microbiology

Istituto Cantonale di Microbiologia del Canton Ticino

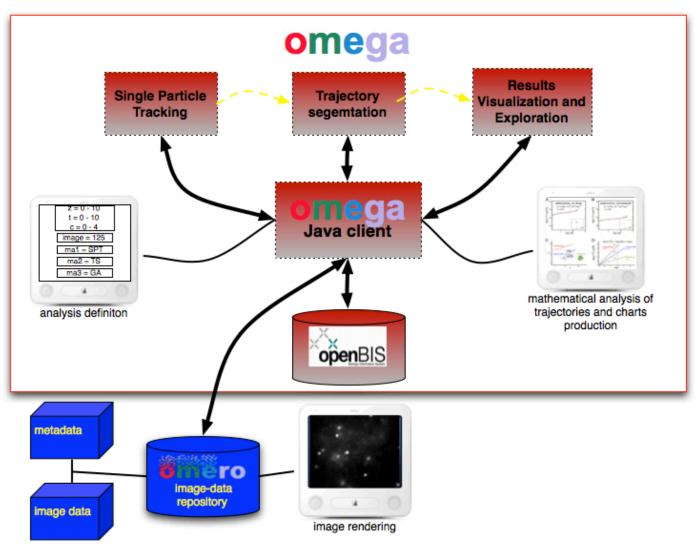


OMERO (2 and 3): viral particle tracking

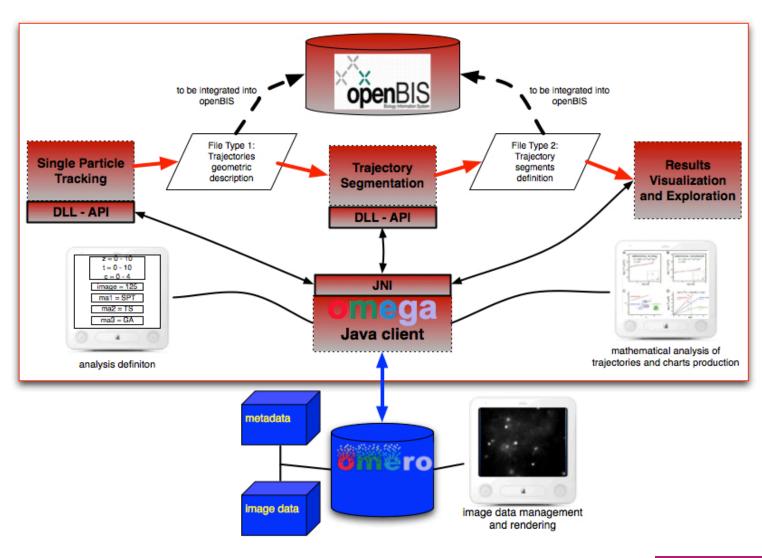
Emory University and SUPSI



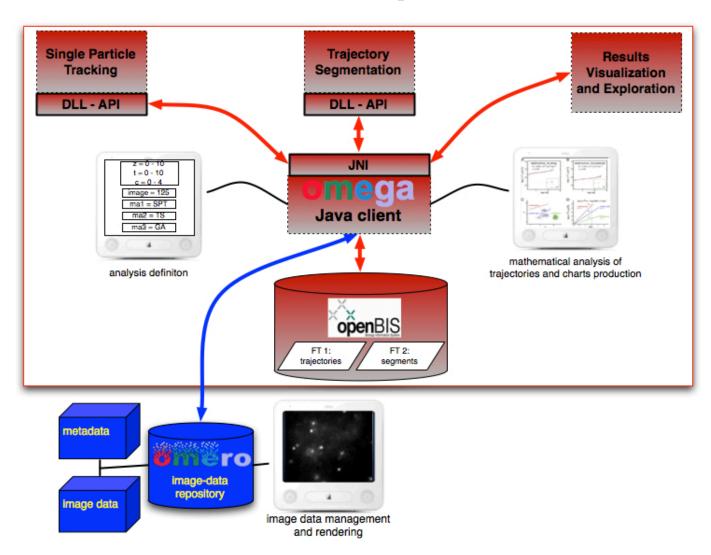
Goals of the OMEGA prototype



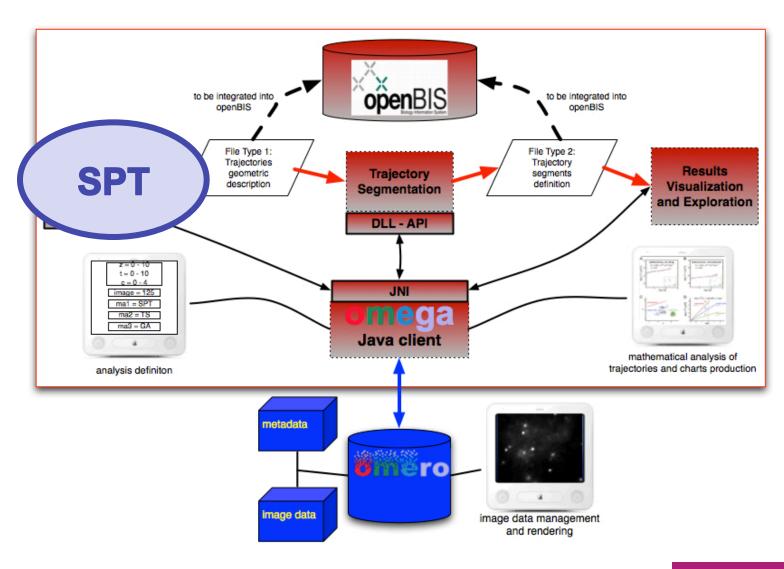
OMEGA prototype: phase I of implementation



OMEGA prototype: phase II of implementation



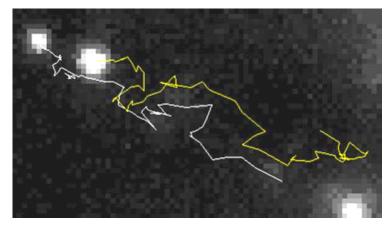
OMEGA: progress report



Single Particle Tracking

Sbalzarini's algorithm

Particle Tracker 2D / 3D



Sbalzarini and Koumoutsakos, J Struct Biol (2005) 151: 182

- •Global-nearest-neighbor approach
- Less accurate in sub-optimal conditions
- Does not allow for trajectory gaps
- Already available in C
- Easier to integrate

☑Re-engineering of sequential application in C → DONE
 ☑Multi-core optimization → DONE
 ☑GPU optimization → NEXT RELEASE

Sbalzarini's SPT: significant speed-ups will facilitate work on lap-top computers commonly available to individual scientists

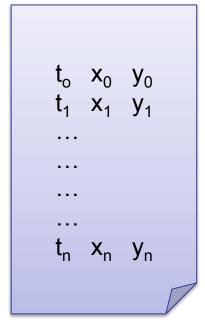
	Particle Radius 6 pixels	
	Run time (s)	Fold Speed-up
Original version (C)	197.65	
Optimized sequential version (C)	133.01	1.49
Multi-cores parallel version (8 cores)	18.18	10.87
Multi-cores parallel version (12 cores)	12.21	16.18
GPGPU parallel version (8 cores)	12.10	16.34

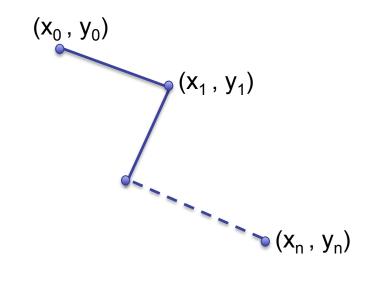
- Software re-engeneering of sequential version afforded significant performance improvements
- Multi-core parallelization on 12 cores achieves 16x speed-up
- The GPU preliminary version achieves 16x improvements already on 8 cores



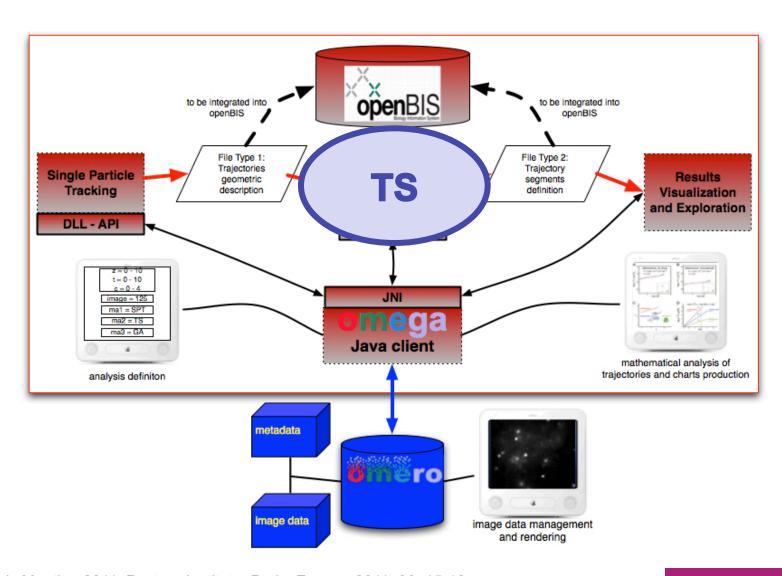
File Type 1: Trajectories geometric description





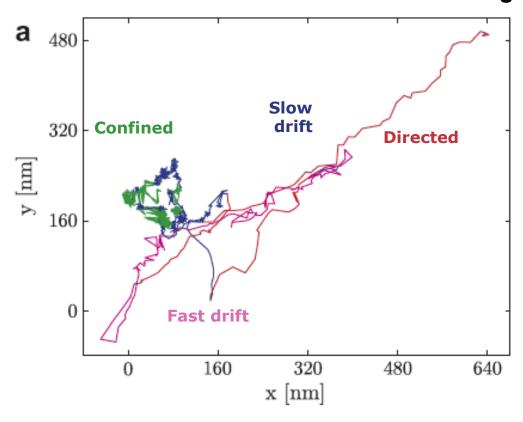


OMEGA: progress report



Trajectory segmentation:

Identifying trajectory segments using machine learning



Burckhardt and Greber. PLoS Pathog (2009) 5:e1000621

Helmuth et al. J Struct Biol (2007)159: 347-358

Trajectory Segmentation Algorithm

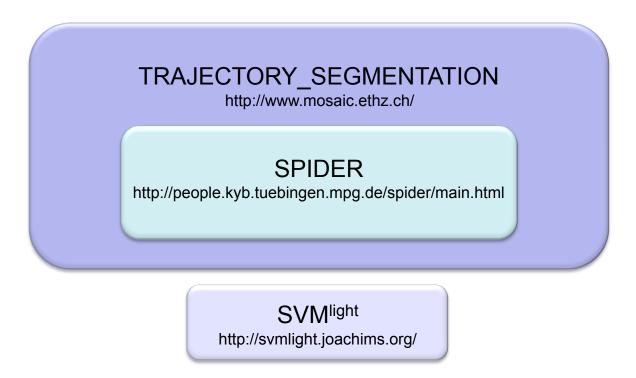
Helmuth and Sbalzarini, 2007

The trajectory segmentation algorithm consists of four steps:

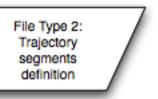
- (1)Trajectory preprocessing: remove position noise
- (2)Feature extraction: A set of 7 features is used to identify the motion type of each segments. Transform trajectory segments to low-dimensional feature vectors.
- (3)Feature classification: evaluate whether a feature vector corresponds to a certain class of motion or not in a binary classification problem. Uses user-trained Support Vector Machines (SVM) as classifiers
- (4)Label assignment: classification ambiguities are resolved and motion types labels are assigned to each individual trajectory segment

Trajectory Segmentation MATLAB TOOLBOX

Jo Helmuth and Ivo Sbalzarini



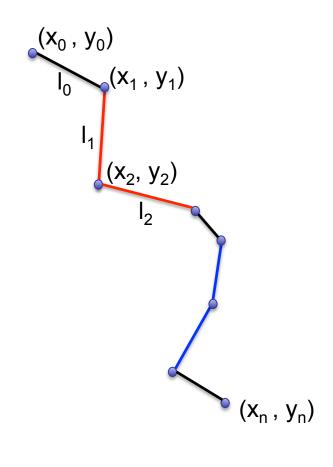
- The engine of the application was ported to C
- •The GUI necessary for supervised machine learning has been integrated into the main OMEGA Java client



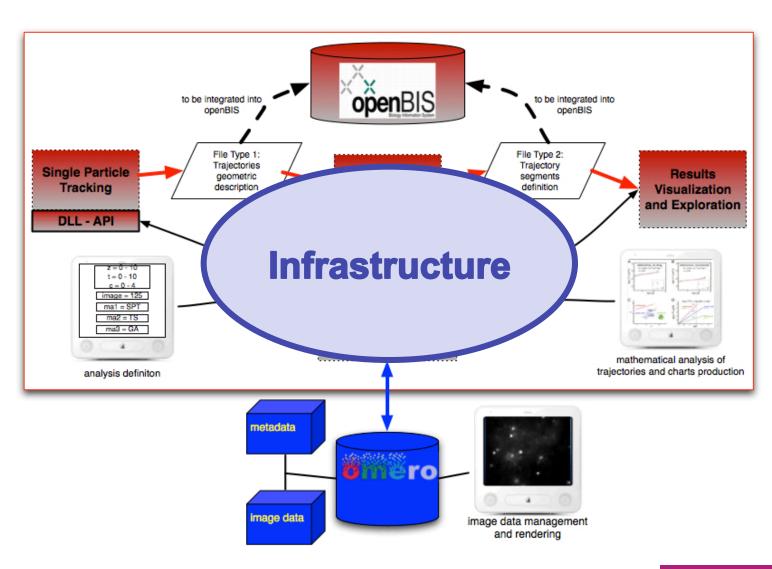
TS output: File Type 2

coord label

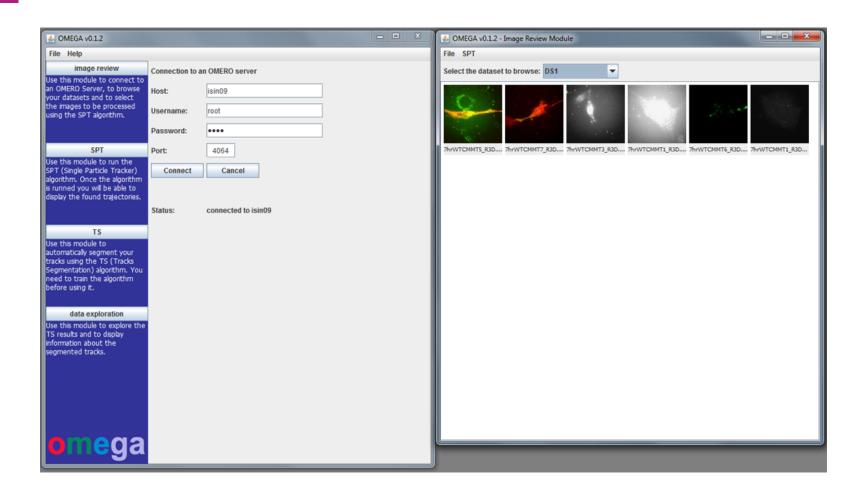
\mathbf{x}_0	y ₀	
x ₁	y ₁	I ₀
x ₂	y ₂	
X ₃	y ₃	l ₂
X ₄	y ₄	l ₃
		l ₄
X ₅	y ₅	l ₅
x ₆	y ₆	l ₆
X ₇	y ₇	l ₇
x ₈	y ₈	
x ₉	y ₉	I ₈
X ₁₀	y ₁₀	l ₉
X ₁₁	y ₁₁	I ₁₀



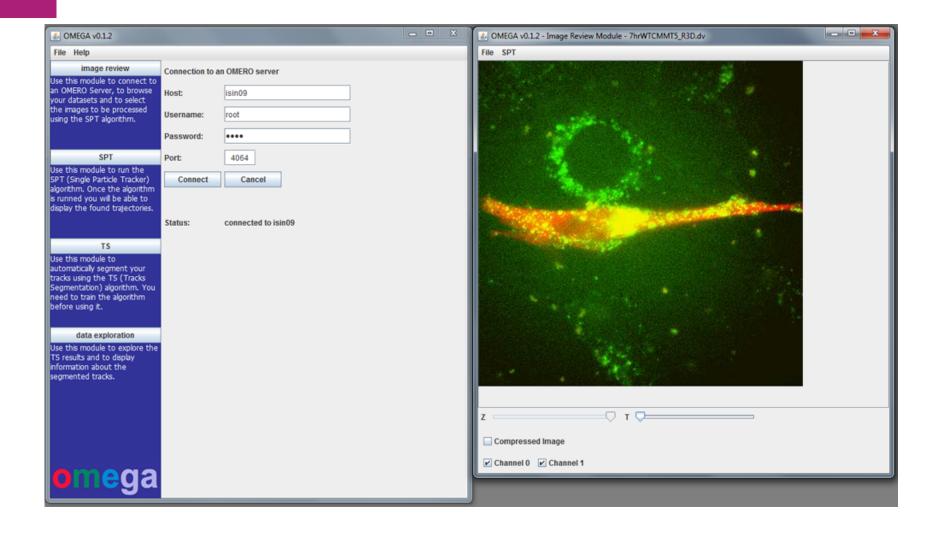
OMEGA: progress report



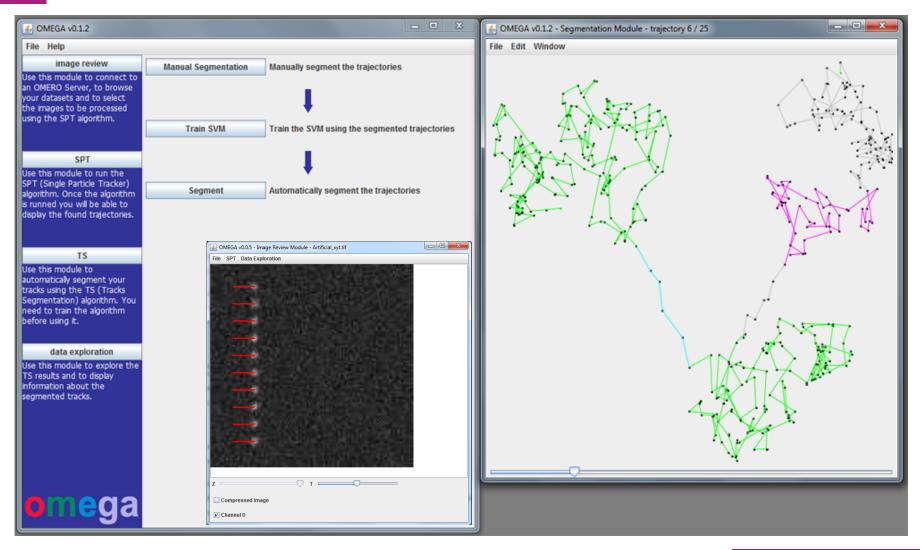
OMEGA: connecting with OMERO



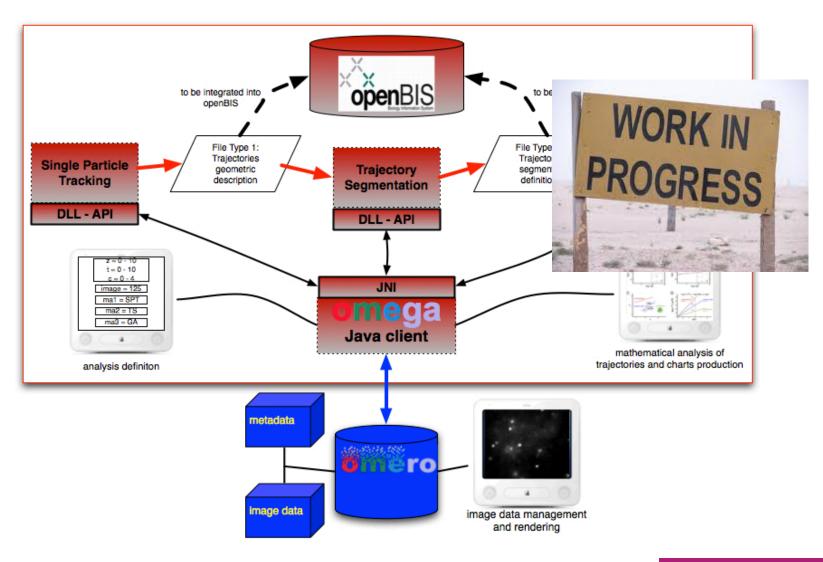
OMEGA: images review



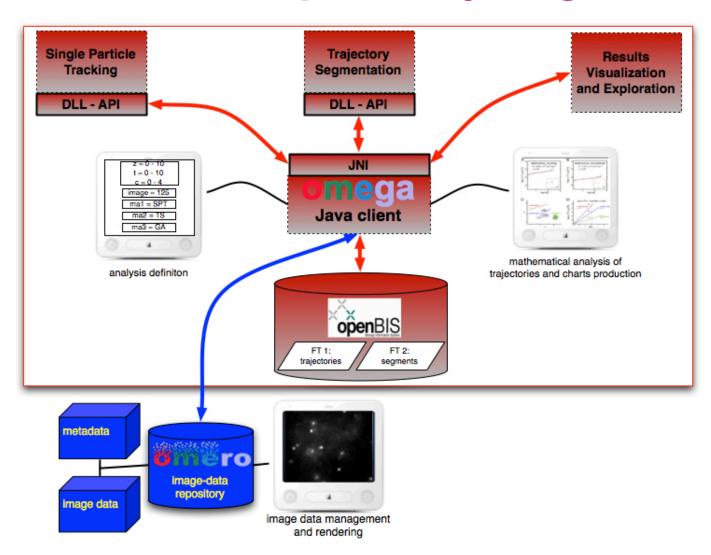
OMEGA: running image analysis



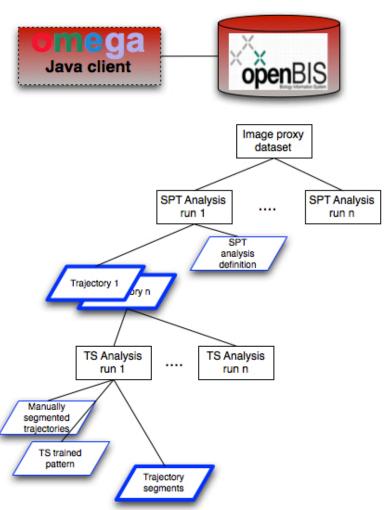
OMEGA: progress report



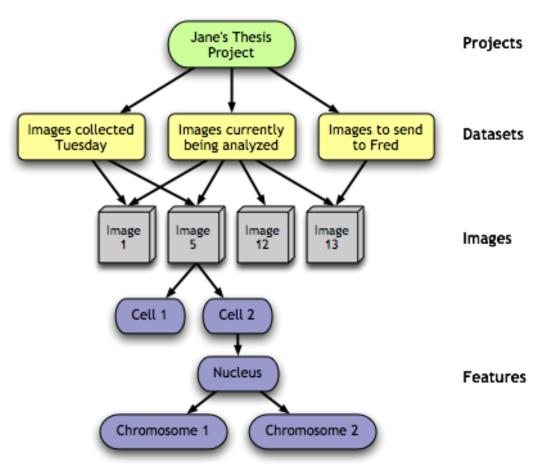
OMEGA: phase II will be completed by August 2011...



OMEGA – openBIS integration: data tree for particle tracking

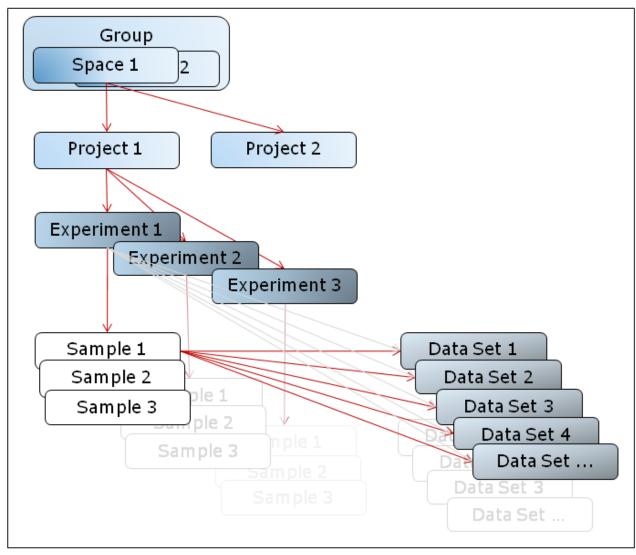


OME data model



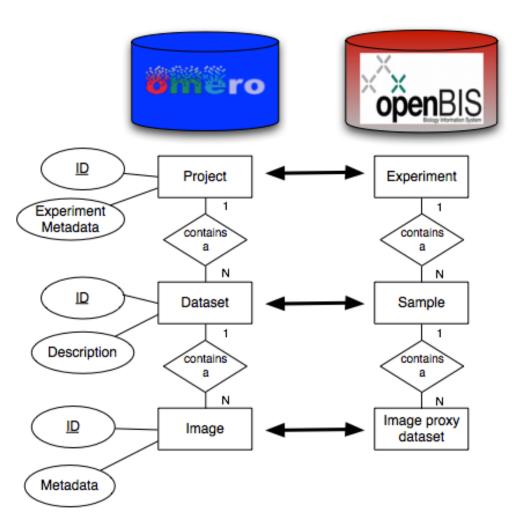
http://www.openmicroscopy.org.uk/site/support/legacy/ome-server/conceptual-framework/object-hierarchy

openBIS data model

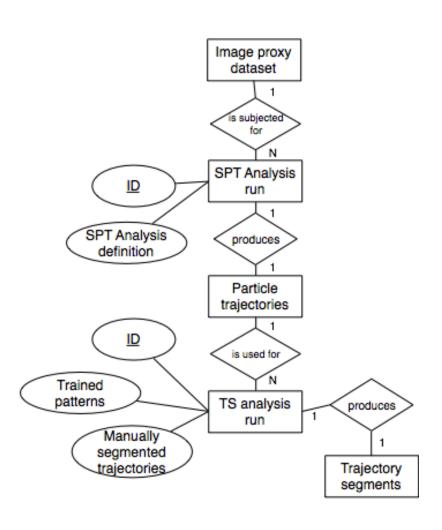


http://www.cisd.ethz.ch/software/openBIS

OMEGA – openBIS integration: mapping between the two models



OMEGA – openBIS integration: particle tracking data model



Where do we go from here? Complexity hinders data integration

- The fundamental complexity of biological systems limits scale and productivity
- •Results are difficult to compare across experimental models, set-ups, data types, time and space
- Each question requires new output data structures
- Data structures are difficult to standardize

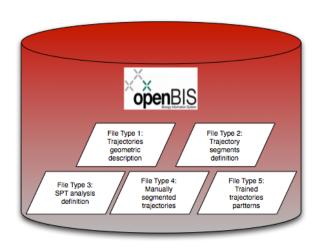
The current "solution" is twofold:

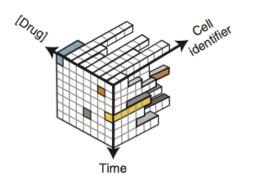
- data bases are used as "static" data repositories, which allow open access to current scientific knowledge and long term archiving of such knowledge
- individual data acquisition centers (i.e. large labs, bioimaging centers, high-through put screening facilities) develop ad hoc solutions for analysis with little cross-integration of results

The size of the problem is staggering...

- 1. Usable systems capable of scaling for increasing amount of data
- 2. Enterprise level systems available in a distributed manner at individual laboratories and facilities
- 3. Systems capable of summarizing results while at the same time revealing the full detail of individual samples
- Compare results between samples, within groups and across multiple levels of complexity
- 5. Integrate several multi-dimensional data types
- 6. Data must be structured, indexed, and annotated

We need to start working on a solution...





Millard et al. Nature Methods 2011

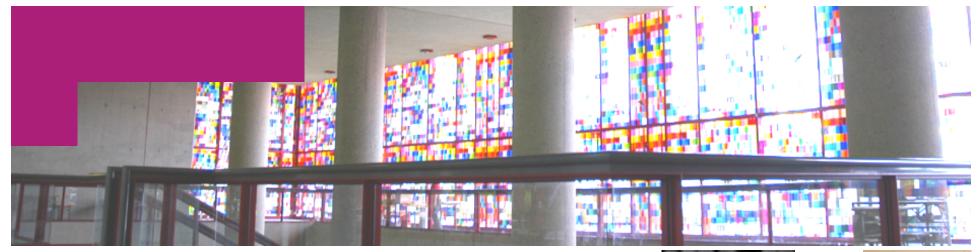
Can DBs be used as "dynamic" data exploration engines capable of integrating results from different sources, different experimental set ups, different experimental models?

Can the data models be abstracted enough without loosing meaning?

Are OLAP cubes a solution?

Would SDCubes that combine HDF + Xml files allow the level of cross-correlation required for capturing biological complexity?

We open to give a meaningful contribution to some of these questions...



GRAZIE!

Caterina Strambio De Castillia



Jasmine Clark

Lello Giulietti

Vanni Galli



Mario Valle

Nicholas Vecchietti



Tiziano Leidi



Federico Santoni



Jeremy Luban