



#OpenScience

#OpenSource

Good practice in scientific programming

Robert Haase

October 2022

Accessible

Reusable

Sustainable

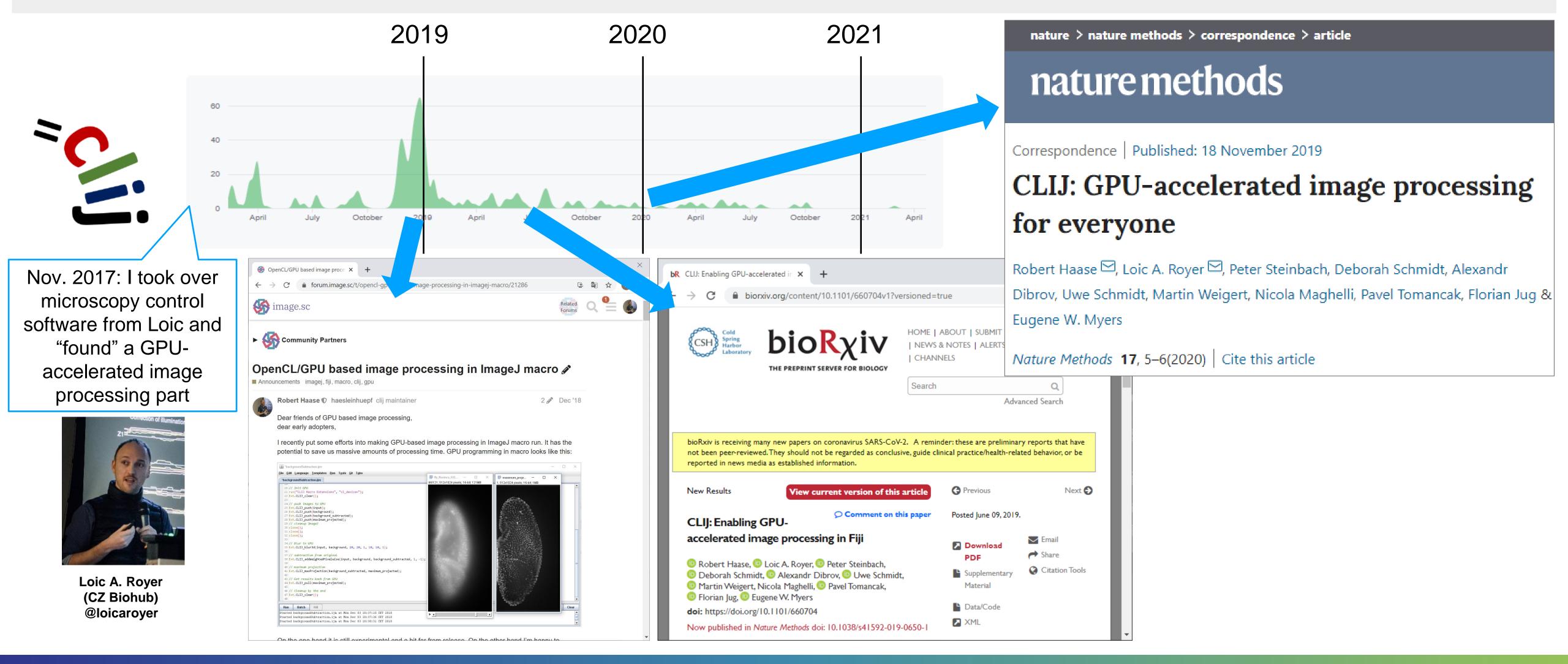






An open science story





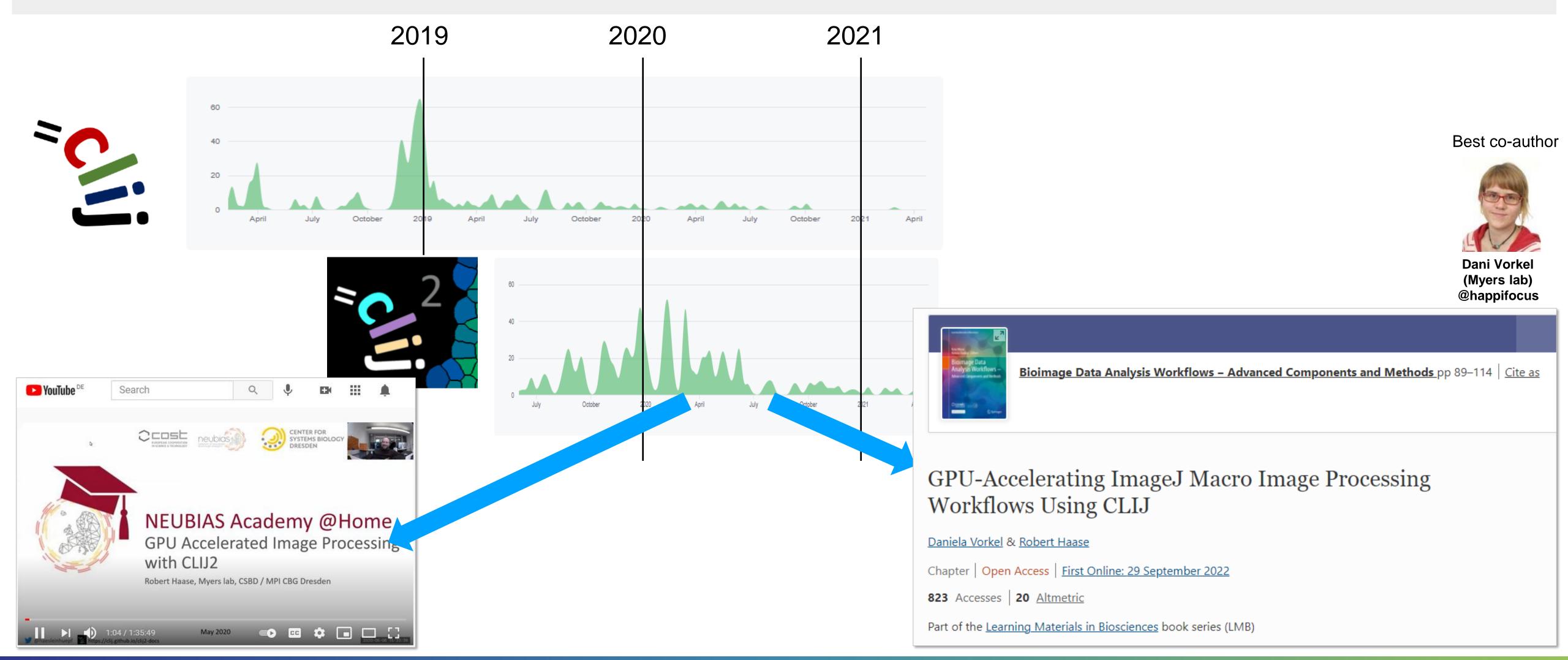






An open science story











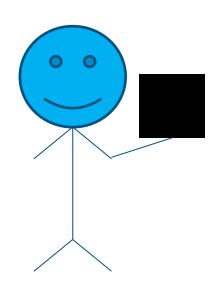
Openness of software / projects



Communication is key!

Choose your project's level wisely, and communicate it clearly

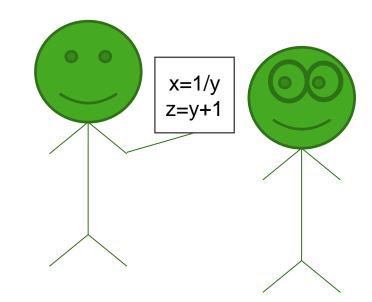
Closed source



- Open to collaborations
- "Black box"
- Compiled code (e.g. C/C++)
- Good for protecting intellectual properties (\$\$\$)

Hardware device drivers

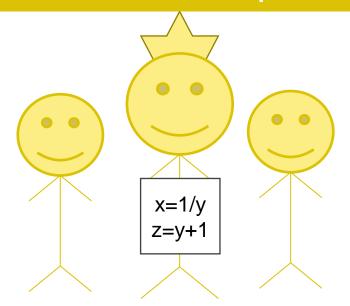
Open source



- Code available to read
- Not necessarily executable code
- No maintenance / support efforts

Custom image analysis scripts

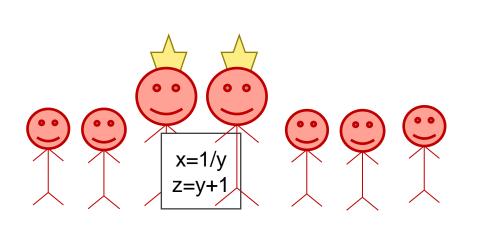
Benevolent dictatorship



- Open to contributions
- Single maintainer, often overwhelmed
- Efficient decision making
- Bus factor ≈1

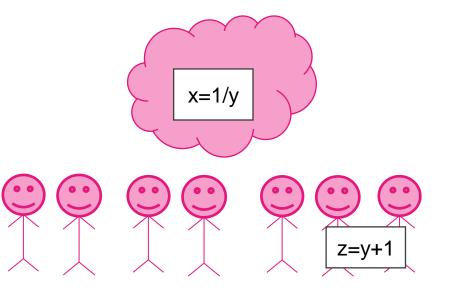
TrackMate, SNT, MorpholibJ, CLIJ

Community driven



- Open to contributions
- Partially democratic
- Board of maintainers (core developers)
- Long-winded decision making

Openly extensible



- Openly extensible; without maintainers involved
- Partially community driven

scikit-image, scipy, OpenCL

ImageJ,
Python, numpy



Read more: https://coiled.io/blog/stages-of-openness/



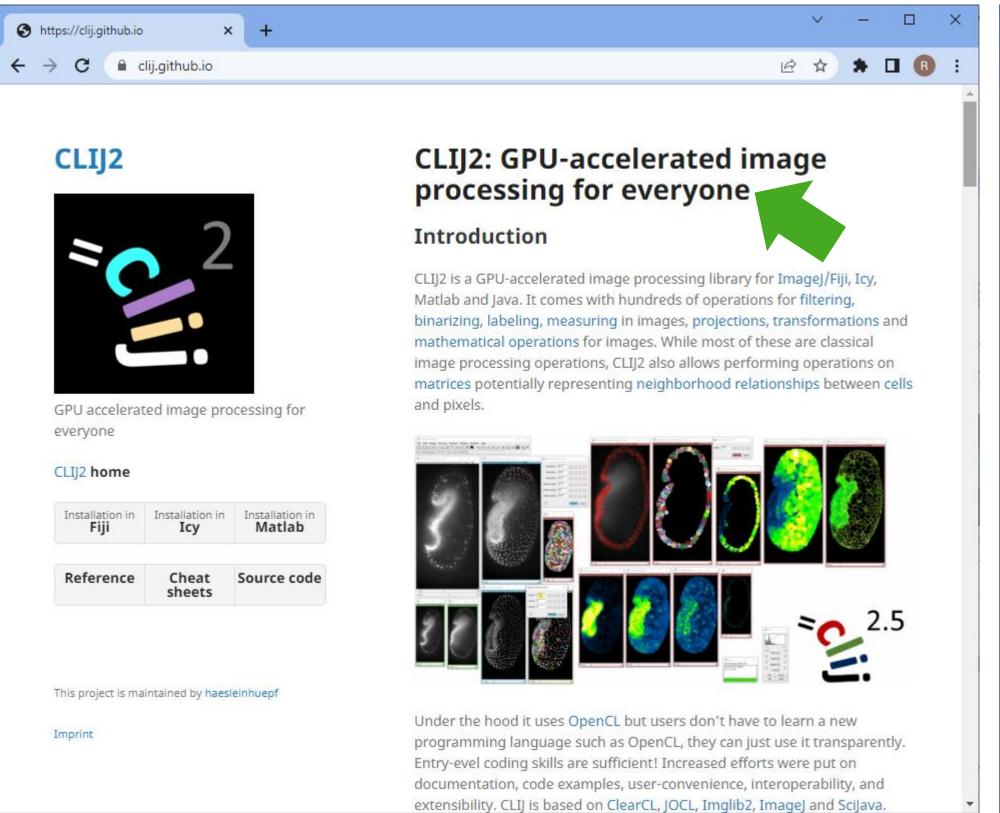


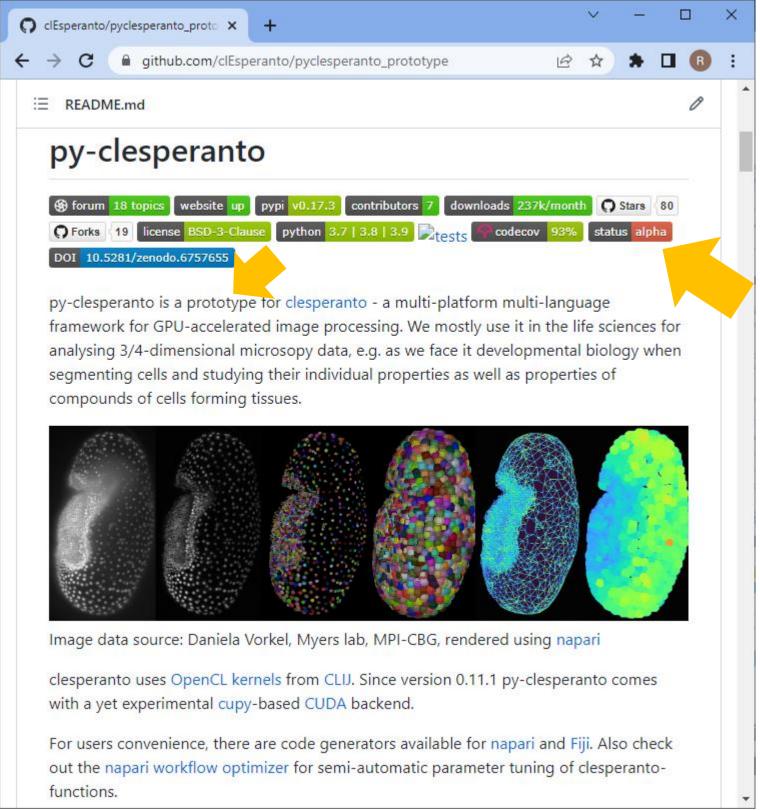
Target audience



Documentation should tell who is the target audience and how far it is developed

Communication is key!





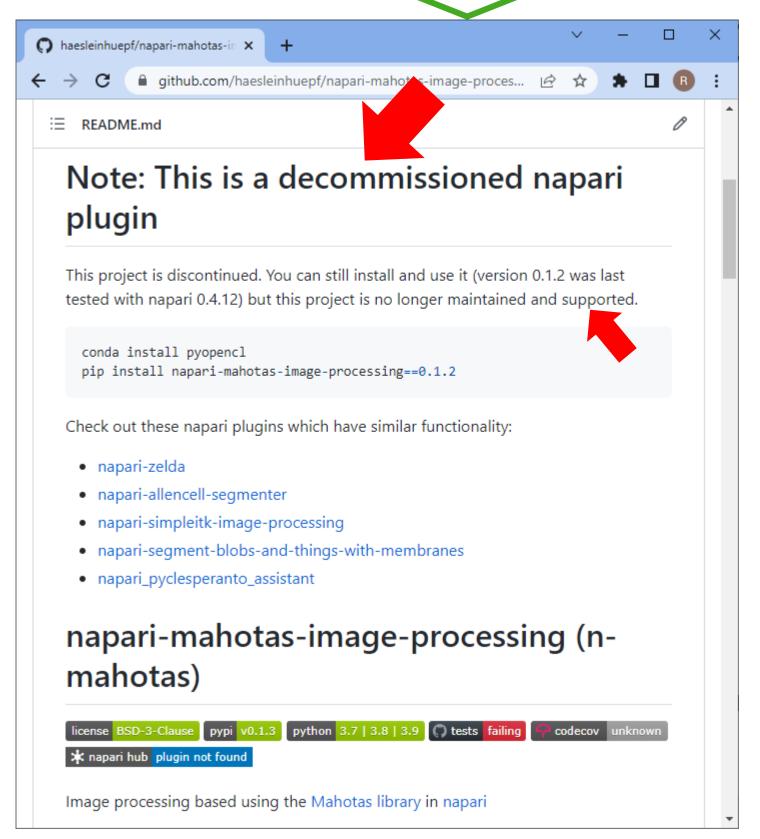
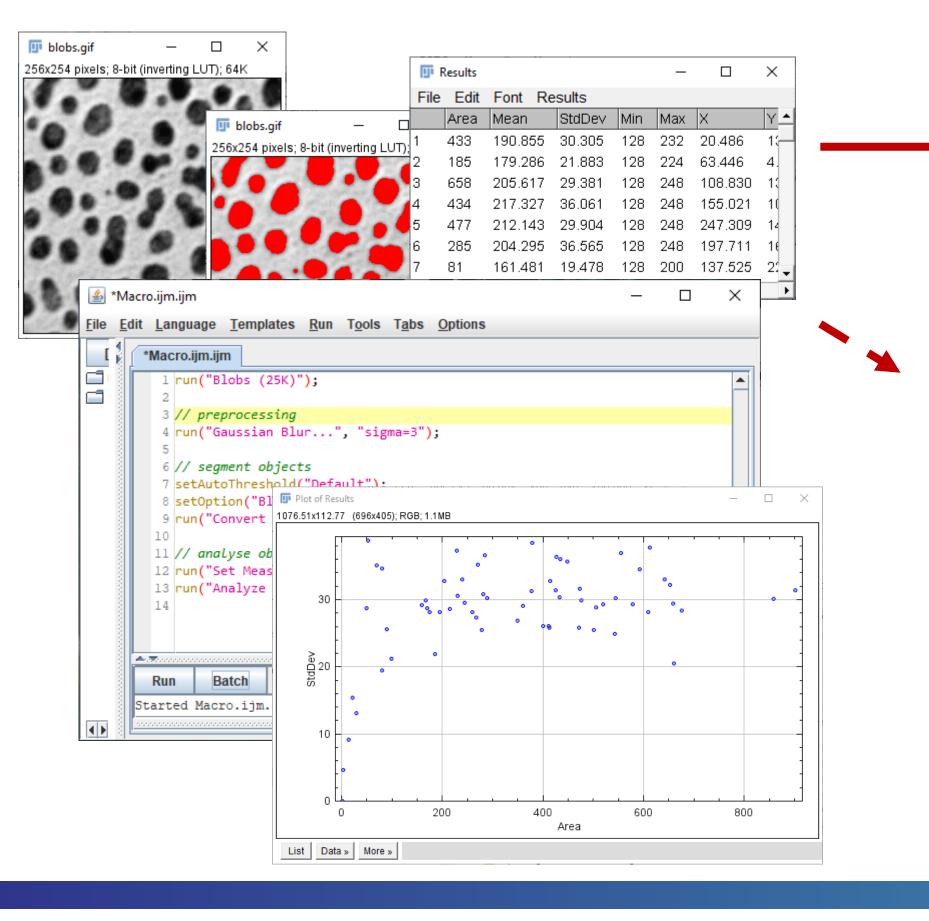


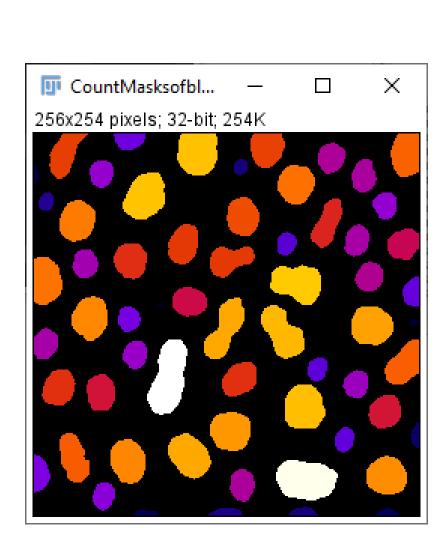


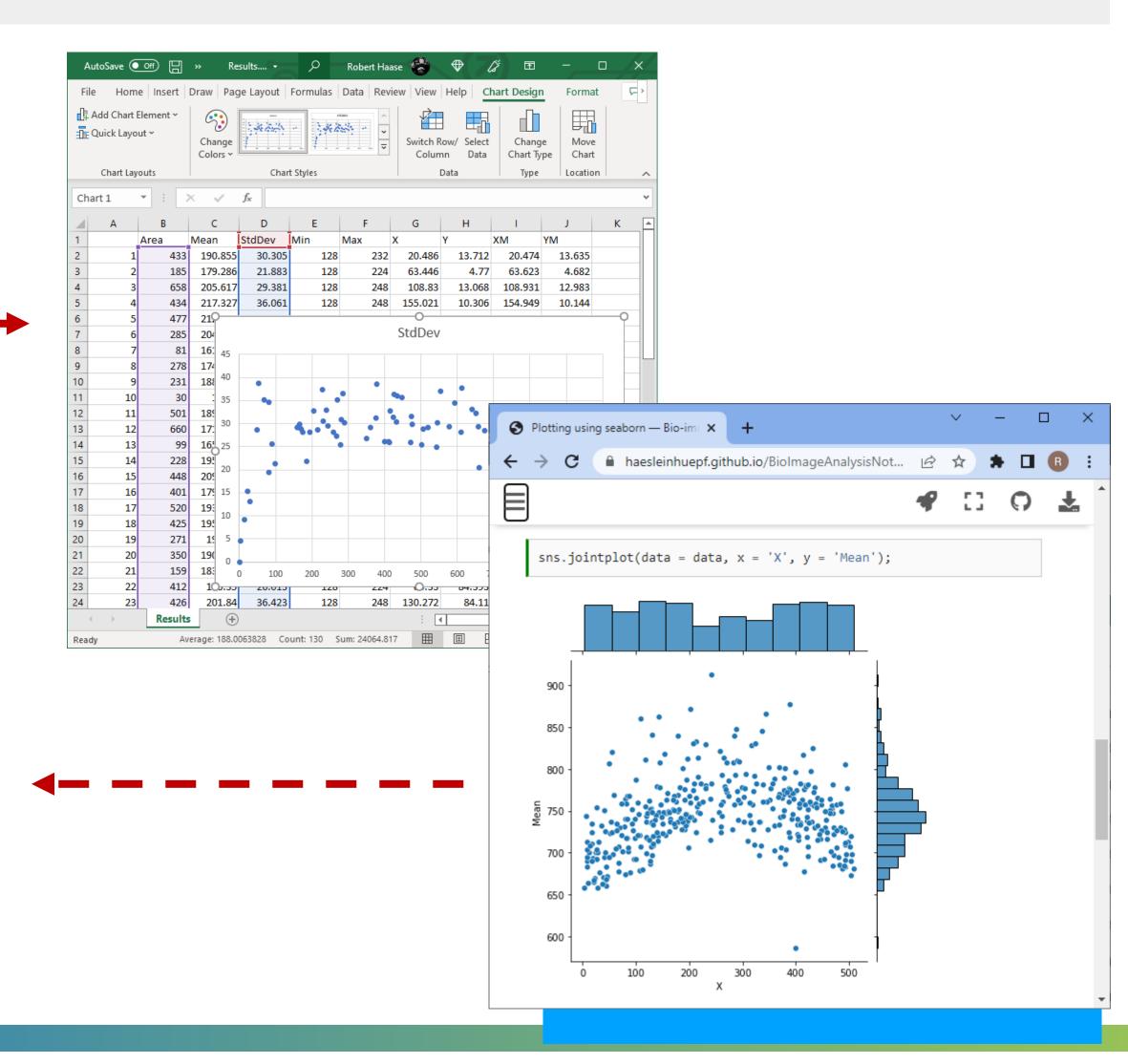
Image data science is cross-disciplinary **P**















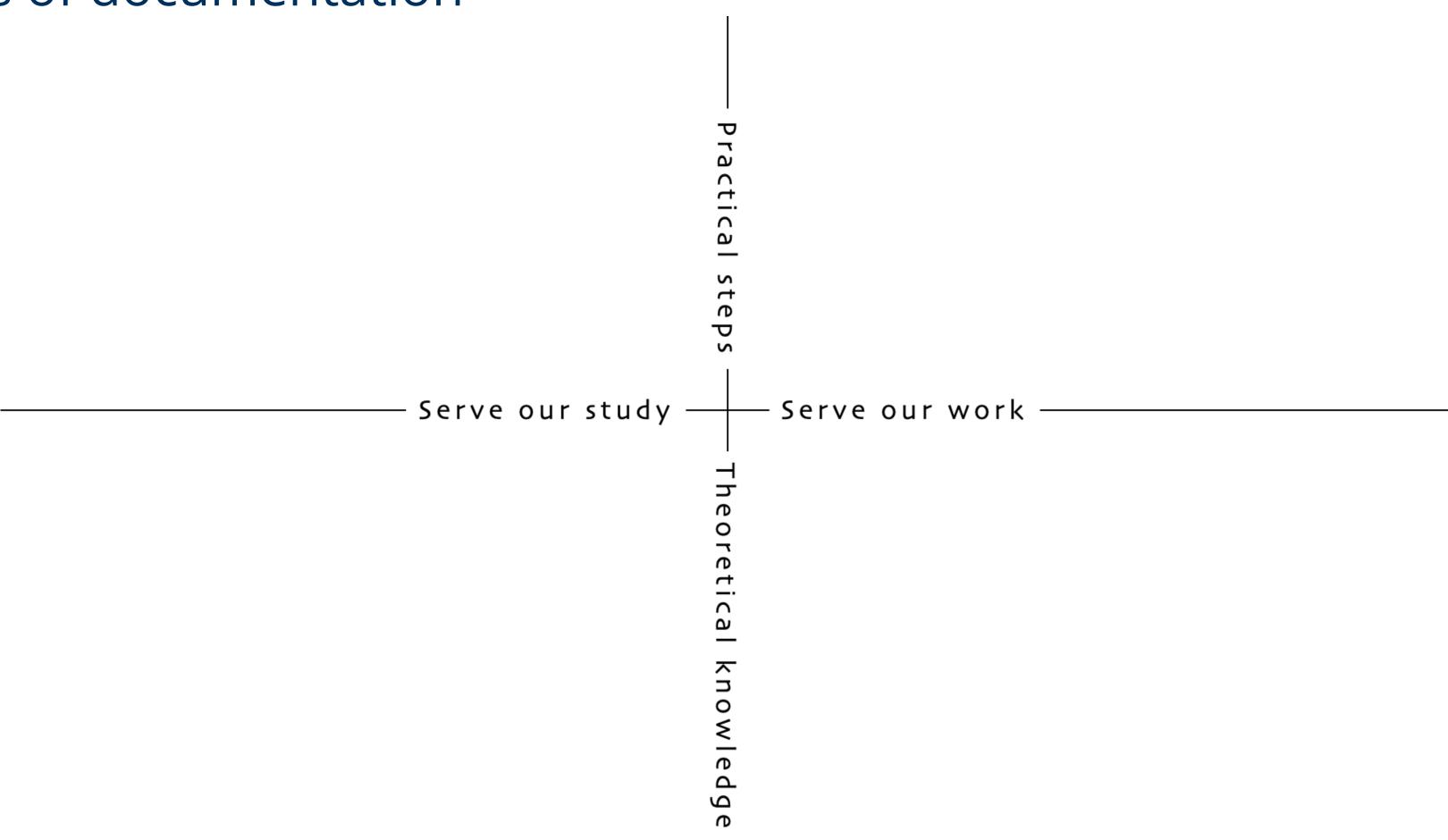


Documentation is key

... to make code reusable



The four modes of documentation

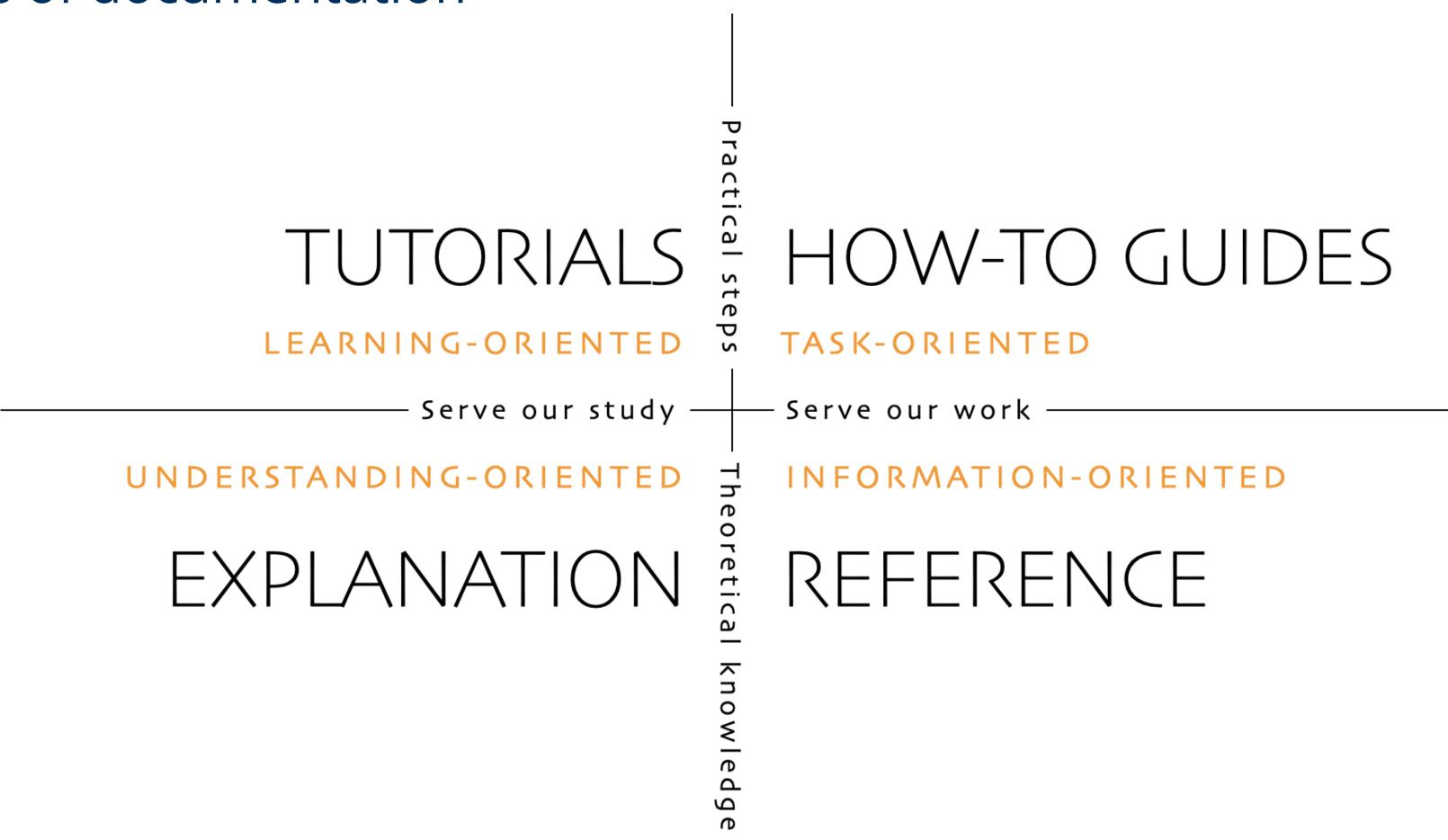








The four modes of documentation



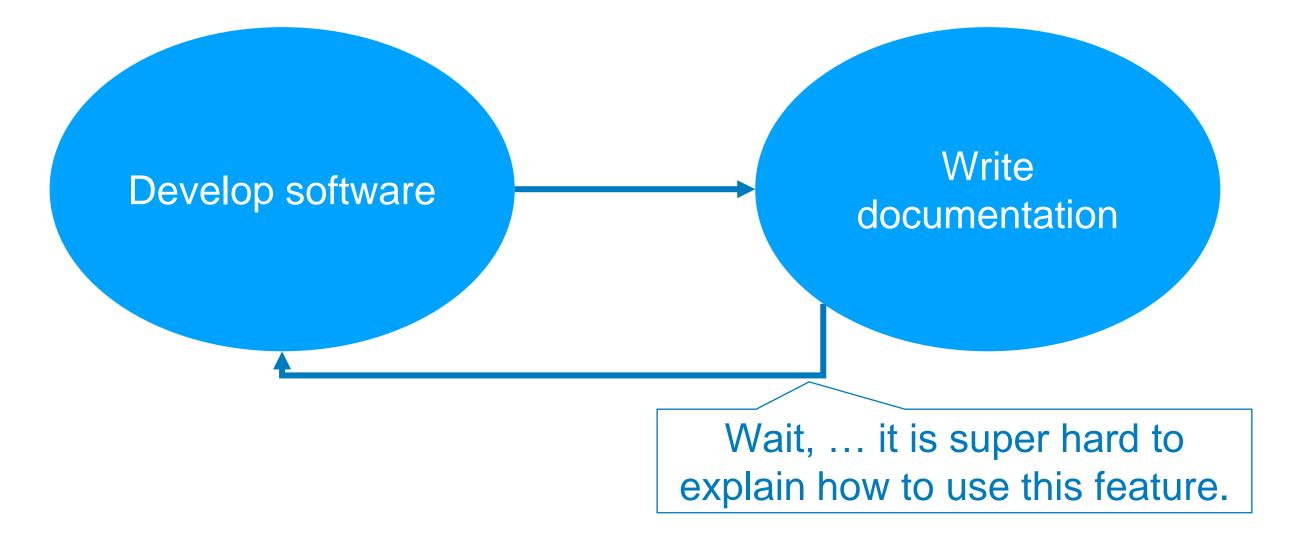








Documentation is a process, not a final step

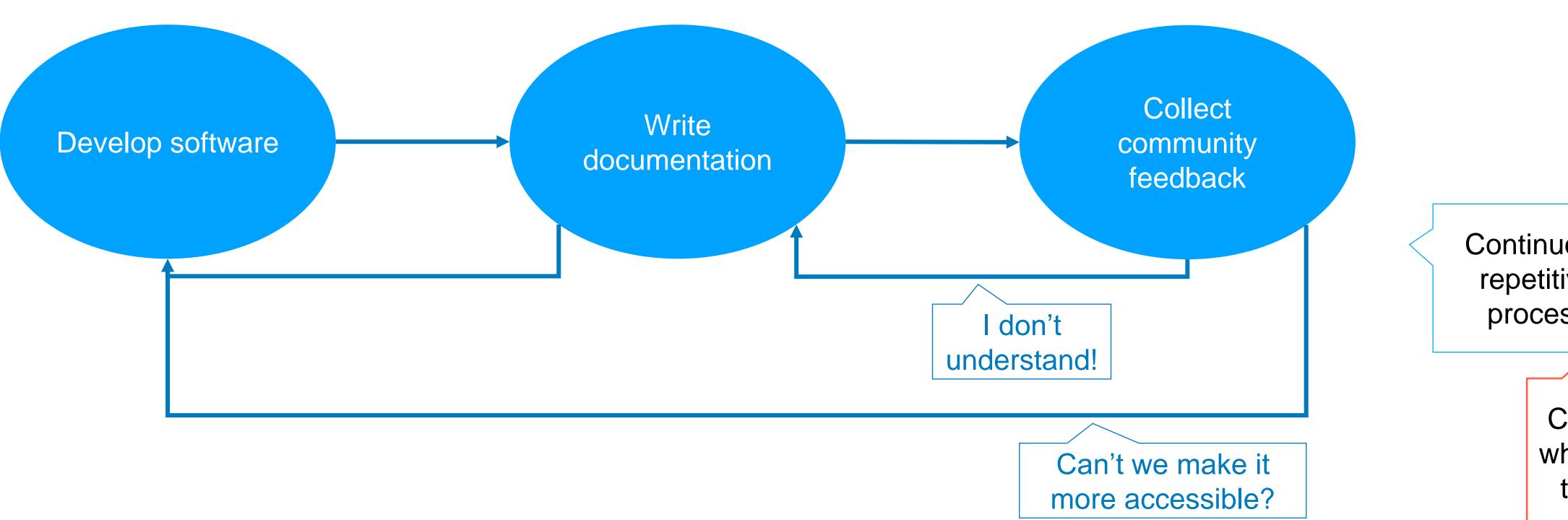








Documentation is a process, not a final step



Continuous repetitive process

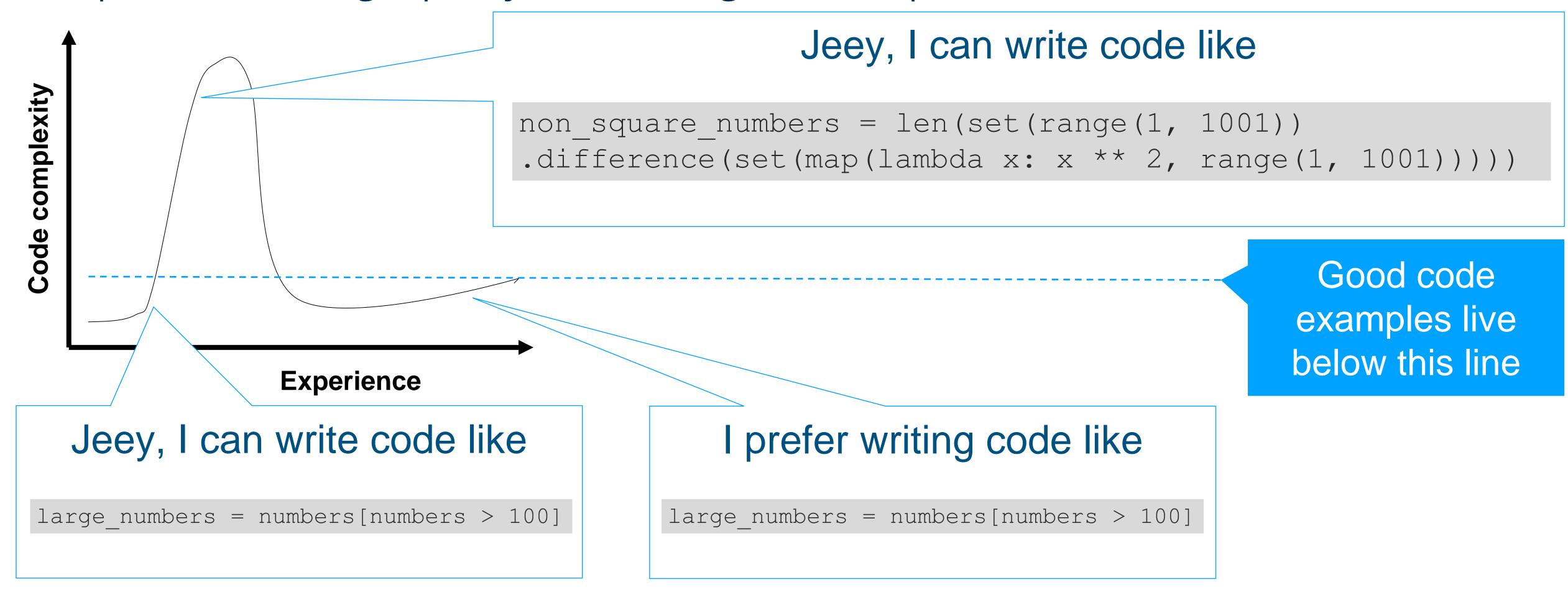
Consider this when planning the project!







Interpretation of "High quality code" changes with experience





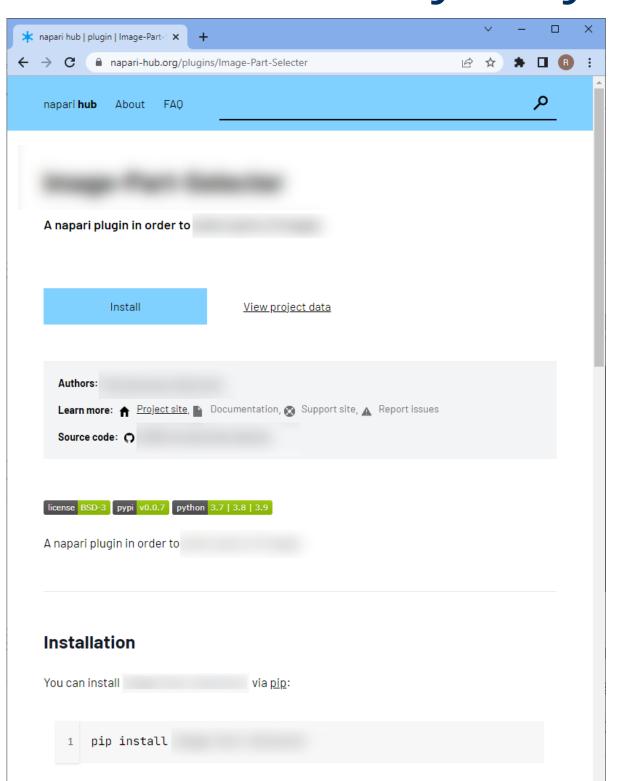


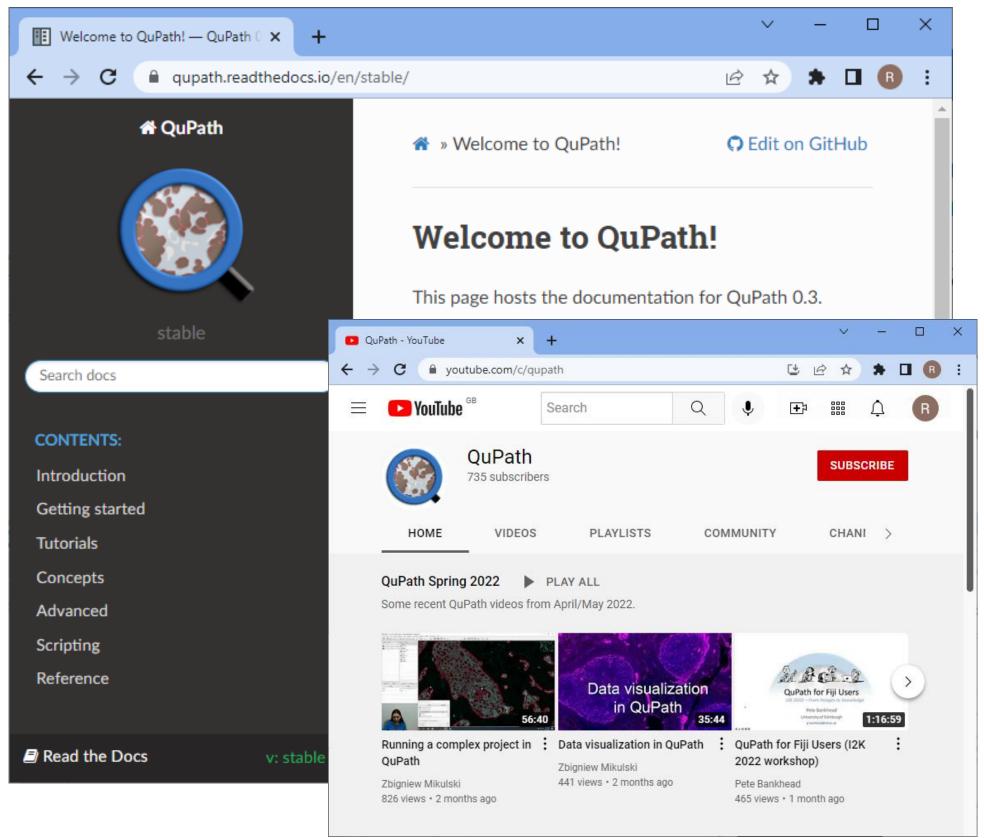


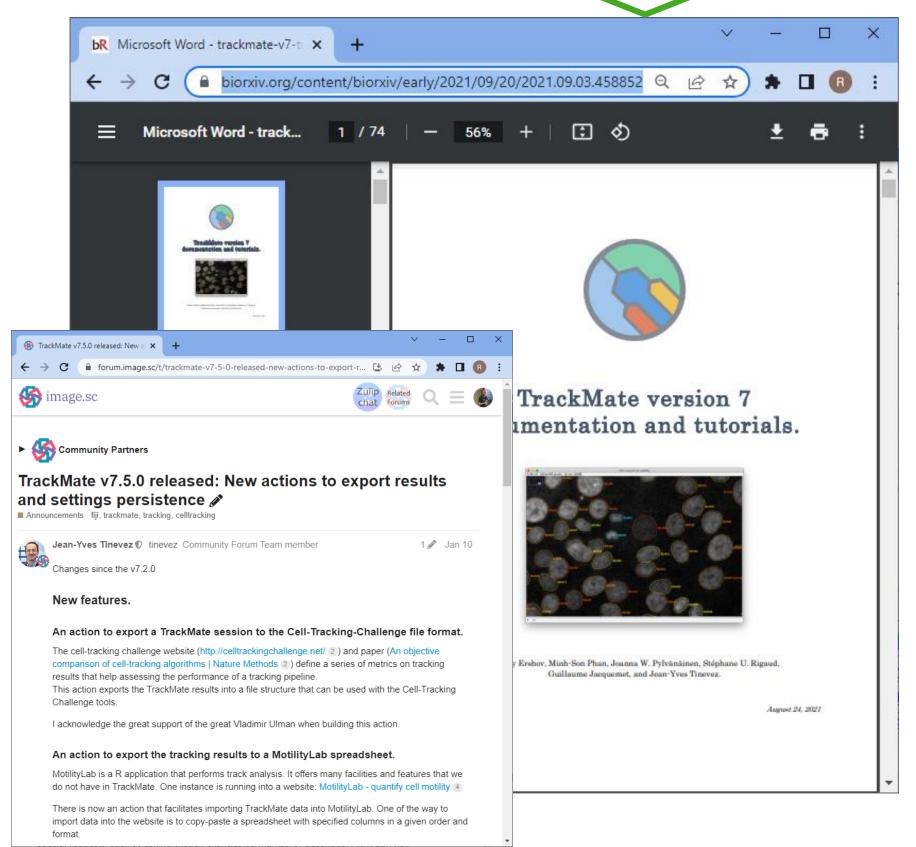


Use your project's website to show others how to use your software. Otherwise, they may not use it.

Communication is key!









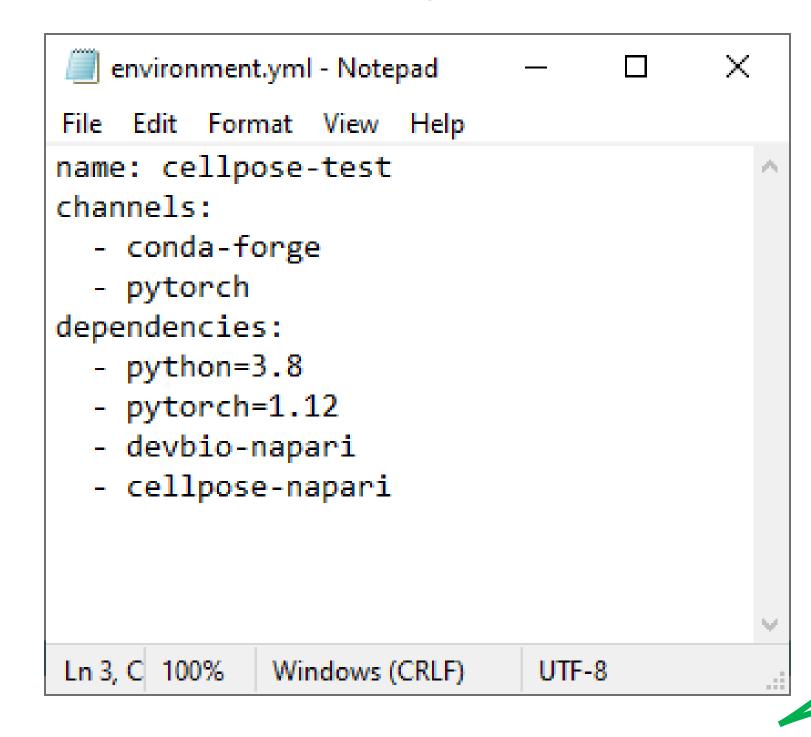


Documenting dependencies



Maintain a document with the dependencies (and versions) you need in your project!

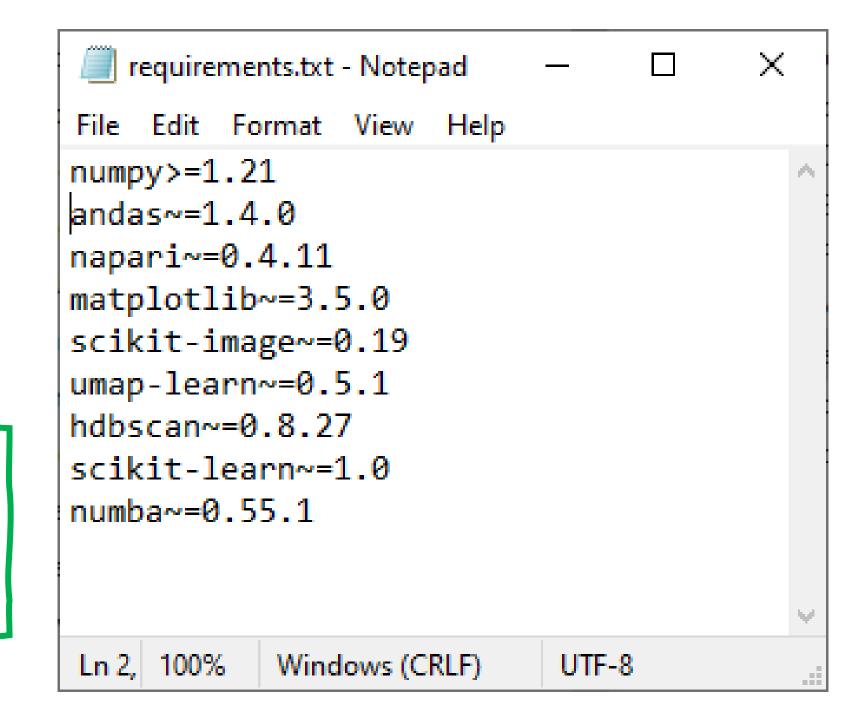
The conda way



In case your environment is screwed up, you can rebuild it any time.

conda env create -f environment.yml

The pip way



pip install -r requirements.txt



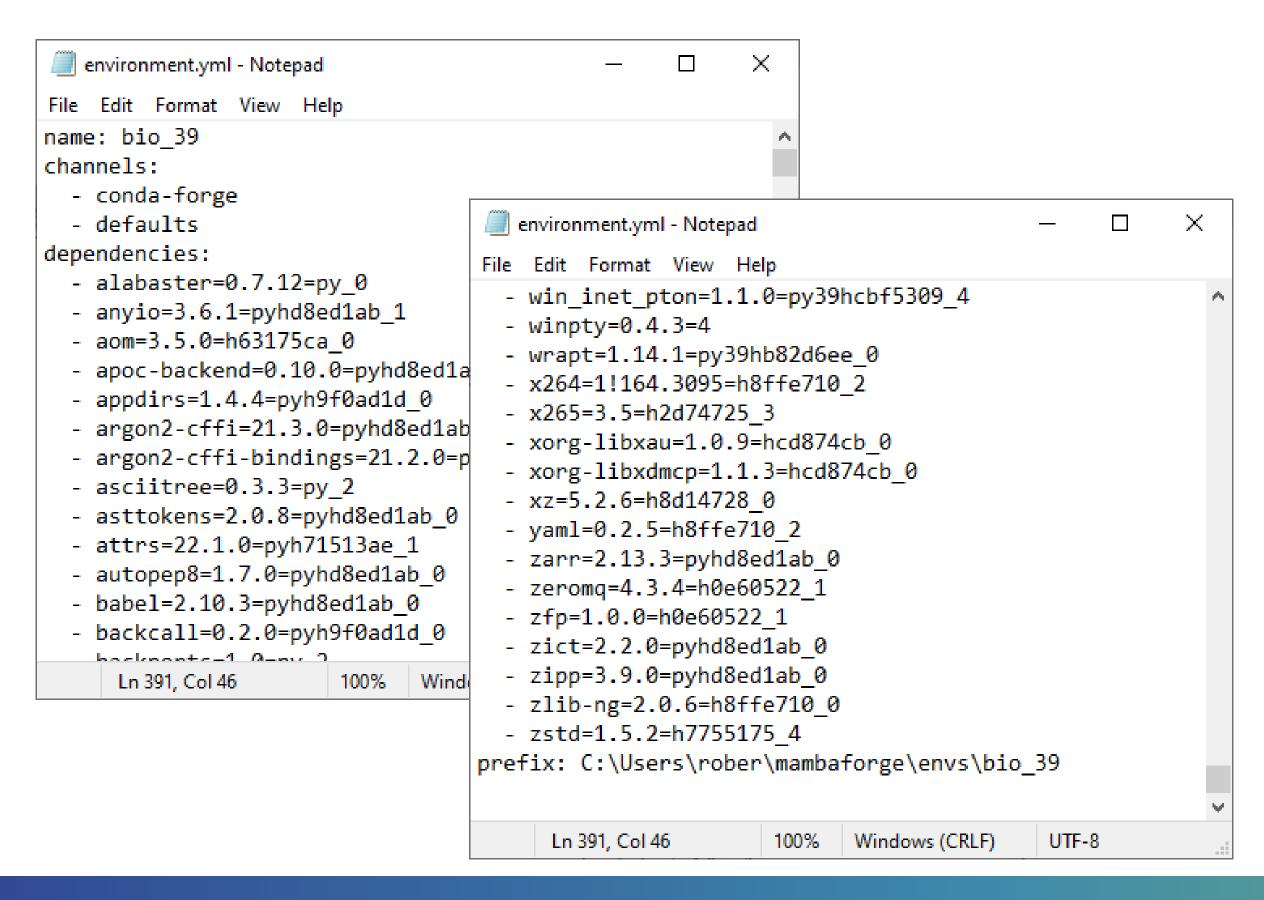


Documenting dependencies



... the complete way.

conda env export > environment.yml



Excellent way to document which dependencies were actually used...

It is *questionable* if recreating an environment from this yml file works.









Shipping scientific software

... so that others can reuse it

Deployment



Method / algorithm / software developers live from making things that are useful for others.

"... is available as user-friendly standalone software."



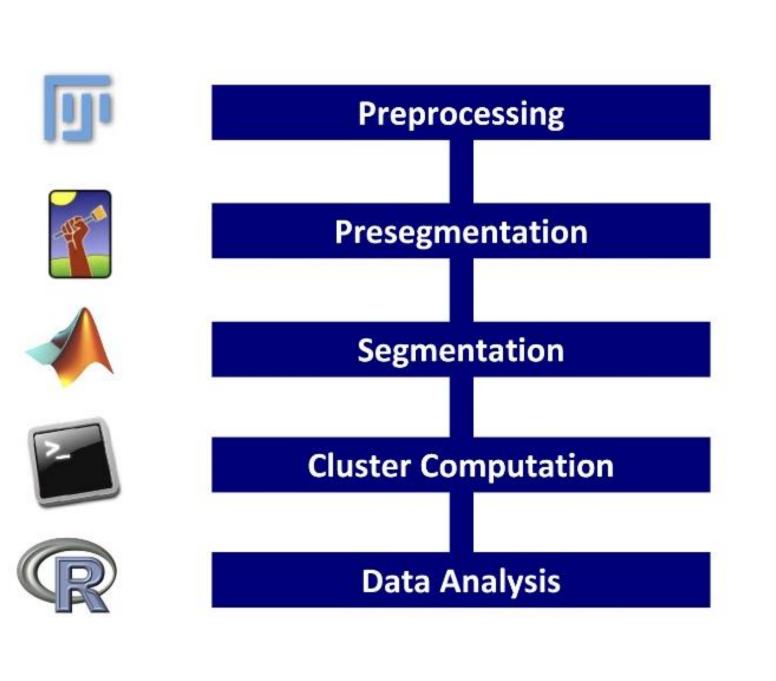
Installation *can* be made easy.



Ensuring interoperability with other software comes with high effort.



Combining standalone software is often painful and not *user-friendly*.



Think of workflow developers, a.k.a. "users".

"... is available as user-friendly Python library."



In life-sciences, a Python library may not be perceived as *user-friendly*.



Installation and dependency management is often tricky.



Interoperability comes basically for free.







Deployment



Ship new methods as plugins for established platforms

minimizing

and

- effort to ensure interoperability with other tools and file-formats,
- entrance-level burden for new users and
- installation efforts.

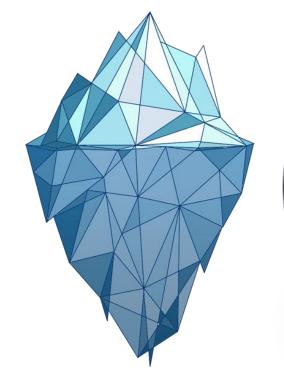
maximizing

- visibility,
- accessibility,
- community support and finally,
- adoption.



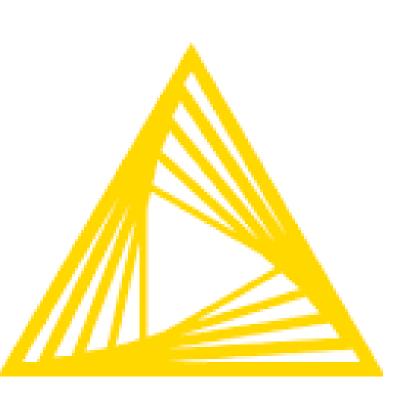




















Interdisciplinary collaboration

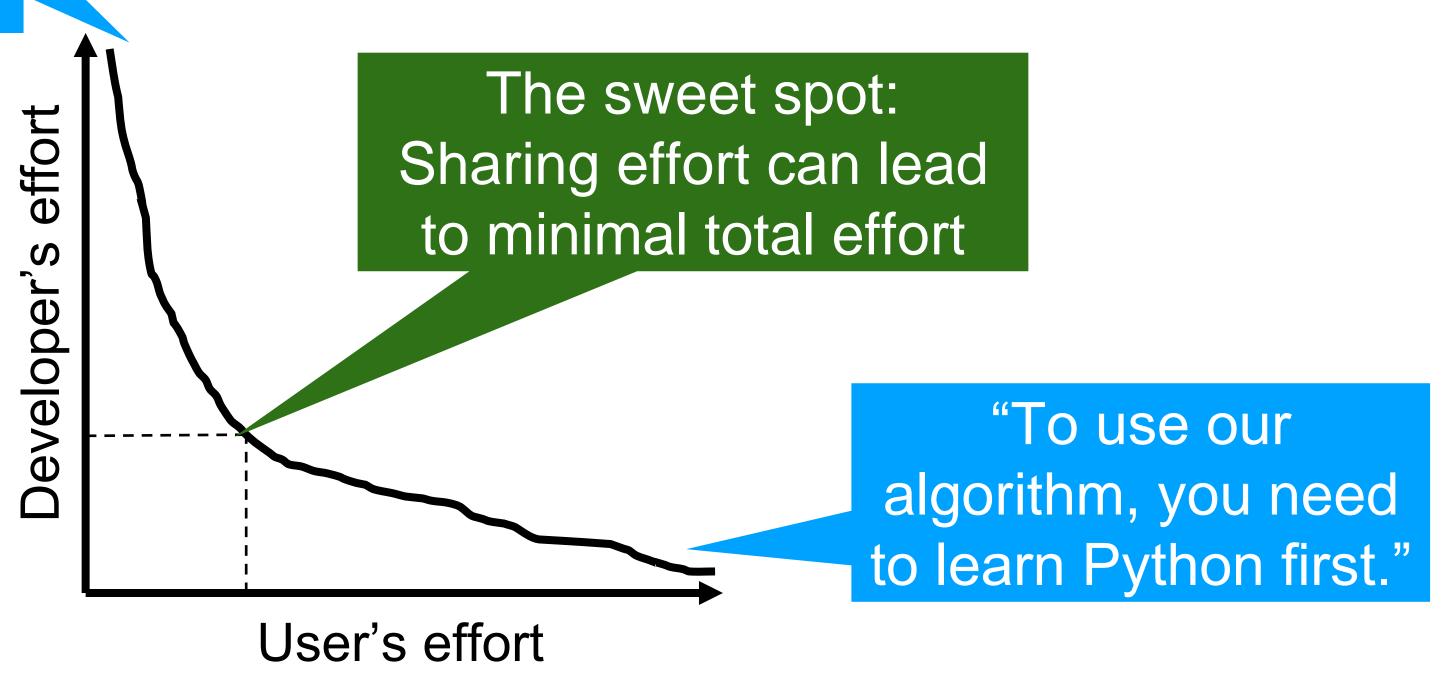
... of life scientists and developers joining forces.

Why we need to work together



• ... to minimize effort

Making the user interface self-explaining and easily accessible









Why we need to work together



• ... to be successful in our projects (and grant applications)

Computer scientist: "I would like to develop software that allows to unravel the underlying principles of embryo development."

Funding agency: "But you're lacking the developmental biology skills to make this happen."

Biologist: "I would like to do quantitative image data science unravel the underlying principles of embryo development."

Funding agency: "But you're lacking the bioinformatics skills to make this happen."



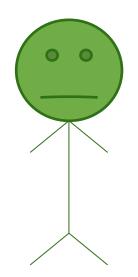




Matching expectations



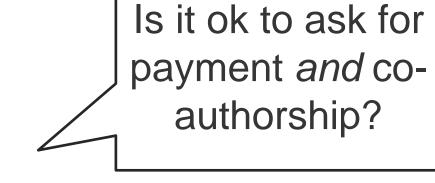
... and meeting in the middle for successful collaborations

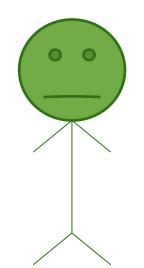


I need somebody who just analyzes the image data. I'm happy to pay for it but can't offer coauthorship.



I need to advance my academic career





I need somebody who analyzes the image data for us. I can't pay but I'm happy to offer coauthorship.



I need to make sure my corefacility has some income.



I wish I was a member of their group. They are paid and coauthors.







Matching expectations

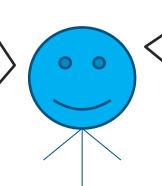


Speak out your expectations and conditions early.

Communication is key!

Hi, my name is Robert, ...

I'm happy to teach you everything I know about image analysis.



When I spend *my time* on making *your research* possible, we are co-authors.

Recommendation of analysis software and tools.

Basic data analysis help and advice.

Constructive data analysis and interpretation.

Creation of complex custom image analysis tools.

Simple acknowledgement

Inclusion of specific facility member on author list

Imaging Facility Guidelines for Acknowledgement

- All publications resulting from the use of instruments within the facility should acknowledge the facility as a whole, e.g. 'the authors gratefully acknowledge the [core facility name] for their support & assistance in this work' and the facility should be informed of the publication.
- Specific grants that have funded the facility instruments used for the work to be published must be acknowledged if the data was acquired during the active period of that grant. Facility staff will advise users of such grant codes.
- Assistance above the technical or routine level, with any facility staff providing scientific input and expertise in experimental set-up, acquisition or analysis, should be recognised through co-authorship on resulting publications. Please discuss acknowledgements with facility staff prior to manuscript submission.

Example scenarios with baseline recommendations:

	Sample Preparation	Fast, routine sample preparation with standard protocol.	Simple acknowledgement
		Development of new sample preparation protocols. Optimisation of existing protocols for specific samples.	Inclusion of specific facility member on author list
	Image Acquisition	Training of users to acquire images themselves. Simple acquisition of raw data.	Simple acknowledgement
		Operational image acquisition with input and decisions dependent on expertise. Design or re-design of experimental conditions.	Inclusion of specific facility member on author list
	Image Analysis	Recommendation of analysis software and tools. Basic data analysis help and advice.	Simple acknowledgement
		Constructive data analysis and interpretation. Creation of complex custom image analysis tools.	Inclusion of specific facility member on author list

Based on the publication policy compiled by Natasha Stephen, Plymouth Electron Microscopy Centre, after discussions with the RMS EM-UK community







Image Analysis



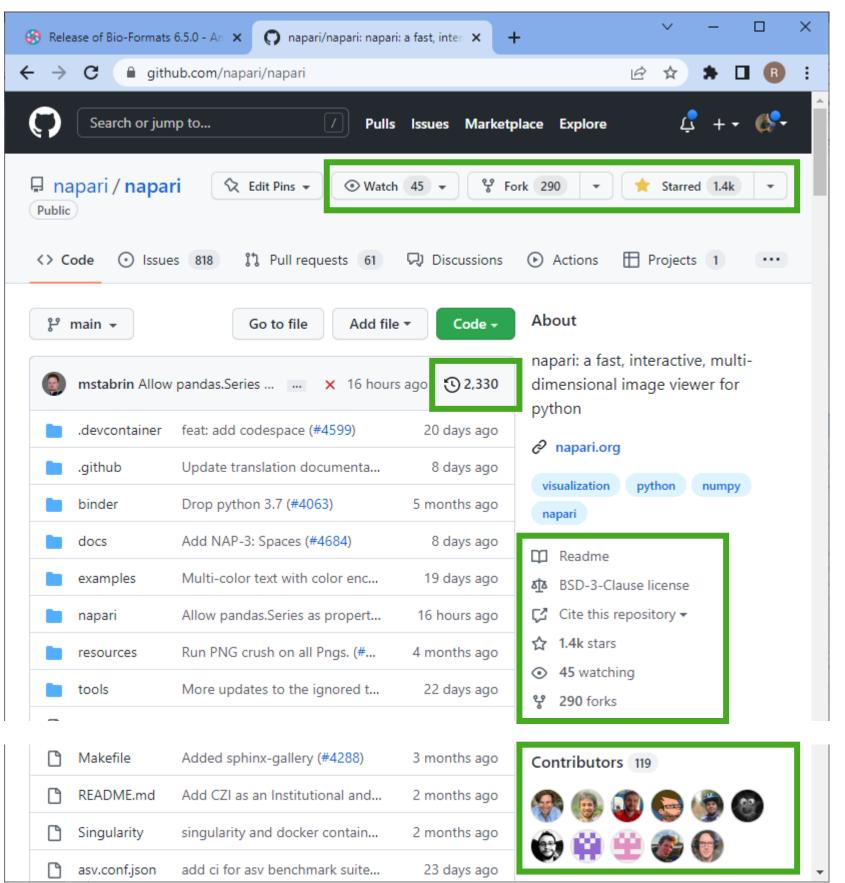




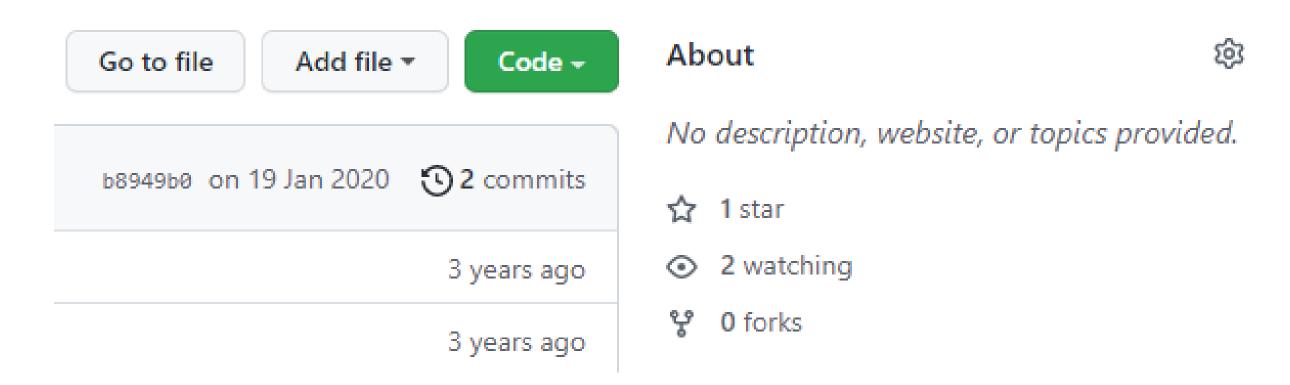
... to differentiate the good stuff from the bad.



Visit the project's github or gitlab page and review indicators.



- Stars: People like software, similarly to tweets on Twitter
- Watching: People receive updates for new releases
- **Forks**: People made a copy of the code, e.g. to contribute to the project
- Contributors: People who contributed to the code
- Commits: Changes to the code



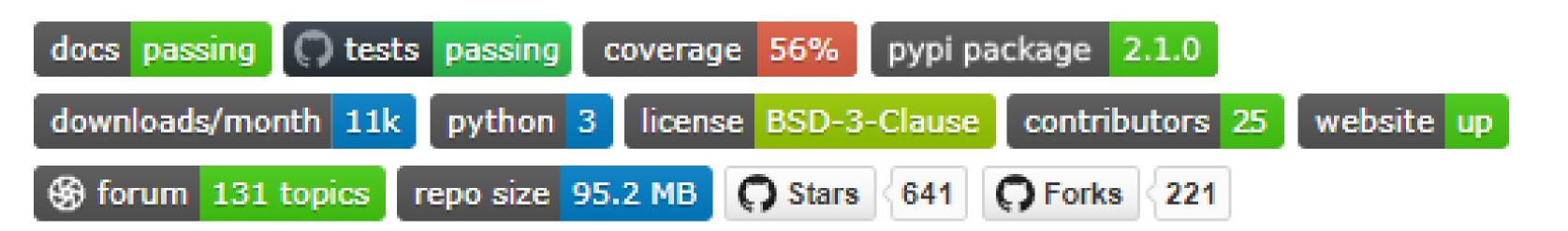








Visit the project's github or gitlab page and review indicators.



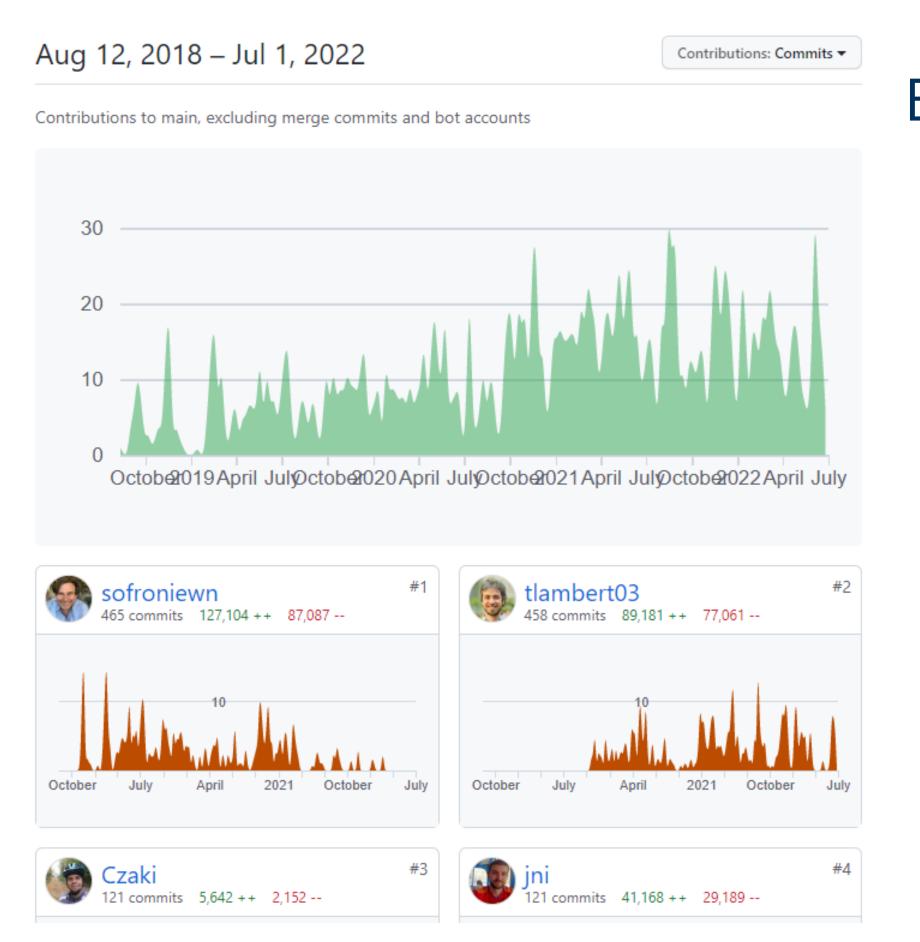






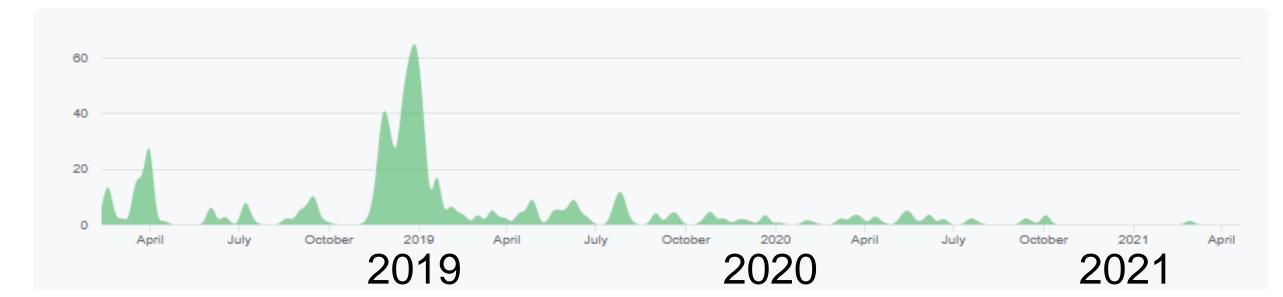


Visit the project's github or gitlab page and review indicators.



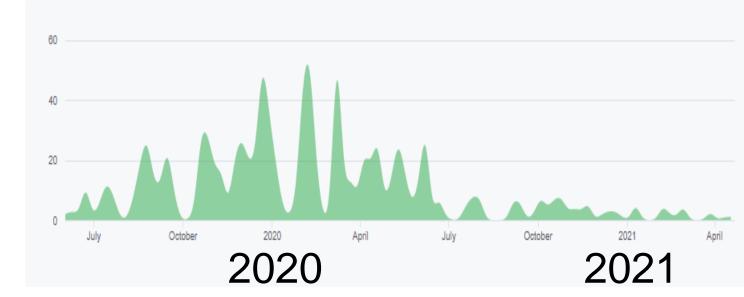
Be a bit careful when development appears to have halted





And check related projects













Crowd sourcing

... ideas, feedback, improvements

Market research



Engage with the community

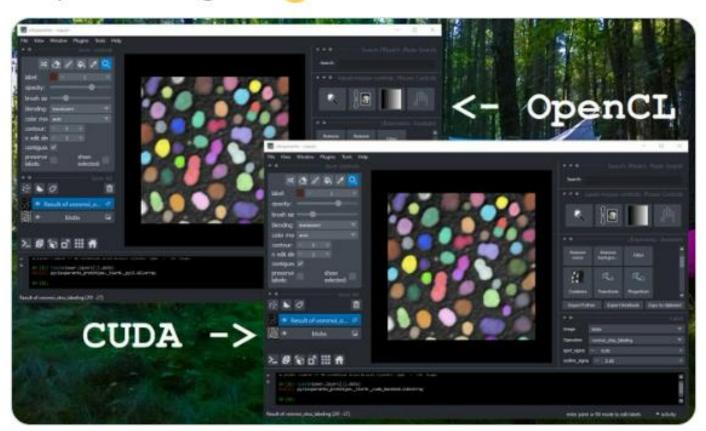
- Use twitter to reach out
- Learn about who needs what



... and we will incorporate #CUDA/@CuPy_Team with #OpenCL/#clesperanto to find new friends who use GPUs differently and learn from them 🚅 💻 🚀

You wonder what will be different between #clesperanto-OpenCL and #clesperanto-CUDA based image processing? 60

I hope not a single bit



5:55 PM · Nov 15, 2021 · Twitter Web App





Communication is key!

My first @napari_imaging plugin is out now! It can measure region properties, plot them, perform kmeans or manual clustering. Huge thanks to #biapol group members for amazing teamwork and all the help @RyanSavill4, @zoccolermarcelo and @haesleinhuepf

napari-hub.org/plugins/napari...



2:56 PM · Nov 15, 2021 · Twitter Web App

74 Retweets 12 Quote Tweets



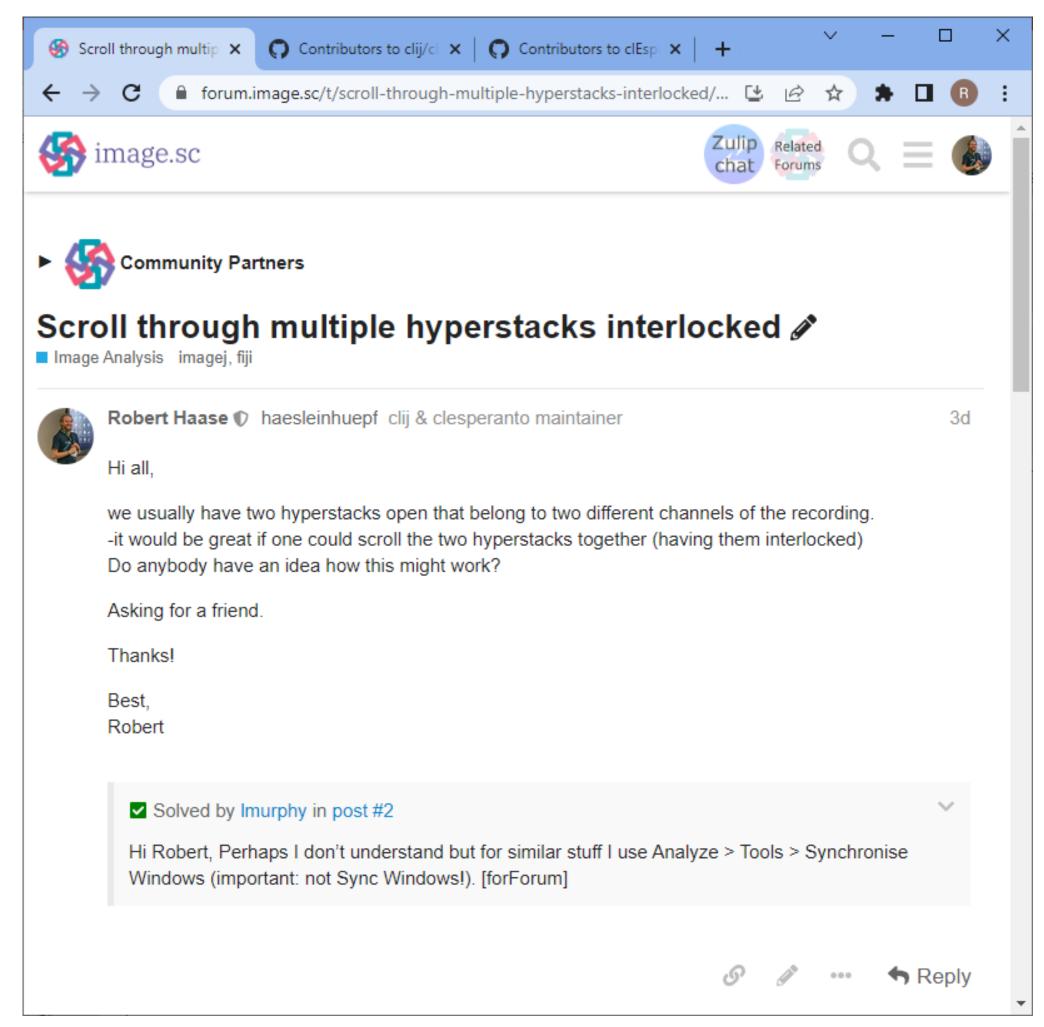


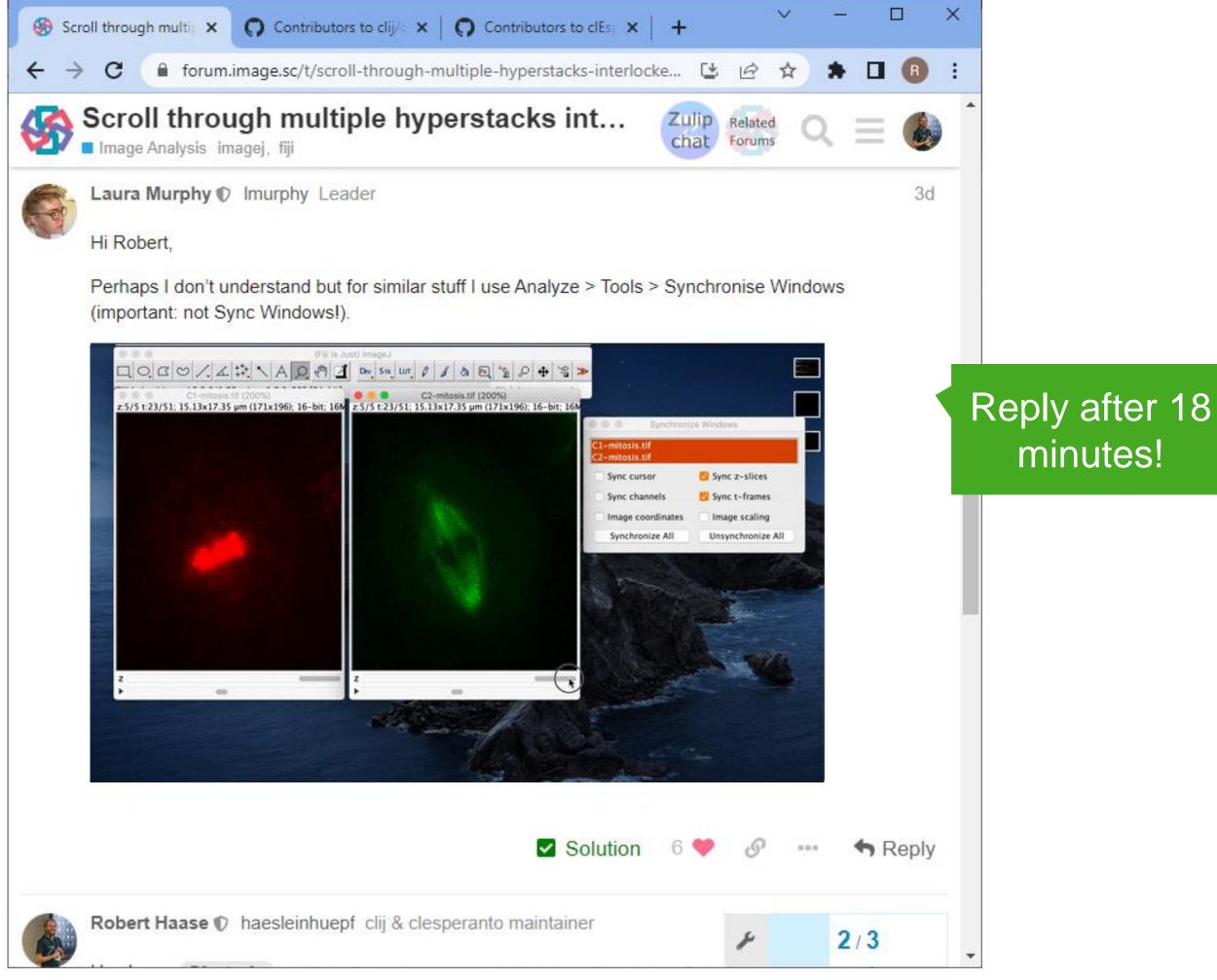




Community involvement





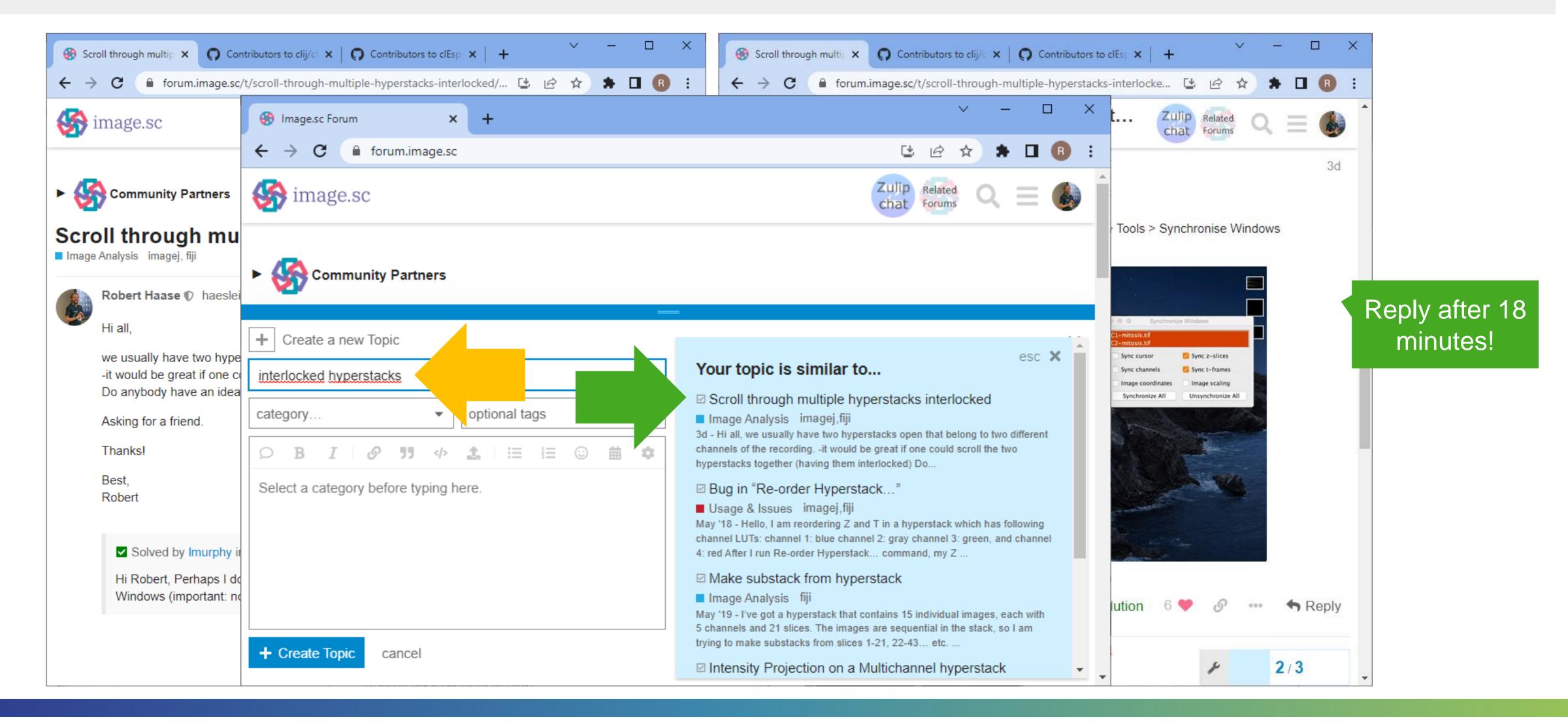






Community involvement









Why public support channels



User's perspective

- Receive feedback from multiple experts
- Also if one expert is on vacation

Developer's perspective

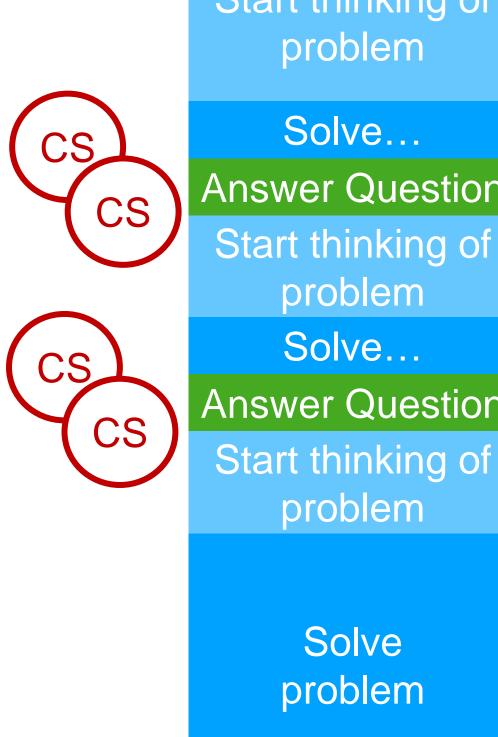
- Reduce repetitive basic questions
- Distribute workload
- Distribute knowledge
- Asynchronous Q&A
- Minimize context-switches

Time

Start thinking of problem

Solve problem

Answer Questions on image.sc



Start thinking of problem

Solve...

Answer Question
Start thinking of problem
Solve...

Answer Question
Start thinking of problem

Start thinking of problem

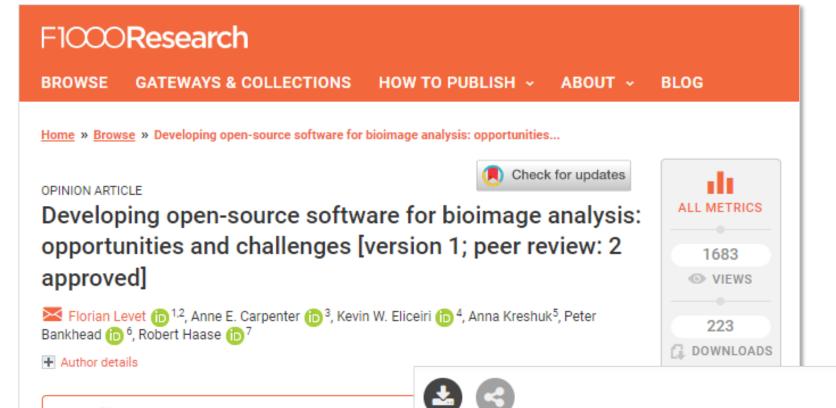






Related publications





PERSPECTIVE article

Cornell University **TYV** > q-bio > arXiv:2110.13951 Quantitative Biology > Other Quantitative Biology [Submitted on 26 Oct 2021] Open microscopy in the life sciences: Quo Vadis? Johannes Hohlbein, Benedict Diederich, Barbora Marsikova, Emmanuel G. Reynaud, Seamus Holden, Wiebke Jahr, Robert Haase, Kirti Prakash

software solutions targeting broad user groups or even non-exp

relevant to cope with the increasing complexity of cutting-edge

We will then extensively discuss the current and future challenge

such, light microscopy has been playing a fundamental role in



hundred years. Fueled by the availability of mass-produced ele icly shared documentation and building instructions, open-source rapid prototyping and 3D printing, and the enthusiasm of contri the concept of open microscopy has been gaining incredible i ophisticated tools to an expanding user base. Here, we will first pen science and open microscopy before highlighting recent pro in open microscopy. We argue that the availability of well-design

py allows observing cellular features and objects with sub-micro Cornell University

TY1V > q-bio > arXiv:2204.07547

Quantitative Biology > Quantitative Methods

[Submitted on 15 Apr 2022]

A Hitchhiker's Guide through the Bio-image Analysis **Software Universe**

Robert Haase, Elnaz Fazeli, David Legland, Michael Doube, Siân Culley, Ilya Belevich, Eija Jokitalo, Martin Schorb, Anna Klemm, Christian Tischer

Modern research in the life sciences is unthinkable without computational methods for extracting, quantifying and visualizing information derived from biological microscopy imaging data. In the past decade, we observed a dramatic increase in available software packages for these purposes. As it is increasingly difficult to keep track of the number of available image analysis platforms, tool collections, components and emerging technologies, we provide a conservative overview of software we use in daily routine and give insights into emerging new tools. We give guidance on which aspects to consider when choosing the right platform, including aspects such as image data type, skills of the team, infrastructure and community at the institute and availability of time and budget

Meeting in the Middle: Towards Successful Multidisciplinary Bioimage Analysis Collaboration

Front. Bioinform., 14 April 2022 | https://doi.org/10.3389/fbinf.2022.889755

🝂 Anjalie Schlaeppi^{1,2*†}, 🌉 Wilson Adams^{3,4†}, 🍢 Robert Haase^{5†}, 🧝 Jan Huisken^{1,6†}, 🎑 Ryan B. MacDonald^{7†}, Kevin W. Eliceiri^{1,8†} and Kevin C. Kugler^{7*†} Morgridge Institute for Research, Madison, WI, United States ²Biolmaging and Optics Platform, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland ³Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, United States ⁴Department of Pharmacology, Vanderbilt University, Nashville, TN, United States ⁵DFG Cluster of Excellence "Physics of Life", Germany and Center for Systems Biology Dresden, TU Dresden, Dresden, Germany Department of Biology and Psychology, Georg-August-University Göttingen, Göttingen, Germany ⁷Faculty of Brain Sciences, Institute of Ophthalmology, University College London, London, United Kingdom



neublos This article is included in the NEUBIAS - the Bio

This article is included in the Bioinformatics ga

This article is included in the Research on Res





⁸Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI, United States

Exercise / discussion



Would you use those libraries?

