

Gateways to analyze and publish High Content Screening imaging data in SystemsX.ch

Mario Emmenlauer, Pauli Rämö, Eva Pujadas, Béla Hullár, Bernd
Rinn and Peter Kunszt

IWSG, Zurich, 05.06.2013

Agenda

- ✕ The specific challenge of high content screening
 - ✕ On the example of the InfectX consortium
- ✕ screeningBee image analysis and object classification modules
- ✕ openBIS database for screening
- ✕ screeningBee workflow manager

InfectX HCS for pathogen entry into human cells

- ✗ Genome-wide screening on multiple pathogens:
 - ✗ What cellular components are involved in pathogen entry?
- ✗ Pathogens:
 - ✗ Adenovirus, Bartonella, Brucella, Listeria, Rhinovirus, Rotavirus, Salmonella, Shigella and Vacciniavirus.

Bacterial entry:

C. Arrieumerlou
BIOZENTRUM
Universität Basel



C. Dehio
BIOZENTRUM
Universität Basel



W.-D. Hardt

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich



P. Cossart

INSTITUT PASTEUR

Viral entry:

A. Helenius

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich



L. Pelkmans

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich



U. Greber

Universität Zürich

Modeling:

C. von Meering



Universität Zürich



D. Iber

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich



N. Beerenwinkel

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich



P. Bühlmann

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich

Proteomics:

B. Wollscheid

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich

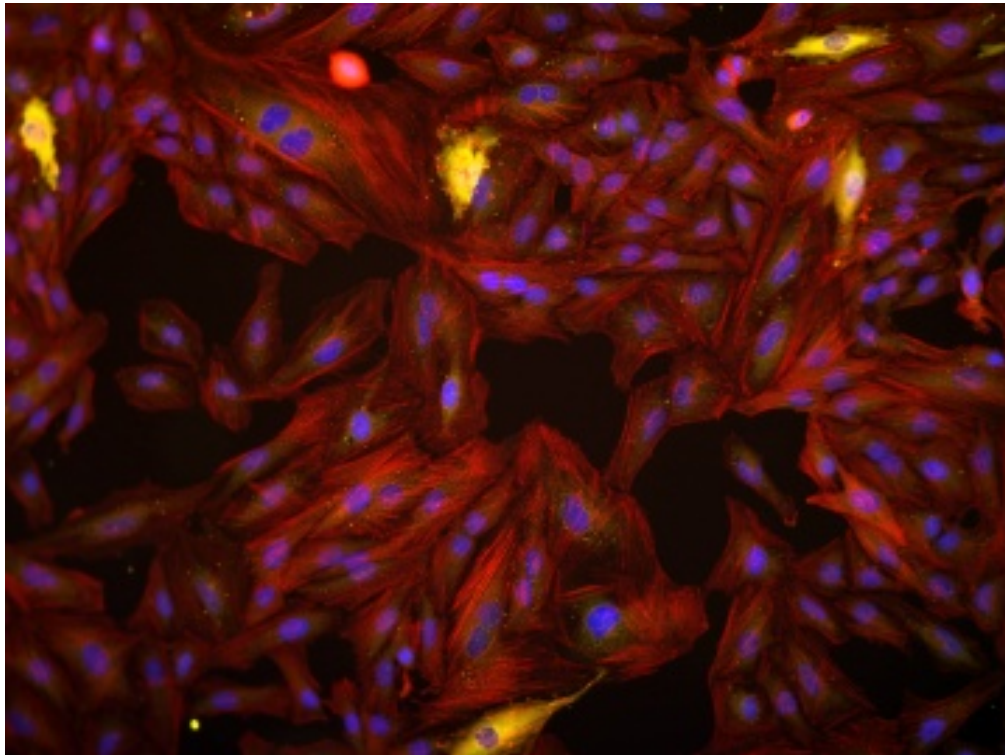


Industrial Partner:

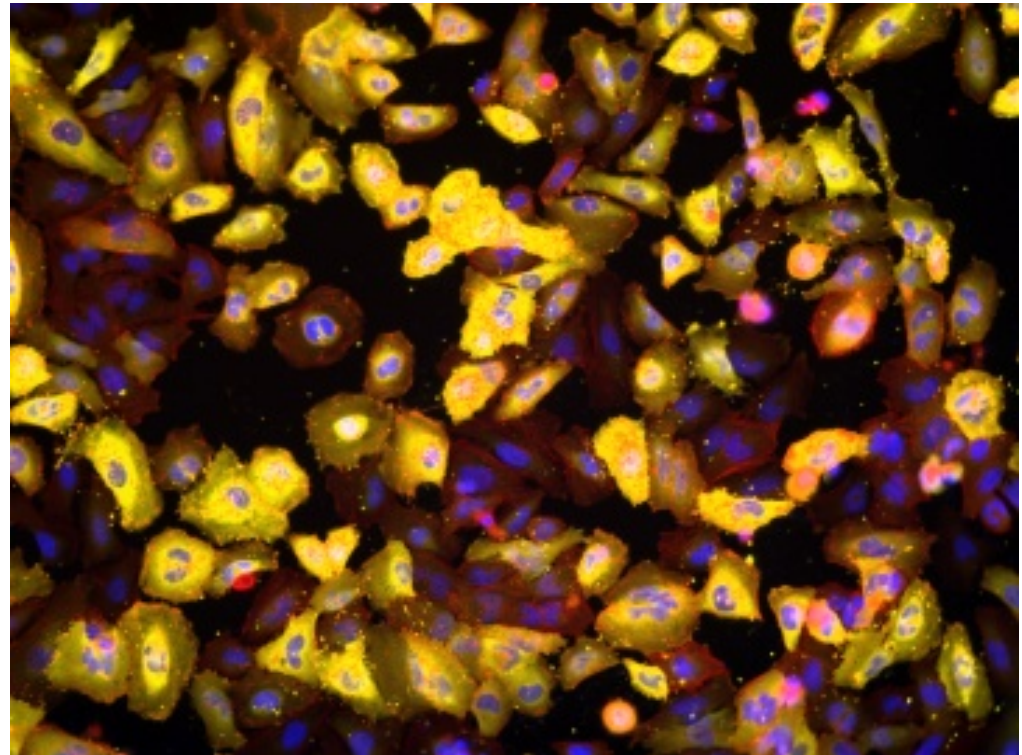


Listeria Infection Phenotypes: Primary Readout

After infection, Internalin-C is secreted in the cell body.



Low Infection

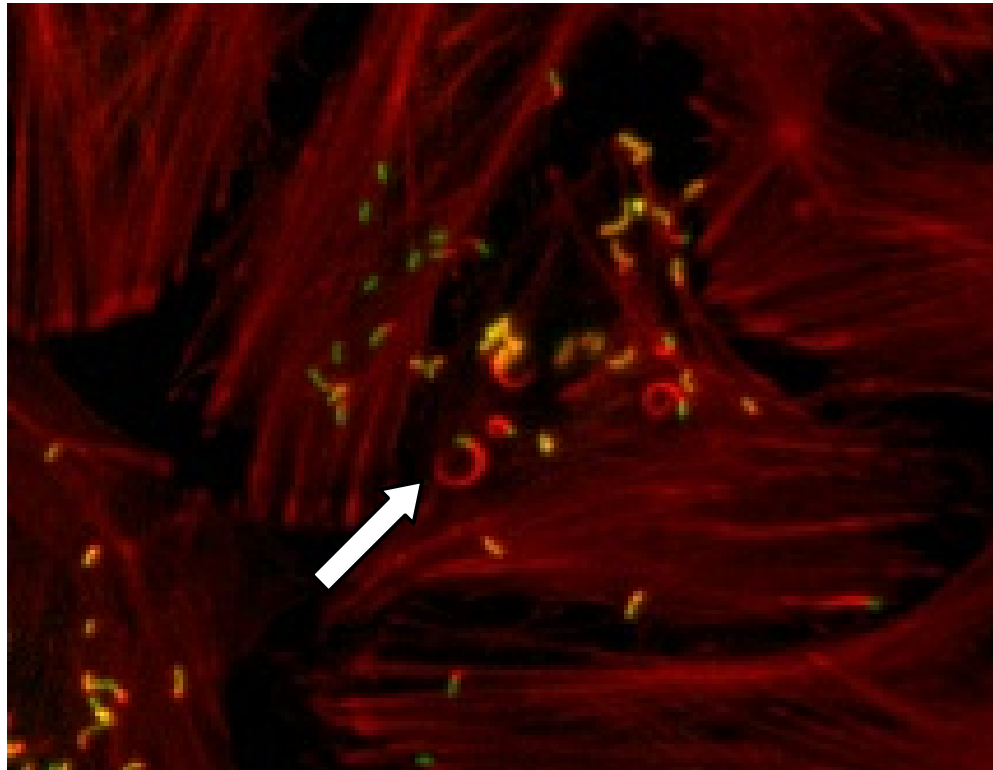


High Infection

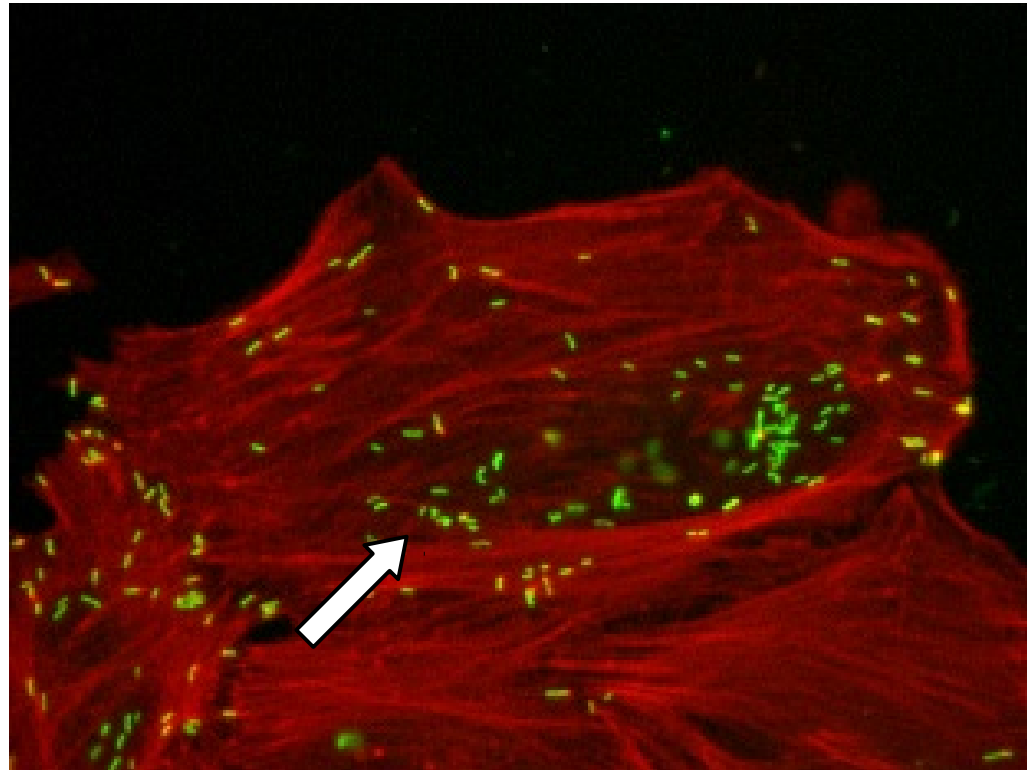
One pathogen out of nine: example assay "Listeria"

Listeria Infection Phenotypes: Secondary Readouts

Actin Comet Tails formation



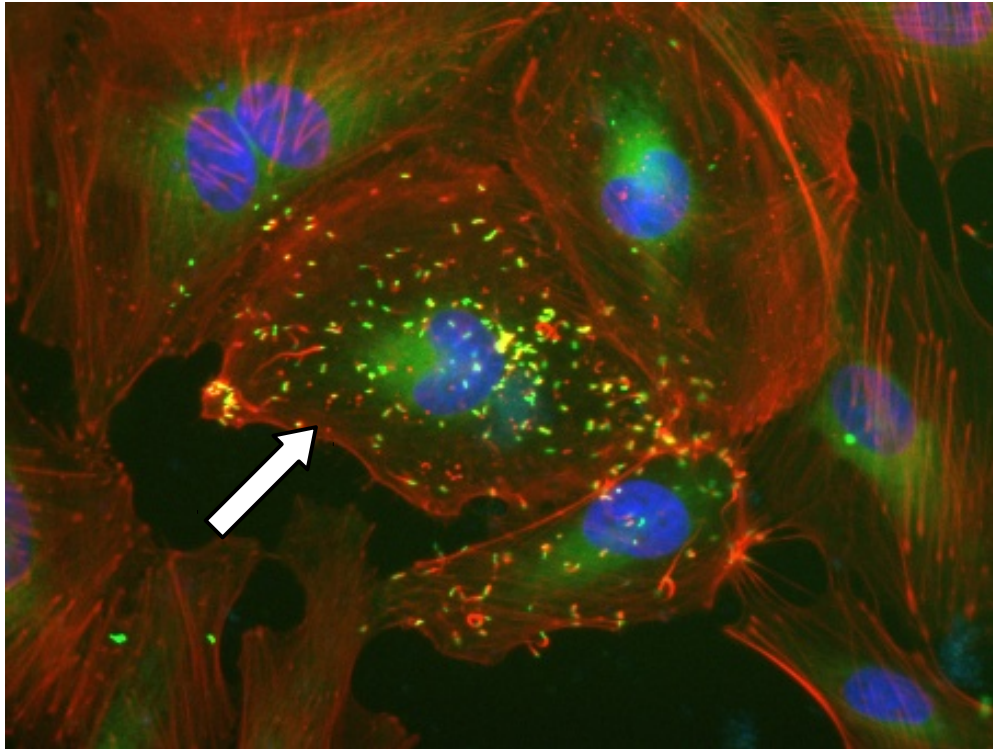
Bacteria detection



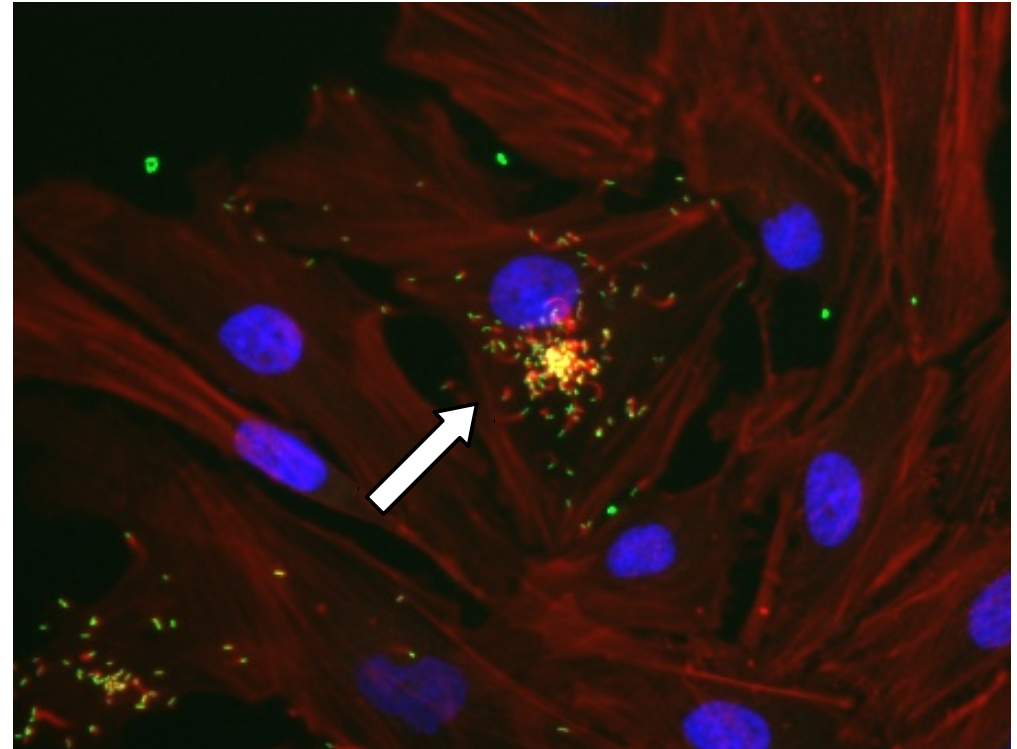
One pathogen out of nine: example assay "Listeria"

Listeria Infection Phenotypes: Secondary Readouts

Listeria randomly distributed

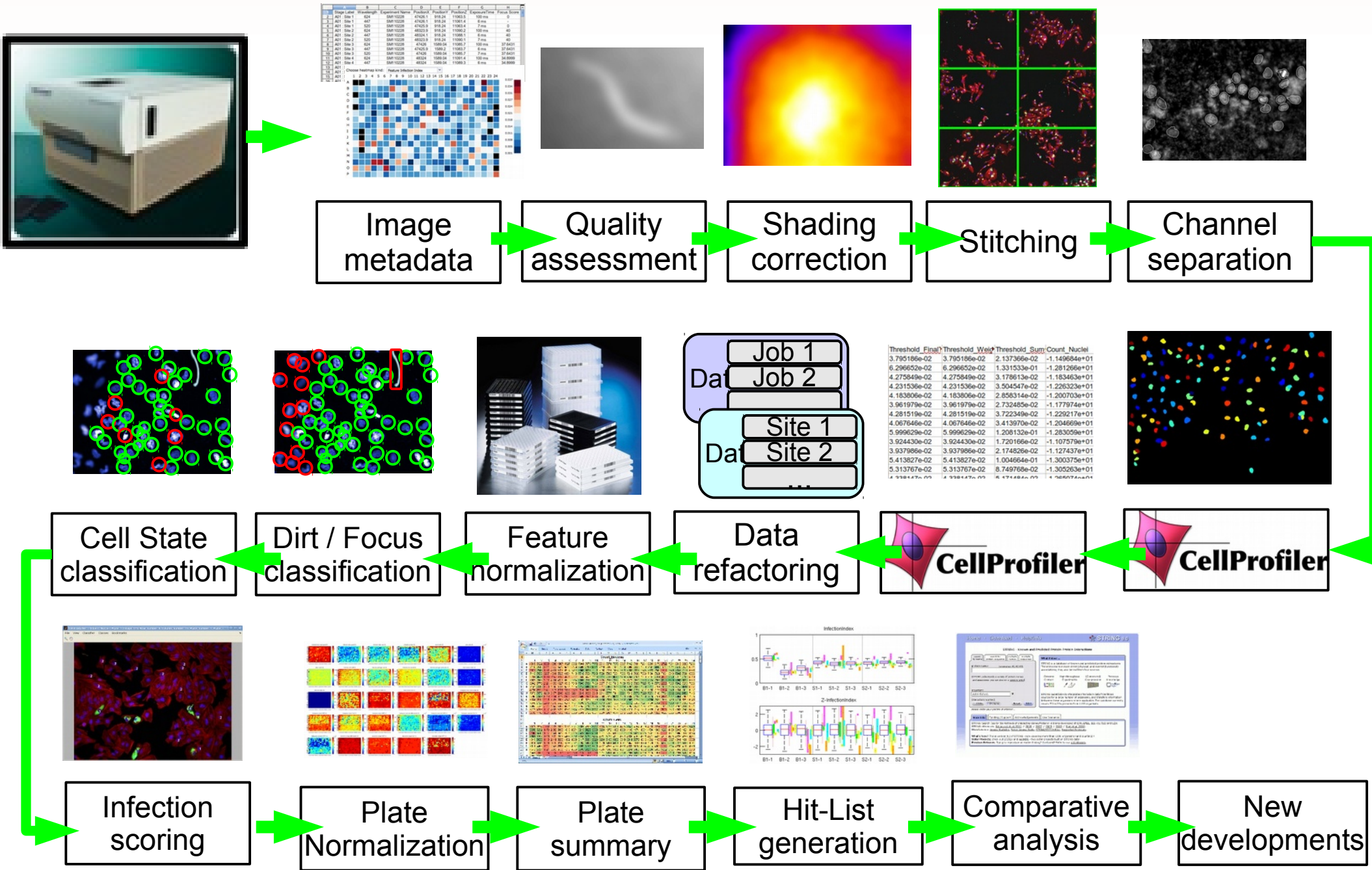



Listeria present in clusters



One pathogen out of nine: example assay "Listeria"

Main Image Analysis Workflow





Specific problems of high content screening image analysis

- ✖ Large input data sizes:
 - ✖ Total data volume InfectX: 3500 plates
 - ✖ Total assays (different experiments): ~120
 - ✖ Typical processing size: 1 plate (20GB size, 10.000 images, 100h CPU)
 - ✖ Typical workflow: 5-10 modules
 - ✖ Typical module: ~8 parallel cluster jobs

- ✖ No standard software exists, continuous improvements:
 - ✖ Use some of 35+ image analysis and classification modules
 - ✖ Continuous improvements: ~25 release versions per module per year

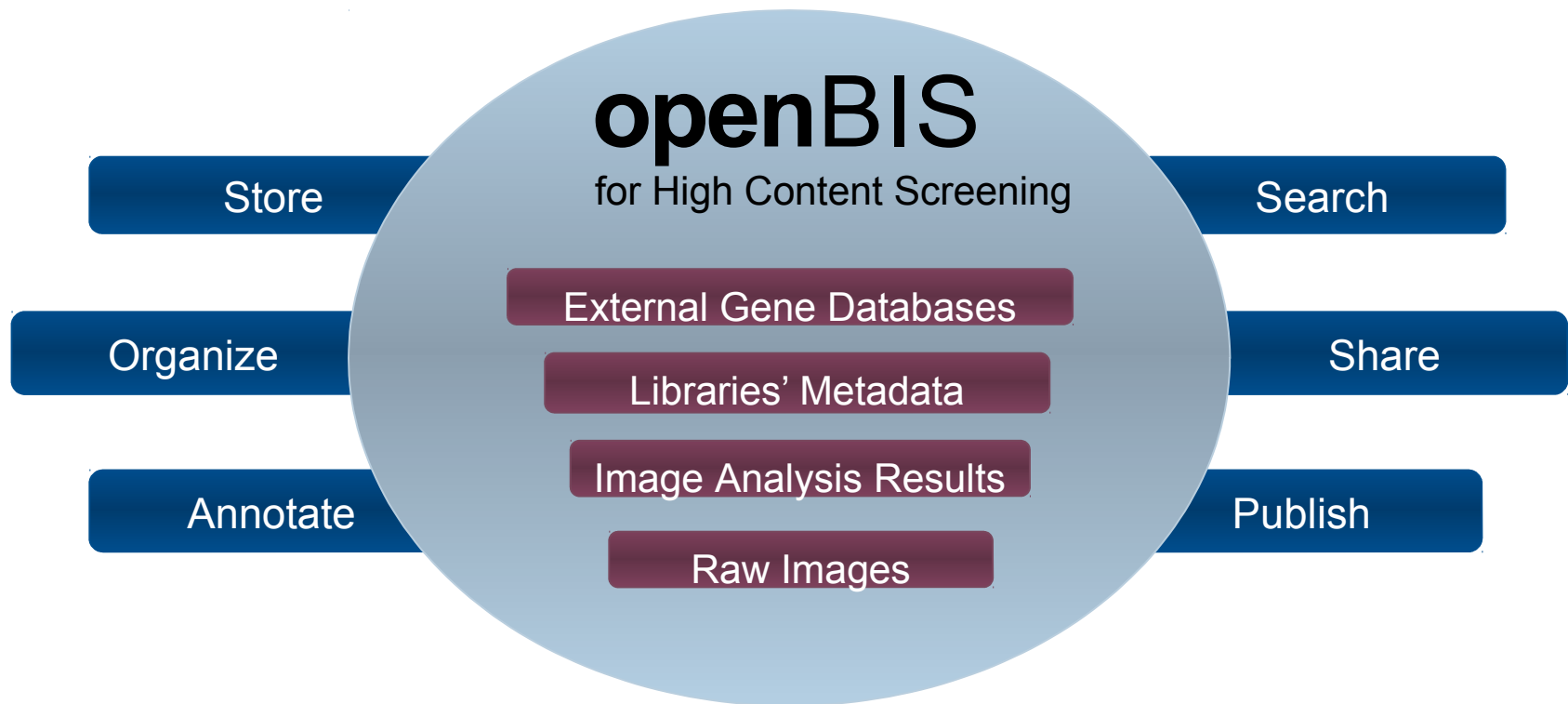
- ✖ Summary:
 - ✖ High processing cost, high chance of errors, and "zoo" of software versions makes manual tracking of processing almost impossible!

Agenda

- ✕ The specific challenge of high content screening
 - ✕ On the example of the InfectX consortium
- ✕ screeningBee image analysis and object classification modules
- ✕ openBIS database for screening
- ✕ screeningBee workflow manager

openBIS database

Comprehensive solution for
HCS Data Management



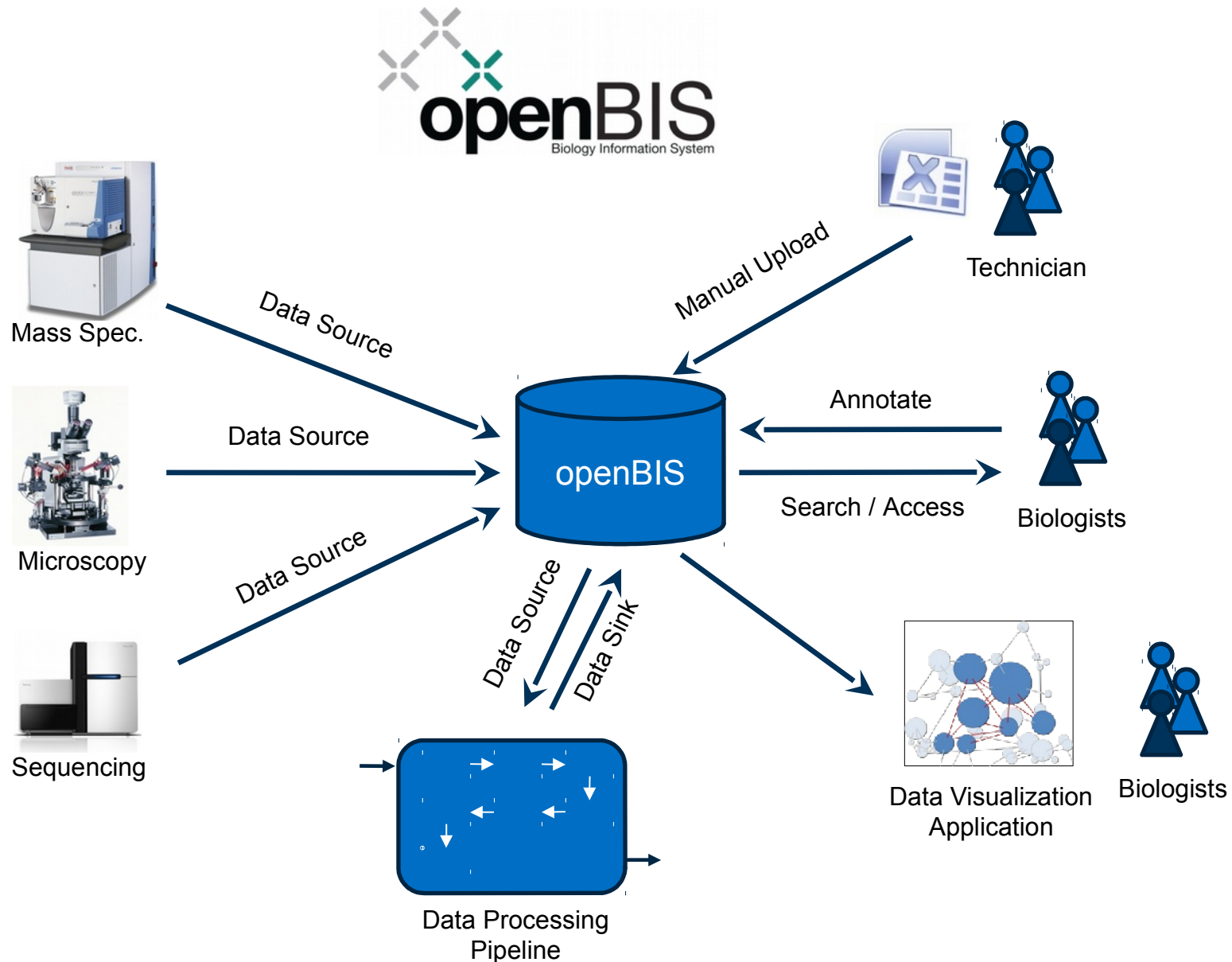


openBIS - a hub for quantitative imaging data

✕ openBIS Features:

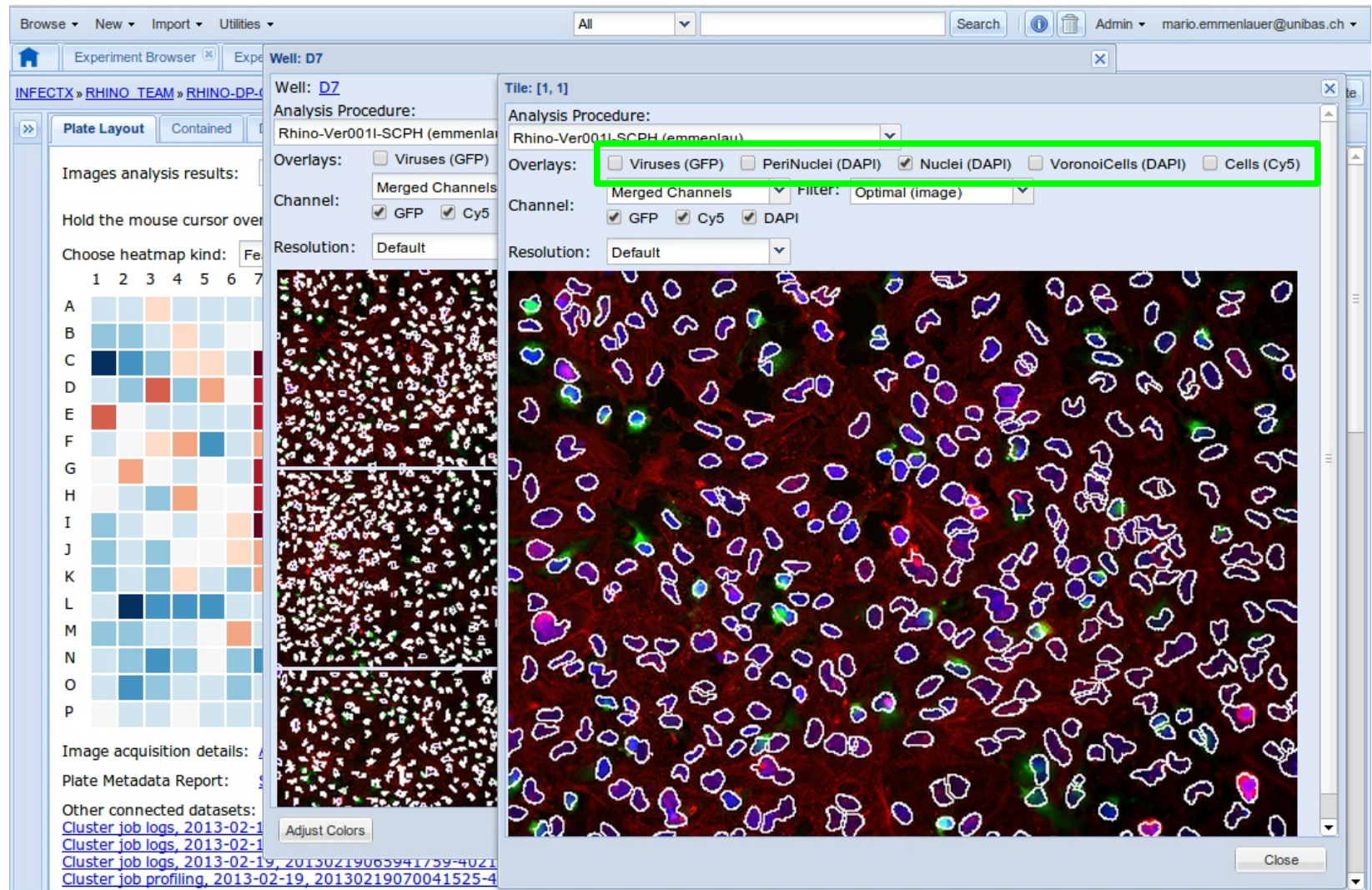
- ✕ Images, image analysis results and biological annotations in one system
- ✕ Organize experiments for different labs in independent “Data Spaces”
- ✕ Strict permission system on Spaces protects data of different users
- ✕ Support for “Data Provenance Tracking”
- ✕ Built-in support for browsing and sharing images, movies, image analysis results
- ✕ Very good automation, i.e. zero manual work for data import
- ✕ Built-in support for “heavy data lifting work”, e.g. compression and archiving of old images, built-in Secure FTP server – can be mounted as NetDrive
- ✕ Publishing of results via guest accounts
- ✕ Extendable in various directions through “Plugin Interfaces”
- ✕ Actively developed since 2007 and available as Open Source

openBIS - a hub for quantitative imaging data



Example: openBIS Display of Cell Segmentation

✖ Original images, superimposed image analysis results:



screeningBee workflow manager



✗ Motivation:

- ✗ Optimized for staging of large datasets (staging to cluster via head node, caching and reuse on cluster scratch, reliable data transfer with resume)
- ✗ Optimized to intrinsically avoid unnecessary processing steps
- ✗ Automatic dependency resolution, automatic parallelization
- ✗ Very easy to use for end-users

✗ Properties:

- ✗ Very lightweight (use openBIS for data storage, visualization, annotation, provenance tracking, data management)
- ✗ Supports LSF, SGE cluster engines, and supports local execution
- ✗ Web-Portal using JSF/IceFaces
- ✗ XML-based workflow description language
- ✗ RESTful API

screeningBee workflow manager



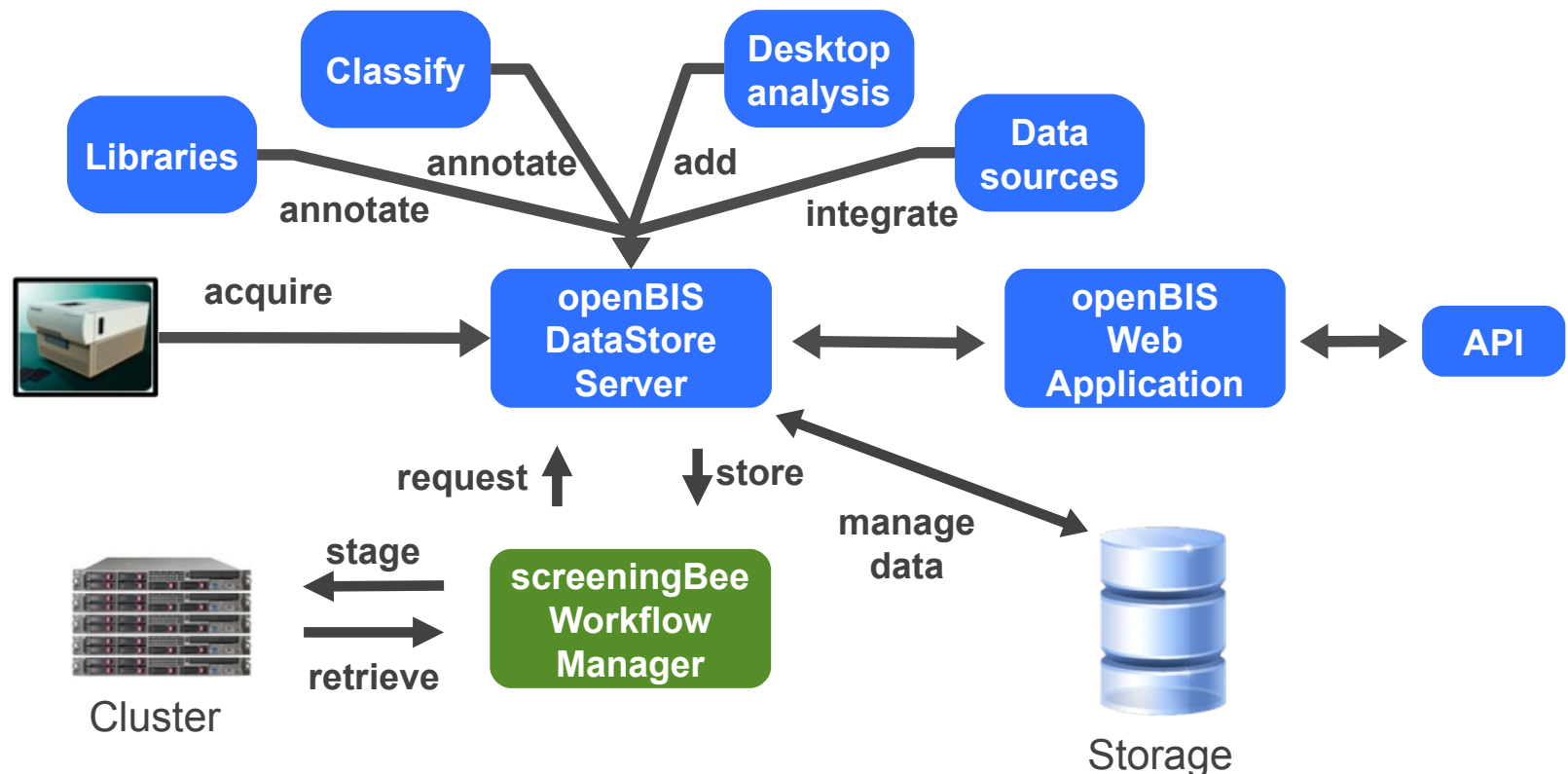
- ✗ Optimized to intrinsically avoid unnessecary processing steps:
 - ✗ Reanalysis of same data with same settings is a null-operation.
 - ✗ Very user friendly: run the newest workflow with newest modules on all datasets: unnessecary processing steps are automatically avoided.
 - ✗ Very user friendly: after potential crash, resume without overhead
- ✗ Based on "equivalence relation" for datasets
 - ✗ Define a software module versioning standard to automatically identify "equivalent" module software versions
 - ✗ Write (and read) provenance tracking information to deduce previously employed modules and settings

Key aspect: Datasets are named "equivalent", if their generation used: equivalent input datasets, equivalent settings, and equivalent software version

screeningBee workflow manager



- ✗ Based on iBRAIN2 components, but very lightweight, generic
- ✗ Tight openBIS integration: use data storage and management



Acknowledgements

Biozentrum, Basel:

- ✗ Christoph Dehio, Houchaima Ben-Tekaya, Alain Casanova, Raquel Conde, Simone Eicher, Shyan Low, Simone Muntwiler, Matthias Truttmann
- ✗ Cécile Arrieumerlou, Christoph Kasper, Christoph Schmutz

University of Zürich:

- ✗ Lucas Pelkmans, Victoria Green, [Berend Snijder](#)
- ✗ Urs Greber, Bettina Cardel, Daria Mudrak, Nina Wolfrum, Artur Yakimovich
- ✗ Christian von Mering, Neha Daga, Andrea Franceschini

ETH, Zürich:

- ✗ Wolf-Dietrich Hardt, Daniel Andritschke, Sabrina Dilling, Saskia Kreibich
- ✗ Ari Helenius, Jason Mercer
- ✗ Niko Beerenwinkel, Edgar Delgado-Eckert, Juliane Siebourg, Ewa Szczurek
- ✗ Bernd Wollscheid, Thomas Bock, Andreas Frei, Karel Novy

SyBIT / SystemsX.ch

Insitut Pasteur, Paris:

- ✗ Pascale Cossart, Andreas Kühbacher, Javier Pizarro-Cerdá

ETH Zürich, Basel:

- ✗ Dagmar Iber, Federico Felizzi, Georgios Fengos, Florian Geier

ResearchIT, Biozentrum (iBRAIN2):

- ✗ [Michael Podvinec](#), [Vincent Rouilly](#), [Stephen Roth](#), [Rainer Pöhlmann](#)

CISD, Basel (openBIS):

- ✗ [Bernd Rinn](#), [R. Chandrasekhar](#), [Franz-Josef Elmer](#), [Pawel Glyzowski](#), [Tomasz Pylak](#), [Jakub Straszewski](#)

LMSC, ETH, Zürich:

- ✗ Gábor Csúcs, Andreas Kauffman, Andreas Vonderheit, Peter Horvath, Kevin Smith, Filippo Piccinini, Csaba Balazs, Karol Kozak, Michael Stebler

InfectX Data Analysis (Bee):

- ✗ [Bela Hullar](#), [Damian Murrezzan](#), [Eva Pujadas](#), [Pauli Rämö](#), [Gabriel Studer](#)

SyBIT, SystemsX.ch:

- ✗ [Peter Kunszt](#)



Questions / Discussion

✕ Questions and Discussion?