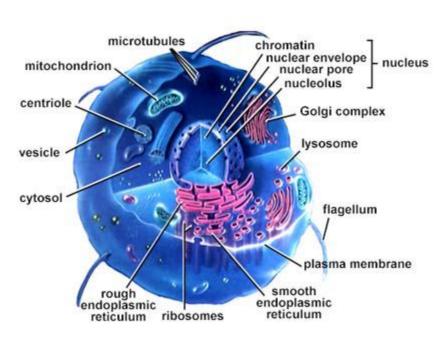
## OMERO.searcher: Context and content-based search

Baek Hwan (BK) Cho

(Mentor: Prof. Robert F. Murphy)

Lane Center for Computational Biology Carnegie Mellon University

- PSLID (Protein Subcellular Location Image Database)
- Mission
  - Providing information on protein subcellular location as determined from microscopy images



#### **Location Proteomics:**

- ➤(dynamic) proteome map to understand normal and abnormal cell behaviors
- >clues for interaction/regulation of proteins
- >clues for protein function

- http://pslid.org
- An open-source image database webapp



Protein Subcellular Location Image Database

## **PSLID - Protein Subcellular Location** Image Database

Go to the PSLID database containing CD-tagged NIH 3T3 cell clones New!

Go to the PSLID Public Database Release 4 (released August 1, 2008)

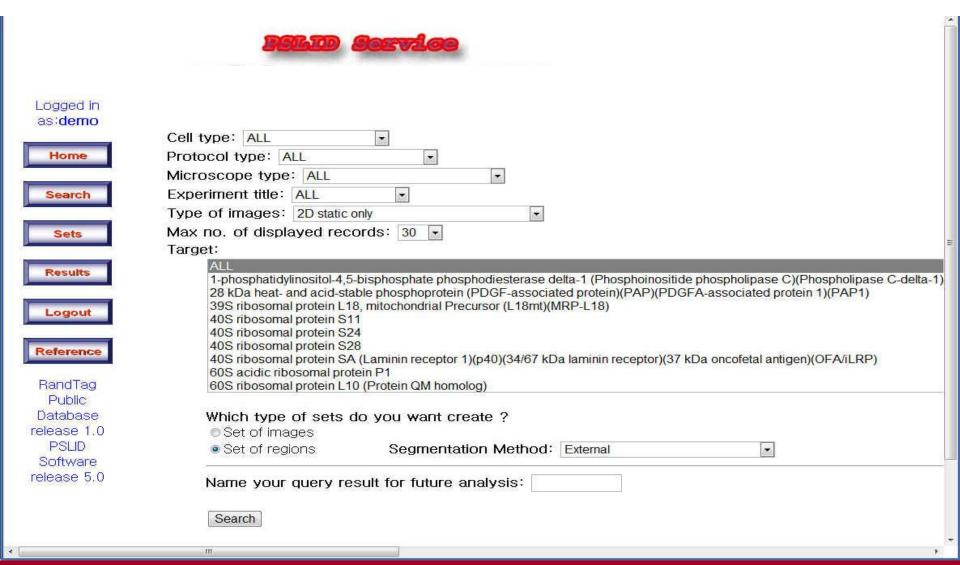
Download datasets contained within PSLID

Download SLIC feature calculation and pattern analysis software used in PSLID and full PSLID software



Supported by the National Institute of General Medical Sciences, Grant GM075205

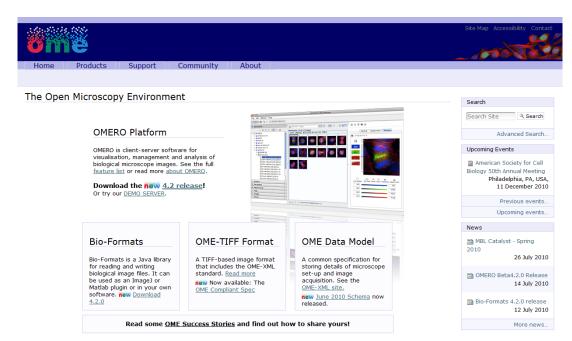
Multi criterion search



- Analysis tools
  - Segmentation/feature calculation
  - Typical image selection
  - Set comparison
  - Feature selection
  - Classifier training/use
  - Stacked/meta classifier training/use
  - Clustering to sets
  - Generative models training/synthesis

### OME

- Open Microscopy Environment (OME)
  - OME data model
    - OME-XML
    - OME-TIFF
  - OMERO



http://www.openmicroscopy.org/

## Motivation

 Let people use PSLID tools on existing OMERO DB

- Use robust infrastructure of OMERO
  - PSLID only accepts a few types of images

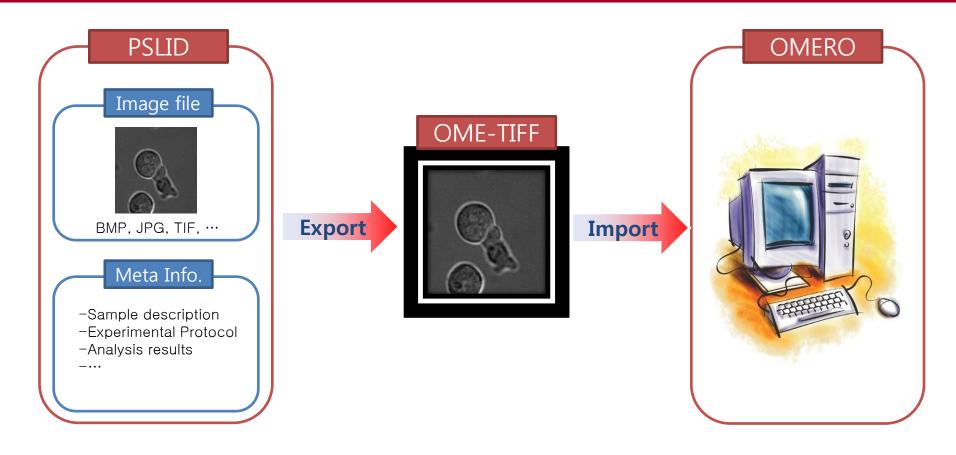
 Avoid maintaining duplicate image storages and retrieval architecture

## Scope of the project

- Extend OME-XML data model
  - Context data model
    - experiment protocol etc.
  - Content data model
    - statistical image features etc.
- Create OMERO-PSLID version
  - In OMERO.insight (using OMERO.script)
  - In OMERO.web

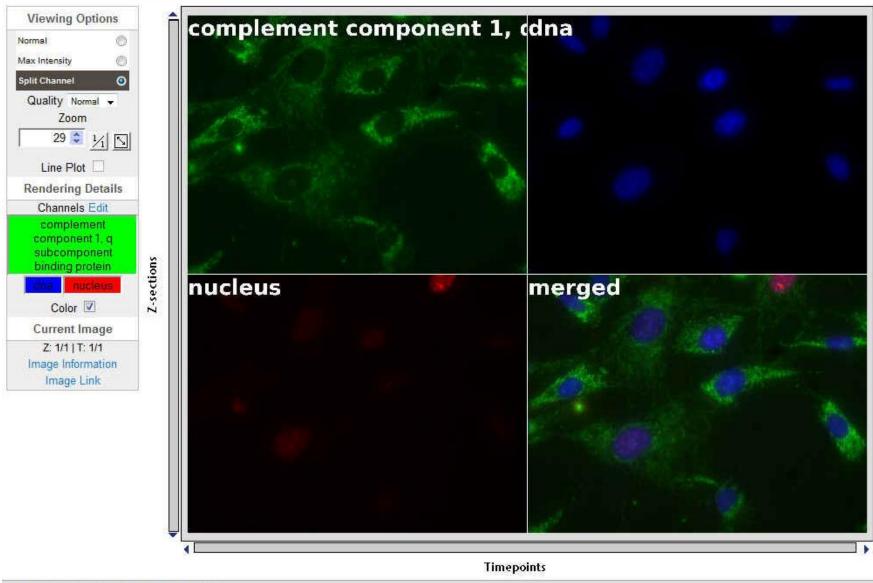
 Implement easy transfer of images and information between OMERO and PSLID

## Converting to OME-TIFF



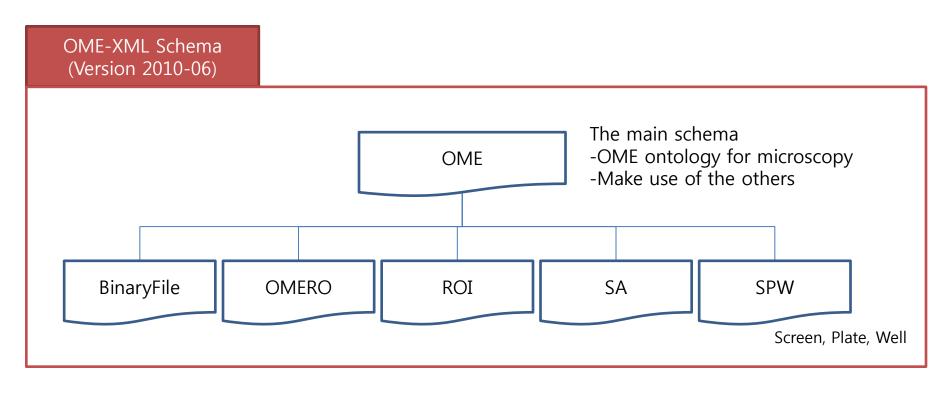
- Automatically map every meta information to either
  - regular OME-XML
  - additional context meta data (will be attached to an image as an external XML file)

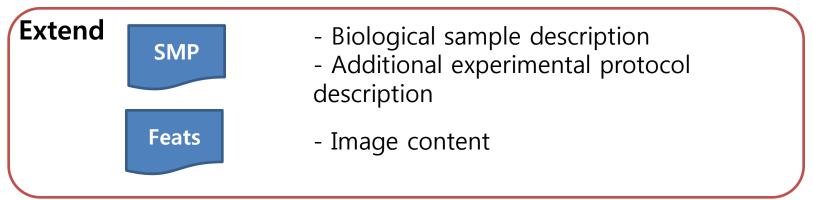
## Converted OME-TIFF image



@2007/2008 Glenoce Software Inc. All rights reserved.

## Extension to current OME-XML data model



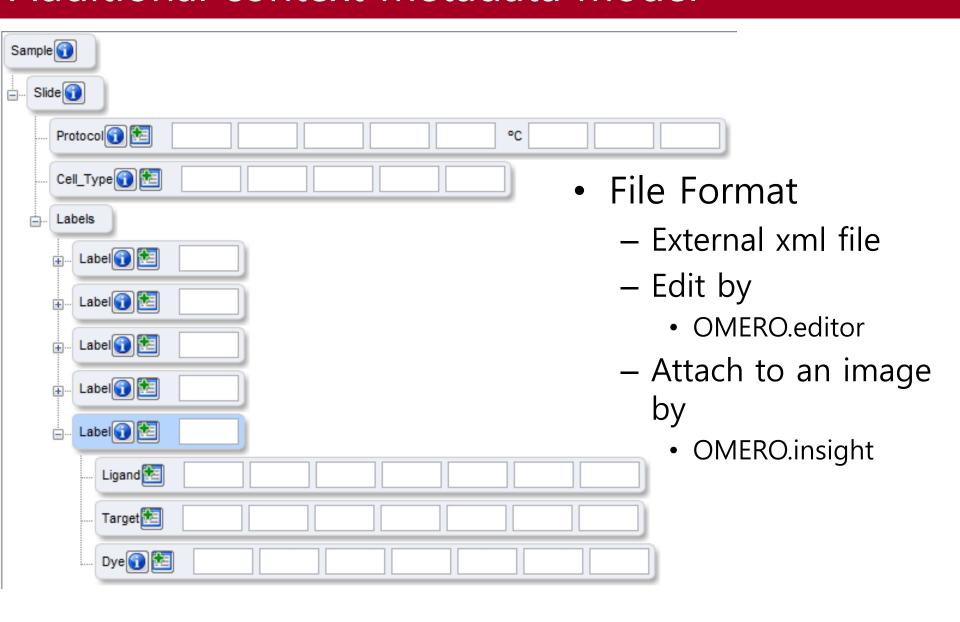


## Initial solution

- Context Meta Data
  - 1) Define an XML Schema for an external XML file
  - 2) Use OMERO to attach the XML file to an image

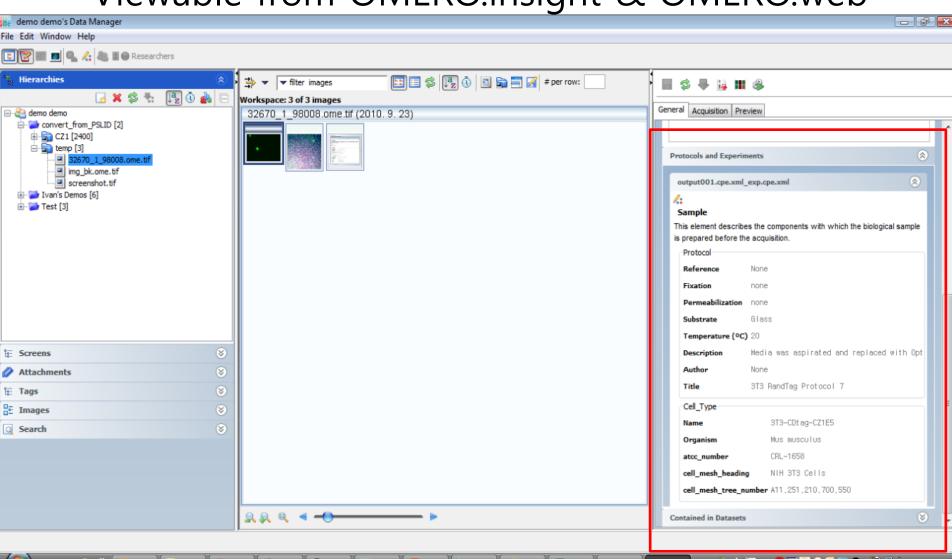
- Content Meta Data
  - 1) Define format for external HDF5 file
  - 2) Attach the HDF5 files to an image

## Additional context metadata model



## Additional context metadata model

Viewable from OMERO.insight & OMERO.web



## Content metadata model

- File Format
  - HDF5 file (1 feature set per a file)

- Examples
  - File Name: iid<4355>\_feature\_<SLF34>\_field.h5

FF_001	FF_002	FF_003	FF_004	FF_005	FF_006	FF_007
10.5	0.7	0.88	0.1	145	20	0.7

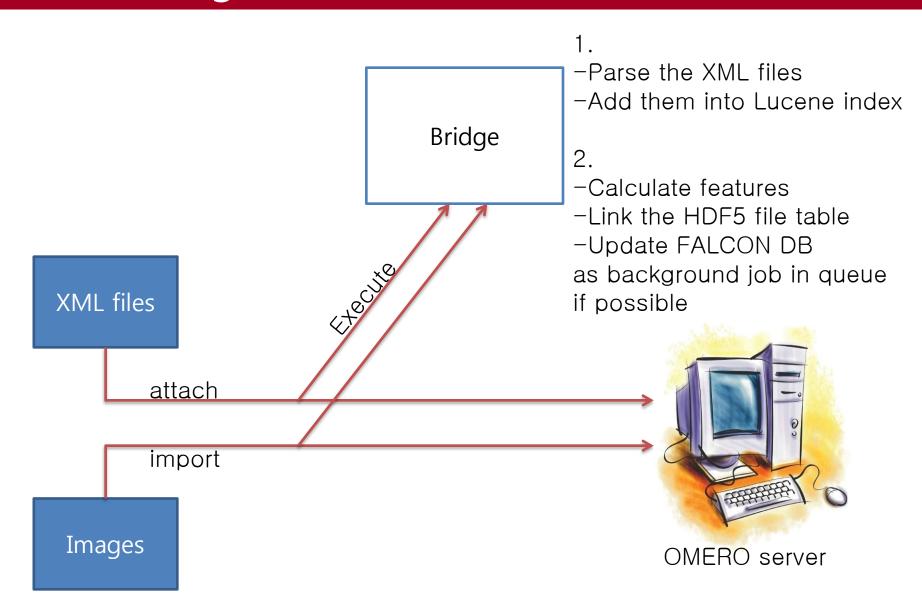
– File Name: iid<2738>\_feature\_<SLF20>\_roi.h5

region	RF_001	RF_002	RF_003	RF_004	RF_005
1	10.5	0.7	0.88	0.1	145
2	13.1	0.1	0.51	1.1	2
3	16.2	1.9	0.12	3.7	30
4	9.5	2.2	0.23	0.9	78
5	11.9	0.2	0.95	0.7	15
6	7.7	0.1	0.42	1.5	86

### Content metadata: feature calculation

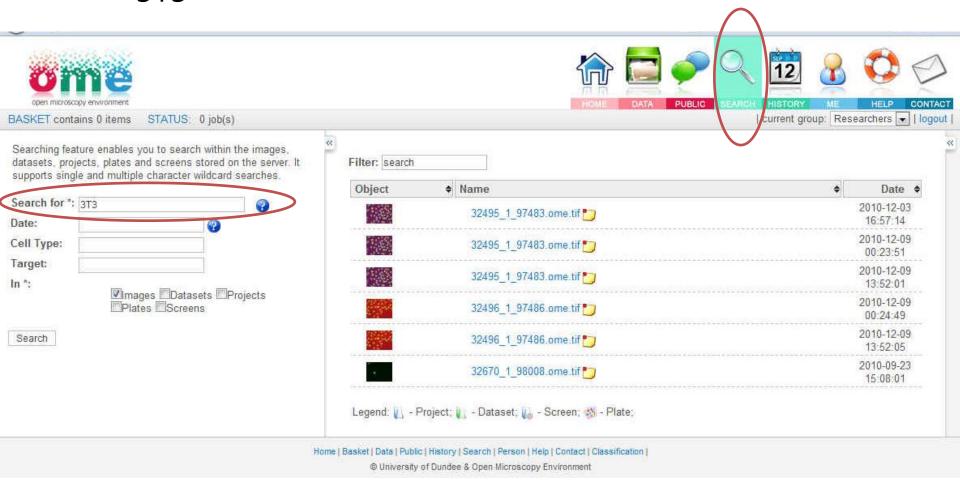
- Implemented feature calculation library
  - Uses PySLIC
    - Texture features
    - Object features
    - Morphological features
    - Overlap features between protein and DNA channel images
  - Supports
    - Feature calculation
    - Linking (attaching) features to an image as an HDF5 file
    - Retrieving features from an HDF5 file

## Custom Bridge



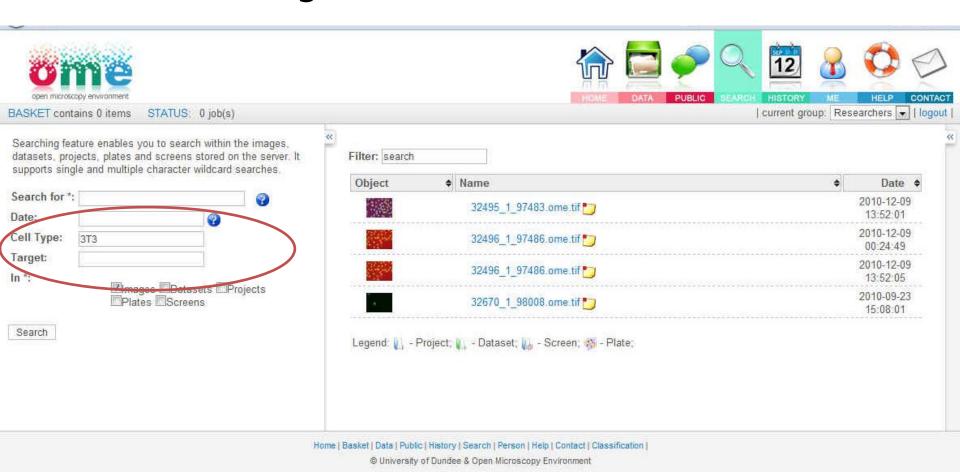
## Search by additional context metadata

- Default search
  - 3T3



## Search by additional context metadata

Specific keyword search (enabled by our custom bridge)



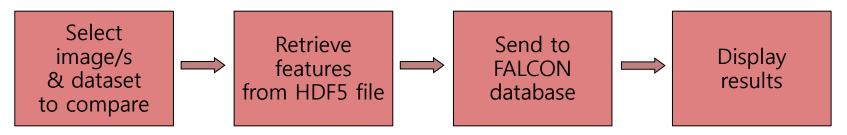
OME users meeting, June 15, 2011, Paris

## Initial use case

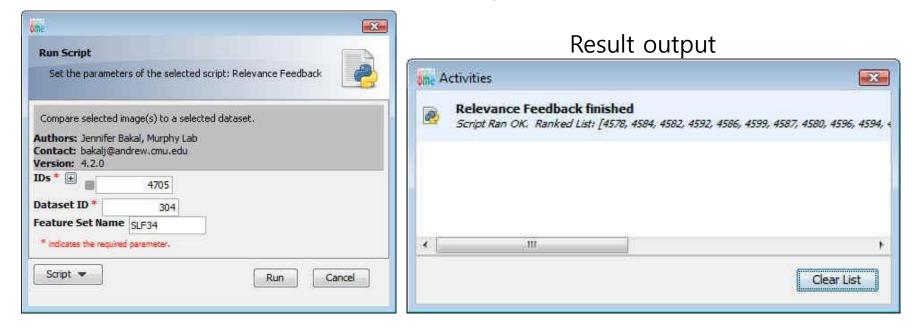
- Search by image content (Relevance Feedback)
  - Give me the images that match best with this/these image(s) in terms of image content
  - Using Murphy lab's subcellular location features (SLFs)
    - 173 features
  - Using FALCON DB

## Initial use case (from OMERO.Insight)

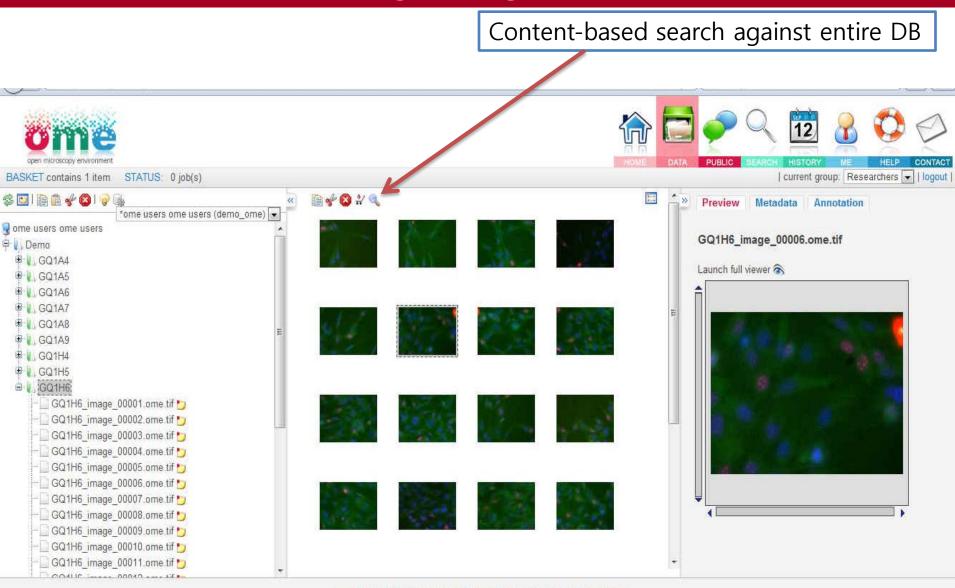
#### Search Process



#### Initial Implementation from OMERO.Insight



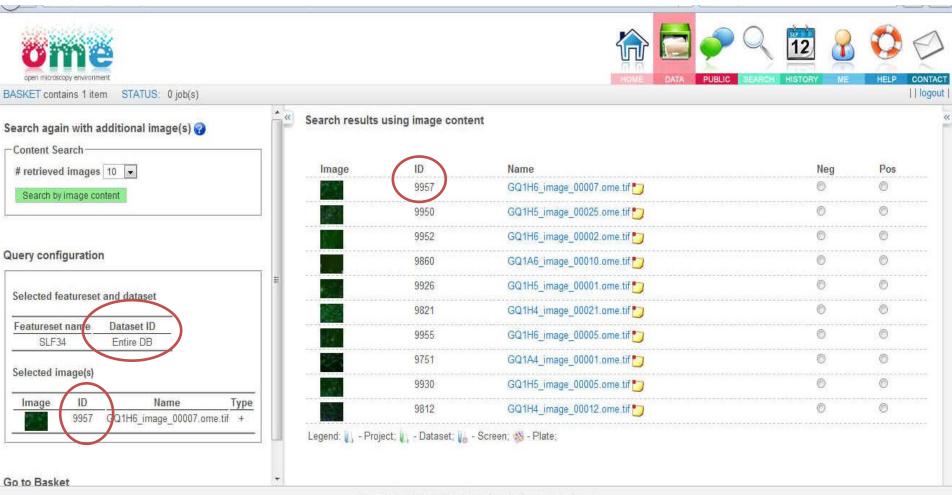
## Screenshot: selecting image(s)



Home | Basket | Data | Public | History | Search | Person | Help | Contact | Classification |

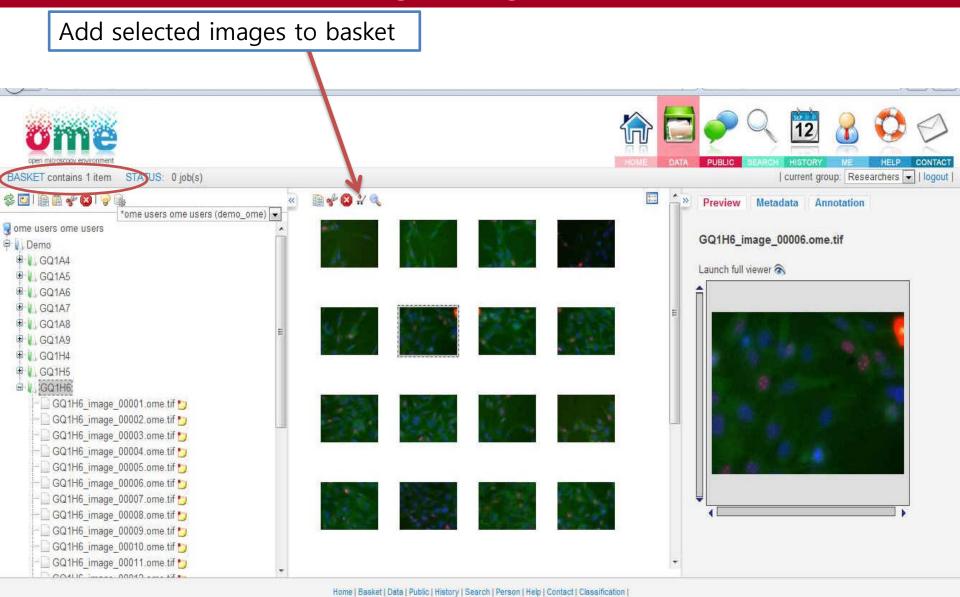
© University of Dundee & Open Microscopy Environment

# Screenshot: search results against entire DB



Home | Basket | Data | Public | History | Search | Person | Help | Contact
© University of Dundee & Open Microscopy Environment

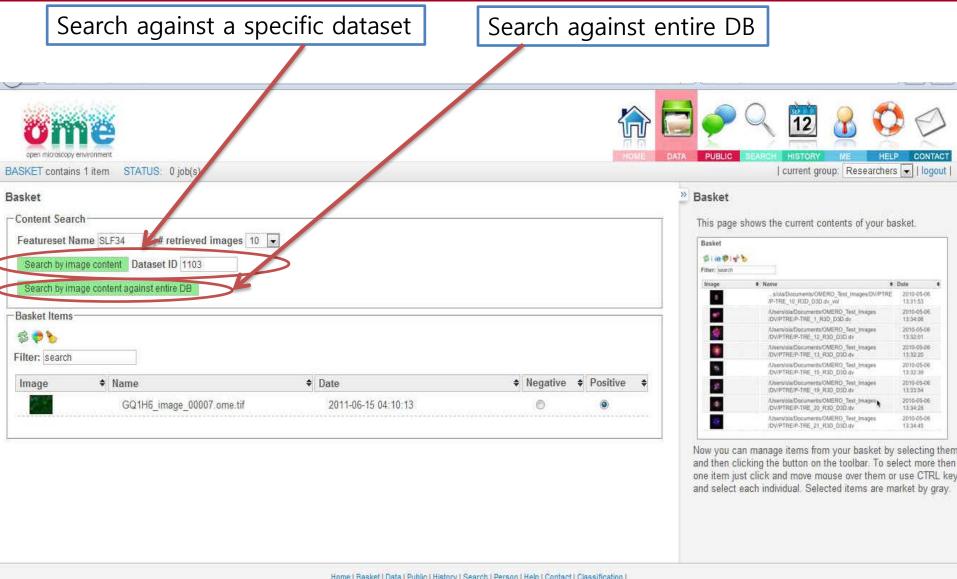
# Screenshot: selecting image(s)



OME users meeting, June 15, 2011, Paris

@ University of Dundee & Open Microscopy Environment

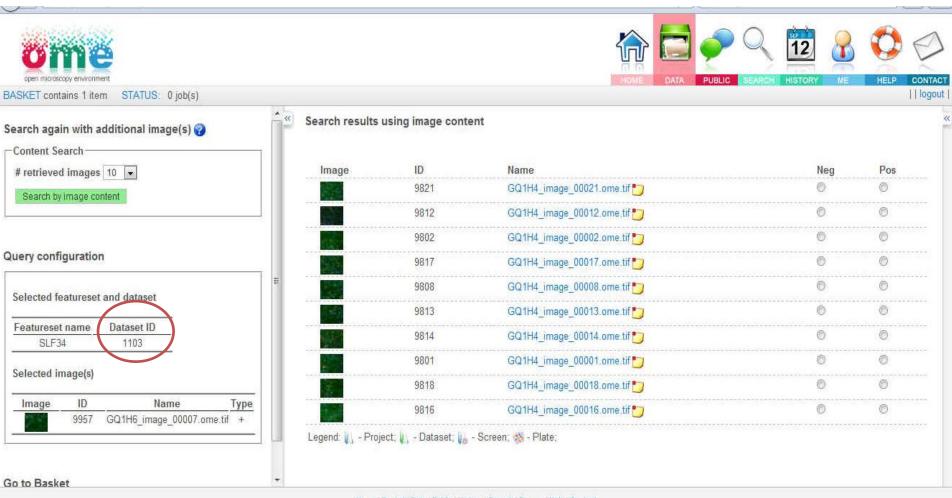
# Screenshot: modified Basket (specific options)



Home | Basket | Data | Public | History | Search | Person | Help | Contact | Classification |

© University of Dundee & Open Microscopy Environment

# Screenshot: search results against dataset 1103



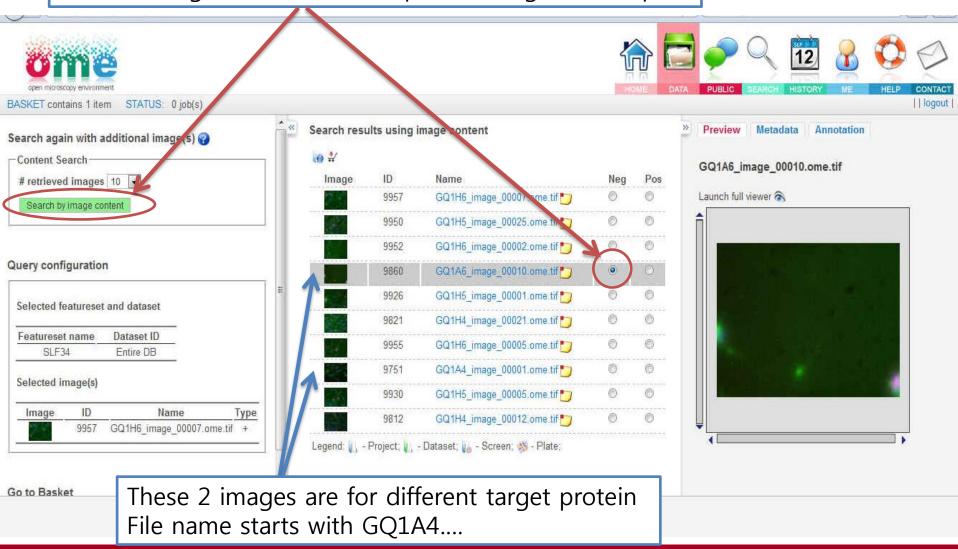
Home | Basket | Data | Public | History | Search | Person | Help | Contact

© University of Dundee & Open Microscopy Environment

## Screenshot: another round of search

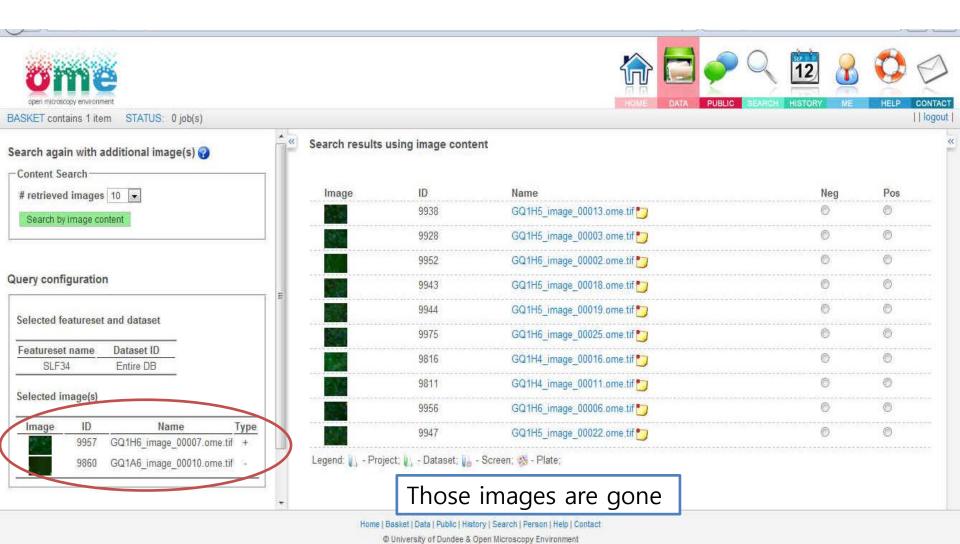
Another round of search

- Select images from results as positive/negative samples



## Screenshot: another round of search results

Keep searching until you are satisfied!

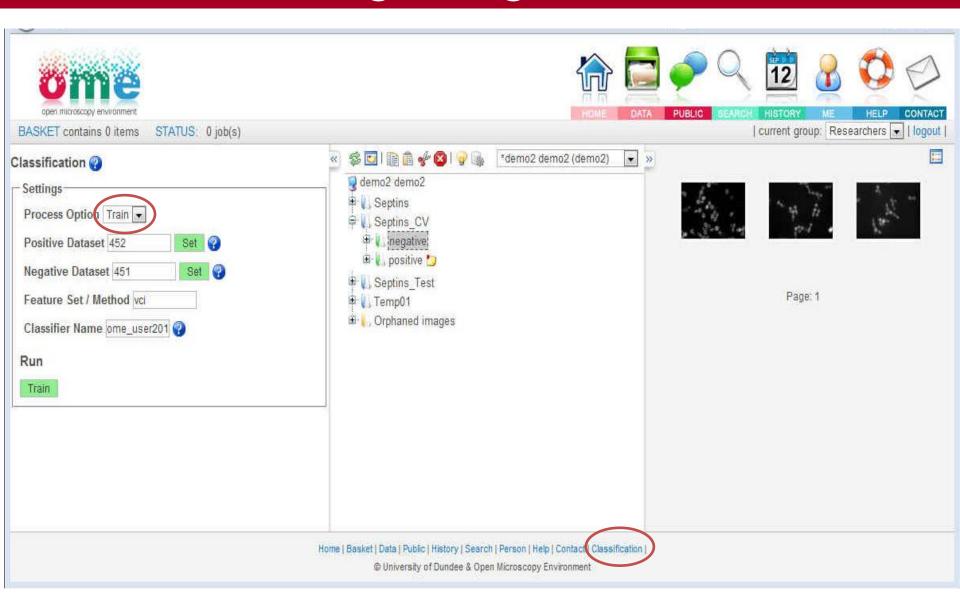


OME users meeting, June 15, 2011, Paris

## Another use case

- Classification
  - Classifying images according to whether they are showing tagged protein or not
    - in order to recognize protein-protein interaction
  - Using texture features and modified histogram features

# Screenshot: training configuration



## Screenshot: training results















BASKET contains 0 items ST

STATUS: 0 job(s)

#### Configurations

Parameter	Value LOOCV	
Cross Validation method		
Dataset ID for positive samples	452	
Dataset ID for negative samples	451	
Saved Classifier	ome user2011.pkl	
Feature set / Method	vci	

#### Cross Validation Results

Parameter	Value
Cross Validation Accuracy	100.0
P-value for positive vs. negative	0.0227501319482

#### Confusion Matrix

Samples	No. of Images	Positive	Negative
Positive	3.0	100.0	0.0
Negative	3.0	0.0	100.0

#### Hide Samples

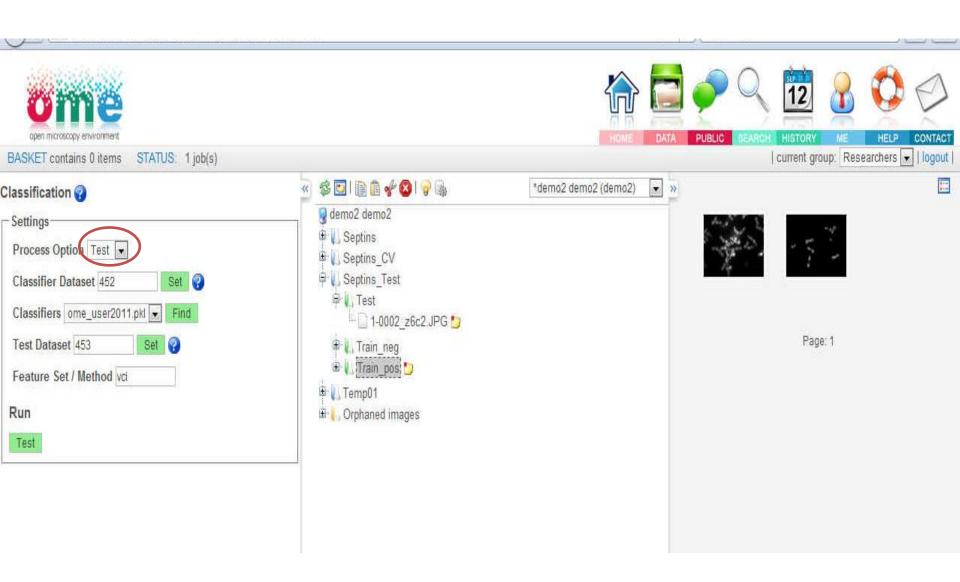
#### Sample(s)

Image	ID	Original Label	Predicted Label	Name
	4781	Positive	Positive	110510 Snf1-GFP Sep7-Snf4-0016_z6c2.JPG
-	4854	Positive	Positive	-0003_z4c2.JPG
	4855	Positive	Positive	110510 Snf1-GFP Sep7-Snf4-0008_z7c2.JPG
	4851	Negative	Negative	0002_z5c2.JPG
W.A.	4852	Negative	Negative	1-0001_z5c2.JPG
	4050	N.Y	M	1.0000 0.0 1000

Home | Basket | Data | Public | History | Search | Person | Help | Contact | Classification |

@ University of Dundee & Open Microscopy Environment

# Screenshot: test configuration



## Screenshot: test results





BASKET contains 0 items STATUS: 1 job(s)

#### Configurations

Parameter	Value TEST	
Cross Validation method		
Dataset ID for Test	453	
Classifier Name	ome_user2011.pkl	
Feature set / Method	vci	

#### Hide Samples

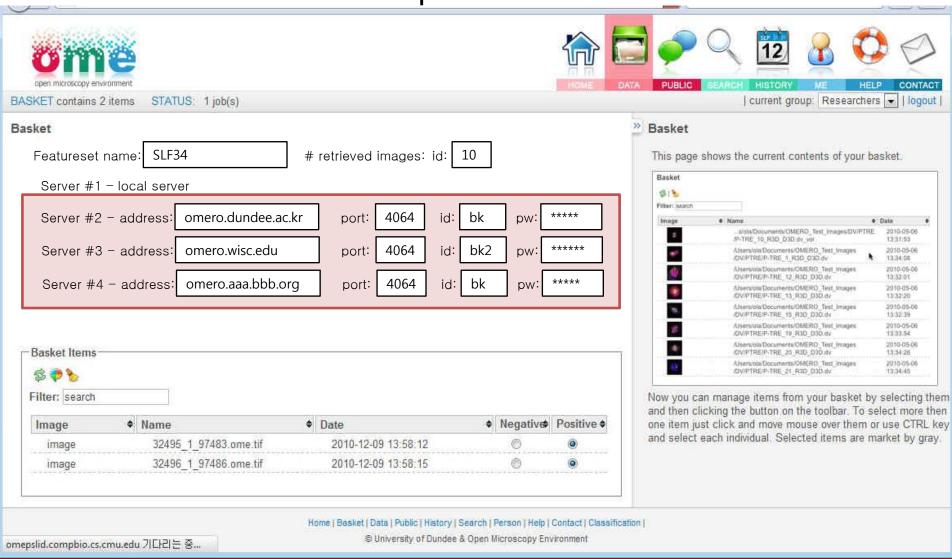
#### Sample(s)

Image	ID	Predicted Label	Name
	4854	Positive	-0003_z4c2.JPG
-51	4855	Positive	110510 Snf1-GFP Sep7-Snf4-0008_z7c2.JPG

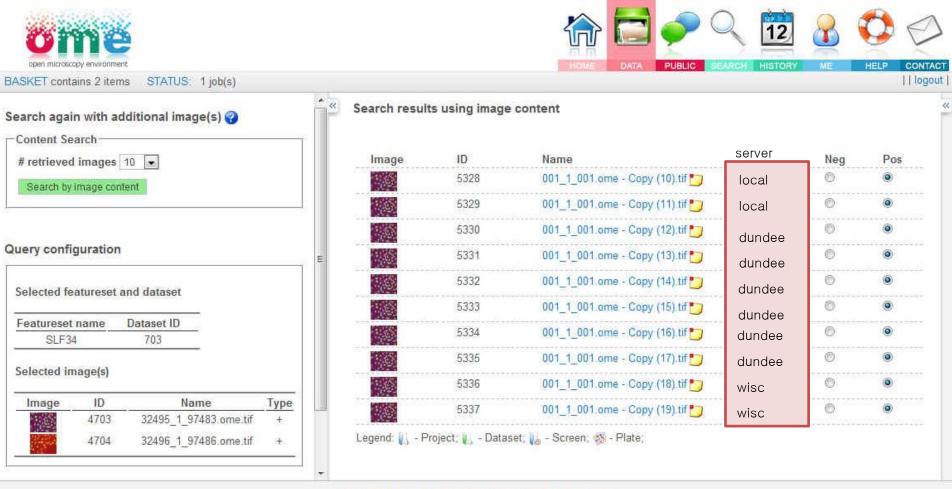


#### Future Work 1

Search across multiple servers



### Future Work 1

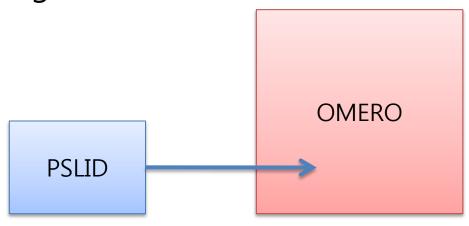


Home | Basket | Data | Public | History | Search | Person | Help | Contact

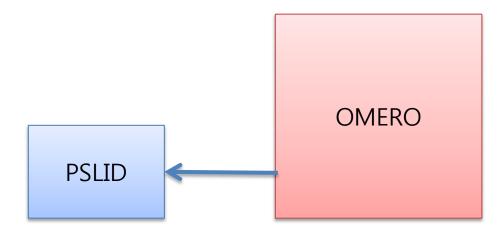
© University of Dundee & Open Microscopy Environment

### Future Work 2

Deploying all PSLID functionalities into OMERO



Export images/metadata from OMERO into PSLID



## Discussion

- Easy to use PSLID functionalities with OMERO's powerful data management
- Easy to save and retrieve additional image context/content metadata

Deploying PSLID functionalities into OMERO becomes faster

 Initial OMERO.searcher is releasing as an update for OMERO.web (version 4.2.1)

## Discussion (cont'd)

- Suggestions to the OMERO team
  - Documentation for OMERO API
  - Smart memory leakage control
    - OMERO.web stops at some point (need to restart OMERO.web)
  - More powerful Bridge
    - Can access to the DB with the administrative privilege
  - Attaching external xml files when importing images
    - Using OMERO.importer

## Acknowledgement

- Murphy Lab @CMU
  - Jennifer Bakal
  - Ivan E Cao-Berg
  - Professor Robert F. Murphy

#### OME

- Josh Moore (Glencoe Software)
- Jason Swedlow (University of Dundee and Glencoe Software)
- Kevin W Eliceiri (University of Wisconsin-Madison)
- Other OME & OMERO developers
- Funding
  - NIH grant R03 EB008516 (Eliceiri)
  - NIH grant R01 GM075205 (Murphy)
  - Lane Center for Computational Biology
- Contact
  - Prof. Murphy: <u>murphy@cmu.edu</u>
  - BK: <u>bhcho@cmu.edu</u>

#### Release

- Need an account for our OMERO server?
  - Send an email to <u>murphy@cmu.edu</u>
- Patch location
  - http://murphylab.web.cmu.edu/software/omeropslid -v0.1.tgz
  - Basic requirements
    - Linux version OMERO server (4.2.1)
    - web on production (Django and Apache)
- Any further questions
  - Send me emails about anything
  - bhcho@cmu.edu