TOWARDS OPEN DATA IN CELL MIGRATION RESEARCH

@pcmasuzzo
paola.masuzzo@vib-ugent.be
STATISTICS SAY I HAVE

60 SECONDS
to capture your attention
So what's your idea of a perfect date?

YYYY-MM-DD

I find other formats a bit confusing

https://t.co/UKmqZQvbHK - tweet by @myusuf3
Diversity and complexity of cell migration data

CMSO: community standards for cell migration research

biotrack: a standard format for cell tracking data
Diversity and complexity of cell migration data

CMSO: community standards for cell migration research

biotrack: a standard format for cell tracking data
Cell migration is involved in many processes

Cell migration is involved in many processes
Cells are able to move following many and diverse strategies

**COLLECTIVE MIGRATION**
- Multicellular streaming
- Collective migration

**INDIVIDUAL MIGRATION**
- Amoeboid (blebs)
- Amoeboid (pseudopodia, filopodia)
- Mesenchymal

Adapted from Friedl, J. Exp. Med., 2010
Many study models are used to elucidate the complex mechanisms behind cell movement.
Many study models are used to elucidate the complex mechanisms behind cell movement.
An imaging-based cell migration experimental workflow has diverse steps.
These steps produce rich data sets represented in various file formats.
Breaking down technical barriers means facilitating data sharing

Image credit: Ainsley Seago, PLoS Biology
Data sharing in turn means more analytical possibilities

1. Data and metadata generation
2. Analysis and interpretation
3. Standardization
4. Submission to repository
5. Global data dissemination
6. Meta-scale and multiscale algorithms

https://multimot.org/
Masuzzo, Trends in Cell Biology, 2015
The first step towards data sharing is standardization

Masuzzo, Trends in Cell Biology, 2016
Diversity and complexity of cell migration data

CMSO: community standards for cell migration research

biotrails: a standard format for cell tracking data
Different information units require specific standardization tasks

- set up experiment: **annotate** investigation, study, assay
- convert raw data & metadata into **standard formats**
- process raw data
- convert processed and derived data into **standard formats**

**analyze data**
Different information units require specific standardization tasks

1. **Set up experiment**: Annotate investigation, study, assay
   - **Experiment**
   - **Raw data & metadata**

2. **Convert raw data & metadata into standard formats**
   - **Process raw data**
   - **Convert processed and derived data into standard formats**

3. **Data re-use**
4. **Data re-purpose**
5. **Data re-analysis**
6. **Analyze data**
CMSO is a young community developing standards for cell migration research

https://cmso.science/
https://github.com/CellMigStandOrg
CMSO is a young community developing standards for cell migration research

https://cmso.science/
https://github.com/CellMigStandOrg
CMSO is re-using and adapting some of the standards available in the community

http://isa-tools.org/
http://www.openmicroscopy.org/site
And it is developing new standards for cell migration-specific information

https://cellmigstandorg.github.io/Tracks/
https://github.com/CellMigStandOrg/biotracks
Diversity and complexity of cell migration data

CMSO: community standards for cell migration research

biotrack: a standard format for cell tracking data
biotracks: tabular data packages for cell tracking files

https://pypi.python.org/pypi/biotracks
https://github.com/CellMigStandOrg/biotracks
biotracks: tabular data packages for cell tracking files
biotrack: tabular data packages for cell tracking files

[Diagram showing the flow of data from segmentation algorithm to linking algorithm, and then to connecting links, resulting in objects, links, and tracks in each frame with a splitting event marked with an asterisk.]

http://frictionlessdata.io/data-packages/
biotrack aims to be for cell tracking data what Bio-Formats is for imaging data

https://commons.wikimedia.org/
http://www.openmicroscopy.org/site/products/bio-formats
biotrack aims to be for cell tracking data what Bio-Formats is for imaging data

https://commons.wikimedia.org/
https://assets.okfn.org/p/data/img/icon-512.png
biotracks integration with pandas enables analysis and visualization routines

\[ y_{it} = \beta^* x_{it} + \mu_i + \epsilon_{it} \]

Data: Marleen Van Troys, UGent

ImageJ/Fiji TrackMate
DoG detector + LAP tracking
163 tracks
A common data format enables quick comparison of tracking algorithms

ICY spot tracking plugin
Spot detector + *Multiple Hypothesis Tracking*

ImageJ/Fiji TrackMate
DoG detector + *complex* LAP tracking

http://fiji.sc/samples/FakeTracks.tif
A common data format enables quick comparison of tracking algorithms

ICY spot tracking plugin
Spot detector + Multiple Hypothesis Tracking

ImageJ/Fiji TrackMate
DoG detector + complex LAP tracking

http://fiji.sc/samples/FakeTracks.tif
Standard data formats alone are not enough
Minimum reporting requirements and CVs improve data verification and accessibility

- ISA
- OME
- biotrack

Data deposition

CV

isa tools

OME

JSON

CSV

MIACME
An open data ecosystem for cell migration research: what are the benefits?

• standardization of data and metadata is essential for reproducibility and unambiguous interpretation of experiments

• data exchange = faster and easier access to results + quality control + integration with in silico modeling

• open data = opportunities for knowledge discovery, data mining, data re-use

• formulation of new hypotheses/principles
CMSO needs community engagement

CMSO workshop in Essen, Germany, 19 and 20 June 2017

The third official CMSO workshop will take place in Essen, Germany, the 19th and 20th of June 2017. The workshop will be hosted by the lab of Prof. Matthias Gunzer.

The three working groups of CMSO will present their progress, and working sessions will take place to advance the activities of the community.

Participation to the workshop is free for all attendees, but that participants will have to pay for their own transport and accommodation. Please note that the accommodation costs at the venue for the first 10 PhD students and postdocs will be covered by the organization!

https://goo.gl/forms/eMs4m8YGZOmWOv0s1
CMSO needs community engagement

CMSO workshop in Essen, Germany, 19 and 20 June 2017

The third official CMSO workshop will take place in Essen, Germany, the 19th and 20th of June 2017. The workshop will be hosted by the lab of Prof. Matthias Gunzer.

The three working groups of CMSO will present their progress, and working sessions will take place to advance the activities of the community.

Participation to the workshop is free for all attendees, but that participants will have to pay for their own transport and accommodation. Please note that the accommodation costs at the venue for the first 10 PhD students and postdocs will be covered by the organization!

https://goo.gl/forms/eMs4m8YGZOmW0v0s1