cytomine

a web platform for collaborative analysis of multi-gigapixel images with machine learning

Raphaël MAREE – Renaud HOYOUX – Grégoire VINCKE

FOSDEM’16
Brussels / 30 & 31 January 2016
Biomedical research and routine pathology

Heavily rely on **semantic annotation & quantification** of tissue slides.
Scientists and pathologist’s daily work
Scientists and pathologist’s daily work

Slide quantifications (annotations?) are usually

- Performed **manually**
- Performed within tissue **subregions** in **small sample groups**
- Created by **isolated experts**
- Stored **locally** (generally no backups)
- **Proprietary** formats

**Hardly repeatable**
Digital histology and pathology
Digital histology and pathology

Multi-gigapixel images

- 15 x 15 mm
- ≈ 0, 20\(\mu\)m/px
- 100K x 100K pixels
- +100Mb → +100Gb
- pyramidal structure
Slide quantifications and annotations could be

- Performed automatically
- Performed in entire sample in large groups
- Shared between experts
- Stored on a cloud
- In Open formats

Easily repeatable
2010 : start of cytominé project
2010: start of cytömine project

Sign in to Cytomine

- Username
- Password
- Remember me
- Sign in

Forgot your username or your password?

Authentification
Roles, permissions, LDAP
<table>
<thead>
<tr>
<th>SHORTCUTS</th>
<th>PROJECTS</th>
<th>IMAGES</th>
<th>ANNOTATIONS</th>
<th>REVIEWED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go to project</td>
<td>3</td>
<td>266</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Go to image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STATISTICS</th>
</tr>
</thead>
</table>
| 3 PROJECTS
Now |
| 266 IMAGES
Now |
| 0 ANNOTATIONS
Now |
| 0 REVIEWED
Now |

<table>
<thead>
<tr>
<th>LAST OPENED IMAGES</th>
</tr>
</thead>
</table>
| TN201190 ISH-BR...
2015-11-19 16h10 |
| 015-TS_13C09246...
2015-11-19 16h10 |
| PGP POUMON PBS...
2015-11-19 16h07 |

<table>
<thead>
<tr>
<th>LAST OPENED PROJECTS</th>
</tr>
</thead>
</table>
| ULG-GENEHUM-BREAST-BRCA1-RNASCOPE
2015-11-19 16h10 |
| ULB-ANAPATH-THYROID-FNAB
2015-11-19 16h10 |
| ULG-LBTD-PGP
2015-11-19 16h03 |

<table>
<thead>
<tr>
<th>YOUR ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Data Available.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROJECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIEW ALL</td>
</tr>
</tbody>
</table>
Important notes!

The maximum file size (per file) for uploads is **100 GB**
Only image files (JPEG, PNG, BMP, TIF, TIFF, JP2K, SVS, SVS J2K, SCN, NDPI, VMS, MRXS, VSI, DICOM) are allowed
VMS and MRXS **must be zipped** (the VMS/MRXS/VSI file and their nested files). One zip per image.
A Zip file containing multiple image is allowed (except for VMS and MRXS)
You can **drag & drop** files from your desktop on the 'Add files...' button with Google Chrome, Mozilla Firefox and Apple Safari.
You are also able to link images with project manually after upload is done.

**Storage**

- gvincke storage

**Project**

- Link automatically uploaded file with selected project
  - ULG-GENEHUM-BREAST-BRCA1-RNASCOPE

**Files**

- Add files...
- Start upload
- Cancel upload
- Hide selected

---

**Direct upload**

**Several file formats supported**

Showing 0 to 0 of 0 entries
<table>
<thead>
<tr>
<th>ID</th>
<th>Preview</th>
<th>Name</th>
<th>Width (px)</th>
<th>Height (px)</th>
<th>Magnitude</th>
<th>Resolution (µm/px)</th>
<th>User an.</th>
<th>Algo an.</th>
<th>Valid an.</th>
<th>Vendor</th>
<th>Created</th>
<th>Status</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>7889452</td>
<td><img src="https://example.com/image1.png" alt="Image" /></td>
<td>PGP POUMON PBS 3 21 - 2012-08-08 11.07.24.jp2</td>
<td>34816</td>
<td>29440</td>
<td>40 X</td>
<td>0.650</td>
<td>13</td>
<td>614</td>
<td>100</td>
<td>Undefined</td>
<td>2012-10-18 10h50</td>
<td>In review</td>
<td>Explore</td>
</tr>
<tr>
<td>7889358</td>
<td><img src="https://example.com/image2.png" alt="Image" /></td>
<td>PGP POUMON PGP7 11 - 2012-08-07 12.07.00.jp2</td>
<td>36864</td>
<td>28416</td>
<td>40 X</td>
<td>0.650</td>
<td>3</td>
<td>130</td>
<td>119</td>
<td>Undefined</td>
<td>2012-10-18 10h50</td>
<td>In review</td>
<td>Explore</td>
</tr>
<tr>
<td>7889229</td>
<td><img src="https://example.com/image3.png" alt="Image" /></td>
<td>PGP POUMON PGP7 41 - 2012-08-07 12.13.50.jp2</td>
<td>36864</td>
<td>28416</td>
<td>40 X</td>
<td>0.650</td>
<td>1</td>
<td>142</td>
<td>142</td>
<td>Undefined</td>
<td>2012-10-18 10h50</td>
<td>In review</td>
<td>Explore</td>
</tr>
<tr>
<td>7889190</td>
<td><img src="https://example.com/image4.png" alt="Image" /></td>
<td>PGP POUMON PGP7 21 - 2012-08-07 12.09.19.jp2</td>
<td>34816</td>
<td>28928</td>
<td>40 X</td>
<td>0.650</td>
<td>6</td>
<td>122</td>
<td>124</td>
<td>Undefined</td>
<td>2012-10-18 10h50</td>
<td>Reviewed</td>
<td>Explore</td>
</tr>
<tr>
<td>7889180</td>
<td><img src="https://example.com/image5.png" alt="Image" /></td>
<td>PGP POUMON PGP7 31 - 2012-08-07 12.11.34.jp2</td>
<td>36864</td>
<td>26624</td>
<td>40 X</td>
<td>0.650</td>
<td>7</td>
<td>271</td>
<td>213</td>
<td>Undefined</td>
<td>2012-10-18 10h49</td>
<td>Reviewed</td>
<td>Explore</td>
</tr>
<tr>
<td>7889170</td>
<td><img src="https://example.com/image6.png" alt="Image" /></td>
<td>PGP POUMON PGP7 41 - 2012-08-07 12.13.50.jp2</td>
<td>36864</td>
<td>32256</td>
<td>40 X</td>
<td>0.650</td>
<td>4</td>
<td>169</td>
<td>151</td>
<td>Undefined</td>
<td>2012-10-18 10h49</td>
<td>Reviewed</td>
<td>Explore</td>
</tr>
</tbody>
</table>
Rich web application

OpenStreetMap-like visualization (tiles)

No extra software to install
Rich web application

Collaborative and semantic annotation of regions of interest
Rich web application

Collaborative and semantic annotation of regions of interest
Manual or semi-automatic analysis

- Generic algorithms of machine learning
- Manual correction and validation (proofreading)
But also

- **Search** similar annotations
- **Sharing** of images and annotations (URL, email)
- **Live broadcasting**
cytomine-based research

a: H&E mice lung cancer research (D. Cataldo's lab, GIGA-Research) @ ISBI 2014.
b: IHC mice lung cancer research (P. Martinive's lab, GIGA-Research) @ Oncotarget 2015.
c: Immunofluorescent mouse ear sponge assays in tumor angiogenesis. @ J. Pathol. 2015.
d: H&E Chondrostoma nasus sexual maturation research (Gennette's lab, CEFRA).
e: in situ hybridization assays in human breast cancer research (C. Josse's lab, GIGA-Research) @ Submitted...
f: Human thyroid cytology (I. Salmon's lab. ULB Anatomical Pathology Department).
g: Danio rerio embryo development (M. Muller's lab, GIGA-Research) @ PLOS ONE 2015.
h: Danio rerio toxicology research (M. Muller's lab, GIGA-Research) @ PLOS ONE 2015.
i: IHC renal ischemia/reperfusion research (F. Jouret's lab, GIGA-Research) @ Am J Transl Res. 2015.
j: IHC in melanoma microenvironment research (P. Quatresooz's lab, GIGA-Research).
Research with cytominé at ULg

- 175 users
- 300 projects
- 20,000 images ($\approx$ 6To)
- 500,000 annotations ($\approx$ 160Go)

a: H&E mice lung cancer research (D. Cataldo's lab, GIGA-Research) @ ISBI 2014.
b: IHC mice lung cancer research (P. Martinive's lab, GIGA-Research) @ Oncotarget 2015.
c: Immunofluorescent mouse ear sponge assays in tumor angiogenesis. @ J. Pathol. 2015.
d: H&E Chondrostoma nasus sexual maturation research (Gennette's lab, CEFRA).
e: In situ hybridization assays in human breast cancer research (C. Josse's lab, GIGA-Research) @ Submitted...
f: Human thyroid cytology (I. Salmon's lab, ULB Anatomical Pathology Department).
g: Danio rerio embryo development (M. Muller's lab, GIGA-Research) @ PLOS ONE 2015.
h: Danio rerio toxicology research (M. Muller's lab, GIGA-Research) @ PLOS ONE 2015.
i: IHC renal ischemia/reperfusion research (F. Joural's lab, GIGA-Research) @ Am J Transl Res. 2015.
j: IHC in melanoma microenvironment research (P. Quatresooz's lab, GIGA-Research).
Research with cytomine at ULg

- 175 users
- 300 projects
- 20,000 images ($\approx 6$ To)
- 500,000 annotations ($\approx 160$ Go)

Teaching with cytomine at l’ULg

- 4,000 users
- 50 projects
- 2,000 images ($\approx 2$ To)
- 200,000 annotations ($\approx 30$ Go)

\[ a: H\&E \]  
\[ b: IHC \]  
\[ c: Immunohistochemistry \]  
\[ d: H\&E \]  
\[ e: In situ hybridization \]  
\[ f: Human \]  
\[ g: Danio rerio \]  
\[ h: Danio rerio \]  
\[ i: IHC renal ischemia/reperfusion research (F. Jouret’s lab, GIGA-Research) @ Am J Transl Res. 2015. \]  
\[ j: IHC in melanoma microenvironment research (P. Quatresooz’s lab, GIGA-Research). \]  
\[ k: H\&E in human breast cancer research (R. Longuespée, GIGA-Research). \]
General architecture
Detailed architecture
Included softwares and libraries

- HTML5
- JavaScript
- Jquery
- Backbone.js
- AngularJS
- Openlayers
- Cytomine-WebUI
- Scikit-learn
- OpenCV
- Redis
- Java
- Python
- Cytomine-DataMining
- REST API HTTP
- NGINX
- Message broker
- RabbitMQ
- REST API HTTP
- NGINX
- Grails
- Groovy
- Java
- Tomcat
- Postgis
- Cytomine-Core
- MongoDB
- GlusterFS
- File system
- Docker
- Cytomine Install/Deploy
- Node.js
- MongoDB
- Cytomine Monitoring
- Cytomine
- Collaborative analysis of gigapixel images
- www.cytomine.be
Docker architecture
## Minimal setups

<table>
<thead>
<tr>
<th></th>
<th>Minimal Some users at the same time and some data</th>
<th>Recommend More users at the same time and more data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CPU</strong></td>
<td>8 core 64 bits</td>
<td>16 core 64 bits</td>
</tr>
<tr>
<td><strong>MEMORY</strong></td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td><strong>Disk</strong></td>
<td>HDD 100GO</td>
<td>HDD with more than 2TO</td>
</tr>
<tr>
<td><strong>OS</strong></td>
<td>Ubuntu / CentOS</td>
<td>Ubuntu</td>
</tr>
<tr>
<td><strong>DNS</strong></td>
<td></td>
<td>1 domain</td>
</tr>
</tbody>
</table>

**CPU**

Cytomine is a set of multithread servers/tools. If you want to support more users, you need to have more core.

**MEMORY**

If you want to store a lot of data or if you have a lot of users, you will need a larger amount of memory.

**Disk**

On average, 1 image = 1 GO and data for 1 image = 50 MB (annotations, activities, ...).
If you plan to store 1000 images => 1 To + 50 GO

⚠️ **This does not include backup space!**

**OS**

Docker needs to be installed
Port 80 needs to be opened (by default, it's open)
You need to have root (sudo) access to the server.
<table>
<thead>
<tr>
<th>Server Count</th>
<th>Server 1</th>
<th>Server 2</th>
<th>Server 3</th>
<th>Server 4</th>
<th>Server 5</th>
<th>Server 6</th>
<th>Server 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 server</td>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 servers</td>
<td>Core PostgreSQL</td>
<td>IMS</td>
<td>Retrieval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MongoDB Retrieval</td>
<td>IIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 servers</td>
<td>Core PostgreSQL</td>
<td>IMS</td>
<td>Retrieval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MongoDB</td>
<td>IIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 servers</td>
<td>Core</td>
<td>IMS</td>
<td>Retrieval</td>
<td>PostgreSQL</td>
<td>MongoDB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IIP</td>
<td></td>
<td></td>
<td>MongoDB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 servers</td>
<td>Core</td>
<td>IMS</td>
<td>Retrieval</td>
<td>PostgreSQL</td>
<td>MongoDB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 servers</td>
<td>Core</td>
<td>IMS 1 IIP 1</td>
<td>IMS 2 IIP 2</td>
<td>Retrieval</td>
<td>PostgreSQL</td>
<td>MongoDB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;= 7 servers</td>
<td>Core</td>
<td>IMS 1 IIP 1</td>
<td>...</td>
<td>IMS N IIP N</td>
<td>Retrieval</td>
<td>PostgreSQL</td>
<td>MongoDB</td>
</tr>
</tbody>
</table>
Welcome to the Cytomine main documentation.

It is recommended to start by reading the main Cytomine overview page for a short introduction on Cytomine main concepts.

Then the documentation is organized in four main sections below for administrators/operators, developers, data/computer scientists, and end-users.

For testing the Cytomine platform, you can either:

1. Install a Cytomine instance at our own site, or

<table>
<thead>
<tr>
<th>Operations</th>
<th>Developments</th>
<th>Data &amp; algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="gear.png" alt="Gear" /></td>
<td><img src="lightbulb.png" alt="Lightbulb" /></td>
<td><img src="play.png" alt="Play" /></td>
</tr>
</tbody>
</table>

**Documentation for administrators/operators:**
- How to install a Cytomine server

**Documentation for software developers:**
- How to install a Cytomine server

**Documentation for data/computer scientists:**
- How to install the Cytomine Data
Acknowledgments

Systems and Modeling GIGA-Research / Montefiore Institute (ULG):

Collaborators at University of Liège (ULg):
GIGA: Didier Cataldo, Natacha Rocks, Fabienne Perin, Christine Fink.
IFRES: Grégoire Vincke, Dominique Verpoorten.
Histologie: Pascale Quatresooz, Valérie Defaweux, le groupe MorphoTIC.
CRIFA: Brigitte Denis, Céline Snoeck.
Students: Julien Confetti, Pierre Ansen, Olivier Caubo, Antoine Deblire.

Other collaborators:
Université Libre de Bruxelles (ULB): Isabelle Salmon, Caroline Degand, Xavier Moles Lopez, Nicky d’Haene.
Medical University Graz: Philippe Kainz.
University of Namur: Patsy Renard, Eric Depiereux.

Research grants of the Wallonia (DGO6):
CYTOMINE (2010-2016) n°1017072
SMASH (2012-2014) n°1217606
HISTOWEB (2014-2017) n°1318185

More information on: http://cytomine.be info@cytomine.be @cytomine