pcdl: PhysiCell data load

For wet-lab scientists and Julia-, Matlab-, R-, Python-bioinformaticians.

Elmar Bucher, Ph.D. Student

Intelligent Systems Engineering
Indiana University

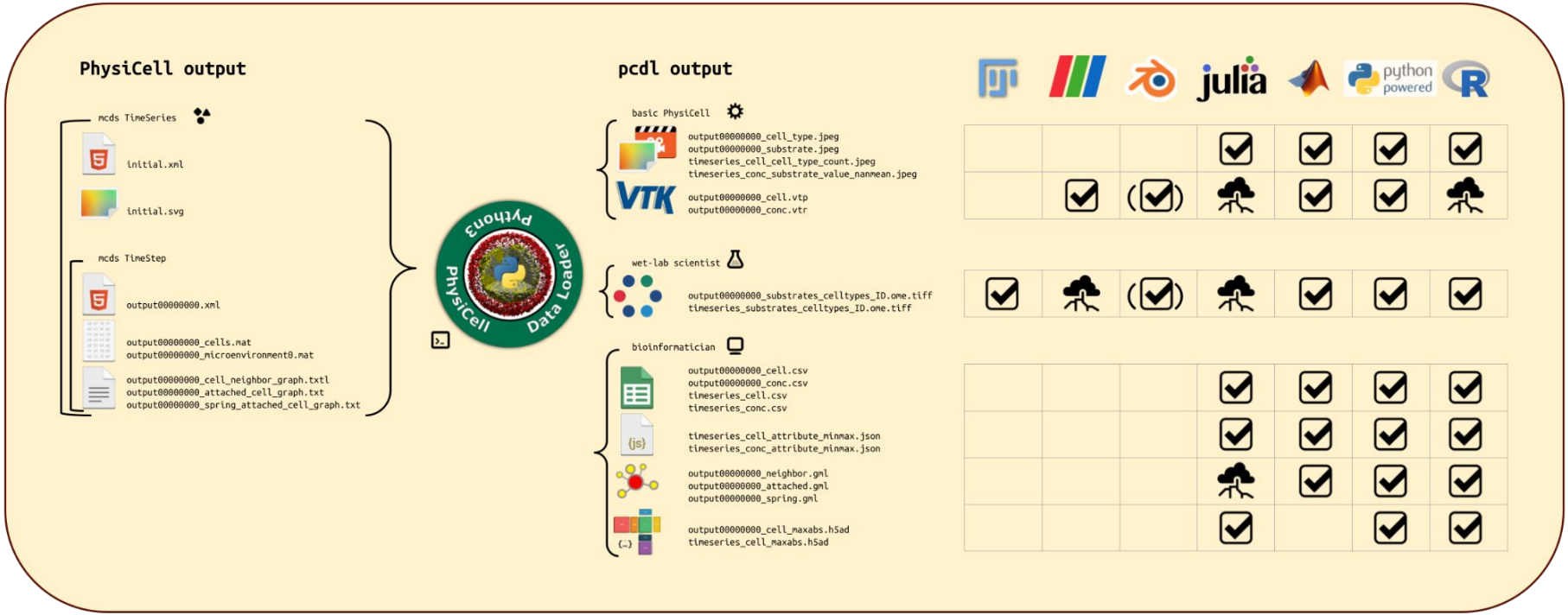
2025-11-14

Concept I

★ pcdl provides an operating system independent (Windows, MacOSX, Linux), set of commands to load PhysiCell output into Python3 or transform PhysiCell output into more widely used data formats.

★ pcdl is not “yet another analysis software”; it is just a powerful interface to analysis software!

Concept II


PhysiCell output
mcds TimeSeries
initial.xml
initial.svg
mcds TimeStep
output00000000.xml
output00000000_cells.mat
output00000000_microenvironment0.mat
output00000000_cell_neighbor_graph.txtl
output00000000_attached_cell_graph.txt
output00000000_spring_attached_cell_graph.txt
Python3
PhysiCell
Data Loader
pcdl output
basic PhysiCell
output00000000_cell_type.jpeg
output00000000_substrate.jpeg
timeseries_cell_cell_type_count.jpeg
timeseries_conc_substrate_value_nanmean.jpeg
output00000000_cell.vtp
output00000000_conc.vtr
wet-lab scientist
output00000000_substrates_celltypes_ID.ome.tiff
timeseries_substrates_celltypes_ID.ome.tiff
bioinformatician
output00000000_cell.csv
output00000000_conc.csv
timeseries_cell.csv
timeseries_conc.csv
timeseries_cell_attribute_minmax.json
timeseries_conc_attribute_minmax.json
output00000000_neighbor.gml
output00000000_attached.gml
output00000000_spring.gml
output00000000_cell_maxabs.h5ad
timeseries_cell_maxabs.h5ad
julia
python powered

Background and history

★ Evolved from PhysiCell-Tools python-loader pyMCDS.py.
- TimeStep class.
- TimeSeries class.
- Command line entry points.

★ Standing on the shoulder of giants: core python, anndata, bioio, matplotlib, neuroglancer, numpy, pandas, (requests), scikit-image, scipy, vtk.

★ pip install pcdl

Run pcdl

- ★ Most commands run on whole time series and single time steps.
- ★ Every command has a docstring that explains all possible input parameters, output options, and the purpose of the function.

Command line:

```
pcdl_get_cell_df output/output00000000.xml

pcdl_get_cell_df -h  # getting help on the fly
```

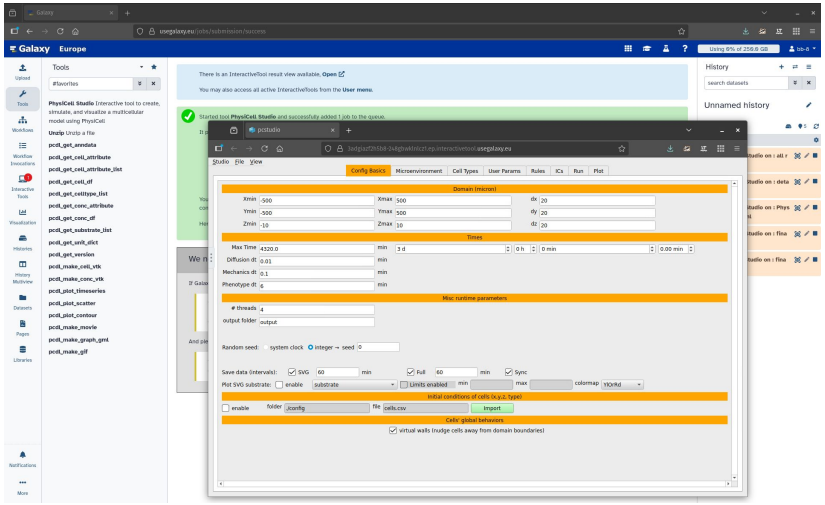
Python3:

```
import pcdl

mcds = pcdl.TimeStep('output/output00000000.xml')
df_cell = mcds.get_cell_df()
df_cell.to_csv('output/output00000000_cell.csv')

help(mcds.get_cell_df)  # getting help on the fly
```

Galaxy (https://usegalaxy.eu/):



PhysiCell 2D and 3D basics

★ Data format:
- 2D vector graphics: svg
- 2D raster graphics: jpeg, png, tiff
- 2D movies: gif, mp4
- 3D vtp visualization toolkit poly data for cells.
- 3D vtr visualization toolkit rectilinear grid for substrate.
  https://docs.vtk.org/en/latest/vtk_file_formats/vtkxml_file_format.html

★ Analysis software:
- default software installed with your operating system.
- physicell studio
- paraview

★ Demo:
- https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_paraview.md
- pcdl_get_version, pcdl_get_unit_dict
- pcdl_plot_scatter, pcdl_plot_contour, pcdl_plot_timeseries
- pcdl_make_gif, pcdl_make_movie
- pcdl_make_cell_vtk, pcdl_make_conc_vtk
- pcdl_get_celltype_list, pcdl_get_substarte_list, pcdl_get_cell_attribute_list

Wet-lab scientist

★ Data format:
- ome.tiff
  https://www.openmicroscopy.org/

★ Analysis software:
- Fiji ImageJ, Icy, Napari, and QuPath.
- Neuroglancer.

★ Demo:
- https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_fijiimagej.md
- https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_neuroglancer.md
- pcdl_make_ome_tiff
- pcdl_render_neuroglancer

Julia bioinformatician

★ Data format:
- jpeg, png, tiff (2D image):
- ome.tiff: https://github.com/tlnagy/OMETIFF.jl/issues/112
- vtk: https://github.com/JuliaVTK/ReadVTK.jl
- csv (tabular)
- json (non-tabular)
- gml (graph modeling language): https://github.com/JuliaGraphs/GraphIO.jl/issues/46
- h5ad (anndata)

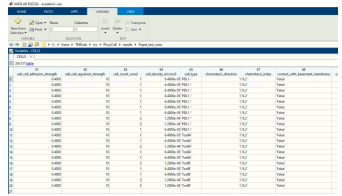
★ Analysis software:
- using FileIO # Images
- using CSV # DataFrames
- using JSON3
- using GraphIO # Graphs
- using Muon

★ Demo:
- https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_julia.md
- pcdl_get_cell_attribute, pcdl_get_cell_attribute
- pcdl_get_cell_df, pcdl_get_conc_df
- pcdl_get_anndata

R bioinformatician

- ★ Data format:
  - jpeg, png, tiff
  - ome.tiff # RBioFormats
  - csv (tabular)
  - json (non-tabular)
  - gml (graph modeling language)
  - h5ad (anndata)
- ★ Analysis software:
  - library("jpeg")
  - library("png")
  - library("tiff")
  - library("RBioFormats")
  - library("jsonlite")
  - library("igraph")
  - library("schard") # anndata
- ★ Demo:
  - https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_r.md
  - pcdl_make_graph_gml neighbor attached spring

Matlab bioinformatician

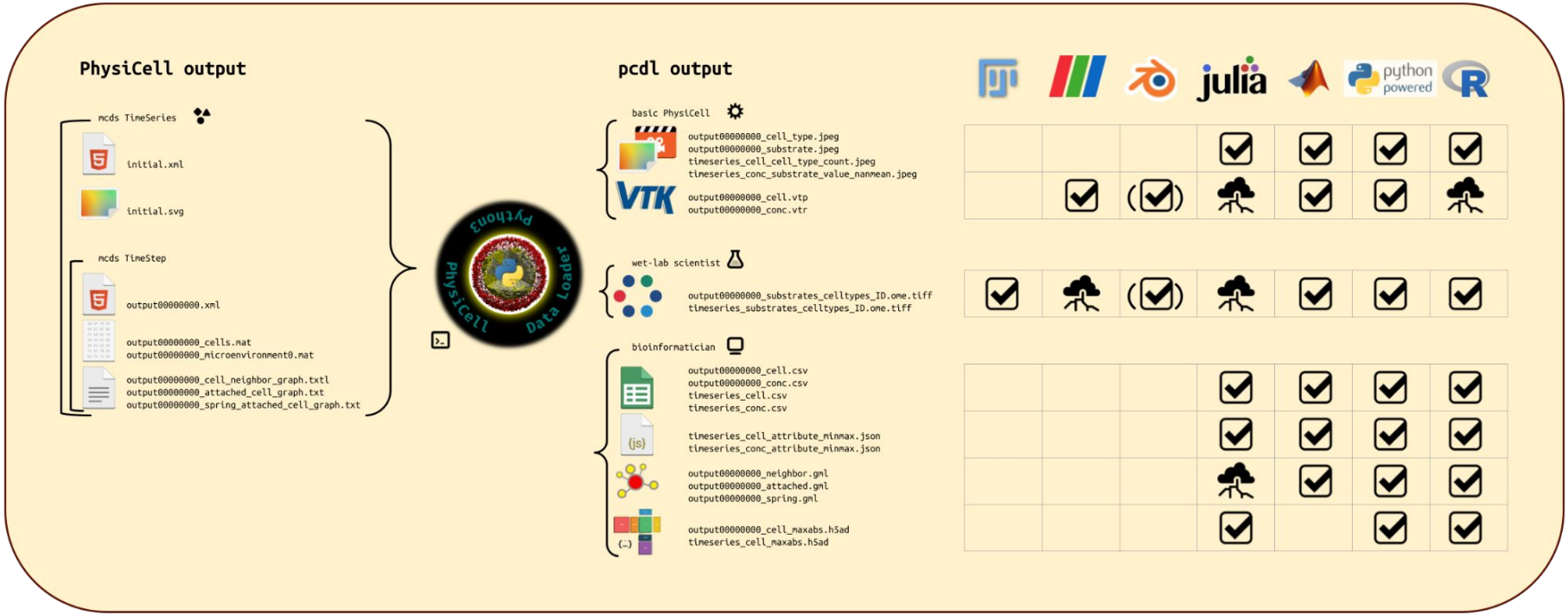


- ★ Data format:
  - jpeg, png, tiff
  - vtk # https://www.mathworks.com/matlabcentral/fileexchange/94993-vtktoolbox
  - ome.tiff # https://docs.openmicroscopy.org/bio-formats/5.7.1/developers/matlab-dev.html
  - csv (tabular)
  - json (non-tabular)
  - gml (graph modeling language) # https://www.mathworks.com/matlabcentral/fileexchange/159001-matlab-igraph

- ★ Analysis software:
  - img = imread("output/timeseries_cell_cell_type_count.jpeg")
  - vtkimg =
  - omeimg =
  - df_cell = readtable("output/timeseries_conc.csv")
  - struct_conc = jsondecode(fileread("output/timeseries_conc_attribute_minmax.json"))
  - g =

- ★ Demo:
  - https://github.com/elmbeech/physicelldataloader/blob/master/man/TUTORIAL_matlab_octave.md

Concept


PhysiCell output
mcds TimeSeries
initial.xml
initial.svg
mcds TimeStep
output00000000.xml
output00000000_cells.mat
output00000000_microenvironment0.mat
output00000000_cell_neighbor_graph.txtl
output00000000_attached_cell_graph.txt
output00000000_spring_attached_cell_graph.txt
pcdl output
basic PhysiCell
output00000000_cell_type.jpeg
output00000000_substrate.jpeg
timeseries_cell_cell_type_count.jpeg
timeseries_conc_substrate_value_nanmean.jpeg
output00000000_cell.vtp
output00000000_conc.vtr
wet-lab scientist
output00000000_substrates_celltypes_ID.ome.tiff
timeseries_substrates_celltypes_ID.ome.tiff
bioinformatician
output00000000_cell.csv
output00000000_conc.csv
timeseries_cell.csv
timeseries_conc.csv
timeseries_cell_attribute_minmax.json
timeseries_conc_attribute_minmax.json
output00000000_neighbor.gml
output00000000_attached.gml
output00000000_spring.gml
output00000000_cell_maxabs.h5ad
timeseries_cell_maxabs.h5ad
julia
python powered

Acknowledgement

★ MathCancer Lab
- Patrick Wall
- Heber Rocha
- Furkan Kurtoglu
- John Metzcar
- Aneequa Sundus
- Randy Heiland
- Paul Macklin

★ UZaragoza
- Raquel Arroyo (matlab)

★ OHSU (ome.tiff)
- Jennifer Eng

★ IU undergrads
- Benjamin Jacobs
- Katie Pletz
- Leena Sohail
- Marshal Gress
- Nick Oldfather
- Thierry-Pascal Fleurant

★ UMB undergrads (neuroglancer)
- Jason Lu
- Viviana Kwong