Gestion du cycle de vie des images en microscopie pour la biologie:

Extraction de connaissances à partir d'une base de données images

Problématique de la gestion de données image en microscopie pour la biologie

- Définition des besoins et contexte.

- Formats et outils de gestion de données disponibles, base de données images ouvertes.

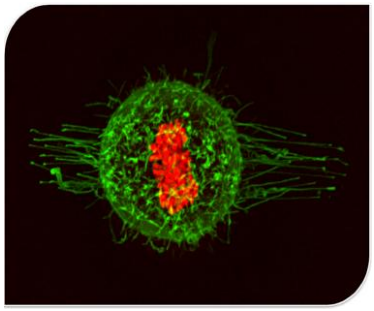
- Création du data center Curie, choix d'une collaboration avec une compagnie.

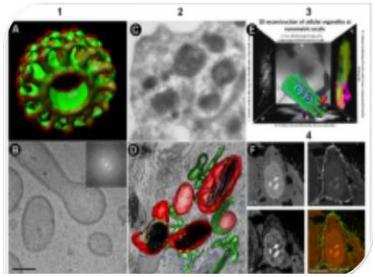
- Exemple d'utilisation et de partage dans le cadre de projet, et dans le cadre de données rendues publiques.

- Exploitation de la base de données

- Questions ouvertes

Une plateforme de données hétérogènes: 200 To par an):
Cell & Tissue Imaging Facility PICT Institut Curie





Photonic

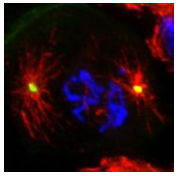
16 confocal laser scanning or spinning disk microscopes (3 equipped with femtosecond pulsed laser for multiphoton excitation, + laser photopertubation modules),
- 8 video-microscopes,
- 3 3D deconvolution microscopes,
- 2 TIRF systems (with FRAP or PALM techniques),
- 2 FLIM systems (TCSPC and phase modulation),
- 2 high-resolution structured illumination microscopes (OMX and Nikon HR-SIM),
-2 optical sectioning Zeiss Apotome systems
-- new in 2015: 2 SPIM, une salle d'imagerie intravitale petit animal (multiphoton)
- several standard widefield microscopes for routine observation or histology.

Electronic

4 electron microscopes (Cryo electron microscopy Cryotomography Cellular tomography, Transmission electron microscopy)
Ion spectroscopy

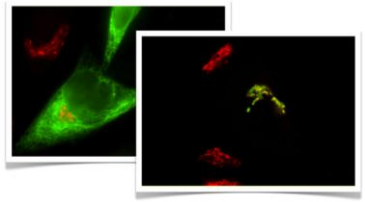
HCS: 2 InCell systems + Compound Libraries

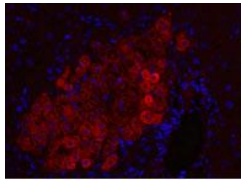
Dans un contexte de questions scitentifiques hétérogenes: ex: UMR144



Biology of centrosomes and Cilia BASTO

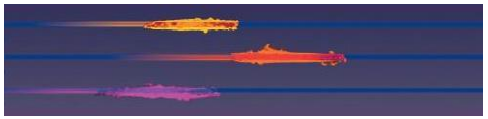
Dynamics of Intra-Cellular Organization PEREZ

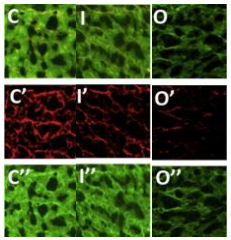




Morphogenesis and intracellular signalling LOUVARD

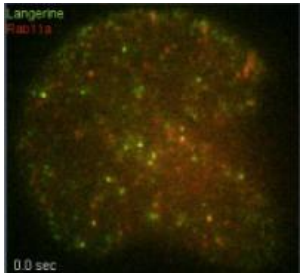
Biologie cellulaire systémique de la polarité et de la division PIEL

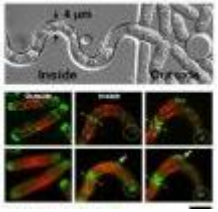




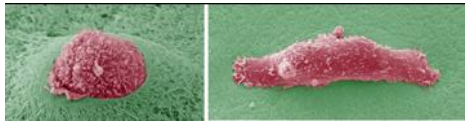
Biophysical and molecular basis of cell adhesion and migration DUFOUR

Space-time imaging of cellular dynamics of organelles and endomembranes SALAMERO

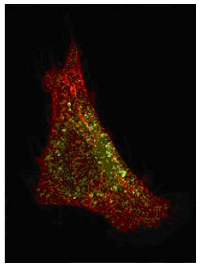




Cytoskeletal Architecture and Cellular Morphogenesis TRAN



Membrane and cytoskeleton dynamics CHAVRIER

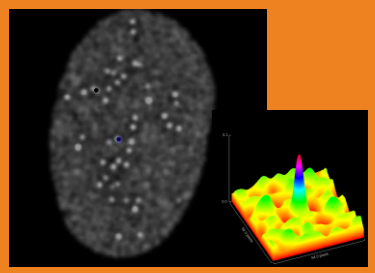


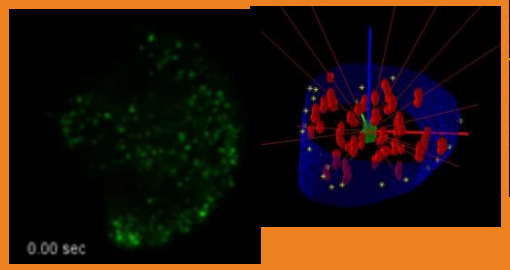
Traffic, Signaling and Delivery JOHANNES

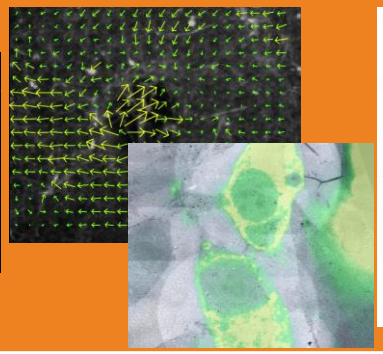
Structure and Membrane Compartments RAPOSO

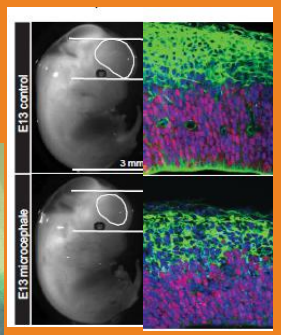
...

Example of Methods using bio image informatics for cell biology from microscopy









Molecular dynamics/ Functional Imaging

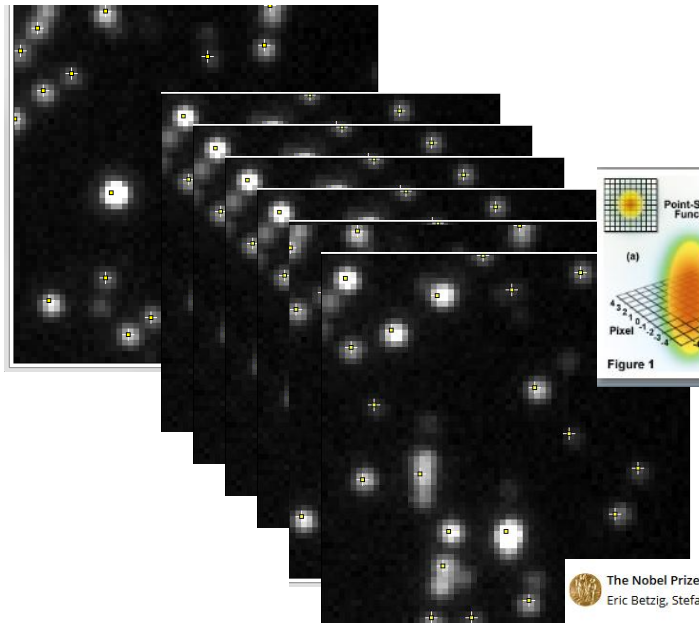
Single cells studies: never single -> cell normalisation

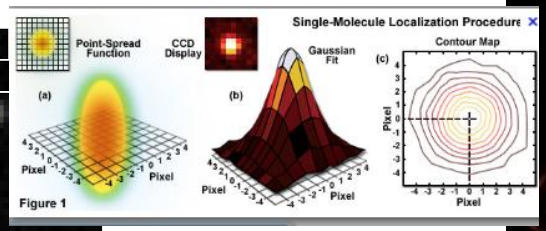
Deformation quantification and registration (Correlative Microscopy)

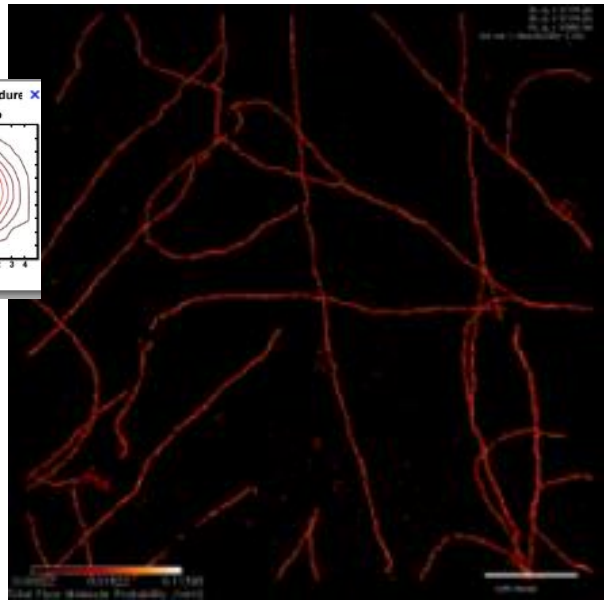
Computational model of embryonic brain development.

Exemple of imaging modalities: HR microscopy

Experimental sequence of 27 529 frames (128x128 pixels) 16 bits for 1 final image : 1 GO







The Nobel Prize in Chemistry 2014
Eric Betzig, Stefan W. Hell, William E. Moerner

Share this:

The Nobel Prize in Chemistry 2014

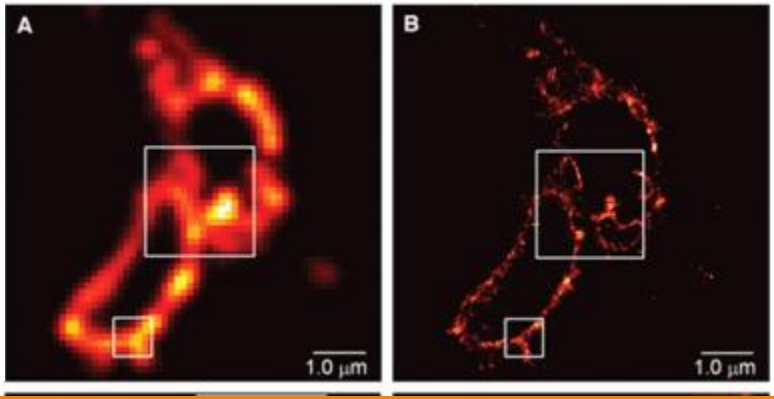
Photo: Matt Staley/HHMI
Eric Betzig
Prize share: 1/3

© Bernd Schuller, Max-Planck-Institut
Stefan W. Hell
Prize share: 1/3

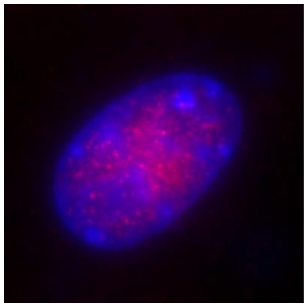
Photo: K. Lowder via Wikimedia Commons, CC-BY-SA-3.0
William E. Moerner
Prize share: 1/3

The Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell and William E. Moerner *"for the development of super-resolved fluorescence microscopy".*

PALM etc.. Based on, Blinking fluorescent proteins

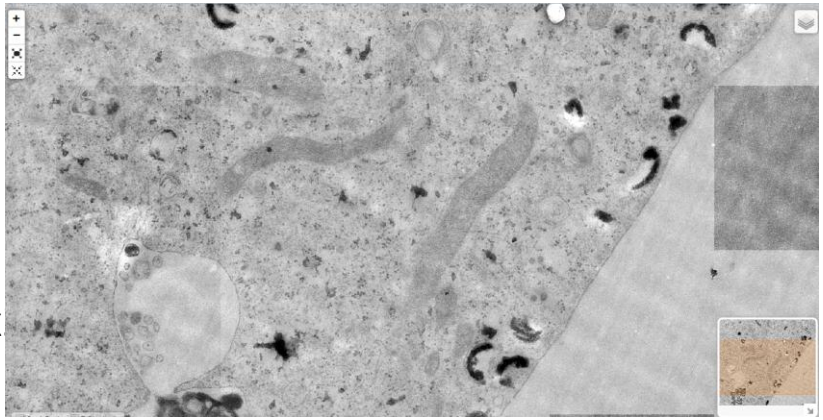


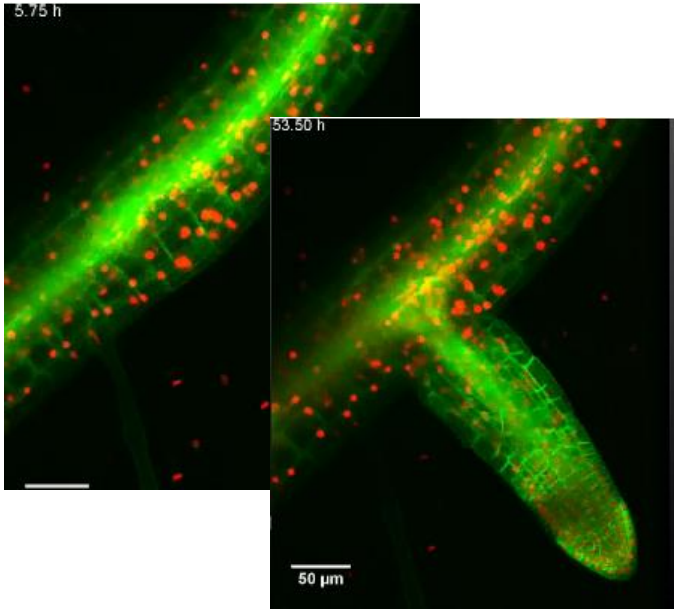
Nd-Images



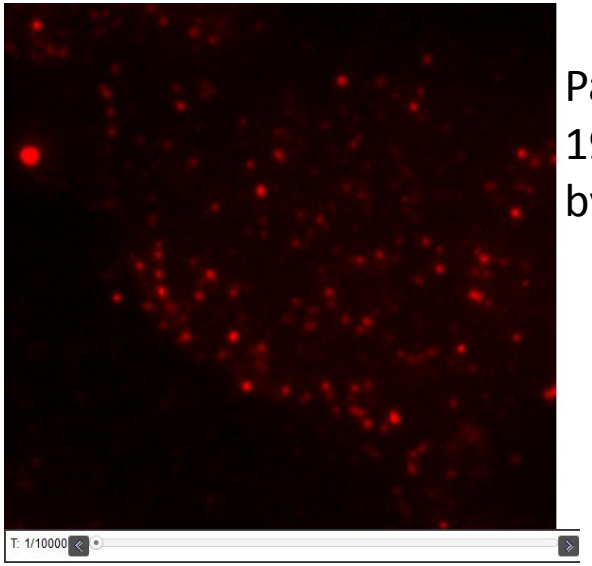
WF 384x384x50x colorsx2 bytes =29.5 MB

10000x10000x1 byte =100 MB



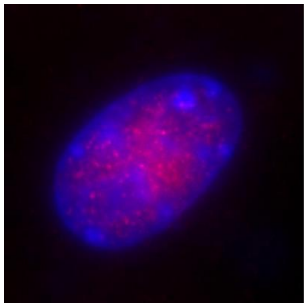


1040x1288x20x300x2 colorsx2 bytes =3.214GB CIL:40982

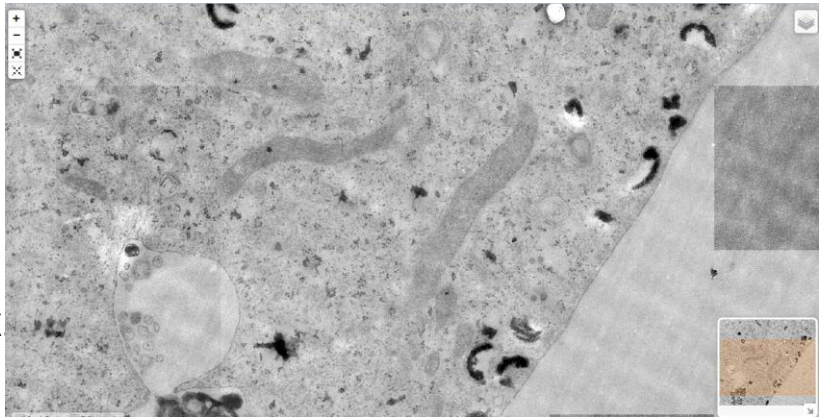


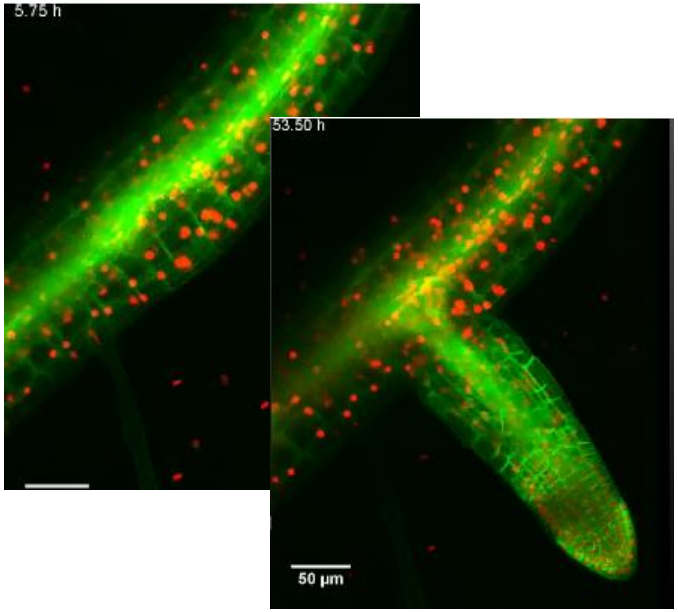
Palm Data 197*188*10000*2 bytes= 740MB

Nd-Images
And co.



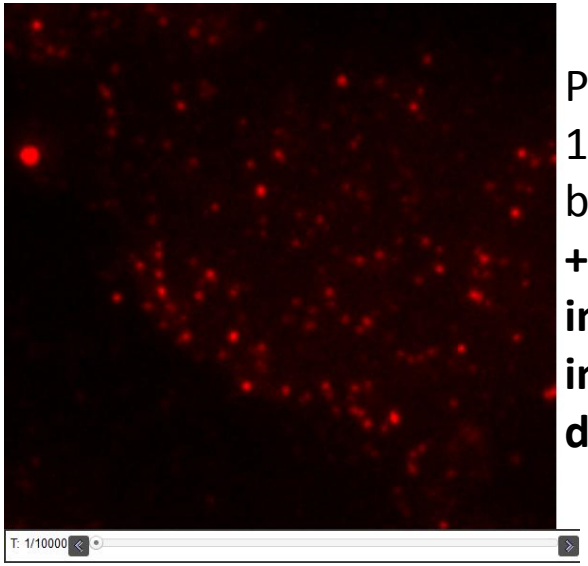
10000x10000x1 byte =100 MB **+ 3D reconstruction+ mosaick**





WF 384x384x50x colorsx2 bytes =29.5 MB **+ 3D deconvolution**

1040x1288x20x300x2 colorsx2 bytes =3.214GB
+ **Reconstructed Spim Data**



Palm Data 197*188*10000*2 bytes= 740MB **+ reconstructed image+ intermediate detection**

-> Un traitement des données indissociables des données brutes.

-> Données mieux résolues, diminution de la phototoxicité, imagerie multimodale (corrélative)

Passage à l'unité de mesure d'archivage au Peta Octet

-> Garantir l'accès aux données 5 ans après publication (responsabilité de l'utilisateur mais aussi de la plateforme?)

2 types de besoins

Donner un accès publique mais sécurisé (cf. Recommendations BioMedBridges)

Gérer la production et l'analyse de données « de travail » produites sur une plateforme

Bioformats

The Bio-Formats Library

Bio-Formats, the solution for reading proprietary microscopy image data and metadata.

Bio-Formats is a software tool for reading and writing image data using standardized, open formats. Bio-Formats is a community driven project with a standardized application interface that supports open source analysis programs like ImageJ, CellProfiler and Icy, informatics solutions like OMERO and the JCB DataViewer, and commercial programs like Matlab.

Bio-Formats is developed by the Open Microscopy Environment consortium, including development teams at LOCI at the University of Wisconsin-Madison, University of Dundee and Glencoe Software. Licensing and citing information is on the OME licensing page.

Bio-Formats currently reads and converts more than 120 file formats to the OME-TIFF data standard.

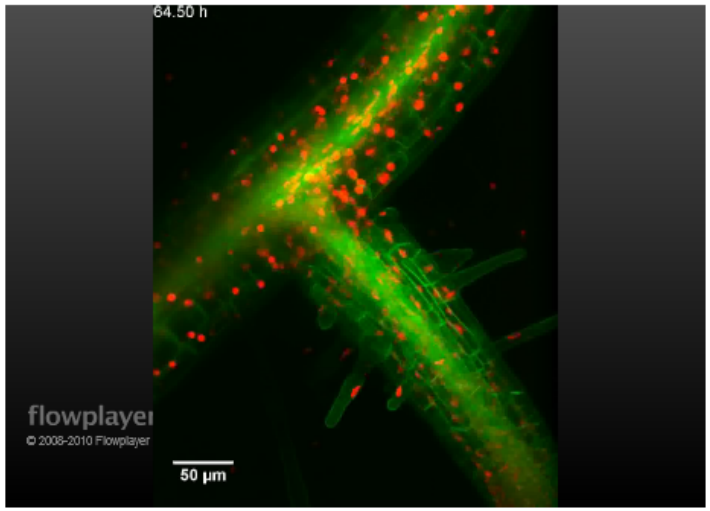
Available resources for biological images

THE CELL™ an image library

Search | Submit | Contributors | Help | Log in

Advanced

Cell Process | Cell Component | Cell Type | Organism | Pivot View



Video Data Download Options...

Licensing

Attribution Non-Commercial; No Derivatives: This image is licensed under a Creative Commons Attribution, Non-Commercial, No Derivatives License. View License Deed | View Legal Code

Tweet 3 Share

Comments:

Add to Photobox

CIL:40982*

Description

A video of Arabidopsis thaliana showing a lateral root growing out of the primary root. Arabidopsis thaliana is widely used as a model organism in plant biology. This movie is created from image stacks recorded every 15 minutes over a period of 75 hours using digital scanned light-sheet microscopy. Second Prize, 2011 Olympus BioScapes Digital Imaging Competition®.

Technical Details

BioScapes INTERNATIONAL DIGITAL IMAGING COMPETITION

Biological Sources

NCBI Organism Classification
- *Arabidopsis thaliana* (thale cress)

Cellular Component
- plant-type cell wall
- male germ cell nucleus

Biological Context

Biological Process
- lateral root formation
- root development

Attribution

Imaging

Image Type
- recorded image

Imaging Mode
- scanned sheet microscopy
- fluorescence microscopy

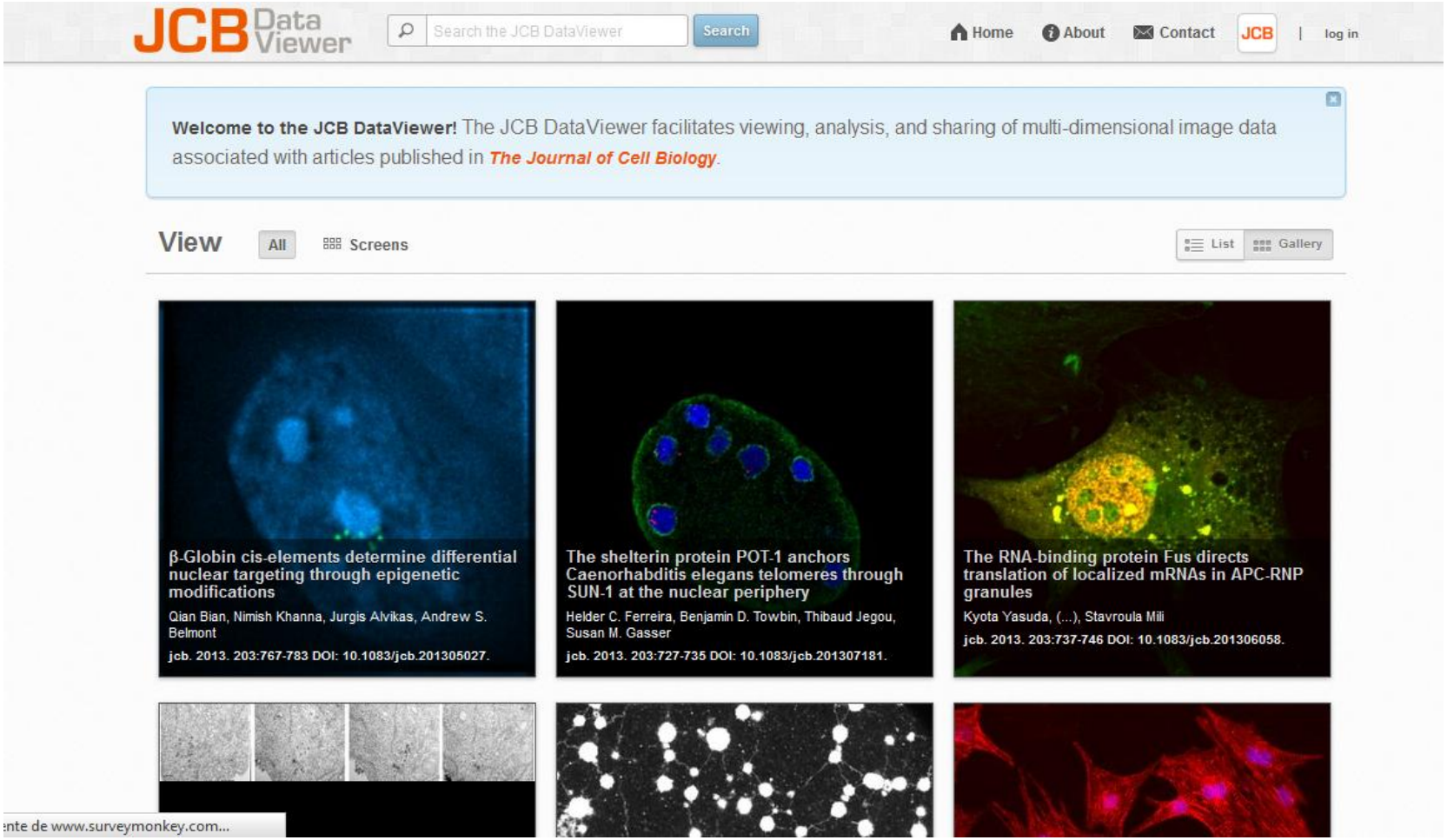
Parameters Imaged
- fluorescence emission

Source of Contrast
- distribution of a specific protein

Processing History
- z-stack reconstruction

Sample Preparation

Available resources for biological images



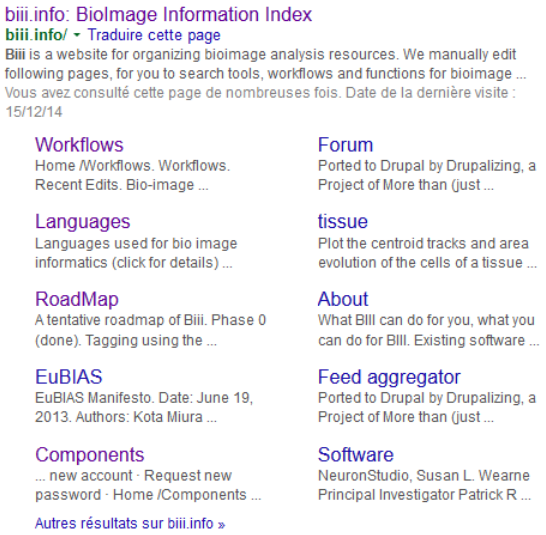
Why do computer scientist needs acess to sample annotated images

material to work/test/assess (CHALLENGES)

ACCESS to what other used (also processing)

BIII.info

+ taggathon



Création d'une base d'outils et workflow logiciels , avec des données tests associées (vers du benchmarking?)

Available resources for biological images

Broad Bioimage Benchmark Collection

Annotated biological image sets for testing and validation

BROAD INSTITUTE

- Introduction
- Image sets
- Benchmarking
- Contribute

LEGEND: KINDS OF GROUND TRUTH
- C Counts
- F Foreground/background
- O Outlines of objects
- B Biological labels

Image sets

Identification and segmentation

Accession	Description	Mode	Fields	Ground truth
BBBC001	Human HT29 colon-cancer cells	Fluorescent	6	C
BBBC002	*Drosophila* Kc167 cells	Fluorescent	50	C
BBBC003	Mouse embryos	DIC	15	C F
BBBC004	Synthetic cells	"Fluorescent"	100	C F
BBBC005	Synthetic cells	"Fluorescent"	19,200	C F
BBBC006	Human U2OS cells (out of focus)	Fluorescent	52,224	F O
BBBC007	*Drosophila* Kc167 cells	Fluorescent	16	O
BBBC008	Human HT29 colon-cancer cells	Fluorescent	12	F
BBBC009	Human red blood cells	DIC	5	O
BBBC010	*C. elegans* live/dead assay (uneven illumination)	Bright field and fluorescent	384	F O B

Home
Information
People
Addresses
Cytometry Development Workshop
FCS API

Research
Projects
Publications
Software
Presentations
Flow Cytometry

Services
PSLID
SLIF
Waldo

Data
Download

Affiliations
Carnegie Mellon University
Lane Center for Computational Biology
Center for Bioimage Informatics
Biological Sciences Department
Biomedical Engineering

Most important	Other Datasets	Supplementary Data
2D HeLa	Raw and processed image collections	ISBI 2006 SImEC2
3D HeLa	2D CHO	ISMB 2007 Yeast Image Classification
2D 3T3 RT Set3	2D 3T3 RT Set1	Cytometry 2007 Generative Models
2D 3T3 RT Set4	2D 3T3 RT Set2	
3D 3T3	3D UCE	
	Hand-labeled image collections	
	2D 3T3 and U20S segmented nuclei	
	Other collections	
	UCSF yeast GFP images	

2D images of 3T3 and U20S cells hand-segmented to show nuclear boundaries

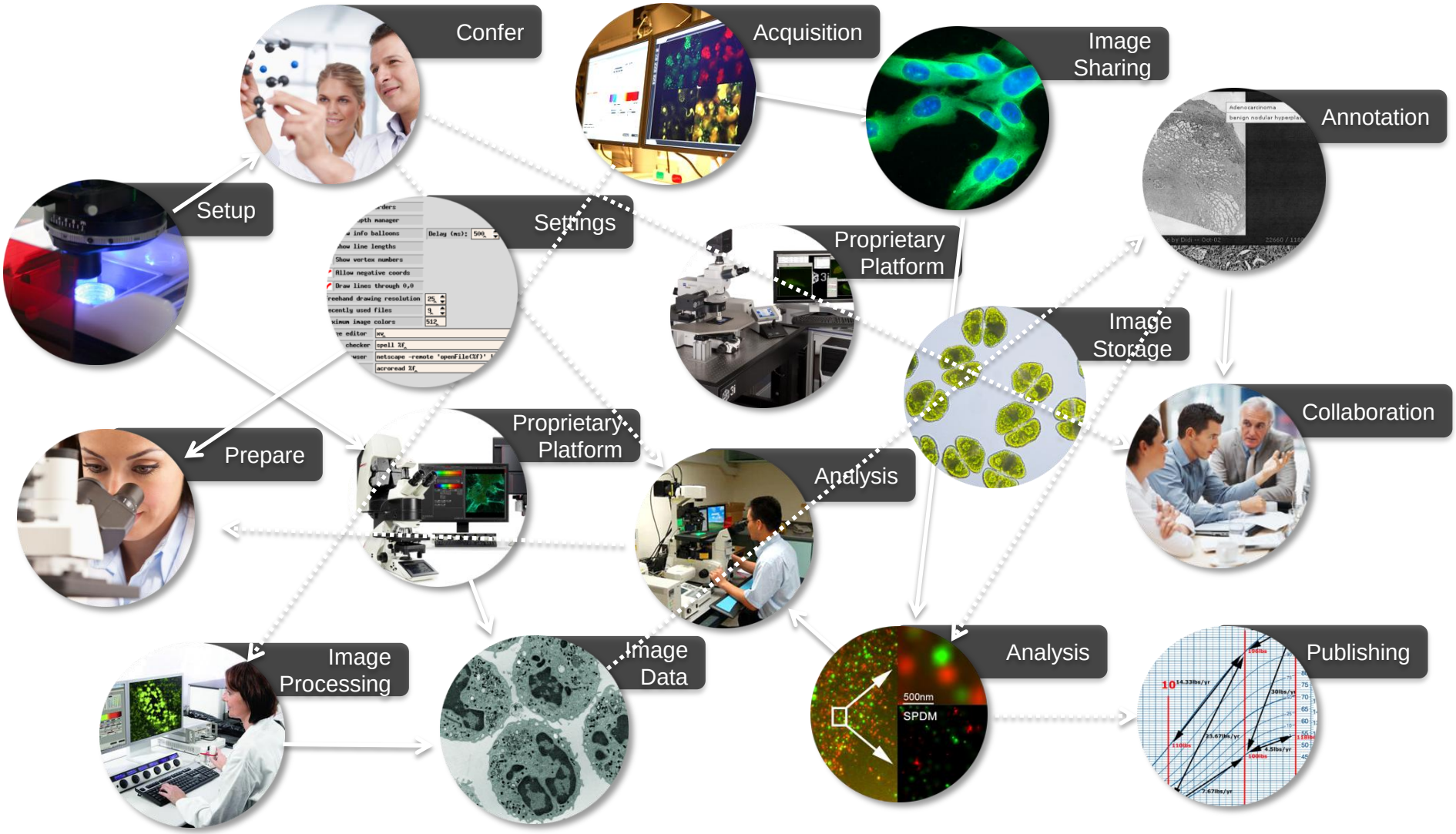
L.P. Coelho, A. Shariff and R. F. Murphy (2009). Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging (ISBI 2009), p. 518-521.

Supplementary Data for Cytometry generative models paper

T. Zhao and R.F. Murphy (2007). Automated learning of generative models for subcellular location: Building blocks for systems biology. Cytometry 71A:978-990.

Supplementary Data for ISMB/ECCB 2007 Yeast Image Classification paper

Researchers deal with huge data and complex analysis workflows



Requirements for Image data base software

FRANCE-BIOIMAGING

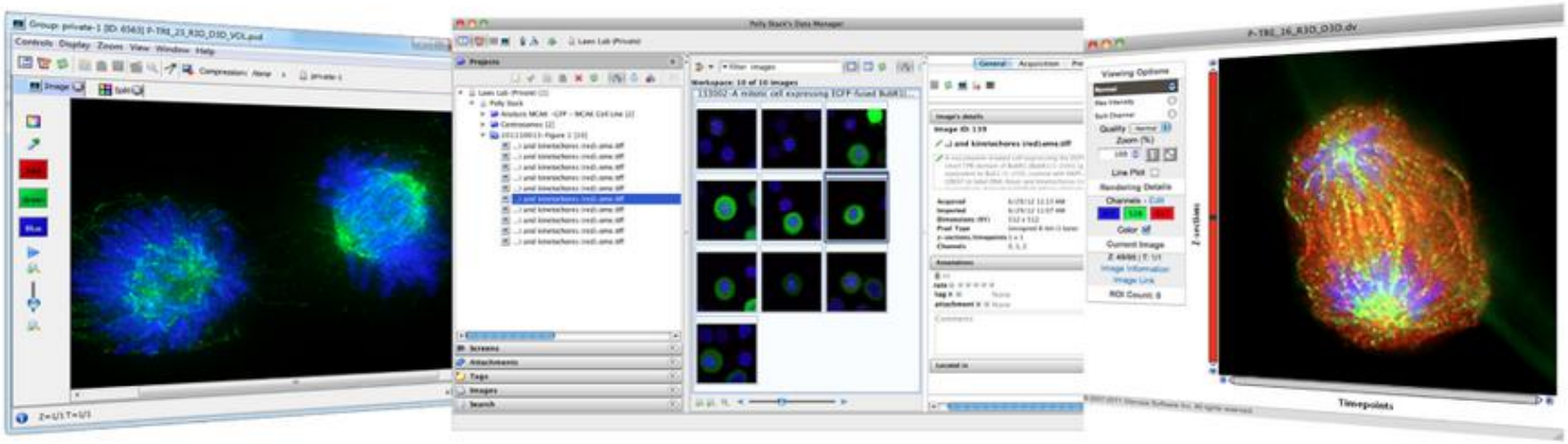
	Montpellier RIO Imaging WIDE	Institut Curie	Institut Interdisciplinaire des Neurosciences, Bordeaux Imaging Center, Bordeaux	IJM	Plateforme Bioemergences	synthèse minimale
Structure Plateforme						
Nombre d'utilisateurs (2011)	environ 600 actifs distribués sur différents sites (12 sites) 10 % venant du secteur privé	400 dont environ 10% externes	environ 100, dont un 20aine en SPT (Gros utilisateurs)	250	20	entre 100 et 1000 dont utilisateurs temporaires
Production annuelle (2011)	> 10 To (devrait augmenter)	>100 To (y compris données traitées)	>50 To (beaucoup de SPT, grosse bande passante : 20Go/h/poste)	>15 To/an pour 10 microscopes (devrais augmenter)	5 To	au minimum 100 to
Production à archiver archivage=archivage à long terme		Seulement un % de la production annuelle: données publiées et Gold standards. Durée de conservation: 4 ans après publication pour données publiées, sûrement plus pour Gold Standard/			100%	Cycle de vie des données: actives, archivées, supprimées.
remarques particulières	Plusieurs sites sur Montpellier	Multi site (Orsay et Paris)			Multi site	Fluidité entre les différents sites. Transparence des sites / projets. Une seule interface pour tous les projets quelque soit le site d'acquisition d'origine. Les données ne sont pas dupliquées et sont au plus près de l'utilisateur (projet) ou des traitements.

OMERO/ OMERO via Glencoe

About OMERO

What is OMERO?

From the microscope to publication, OMERO handles all your images in a secure central repository. You can view, organize, analyze and share your data from anywhere you have internet access. Work with your images from a desktop app (Windows, Mac or Linux), from the web or from 3rd party software.



Import

Over 100 file formats supported, including all major microscope formats.

Organize

Organize by 'Project', 'Tags' or acquisition date. Annotate with attachments, comments, ratings.

View

Move through multiple dimensions, copy and apply rendering settings, zoom and pan 'Big' images.

Analyze

Draw and measure regions of interest, write Python scripts in OMERO or connect to OMERO from your favorite analysis software.

Export

Export your images for analysis or publication. Save images as 'figures' ready for presentations.

BISQUE

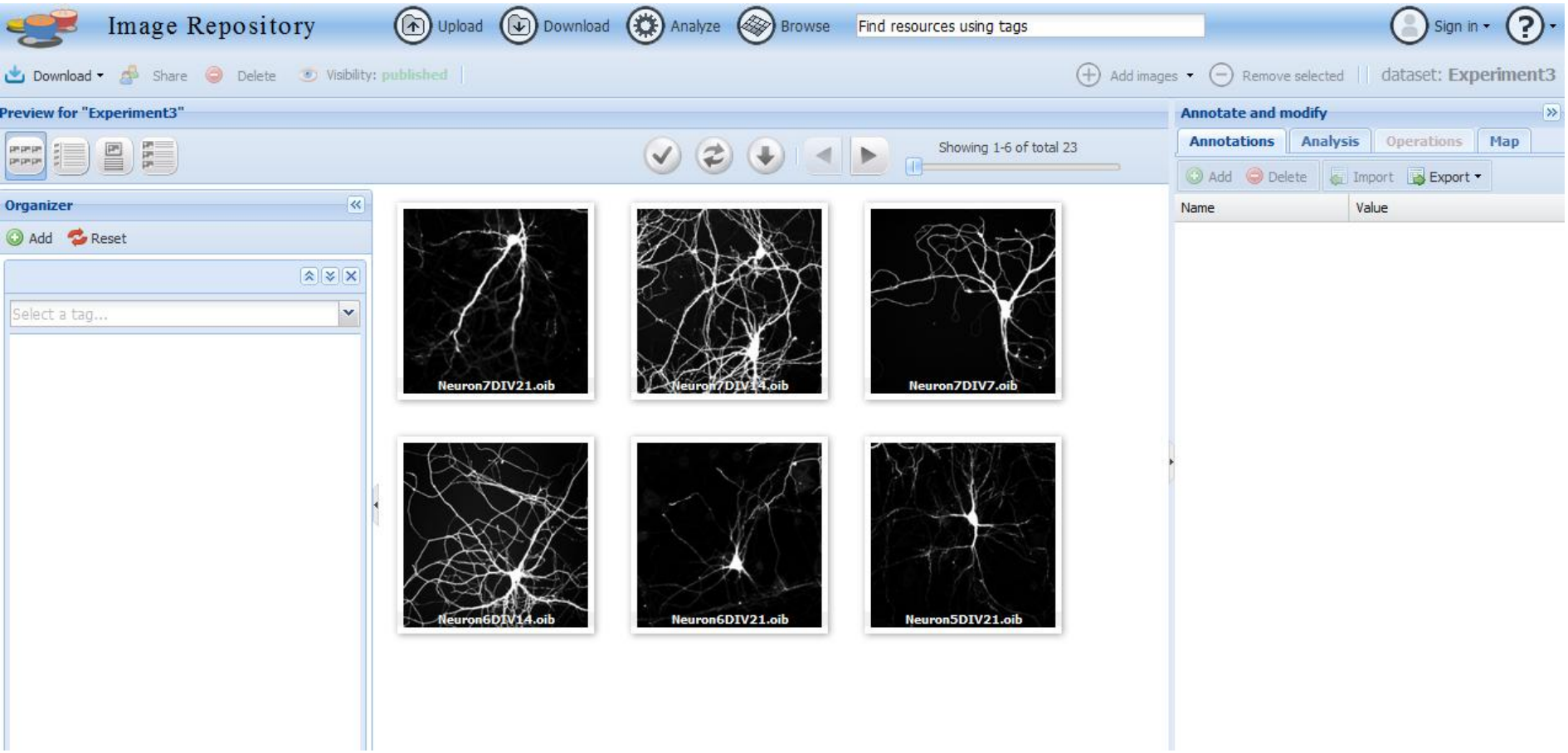
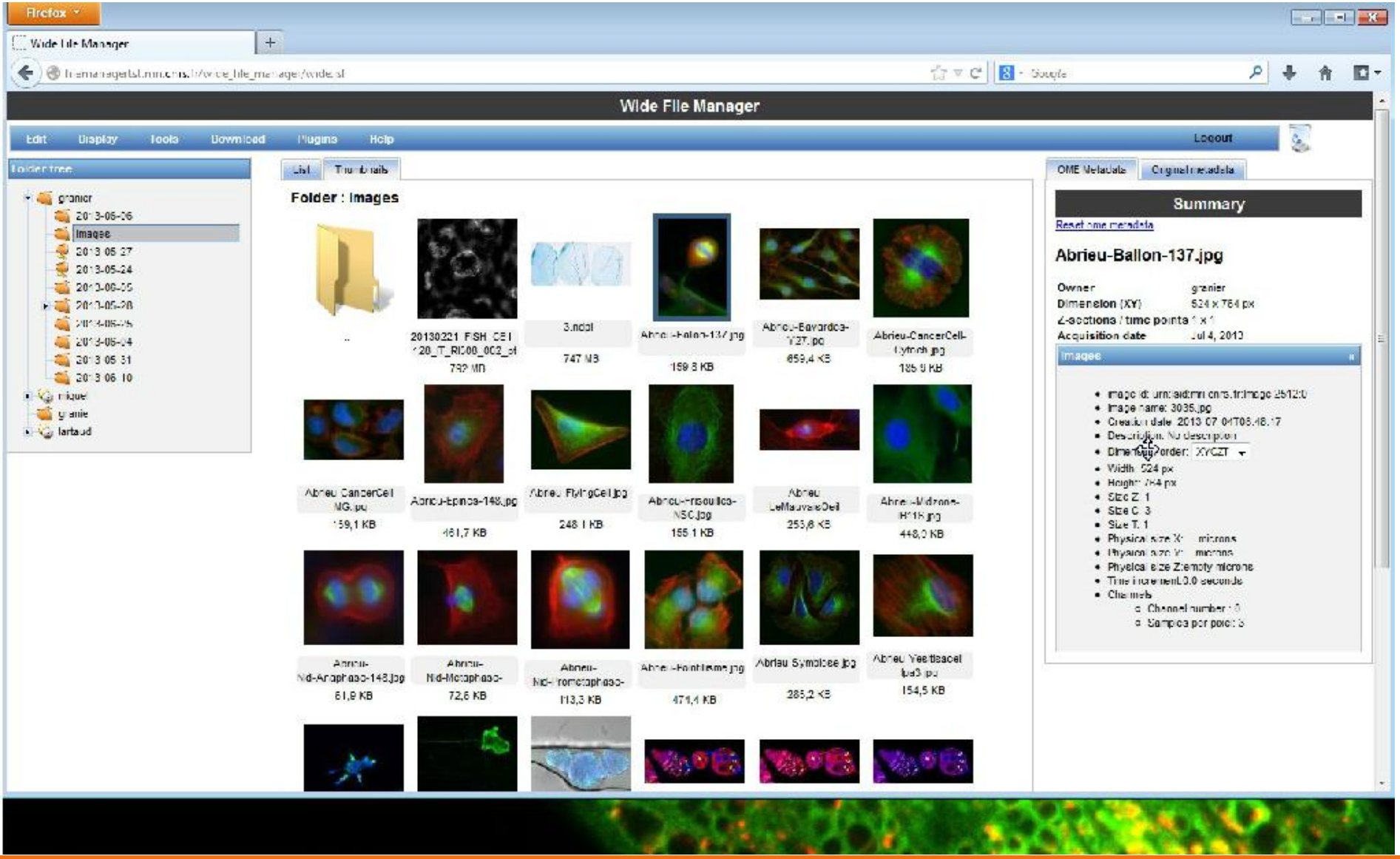
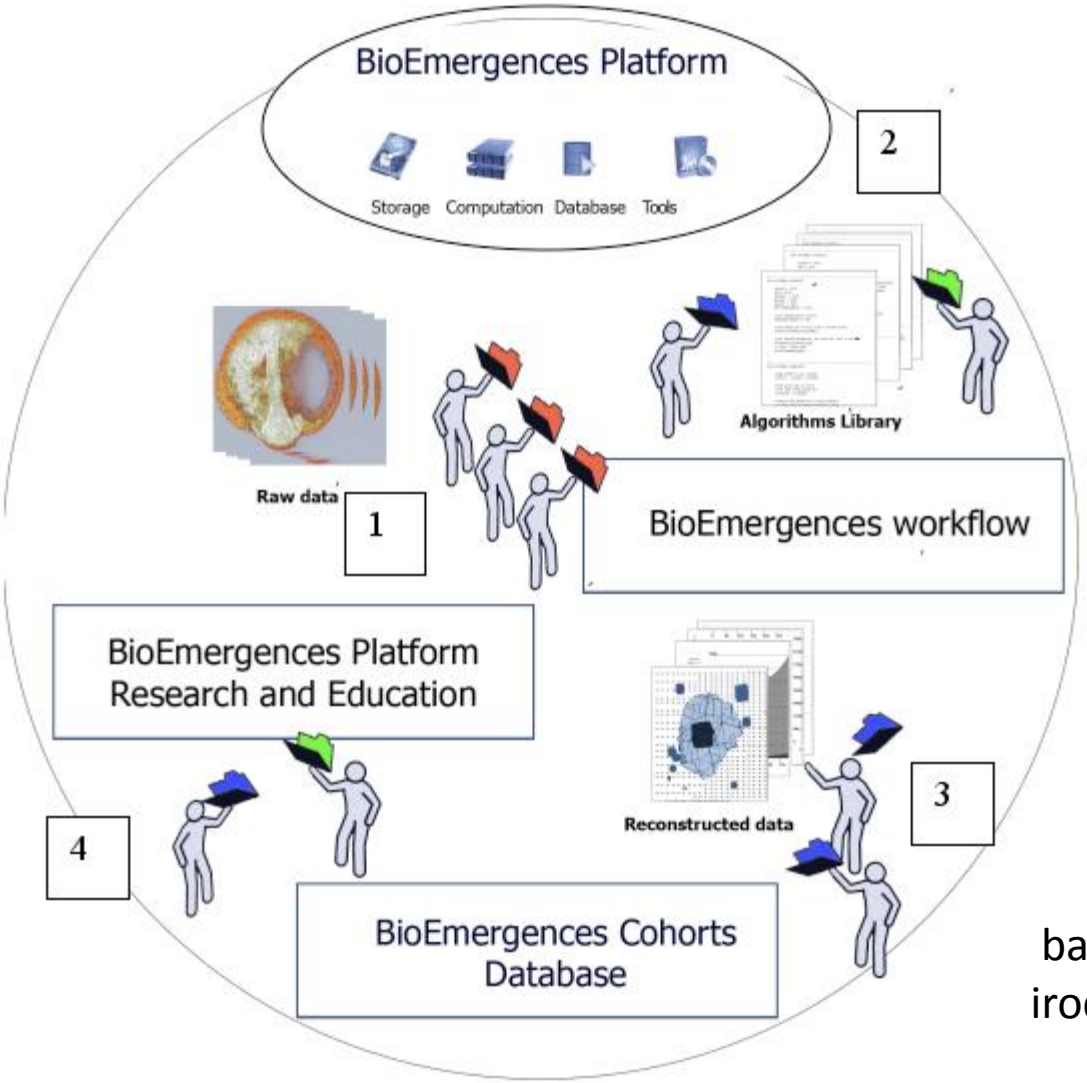


Image Repository
Upload
Download
Analyze
Browse
Find resources using tags
Sign in
Download
Share
Delete
Visibility: published
Add images
Remove selected
dataset: Experiment3
Preview for "Experiment3"
Showing 1-6 of total 23
Organizer
Add
Reset
Select a tag...
Annotate and modify
Annotations
Analysis
Operations
Map
Add
Delete
Import
Export
Name
Value
Neuron7DIV21.oib
Neuron7DIV14.oib
Neuron7DIV7.oib
Neuron6DIV14.oib
Neuron6DIV21.oib
Neuron5DIV21.oib

Wide


Firefox
Wide File Manager
Wide File Manager
Edit
Display
Tools
Download
Plugins
Help
Logout
Folder tree
granier
Image
List
Thumbnails
Folder : Images
OME Metadata
Original metadata
Summary
Abrieu-Ballon-137.jpg
Owner granier
Dimension (XY) 524 x 764 px
Acquisition date Jul 4, 2013
Images

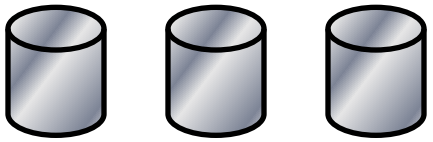
BioEmergences

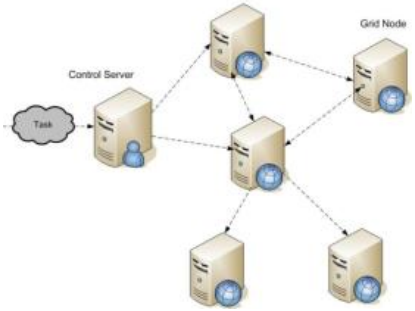


based on irods

Image data base

STORAGE





Processing batch of images on cluster (denoising, deconvolution, …)

Images Server+ Metadata +annotations (manual or analysis results)/ attachments (publications,xls file…)

Acquisition Client

Web Client

Interface

Web admin for project managing

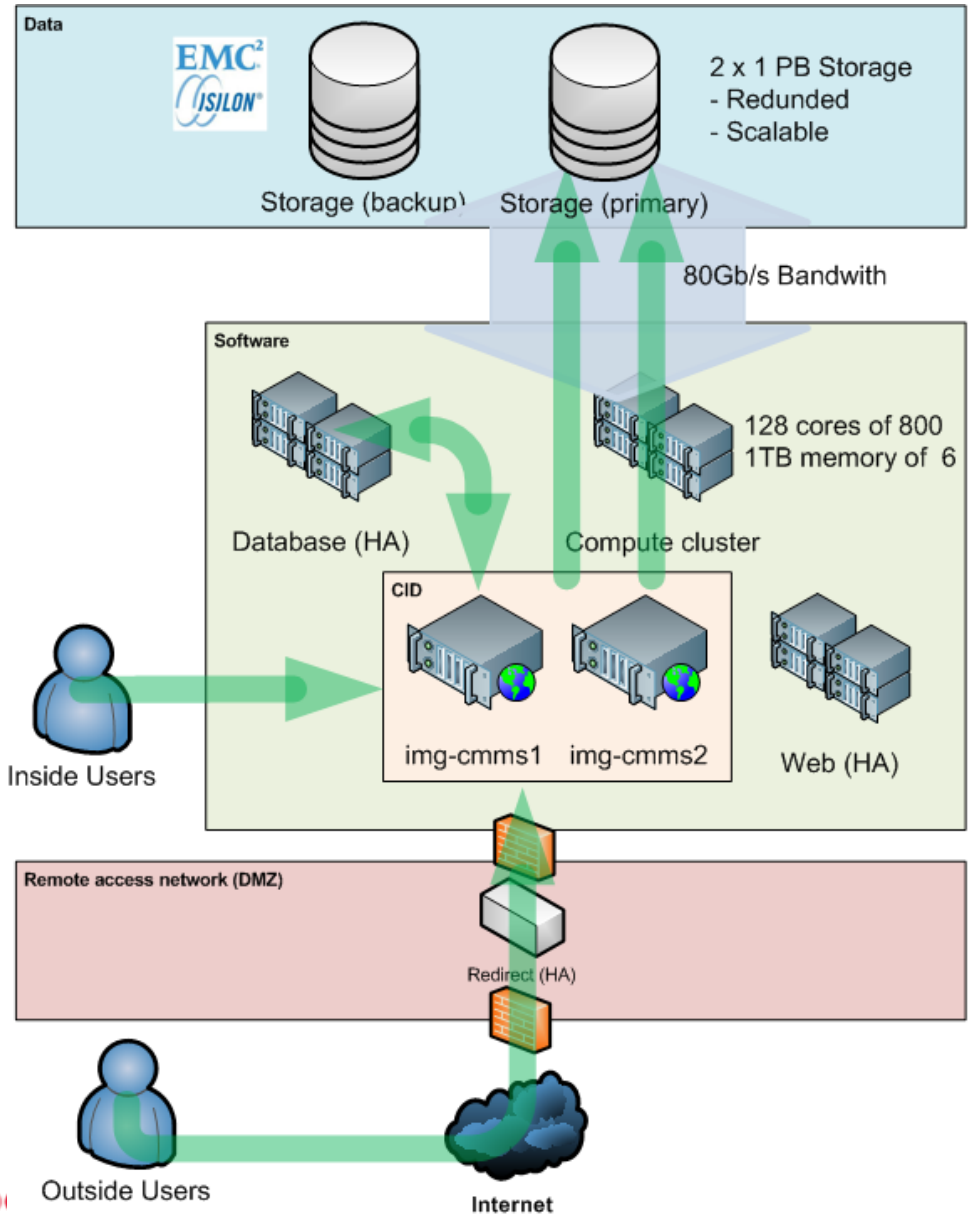
Automatic analysis without full download, Data fusion, advanced visualisation EXEMPLE

Dynamic Organisation, Visual search or advanced search functionalities

Metadata (pixel size, acquisition time,…) annotations,
Parsing nD images
Insuring reproducibility by storing all processing

Strand Life Sciences
Algorithms for Life

CID now



True Collaborative tool :

- All users are accessing the same infrastructure with **hierarchical rights**.
- All images are stored on a **central secured place,** and are **indexed**

Security & Scalability :

We have an overall **High Availability infrastruture**, with many **parts shared between platforms** (NGS/ Bioinfo/Imaging/...) :

- Network
- Storage
- Database
- Computing

What next ? :

Migrate CID on the Web shared infrastructure.

Take profit of « big data » processing.

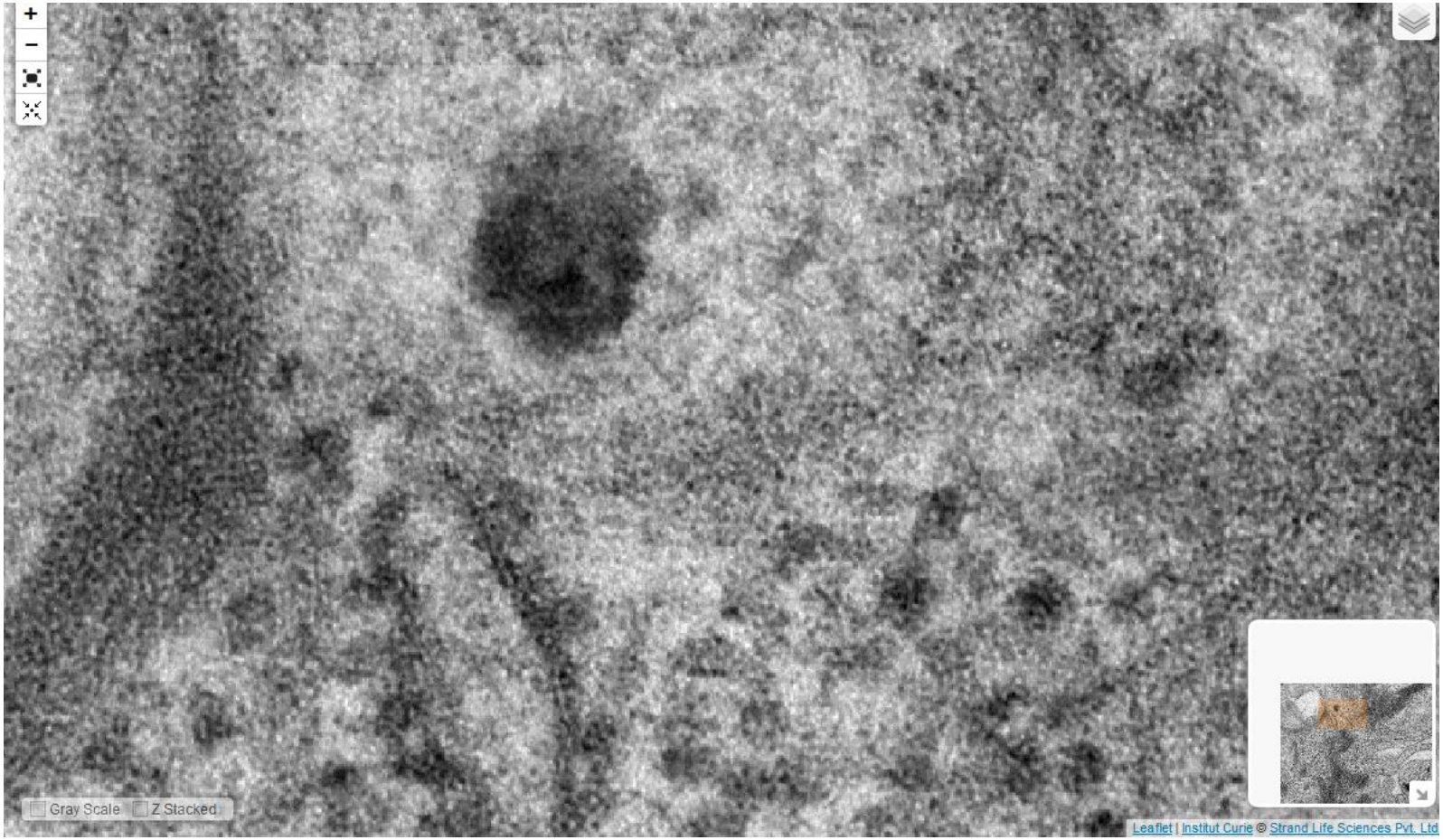
Some key numbers :

- **800 cores & 6TB compute Cluster**
- **1 PB scalable & redondant storage**
- **High througput network**

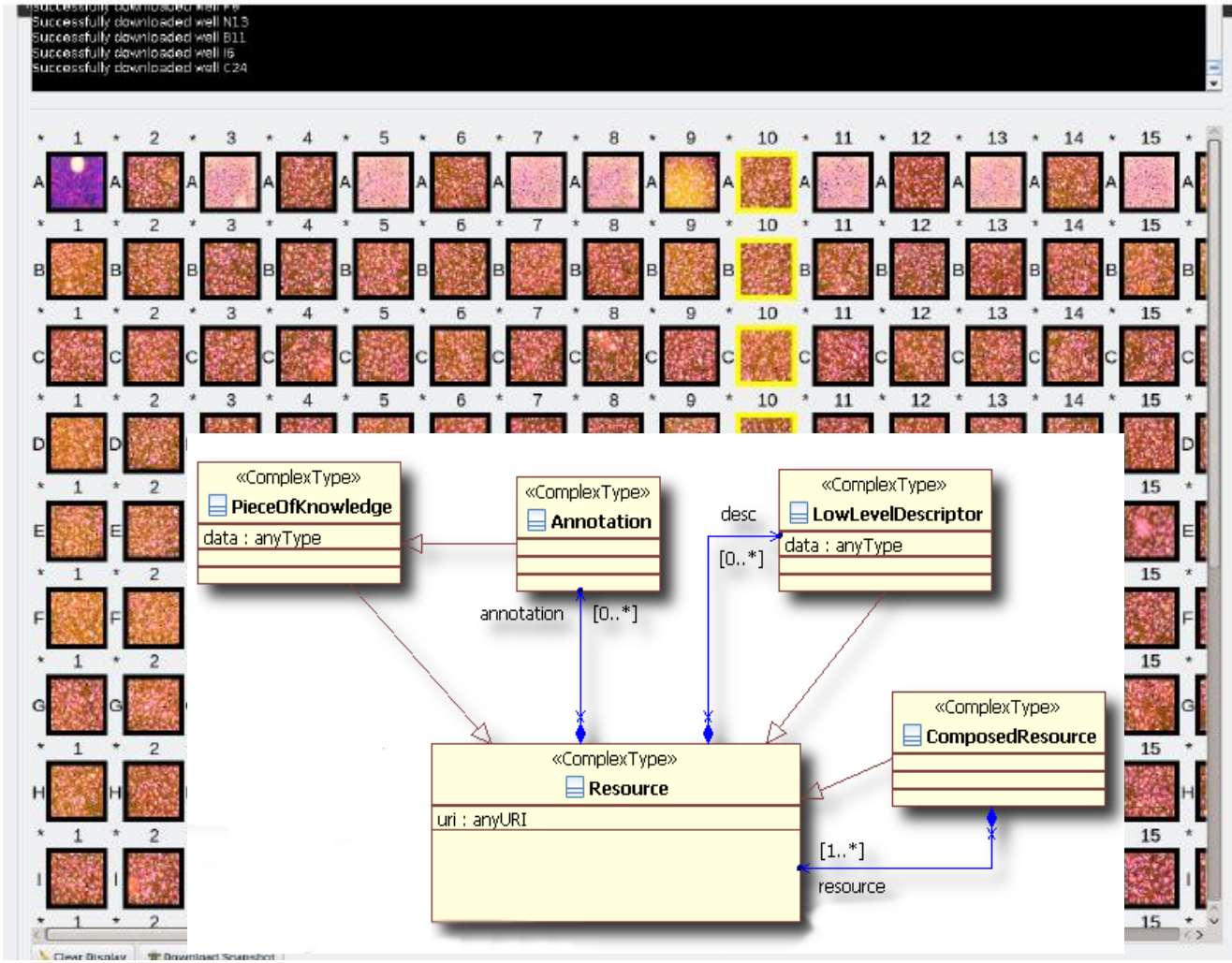
Big Data processing ready ! ☺

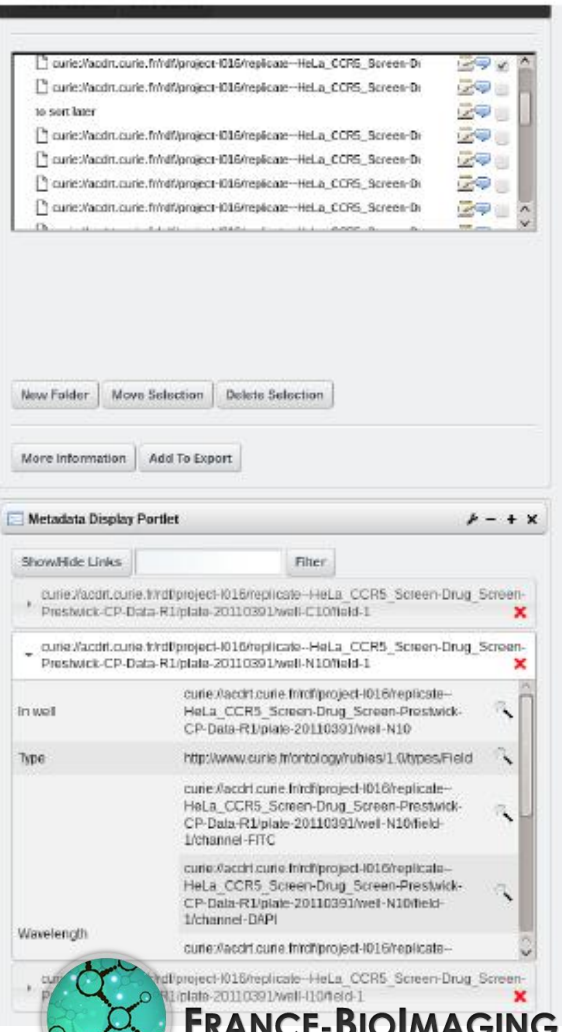
Exemple of additional clients: Tile Viewer for big data through a url invokation

Search: imageWidth:[5000 *]



Exemple of additional clients: RUBIES: specific for HCS data (compound librairies management etc., hits…) (WebLab based)





Dmitri Voitsechovitch

FRANCE-BIOIMAGING

Organisation

Have access with different levels of rights

Users (administrator/Facility Staff, Team Leader, project Manager, Team Member)
Active
Suspended
Deleted

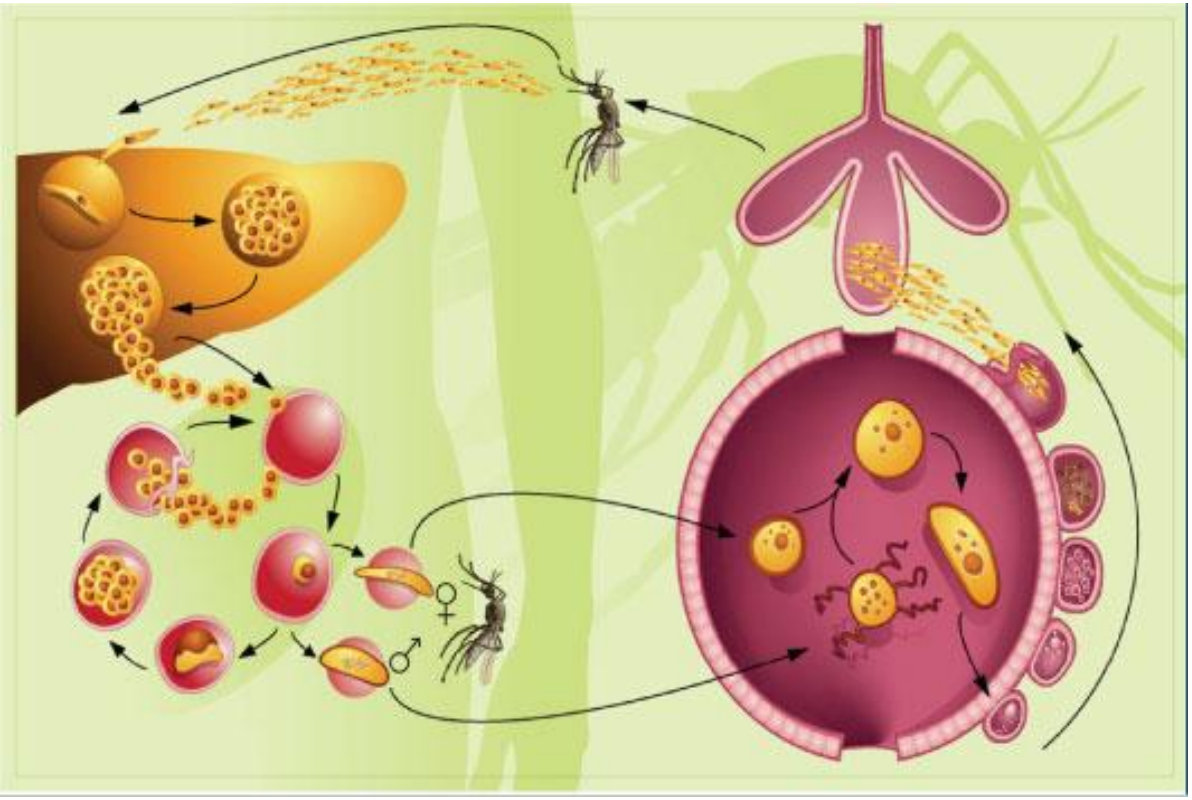
Access to records are defined by access to projects

Projects (=// rdf graphe nommé?)
Active
Archived

Are part of one or several projects (without duplication)

Records (=imagedata) +Metadata, annotations, attachments+ history and links (SQL + noSql)

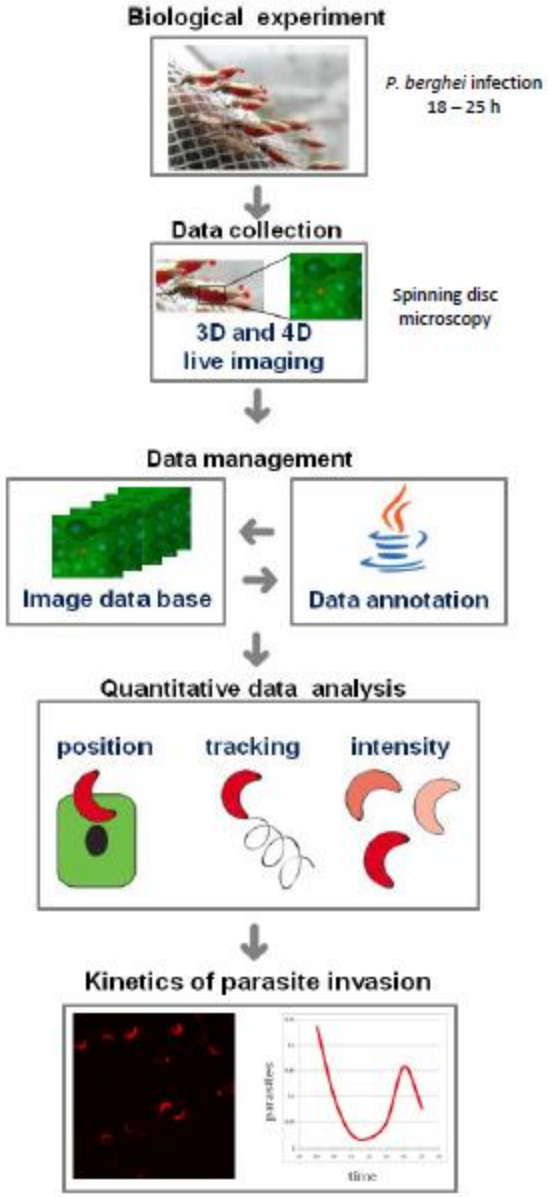
Example of on-going project (under submission).

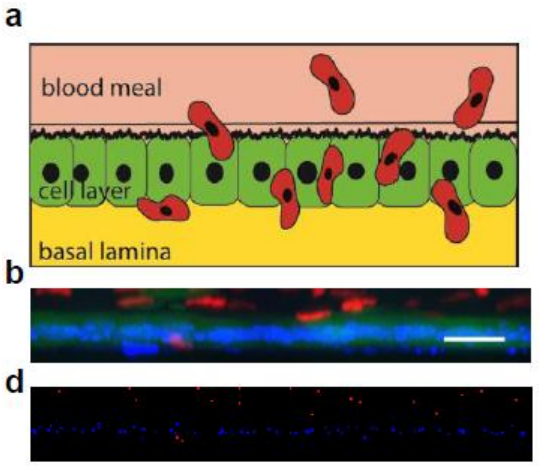


Malaria parasite invasion in the mosquito tissues

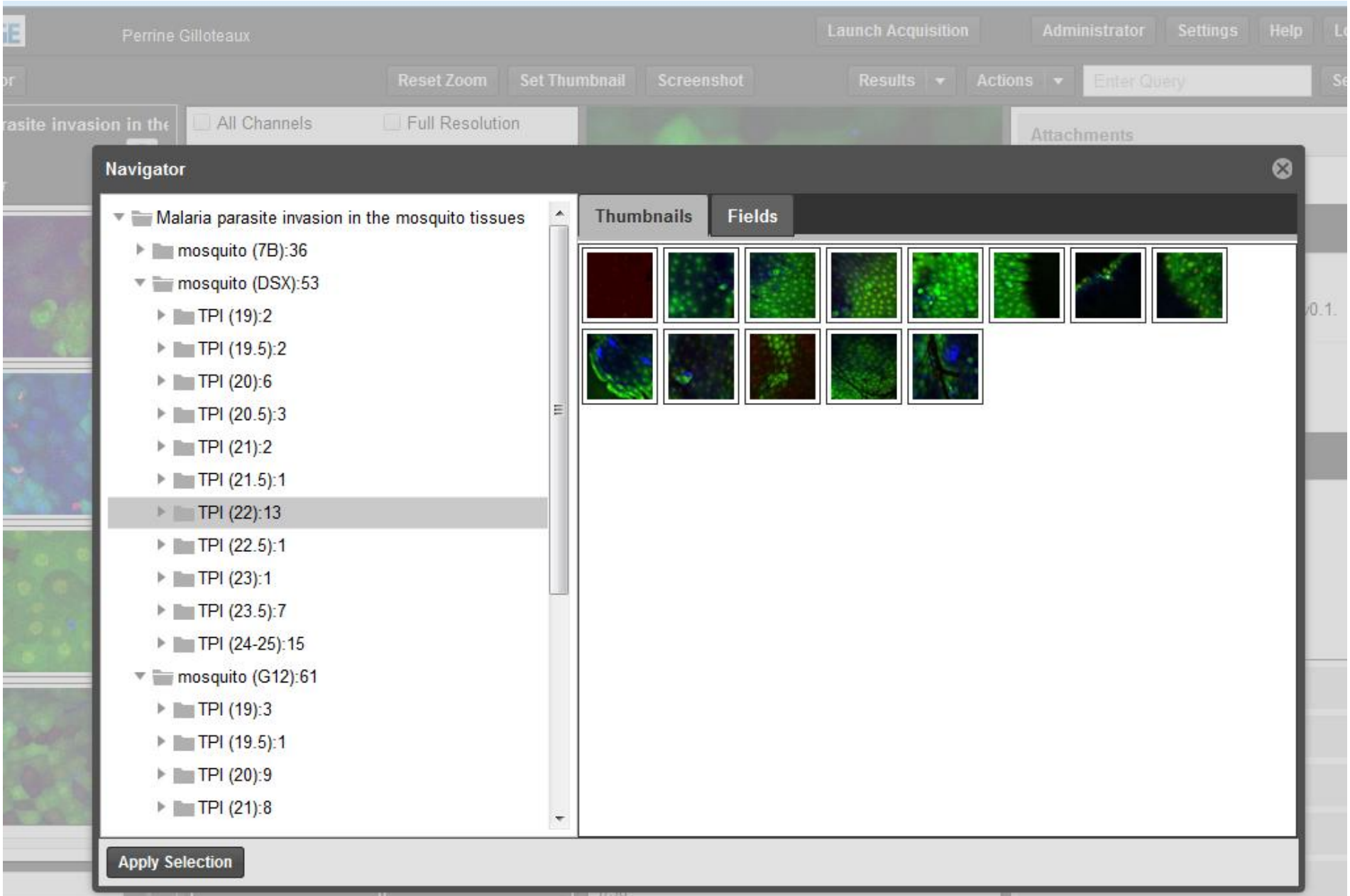
Gloria Volohonsky
IBMC Anopheles group
Unistra
Elena Levashina
Max Planck Institute for Infection Biology

malaria parasite invasion in the mosquito tissues

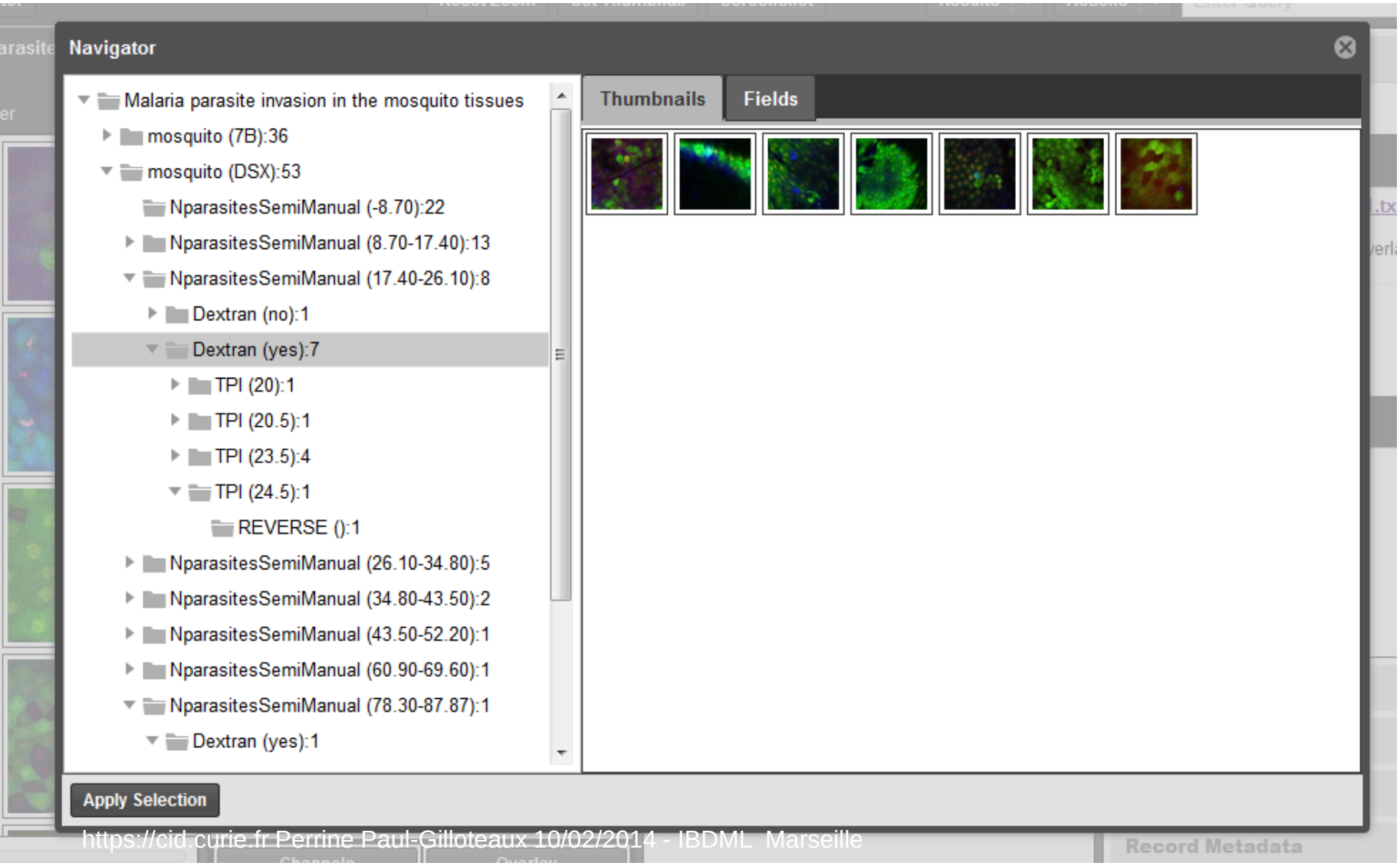




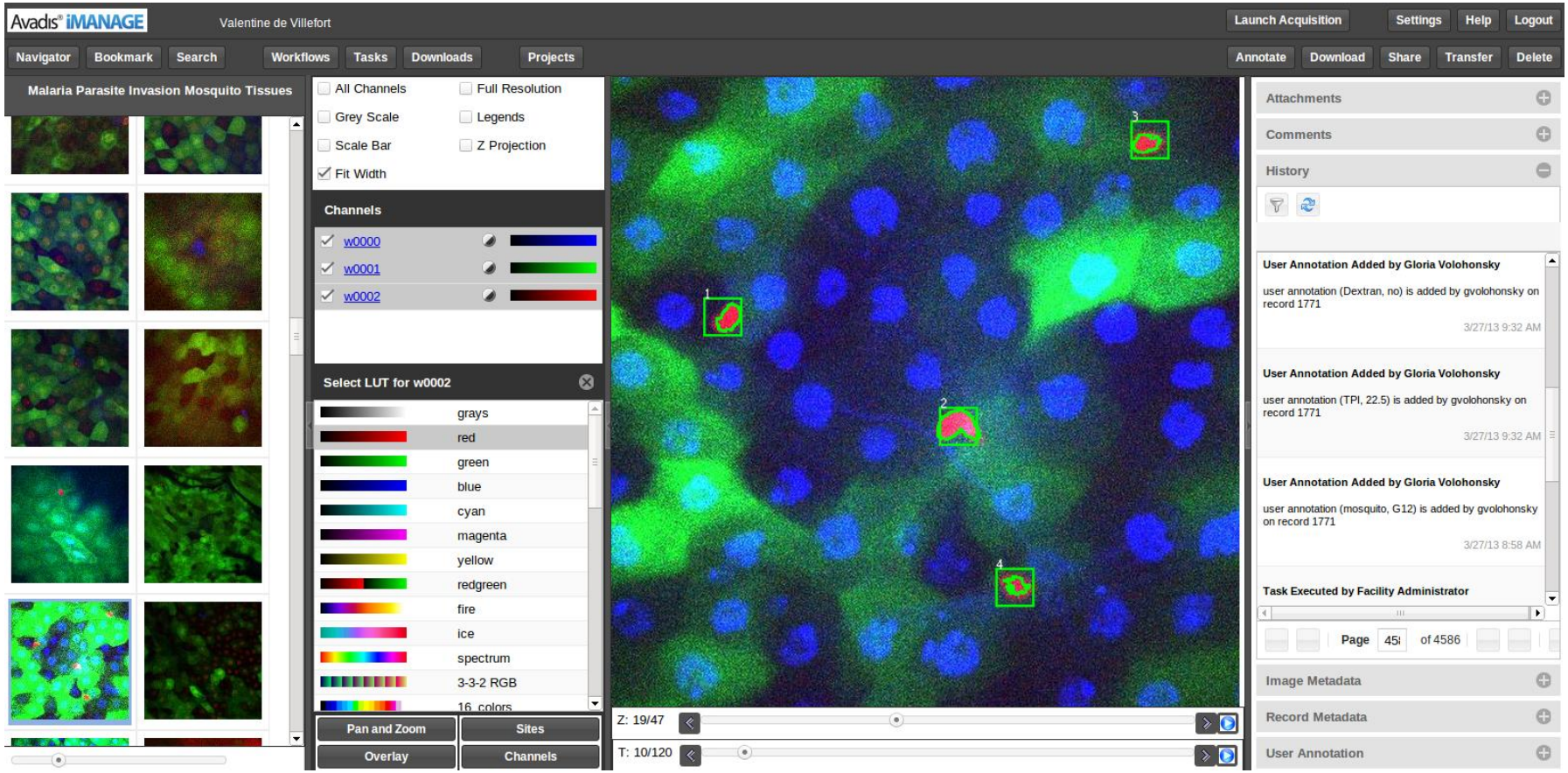
Data are annotated



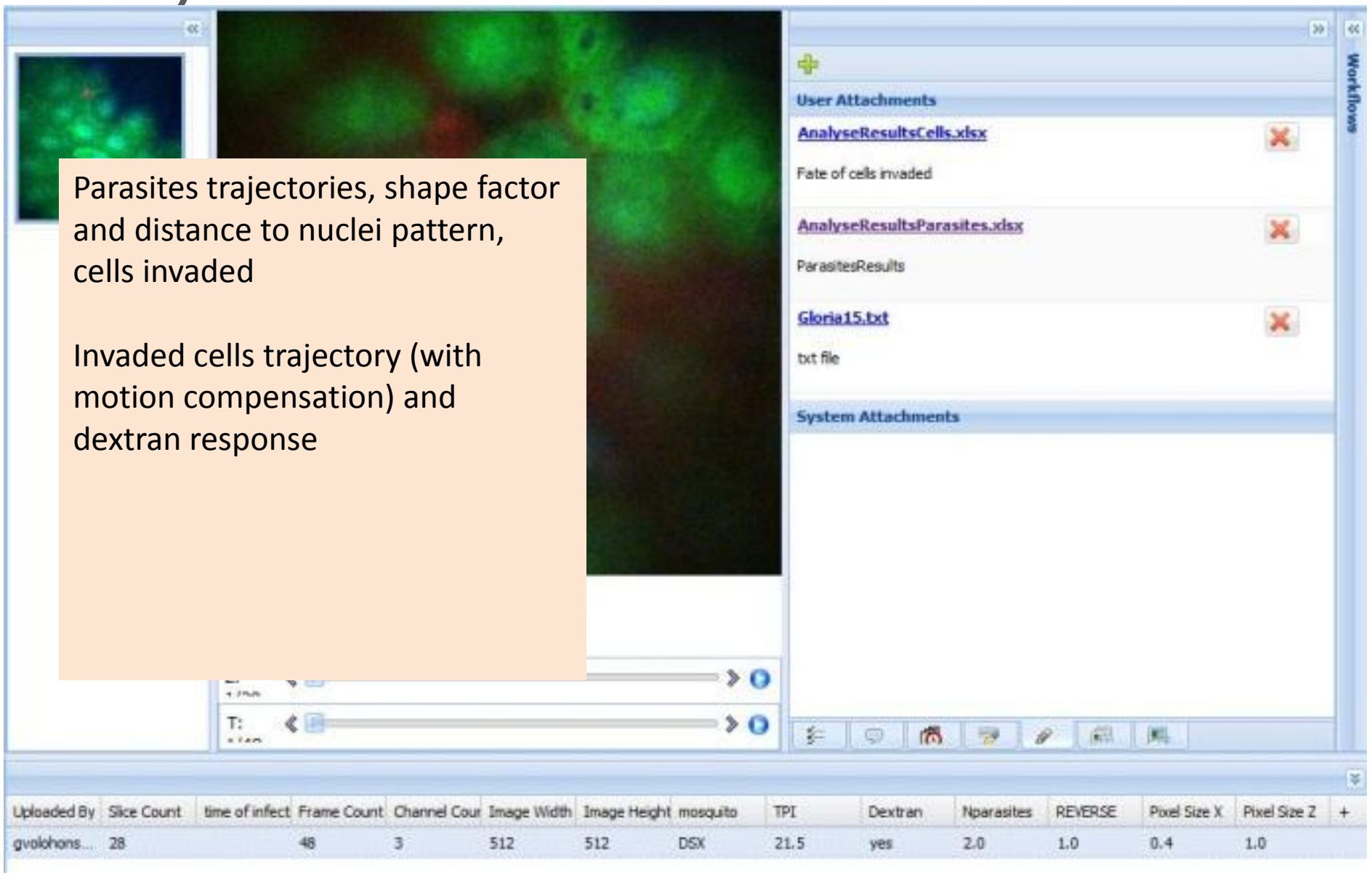
Data are annotated. Results of processing at the record level are uploaded by the processing algorithm



Annotation automatically created are also be visual.



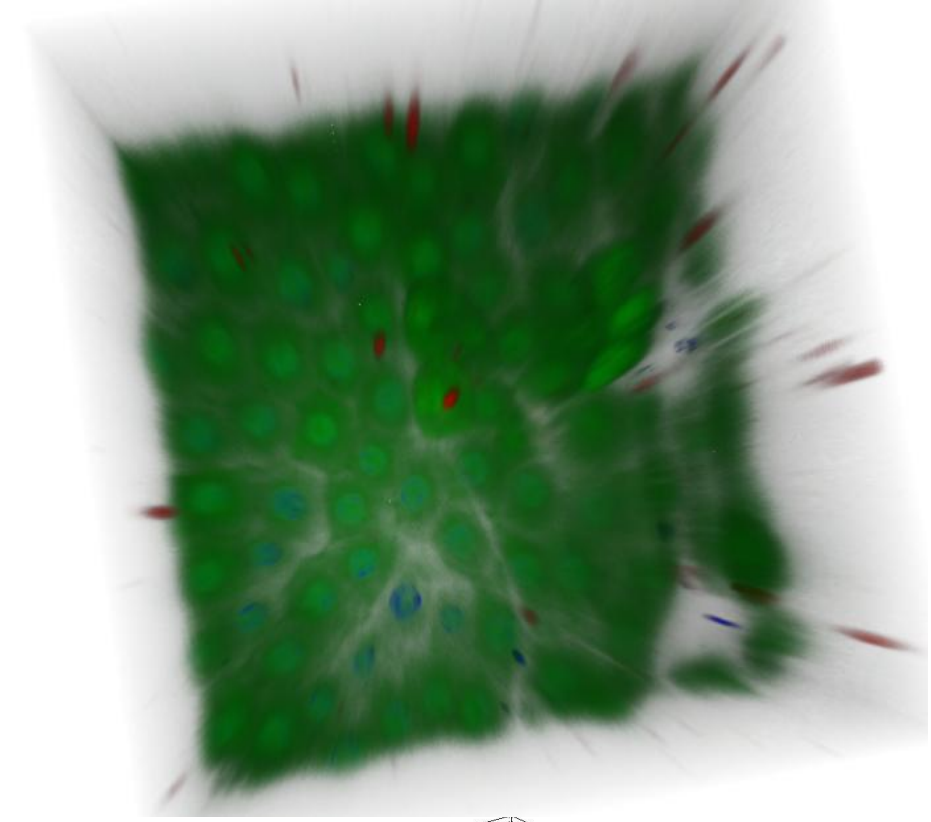
Information at the object level (here parasites and cells)

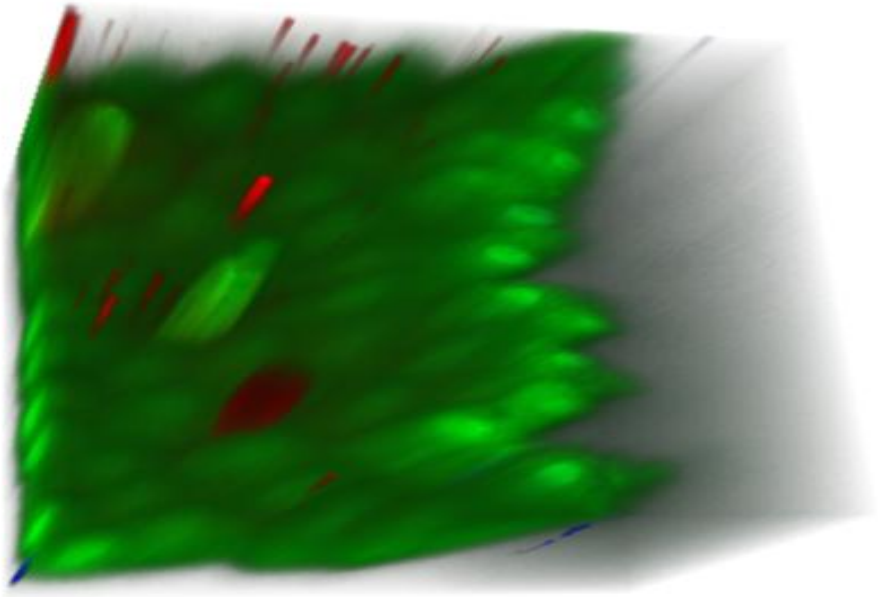


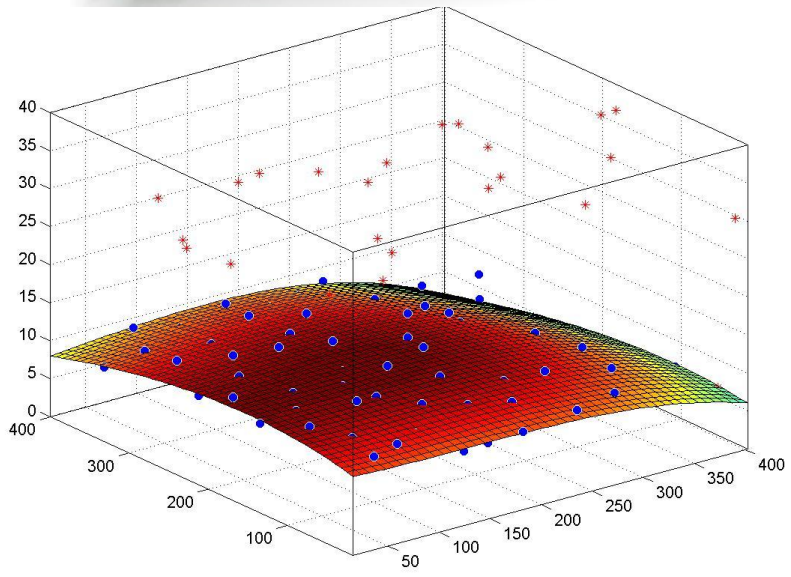
Example of analysis: proportion of parasites crossing the cell layer against time post infection

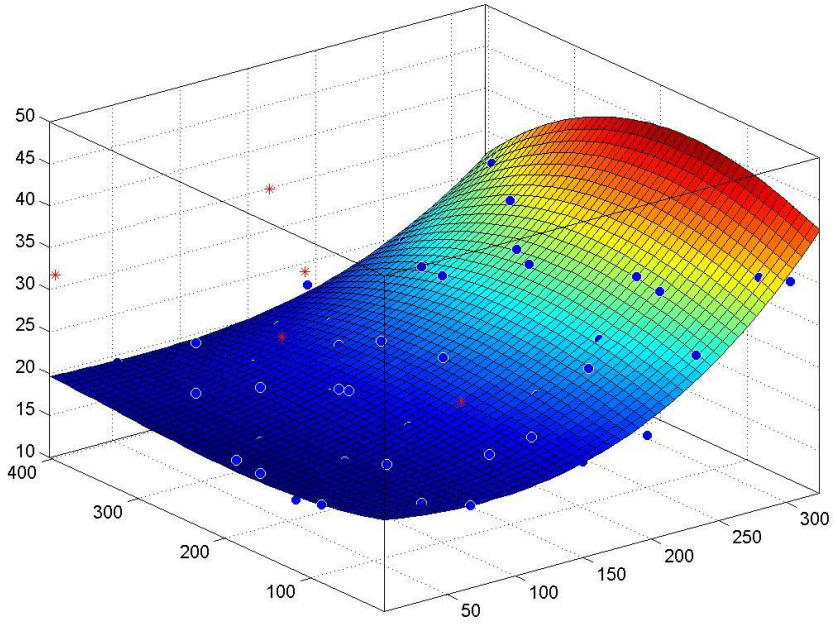
Script in Matlab: surface fitting by a global 3rd degree polynomial in x and Y of the nuclei position to re expressed all parasites coordinates relatively to the cell layer

```
ispace.login(true,Server,443,AppID, password);  CONNECTION

projectList = ispace.getActiveProjects();   SEARCH DATA
is=ImageSpaceObject.getImageSpace();
projectSpecGuids=is.search('DSX', 'Malaria parasite invasion
in the mosquito tissues', conditions, 100);

for i=1:length(projectSpecGuids)   GET DATA AND PROCESS DATA
    test=rec.getAttachments();


 Upload Results and fill History
end
```





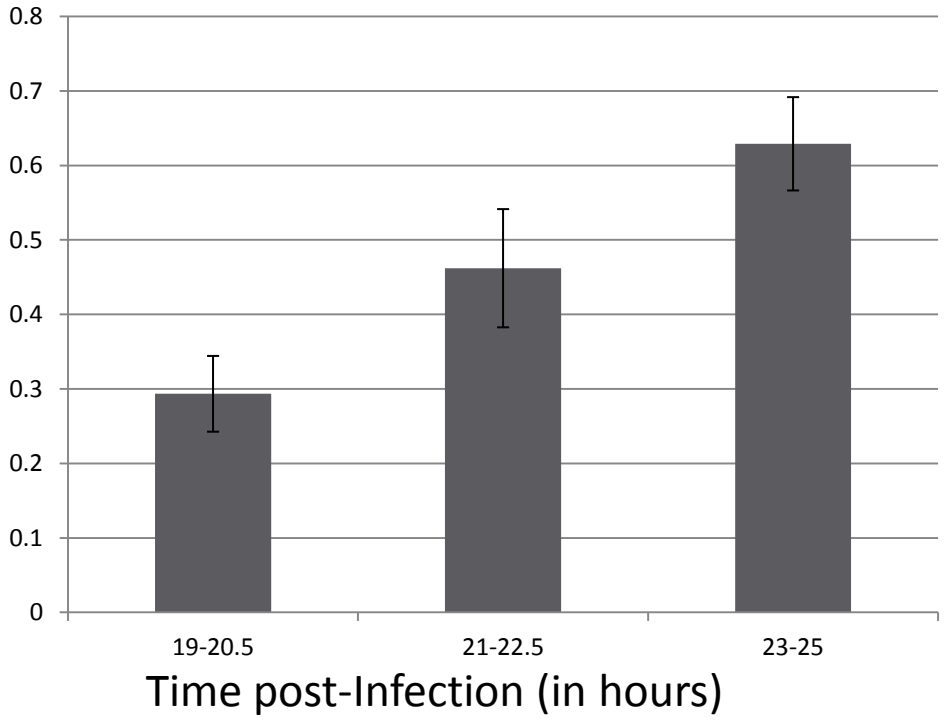

40
35
30
25
20
15
10
5
0
400
300
200
100
50
100
150
200
250
300
350
400


50
45
40
35
30
25
20
15
10
400
300
200
100
50
100
150
200
250
300

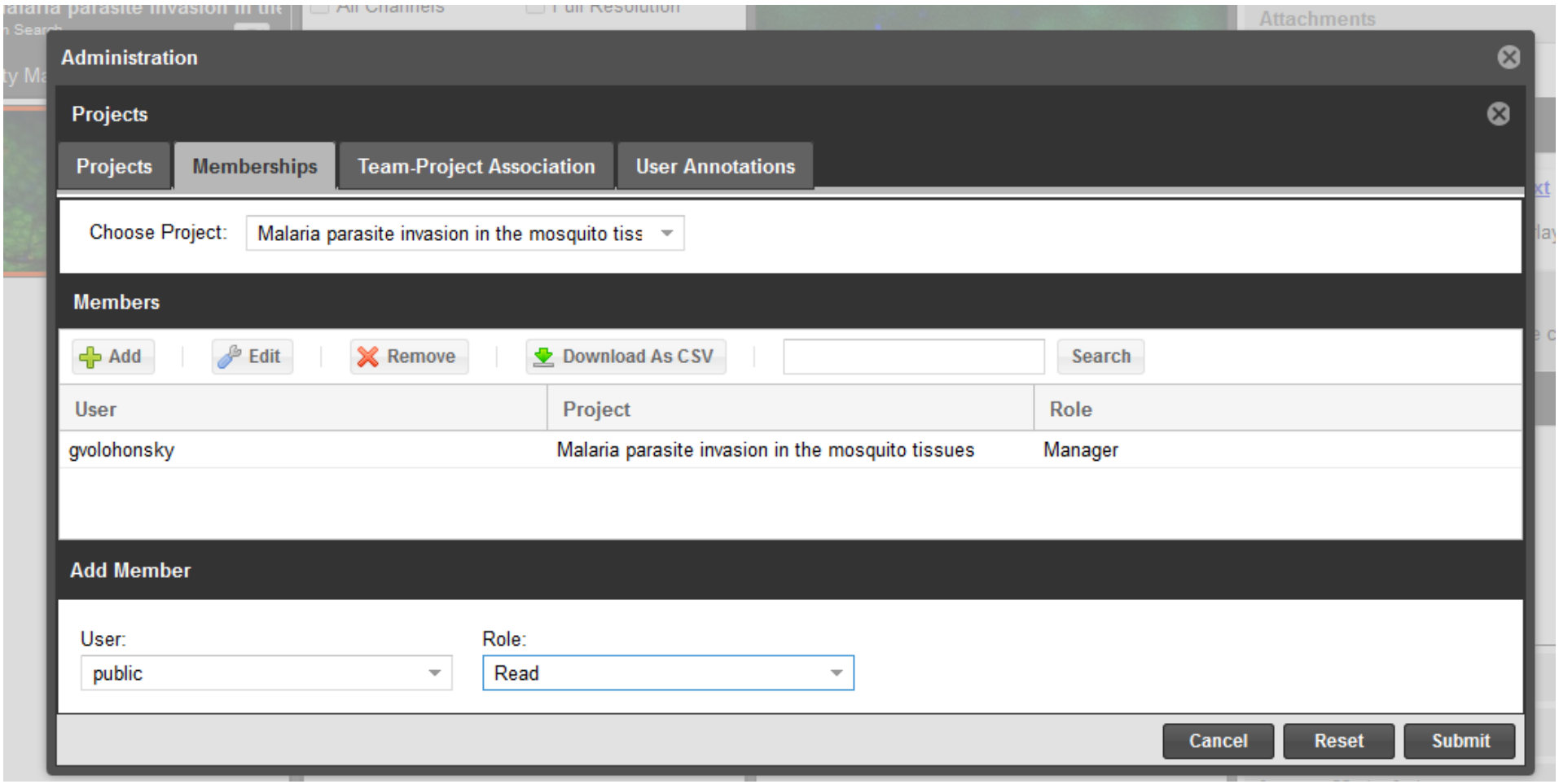
Example of request done from Matlab to the database: for Anopheles Gambiae (DSX)

Probability that the parasites can NOT cross the midgut epithelial tissue

computed as the ratio of parasites not having crossed (inside or under the cell layer) against total number of parasites)



Once ready for publication: Making the dataset public or creating an account for reviewer:



Questions ouvertes

-L'interopérabilité plutôt que la centralisation

Interopérabilité déjà possible (standardisation de la definition des ROI, API commune (proposition faites de webservices 'de traduction', web semantique pour inferer les demandes: ex requete: image reponse image et api dispos) inférence de type?.

- Intégration de données hétérogènes:

Un seul serveur ou des bases liables par API? client tiers de liage (Exemple projet Shiva KDI Bioinfo Curie)

- Réduction des volumes: Données à conserver? Nouveaux formats éparses? Nouveaux format de compression?

Acknowledgments

Conception et spécifications :CID iManage PICT IBISA Curie, Curie IT department, Strand Life sciences

Plugins add ons: Pict IBISA, Serpico INRIA

Funding: France Bioimaging, Cancéropole Ile de France.

Réseaux: France BioImaging (Bio Image Informatics Node (ex IPDM)), EUBIAS/ BIAS, RtMFM/MIFOBIO, GDR2588, EuroBioimaging/EuLife, Réseau RBDD

Need (1/2)

No image data conversions, no duplication

-Uploads data to a secure server in the original format

- Unique id for data (check sum like)

Supports sharing and collaboration with access control

- Allows users to upload, view, update or download data based on their access privileges

Supports multiple ways of attaching meta-information

- Annotations, comments and file attachments

-Analysis results as query-able visual objects

-**Supports Archiving** (data moving to another long-term storage but still searchable)

Need (2/2)

Facilitates custom visualization and analysis

- Access data from preferred analysis and visualization tools
- Access relevant bits of data to build efficient web and mobile Applications

Facilitate quota of storage and facility administration

Facilitate easy access to analysis and visualization applications hosted on other servers

- Run analysis on dedicated compute clusters
- Access applications hosted and published by other users

Highly Scalable

- Supports on-the-fly addition of server nodes to scale concurrent usage