

A framework for image analysis in plant breeding

Paul van Schayck

Paul van Schayck



Wageningen University
MSc Biotechnology
(molecular life science)

p.vanschayck@maastrichtuniversity.nl

A banner for Maastricht University featuring the university logo, the text 'Maastricht University' and 'Leading in Learning!', a photograph of a cityscape along a river, and the 'M4i' logo. Below the photo, the text 'Nanoscopy' and 'Data management' is displayed.

Maastricht University *Leading in Learning!*

M4i

Nanoscopy
Data management

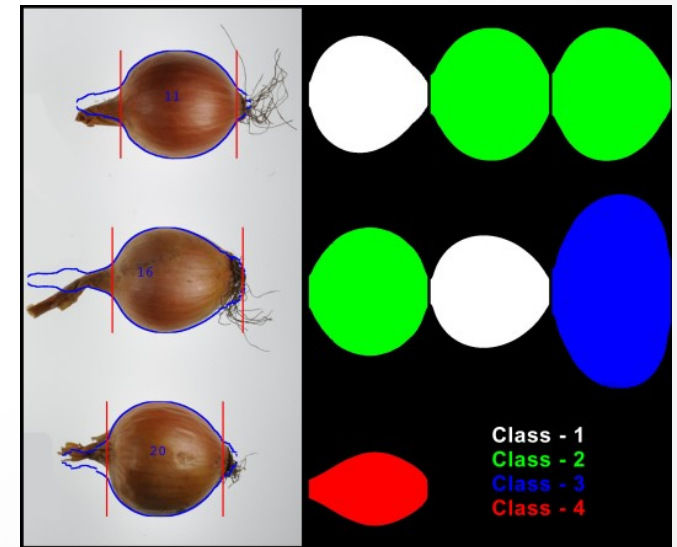
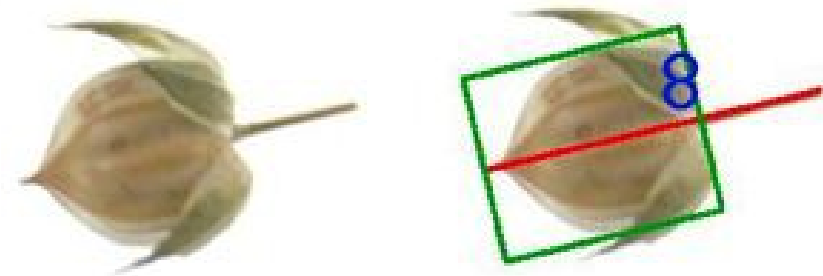
Image analysis in plant breeding

- Marker assisted breeding
- Markers:
 - Molecular
 - Biochemical
 - **Morphological**
- Shape, color and quantitative
- Tool for breeding



From melon to cucumber

Examples of image analysis used in plant breeding



Existing practice of image analysis

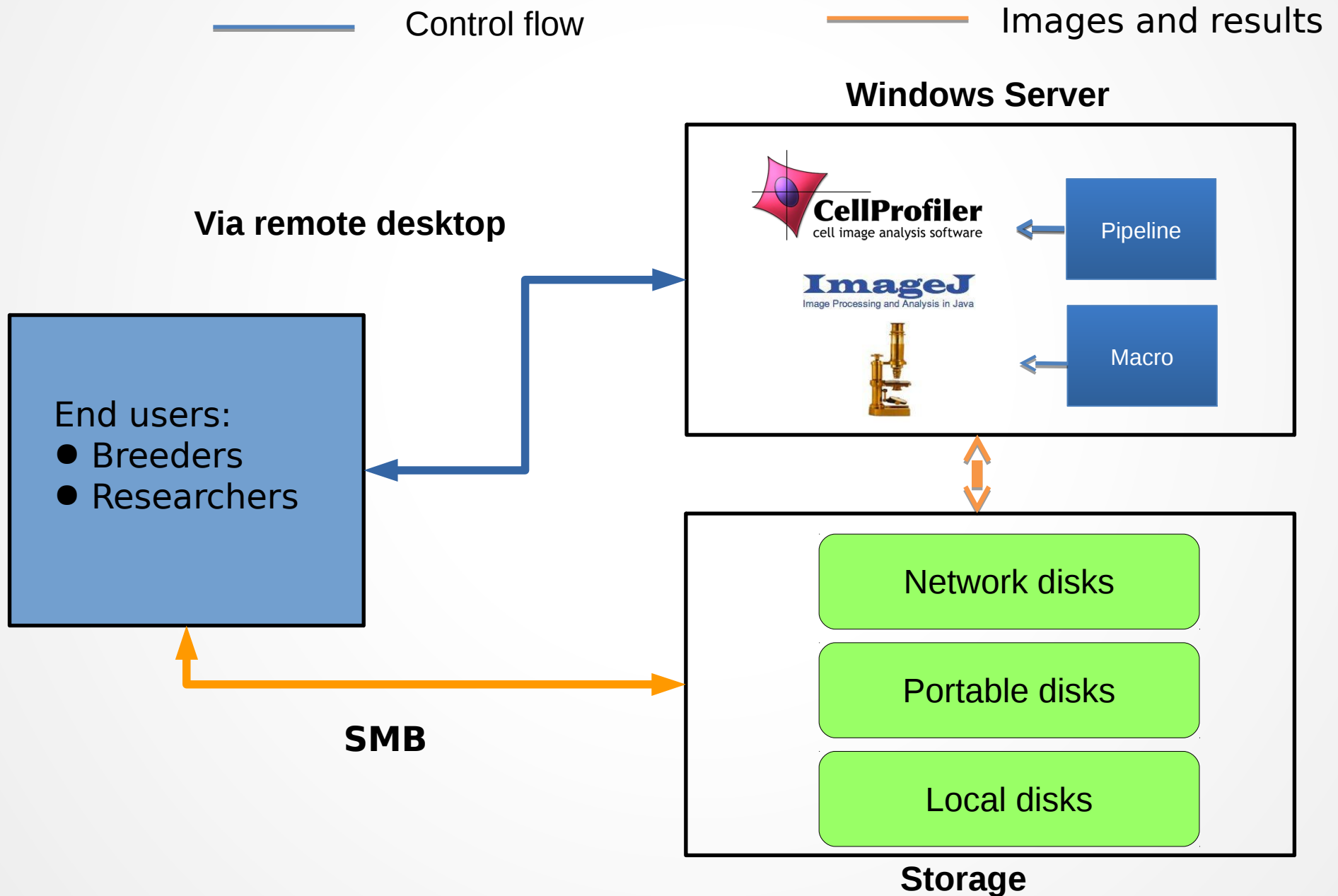
Use of popular open source tools



ImageJ
Image Processing and Analysis in Java



Current work flow of image analysis



Current challenges

- Increasing size of datasets
- Data storage decentralized
- Little structure in metadata:
 - Identification
 - Results
- Quality assurance
- User interface

Current challenges

The screenshot displays the CellProfiler 2.1.1 software interface. The main window shows a pipeline of analysis modules on the left, including LoadImages, ClassifyPixels, Morph, ApplyThreshold, IdentifyPrimaryObjects, ExpandOrShrinkObjects, Crop, ClassifyPixels, and ApplyThreshold. The central area shows a preview of an image with a red box indicating a region of interest. The bottom left shows a 'Local Docs library' with a grid of image thumbnails. The bottom right shows a 'Results' table with columns for Label, Area, Mean, StdDev, Mode, Min, and Max.

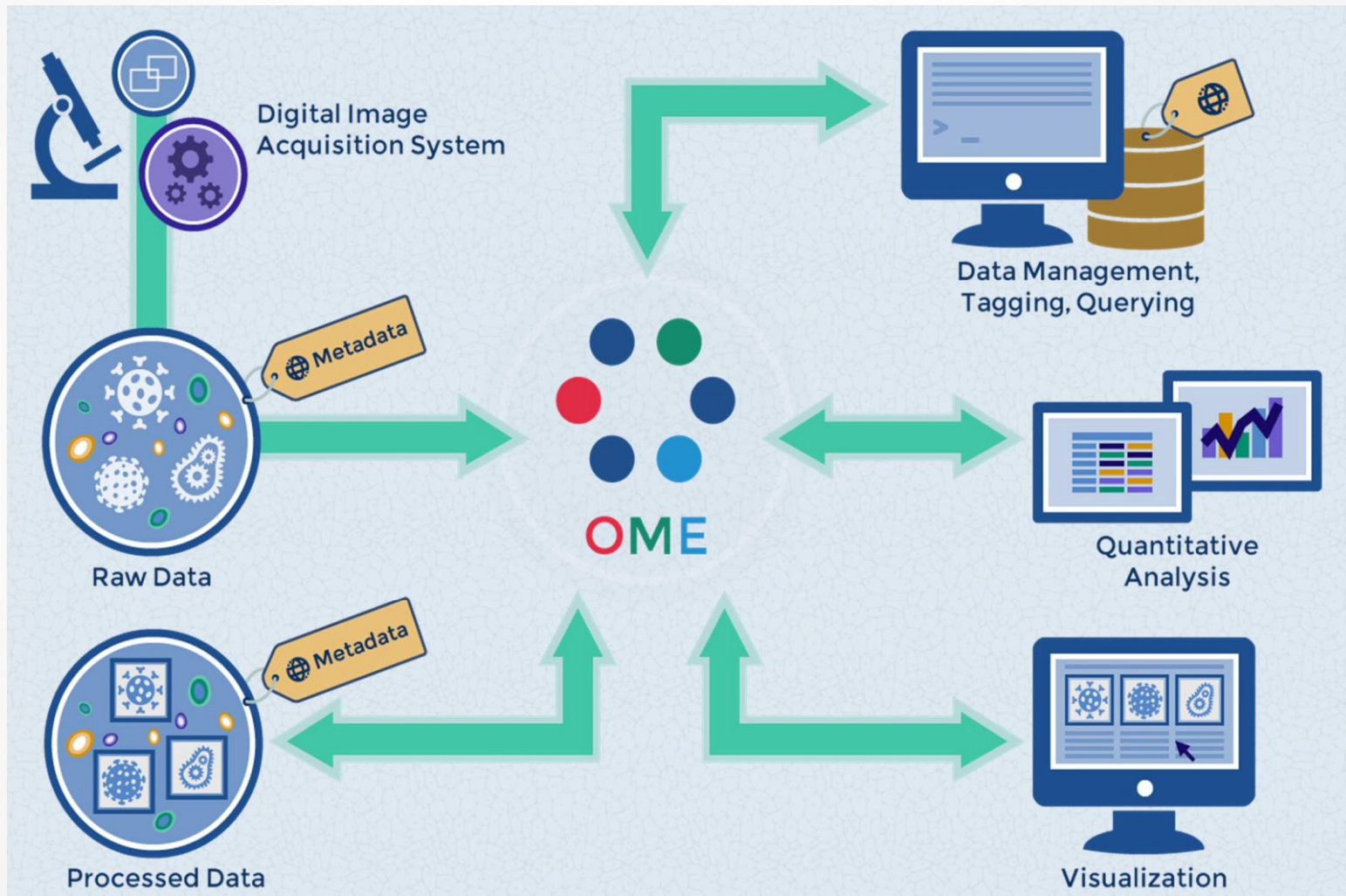
Label	Area	Mean	StdDev	Mode	Min	Max
wortel_carrot_01_cr_Para_Lab.png	1	94	0	94	94	94
wortel_carrot_01_cr_Para_Lab.png	2	95	17.0	83	83	107
wortel_carrot_01_cr_Para_Lab.png	1	87	0	87	87	87
wortel_carrot_01_cr_Para_Lab.png	2	92	12.7	83	83	101
wortel_carrot_01_cr_Para_Lab.png	1	84	0	84	84	84
wortel_carrot_01_cr_Para_Lab.png	4	85.2	3.9	82	82	90
wortel_carrot_01_cr_Para_Lab.png	1	82	0	82	82	82
wortel_carrot_01_cr_Para_Lab.png	57	131.8	32.6	83	83	196
wortel_carrot_01_cr_Para_Lab.png	4	90	6.7	82	82	98
wortel_carrot_01_cr_Para_Lab.png	8	93.6	6.6	96	83	102
wortel_carrot_01_cr_Para_Lab.png	1	85	0	85	85	85
wortel_carrot_01_cr_Para_Lab.png	12	100.5	10.8	88	84	118
wortel_carrot_01_cr_Para_Lab.png	13	96.2	7.6	90	84	107
wortel_carrot_01_cr_Para_Lab.png	9	96.2	7.9	87	87	109
wortel_carrot_01_cr_Para_Lab.png	4	91.8	5.1	86	86	98
wortel_carrot_01_cr_Para_Lab.png	14	102.8	15.1	84	84	129
wortel_carrot_01_cr_Para_Lab.png	14	104.2	15.0	91	82	125
wortel_carrot_01_cr_Para_Lab.png	7	84.7	3.4	83	82	92
wortel_carrot_01_cr_Para_Lab.png	7	97	6.2	89	89	104
wortel_carrot_01_cr_Para_Lab.png	4	85.5	2.1	83	83	88

Difficult end-user experience

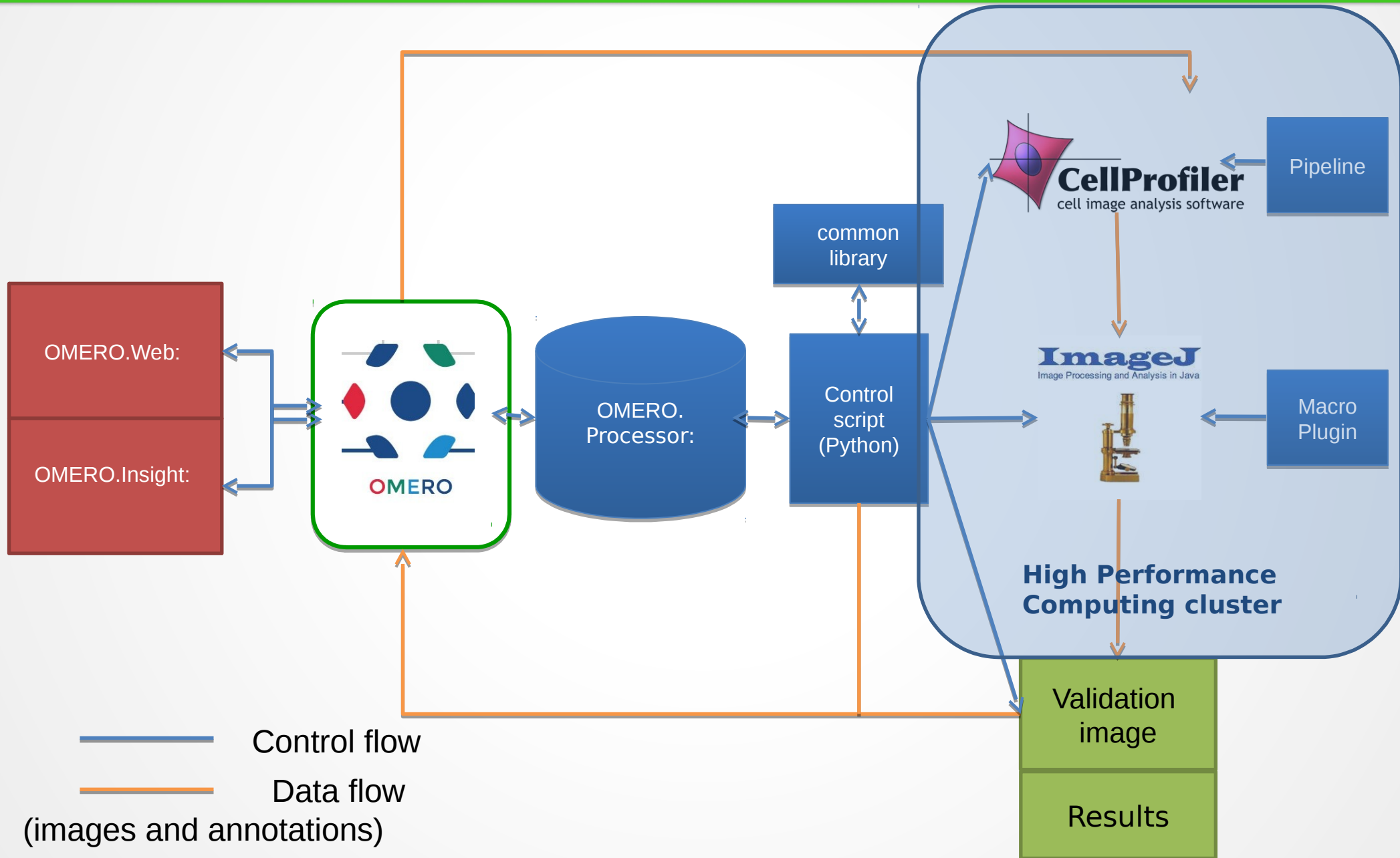
Aim of project

- Improve end-user experience
- Store metadata of results and images
- Improve data storage capabilities
- Improve performance scalability
- Improve quality assurance

Discovering OMERO



Pipeline of image analysis



Progress monitoring

The screenshot displays the OMERO Script Monitor interface. The top navigation bar includes 'Data', 'History', 'Help', 'Script Monitor', 'RZ Help', 'Importer', and 'Admin'. A search bar is located on the right. The left sidebar shows 'Script Monitor' with sub-options for 'OMERO Script Jobs' and 'HPC Jobs'. The main area is titled 'OMERO SCRIPT JOBS:' and contains a table of the latest 10 jobs. The first job, ID 413001, is highlighted in orange and is in a 'Running' state. Below the table, a detailed view for job 413001 is shown, including job details, status, and logs.

JobID	User	Started at	Finished at	Running for	Status	Message
413001	frans	2015-06-01 17:24:13	N/A	0:00:45	Running	PID:20484 Progress:0.25 Message:Extracting tiles using CellProfiler
412515	MDV	2015-05-12 15:31:38	2015-05-12 15:31:48	0:00:10	Finished	
412366	pvl	2015-05-11 09:05:57	2015-05-11 09:19:42	0:13:45	Finished	PID:15890 Progress:1.00 Message:Finished
411515	MRKO	2015-05-08 11:24:51	2015-05-08 11:34:31	0:09:40	Finished	PID:26278 Progress:1.00 Message:Finished
405649	MDV	2015-05-01 14:40:42	2015-05-01 15:20:49	0:40:07	Finished	PID:4264 Progress:1.00 Message:Finished
399694	frans	2015-04-28 11:41:08	2015-04-28 11:43:22	0:02:14	Finished	PID:23795 Progress:1.00 Message:Finished
394049	MDV	2015-04-28 10:36:31	2015-04-28 11:17:56	0:41:25	Finished	PID:11475 Progress:1.00 Message:Finished
394048	MDV	2015-04-28 10:02:41	2015-04-28 10:15:18	0:12:37	Finished	PID:6662 Progress:0.00 Message:Stopped with scripting error: Job 186483 has a failed status
386384	pbo	2015-04-22 10:51:02	2015-04-22 10:54:30	0:03:28	Finished	PID:4539 Progress:1.00 Message:Finished
386099	MDV	2015-04-20 07:39:35	2015-04-20 07:53:55	0:14:20	Finished	PID:30051 Progress:1.00 Message:Finished

Click a job for details

JOB DETAILS
Job ID: 413001
Running for: 0:00:45

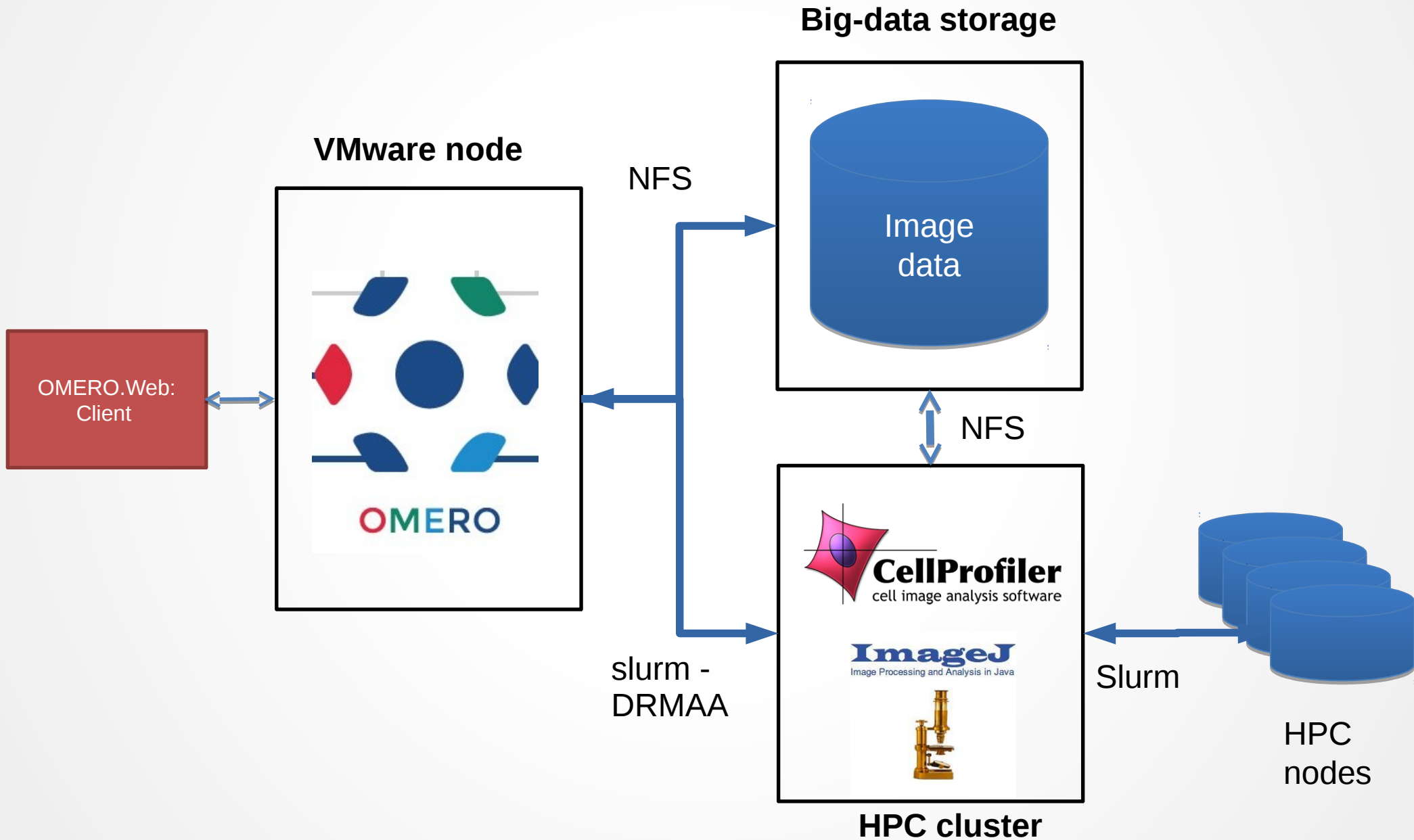
JOB STATUS:
Progress: (25) %
Latest status message: Extracting tiles using CellProfiler

JOB LOGS:
[Log entry]

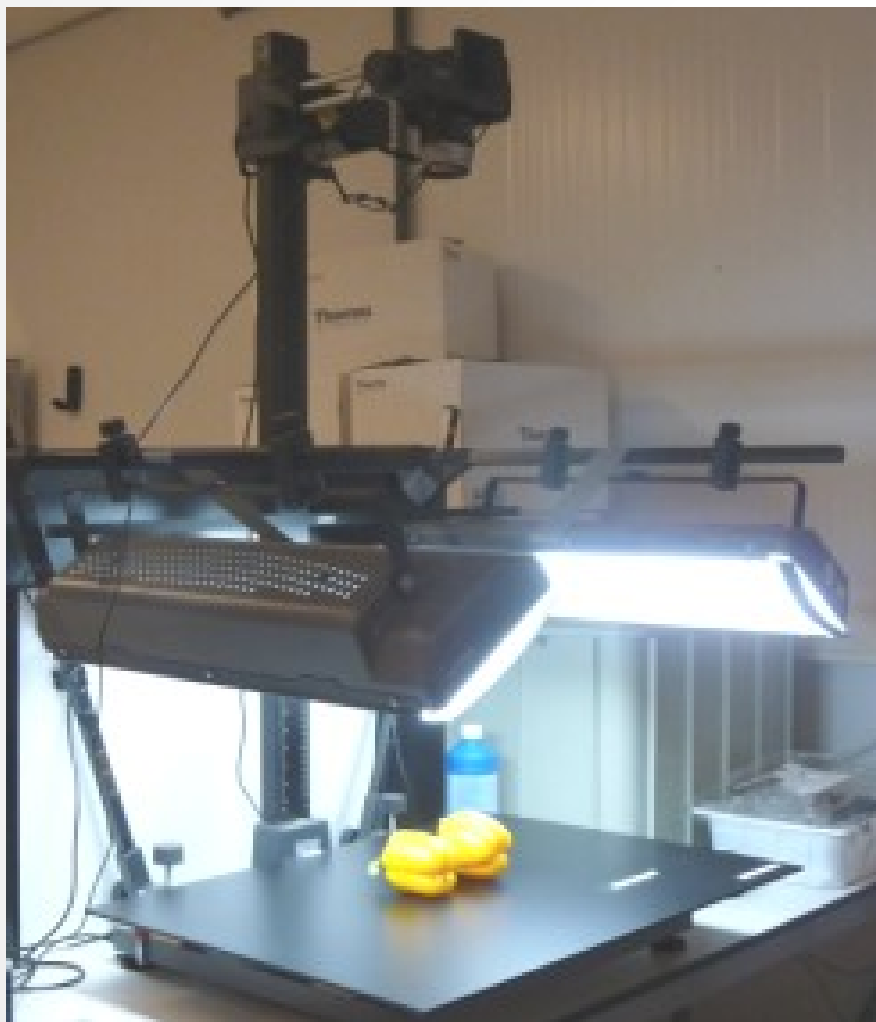
JOB CONTROL:
Cancel job: ✕

Making use of `jobHandle.setMessage()`

Infrastructure of the framework



Acquisition of images into OMERO



OMERO.
Importer



NAS storage

- Identification metadata
- Barcode scanner

Quality Assurance

- Version control (SVN)
- Doc generation (Sphinx)
- Automated testing
 - Functional and acceptance (Python nose)
- Nightly tests. Reporting to monitoring system (PRTG).
- Testing environment
- Monitoring system availability and regressions



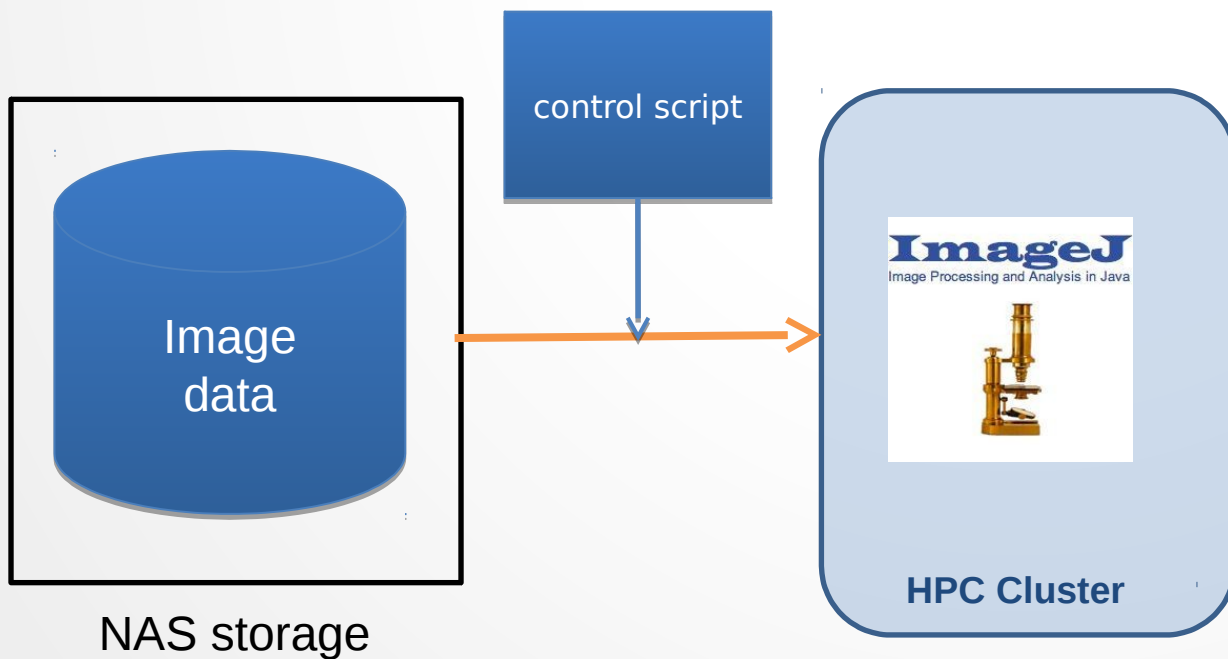
Test case: Shape parameters of carrot

- Determine different shape parameters
- Crop selection



Step 1: Acquisition and calibration

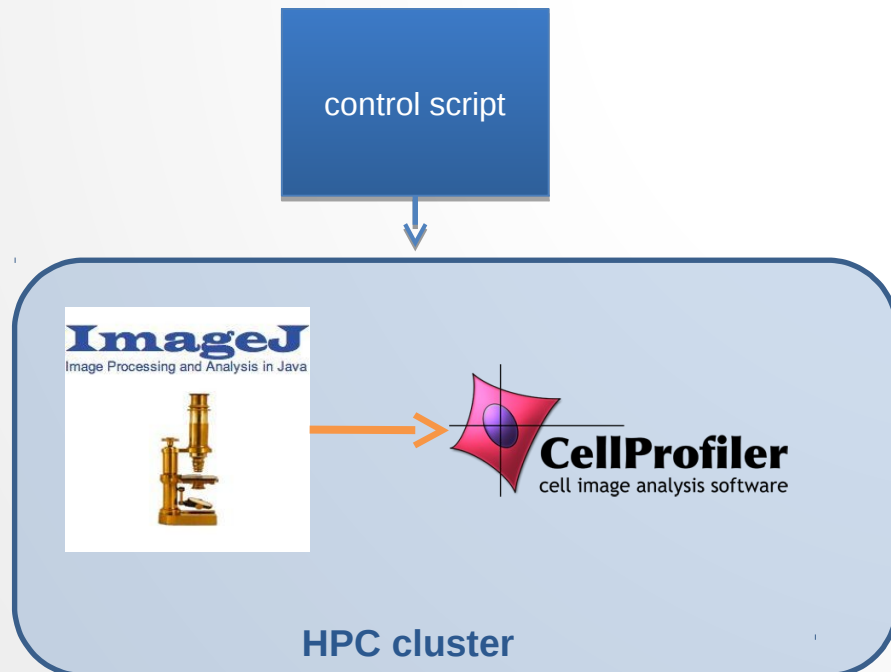
- Batch-wise import
- Color and light



Before

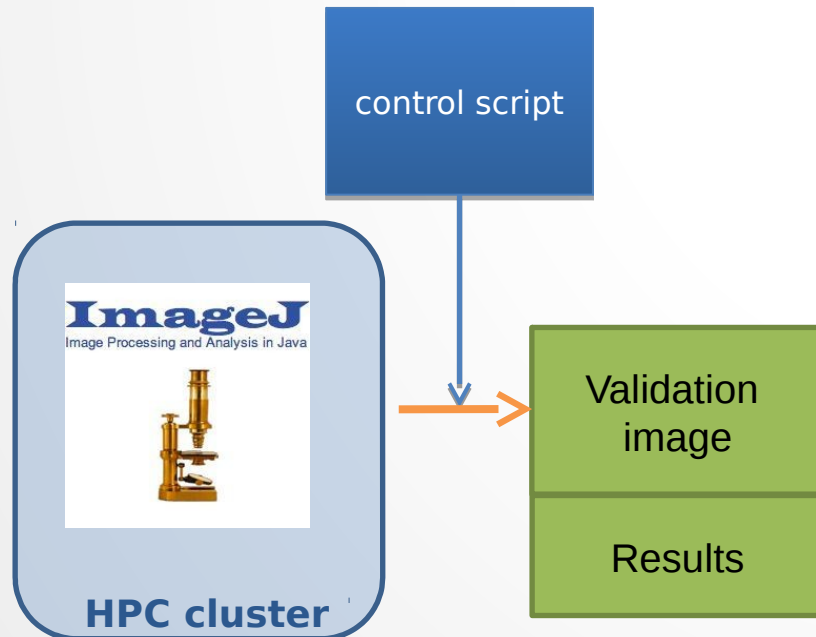
Step 2: Classification and segmentation

- Crucial step
- Time-consuming calculations



Step 3: Measurement of parameters

- Results in CSV format
- FileAnnotation



Conclusions of the carrot test case

- Further implementation:
 - Validation
 - Integration



Conclusions



Improved end-user experience

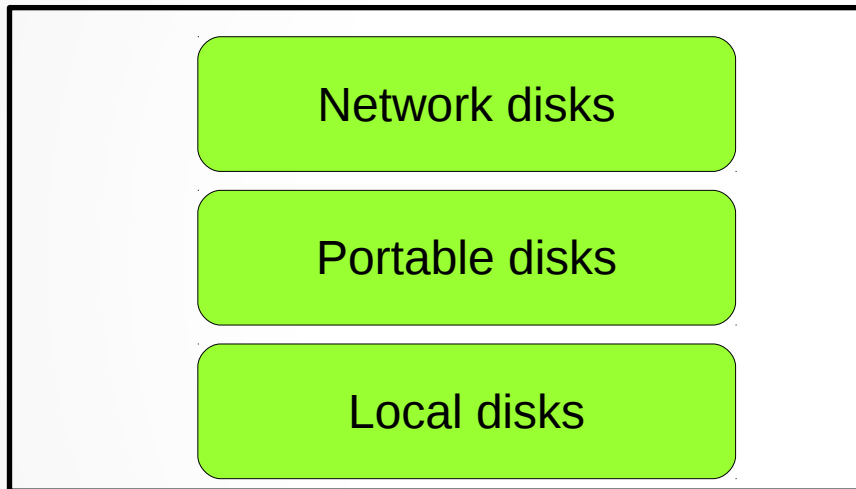
The screenshot displays the OMERO web interface. On the left, a file explorer shows a tree structure under 'Paul van Schayck' with folders like 'Carrot serie 2', 'Carrot serie 3', and 'colorCorrected'. The main area shows a grid of carrot images. A modal window titled 'Run Carrot Shape.py - Mozilla Firefox' is open, showing the script's configuration. The script is 'Carrot Shape.py' and its description is 'This tool measures different shape aspects of the carrot.' The configuration includes 'Data Type: Dataset', 'IDs: 1156', 'Ignore White Plate Check: []', 'Run Image Correction: []', 'Classifier: selection', and 'Tag Bad Input: []'. At the bottom of the modal are 'View Script', 'Cancel', and 'Run Script' buttons. On the right, a metadata panel for 'colorCorrected' shows 'DATASET ID: 1156', 'Owner: Paul van Schayck', 'Creation Date: 2014-08-25 12:55:35', 'RATING: No ratings', 'TAGS: +', 'ATTACHMENTS: +', 'OTHERS: Double: 113.25', and a 'COMMENT:' section with an 'Add Comment' button. A comment by 'Paul van Schayck' at 2014-08-25 12:55:35 is visible, detailing the 'RZ Image Correction v2.0.0' parameters: 'White plate check: False', 'Removed outliers: yes', 'Resize factor: 1.00', 'Color chart: Small A6', 'Pixel size: 113.25 (Px/cm)', 'Color correction: yes', and 'Rotation angle: 0'.

Improved end-user experience

Conclusions

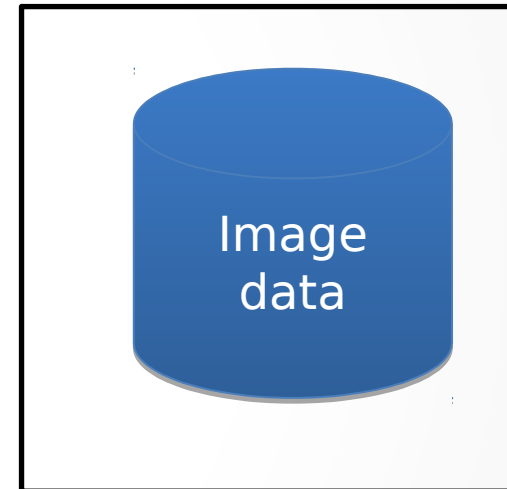
- ✓ Improved end user experience
- ✓ Centralized storage

Before



Storage

After



Centralized NAS storage

Centralized storage

Conclusions

- ✓ Improved end user experience
- ✓ Centralized storage
- ✓ Storage of metadata

The screenshot displays a file management interface for a file named "shape-measurements.csv (5.4 KB)". The interface includes sections for "Add Description", "Owner" (Paul van Schayck), "Creation Date" (2014-08-26 15:00:38), "RATING" (No ratings), "TAGS", and "ATTACHMENTS". A tooltip is visible over the file name, providing the following metadata:

- Annotation ID: 1711
- Owner: Paul van Schayck
- Linked by: Paul van Schayck
- On: 2014-08-26 15:02:22
- Description: Results
- Namespace: shapeMeasured
- File ID: 31685

Below the file list, there is an "Add Comment" button and a comment by Paul van Schayck from 2014-08-26 15:02:22, which reads: "Measured Carrot Shape", "Revision: \$Rev: 159 \$", and "Classifier: Carrot-Paul-v120.h5".

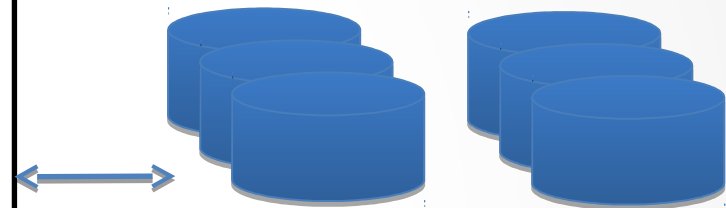
Storage of metadata

Conclusions

- ✓ Improved end user experience
- ✓ Centralized storage
- ✓ Storage of metadata
- ✓ Performance scalability



NAS storage



HPC nodes

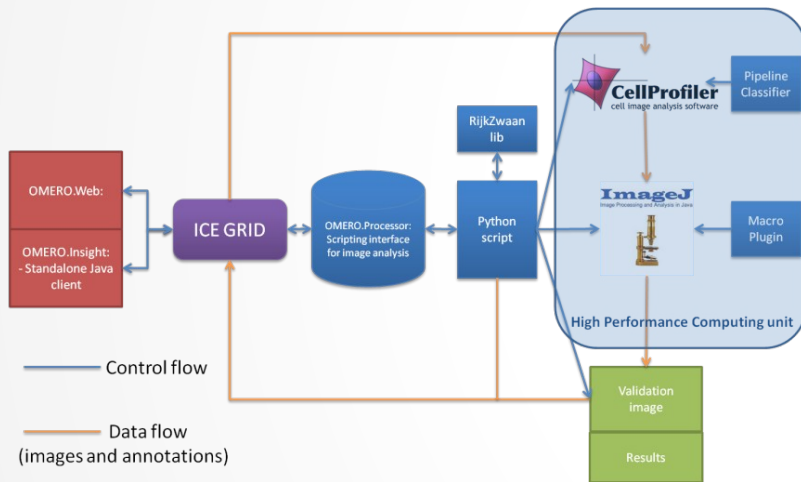
HPC cluster

Performance scalability

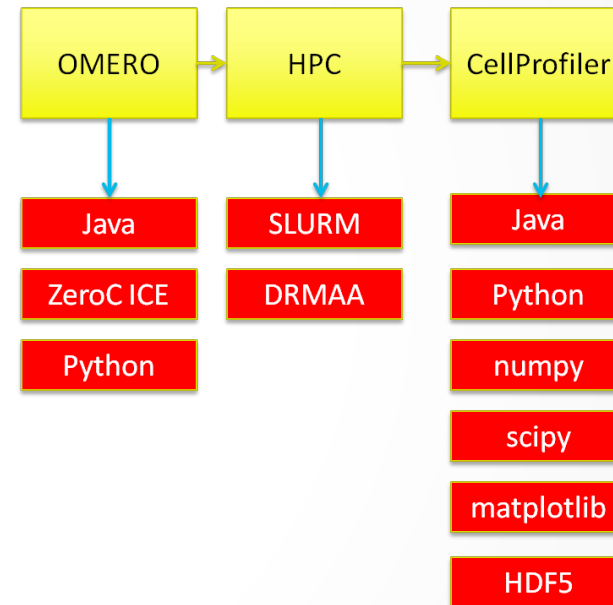
Conclusions

- ✓ Improved end user experience
- ✓ Centralized storage
- ✓ Storage of metadata
- ✓ Performance scalability
- ✗ Complexity for developers
- ✗ Software dependencies

At the cost of



More steps required by developers



Software dependencies

Conclusions

- ✓ Improved end user experience
- ✓ Centralized storage
- ✓ Storage of metadata
- ✓ Performance scalability
- ✗ Complexity for developers
- ✗ Software dependencies
- ✓ Quality assurance



Quality assurance

Acknowledgements



Raimond Ravelli
Peter Peters



Jan-Willem Borst