OMERO: source for image analysis data on the cloud

NEUBIAS 2023

OME Team
Motivation

Have you ever analyzed large, heterogeneous image data?

Have you ever wanted to find existing data to reuse?

https://www.xxxx.images
Analysis landscape - local

Environment

Image analysis
Software, dependencies, Jupyter…

Data

Environment

Image analysis
Software, dependencies, Jupyter…

Data
Analysis landscape – remote data

Environment

Image analysis
Software, dependencies, Jupyter…

Data

Results

Data

NEUBIAS 2023
Analysis landscape – remote

Environment

Data

INTERNET
Analysis landscape – remote

Environment

CO & binder

Data

Internet

Results

Data
Analysis landscape – remote
OMERO is a server with clients
Metadata

- Stored in relational database
- Acquisition metadata
- User-added annotations
- ROIs / Segmentations
OMERO.tables

- HDF5 tables for analysis results
- Flexible columns
OMERO API

Python | Java | Matlab | R
Integration

- TensorFlow
- OME RO
- CellProfiler
- MATLAB
- PyTorch
- R
- Scikit-image
- Learn
- VisionPharm
- Pandas
Analysis using OMERO API

**Data stored in IDR as .tiff with OMERO ROIs**

Binary image

Segmentation labels

**Python-based analysis environment**

**Notebook with StarDist**

---

**Load Image with labels from IDR, analyze using StarDist and compare results**

The notebook shows how to load an IDR image with labels.

The image is referenced in the paper "Neosys: a novel method for accurate nuclear segmentation in 3D" published August 2019 in PLOS Biology: https://doi.org/10.1371/journal.pbio.3000089 and can be viewed online in the Image Data Resource.

In this notebook, the image is loaded together with the labels and analyzed using StarDist. The StarDist analysis produces a segmentation, which is then viewed side-by-side with the original segmentations produced by the authors of the paper obtained via the loaded labels.

---

In [0]: from omERO import masks

In [1]: roi_service = omero.getRoiService()
result = roi_service.find

dims = (image.getSize3D()),
shapes = []
for roi in result.ROI:
    shapes.appendROI(roi)

saver = masks.MaskSaver()
labels, fillLabels, proper

Segmentation labels from original

Segmented by StarDist in the Notebook

---

Notebook segments images using StarDist and produces new segmentation labels

Notebook compares the labels
The Image Data Resource (IDR) is a public repository of image datasets from published scientific studies, where the community can submit, search and access high-quality bio-image data.

Choose search field (optional)

2342 Images matched Gene Symbol: pax6
991 Images matched Antibody: pax6
9 Images matched HGNC Gene Symbol: pax6
6 Images matched Nearest Human Homologue: pax6
4 Images matched Original Gene Target: pax6
3 Images matched Gene Symbol Synonyms: eyeless tumor-head-53 l(4)102cdh ey eyel pax-6 ok107 ey/pax6 eye l(4)33 pax6 ey dpax-6 l(4)102cdr pax6 cg1464

118 Studies
13,666,719 Images
353 TB
The Image Data Resource (IDR)

- Public access
- **Reference datasets** - complete datasets containing molecular and functional annotations, associated with an existing or upcoming publication.
- **Study integration** - integrating studies or datasets with other datasets via genes, compounds or phenotypes.
- Curated metadata
- Cloud re-analysis
Which diabetes related genes are expressed in the pancreas?

**HumanMine**
An integrated database of Homo sapiens genomic data

[https://www.humanmine.org/humanmine](https://www.humanmine.org/humanmine)

**IDR MULTIOMICS API**

Genes found:

Images linked to gene PDX1:
- CS16
- CS21
- 9PCW
- 15PCW

Developmental stage

[https://workflowhub.eu/workflows/242](https://workflowhub.eu/workflows/242)
## Access possibilities of images

<table>
<thead>
<tr>
<th>Access via S3</th>
<th>IDR API Access</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No, only per file</td>
</tr>
</tbody>
</table>

- **Lazy loading**: No
  - **OME-Zarr Access via S3**: Yes. Use Dask collections: `da.from_zarr(endpoint_url)`

- **Easily analyze in parallel**: No, depends on file format which may require a translation library.
  - **Difficult due to the transfer protocol used (zeroc-ice)**
  - **Yes. Use Dask schedulers**: `dask.delayed(analyze)(t, c, z)`

---

Adapted from Table 3, Josh Moore et. al.: OME-Zarr: a cloud-optimized bioimaging file format with international community support, bioRxiv 2023.02.17.528834
What is OME-NGFF?

OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies

Josh Moore, Chris Allan, Sébastien Besson, Jean-Marie Burel, Erin Diel, David Gault, Kevin Kozlowski, Dominik Lindner, Melissa Linkert, Trevor Manz, Will Moore, Constantin Pape, Christian Tischer & Jason R. Swedlow

Nature Methods 18, 1496–1498 (2021) | Cite this article

https://ngff.openmicroscopy.org
https://ngff.openmicroscopy.org/data
OME-NGFF uses Zarr

Multiscale image representation

Segmented mask representation

 TIFF
OME-XML

HDF5
Binary HDF5 files contain a hierarchy of ND arrays with associated metadata.

Zarr
NGFFs folders store the same information as HDF5s spread across many, smaller files.

Binary TIFF files contain a metadata block and a linear sequence of 2D tiles of varying size.

Equivalent metadata
chunk = file = object
Remote cloud storage

S3 Bucket
This workshop - OMERO

Environment

Data

Results

Data

INTERNET
This workshop – OME-NGFF

Environment

CO™ binder

Data

S3 Bucket

OME-NGFF

Results
Suggested workflow

- Find data of interest in IDR
- Set-up analysis environment
- Run examples to access IDR data
- Edit the code to add your own analysis

YouTube openmicroscopy channel:
  1. Exploring IDR
  2. IDR Analytical workflows playlists
Materials

- Link to the walkthrough pdf
  https://downloads.openmicroscopy.org/presentations/2023/Neubias

- Videos
  - openmicroscopy YouTube channel
  - IDR homepage
  - I2K video (minute 28) for NGFF
Thank you
Parallel analysis

- Natively scales Python
- Advanced parallelism for analytics
- See https://dask.org/index.html
This workshop – IDR

Environment

Data

Results

Data

CO  binder

IDR
This workshop – local env

Environment

Image analysis
Software, dependencies
Jupyter…

Data

S3 Bucket

OME-NGFF
This workshop – local env

Environment

Image analysis
Software, dependencies,
Jupyter…

OMERO

Data

Results

Data
This workshop – local env

Environment

Image analysis
Software, dependencies
Jupyter…

Data

IDR

Data