

H3D Foundation and Ersilia present

Bringing data science and AI/ML tools to infectious disease research

Event Sponsors



CS&S

Code for
Science
& Society





Session 4: Using open source ML assets

Skills Development

Gemma Turon, @TuronGemma, gemma@ersilia.io

Miquel Duran-Frigola, @mduranfrigola, miquel@ersilia.io

Ersilia Open Source Initiative, @ersiliaio, <https://ersilia.io>

30th September 2022



What is Git and GitHub?



Git is a software that tracks changes to files. Changes include creating new files, deleting files or folders or editing the content of a file.

When can Git be useful?



GitHub is an internet hosting service for software development and version control using Git.

When can GitHub be useful?



GitHub interface

Many organisations and laboratories working with coding have a GitHub Organisation profile

<https://github.com/ersilia-os>

Organisation Page

Repository 1

Repository 2

Repository 3

Repository 4

Repository 5

Repository 6

1 Repository =
1 Project



Ersilia's backend is in GitHub

An example of repositories we have been using:

- Ersilia: <https://github.com/ersilia-os/ersilia>
- Retrosynthetic Accessibility: <https://github.com/reymond-group/RAscore>
- Antibiotic activity:
<https://github.com/chemprop/chemprop>



Important files in a GitHub Repo

<https://github.com/ersilia-os/event-fund-ai-drug-discovery>

- README File: basic information about the repository, shows up on the repository landing page
- LICENSE: it is essential to check the license of the code before using it. Common Open Source Licenses are:
 - MIT, GPLv3, Apache, Mozilla ...



Important sections in a Repo

<https://github.com/ersilia-os/event-fund-ai-drug-discovery>

- Contributors: let's you see who is contributing to this code
--> great for contacting people
- Issues: a place where everyone can drop questions or problems they have encountered when using the code in the repository, and hopefully get help from the community or the developers.

How to use the Ersilia Model Hub with my own data



Download a dataset from ChEMBL

- chembl assay id: 3882128
- Select all molecules and start download
- Unpack ZIP file and extract .csv
- Save .csv to a folder in Drive



Let's jump on the Notebook for Session 4 Skills Development

- The Notebook contains an easy to use interactive tool to run models from the Ersilia Model Hub.
- Run all the cells in order! If the notebook disconnects for some reason, run everything again
- Input the full paths to the files:
drive/MyDrive/foldername/filename.csv
- Check the cells that require input (✍️) are properly filled

Contact us



<https://ersilia.io>



hello@ersilia.io



[@ersiliaio](https://twitter.com/ersiliaio)