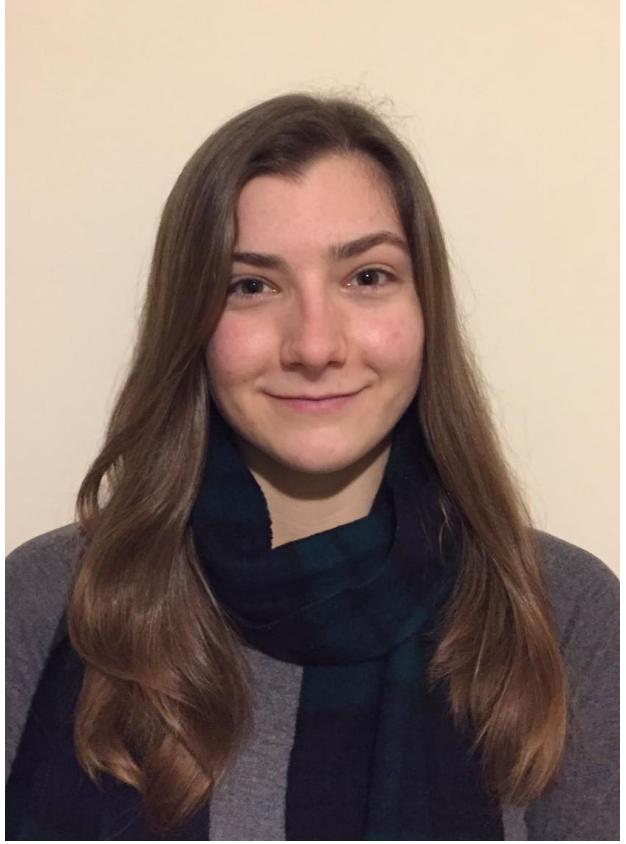
The Benefits of an Open-Science Approach in Student Research Projects

Emma MacKenzie, Sophie Winterbourne, Edward Wallace, Flic Anderson

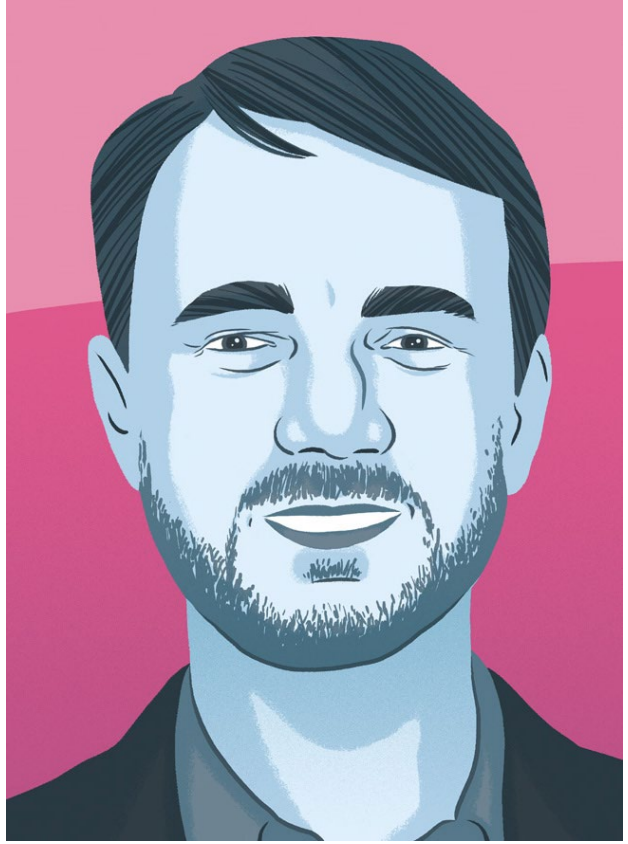
@ Edinburgh Open Research Conference

27/05/22

Emma MacKenzie
Student



Sophie Winterbourne
Student



Dr. Edward Wallace
PI

Felicity Anderson
Day-to-Day Supervisor

Why we're here...

Co-authored an article in The Biochemist:

The benefits of an open-science approach in student research projects (2021) https://doi.org/10.1042/bio_2021_198

Movement to **allow scientific information, data and outputs to be more widely accessible and reusable**. This includes within research groups, where participants share their data, analysis code, ideas and feedback.

riboviz

riboviz is open source software which processes & analyses ribosome profiling (sequencing) data

Developed from 1.0 to 2.0 in **collaborative project** BBSRC-NSF funded project - working with experts in UK & USA

Aims: use **software engineering** techniques to create more **robust** & **reliable** code
-> *more researcher time for biological questions*

Ribosome profiling data helps **unlock details of translation**: how the cell creates proteins from genetic code (DNA, RNA)

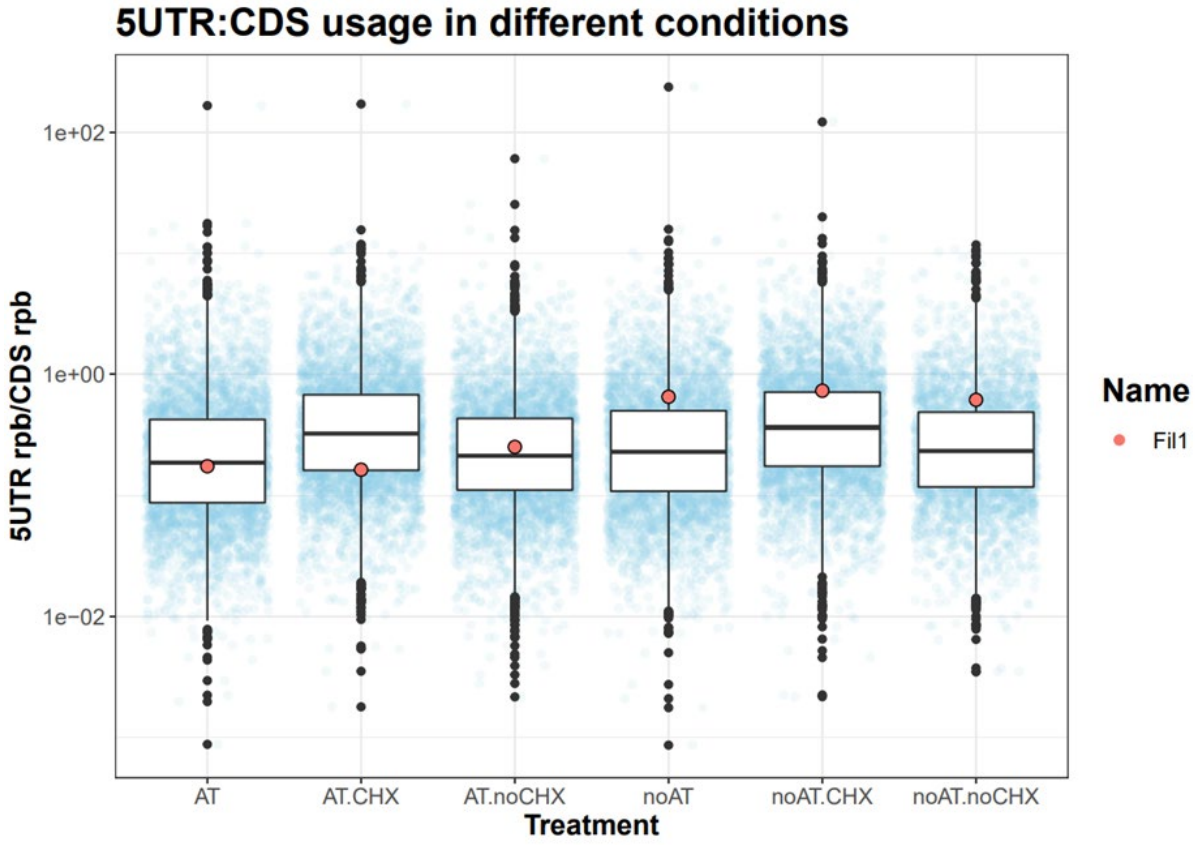
riboviz helps researchers check **quality control** & understand the **characteristics** of the experimental data - ie. do different experimental conditions change how the cell translates proteins from genes?

riboviz 2.0 publication: https://doi.org/10.1093/bioinformatics/btac093 (2021) Bioinformatics
riboviz on GitHub: https://github.com/riboviz/riboviz

Student Project Background

- Wallace lab uses riboviz to analyse translation dynamics in **fungal datasets**
- Several successful student projects so far using riboviz!
- 2(+) heads are better than one:

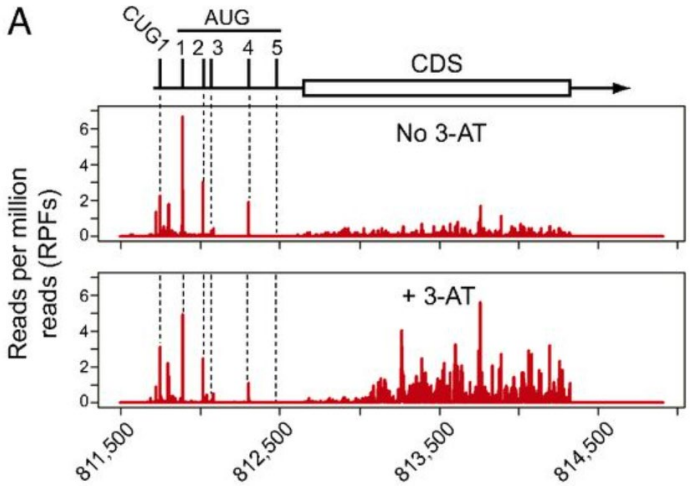
  similar projects = peer learning!
- **Adjusting scopes** as we gain more experience in successful open (software) projects.



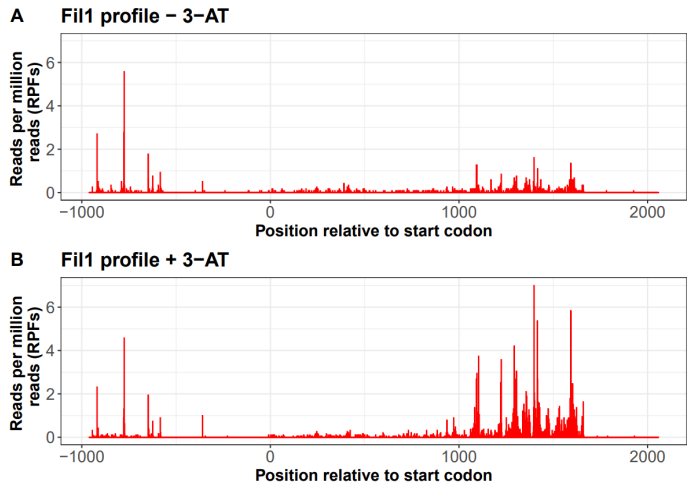
Thesis image. Data reanalysed from Duncan et al. 2018. https://www.pnas.org/doi/full/10.1073/pnas.1713991115

What Makes a Good Student Project?

Published data



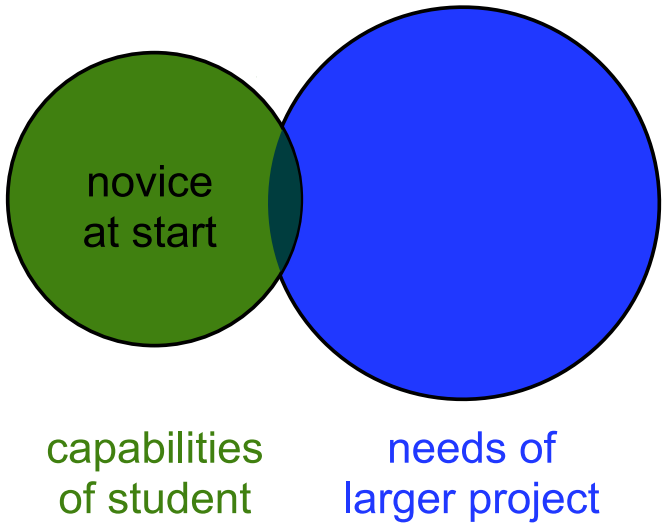
Student data



- **Recreating published results** works well
- Adding **new features** to the codebase to address specific research questions
- Comparing datasets from **different organisms**
- Developing **new visualizations** to explore data
- **Getting credit** is important: all previous students co-authored riboviz 2 publication for contributing to the software

Published data from Duncan et al. 2018 – figure 5. https://www.pnas.org/doi/full/10.1073/pnas.1713991115

Student Skills & Project Requirements (At Kickoff)



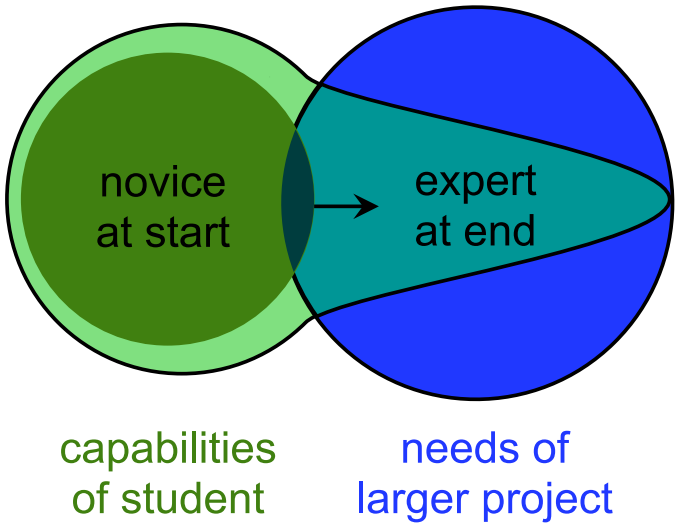
- Students are **expert novices** – let them identify and combat your 'expert blind-spots'!
- They have **time and motivation** to contribute to code, documentation and publications (and get credit for these!)
- **Overlapping goals**: aim for a project where project tasks help progress lab goals
- Plan a **structured pathway** from novice to expert for students, using authentic research tasks...

Structured Skills Pathway: Bioinformatics Example

Skill Level	Example Student Task	Learning / Skills Gained
Novice	**Installing** riboviz software and running built-in 'vignette' small **test dataset**	Installation, setup, understanding output files, troubleshooting
Confident Beginner	**Run** existing full-size example dataset & improve **documentation** as needed	Data size problems, navigating directory structures, contributing to documentation
Competent Practitioner	**Analyse a published dataset** from an existing annotation (organism-specific files already exist)	Learning about adapter removal in sequencing data, other common bioinformatics tasks
Expert	**Identify, adapt and run** a new dataset with a **new genome annotation** (create new organism-specific files)	Tricky task with organism-specific quirks, potential data availability or format issues

Gradually builds skills, confidence and understanding needed to successfully undertake expert-level tasks

Student Skills & Project Requirements (By Wrap-Up)



- Student has **expanded their capabilities** generally...
- ... and is **expert** in their specific project skills!
- Student project has **contributed significantly** to **overall project needs** (Happy PI time!)
- 'Open science' project has **contributed to lab culture** (e.g. skill-sharing, demonstrating new tools, encouraging asking for help when stuck)

Open Tools Help Develop Experts: Invest Early!

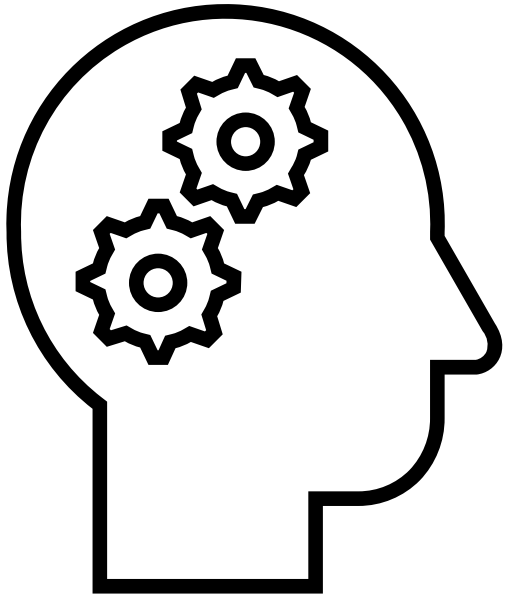
Tool	What's It Good For?
Hackathons	Build momentum, progress leaps
Stand-ups (daily update)	Check-ins & day-to-day planning
Slack (chat-like service)	Informal discussion, questions, sharing successes
Videocall Supervision Meetings	Review progress, in-depth explanations, code demos, troubleshooting, pair-programming
Open Documentation & Courses	Accessible any time, pro-active solution-finding and learning (e.g. Carpentries courses, lab guides, documentation)
Git (version control)	Developing & managing code, pinging scripts back and forth for checking, testing, review, traceability
Issue Tickets @ GitHub	Assigning/investigating issues, asking questions
Kanban Boards @ GitHub	Prioritising/managing project tasks & tracking progress

Professional Skills: Day-To-Day Supervisor

- **Develop teaching/supervisory skills** while progressing my work; students bring new insights & overlapping goals
- **Open Tools are great supervision tools** – easily track what students are working on / stuck with, train and support progress
- Working in (temporarily) larger team, **improve my collaboration and 'open science' skills**
- **Visible skills** - my commits / comments are visible to future employers or collaborators!
- **Communication**: publications, posters, lab meetings, presentations...

'Open' Mindset: Day-To-Day Supervisor

- Keeping an 'Open' Mind:
  - Visibly & frequently **ask questions & seek help**
  - **Sharing ideas:** less reluctant to put own ideas forward
  - **Errors** – showing/explaining avoids reoffending
- **Stand-ups for research:** sharing priorities, achievements and problems helps project planning, brings support
- **Open & inclusive lab culture** – cooperative, welcoming and positive environment encourages **academic and personal growth**



'Open' Mindset: Student's View

- **New** and scary
- **Uncomfortable** admitting mistakes and seeking help
- Only markers saw our work in previous assignments
- Errors or unexpected results were not exciting due to expected outcomes in lab work
- Had to **develop** over the project
- Development is clear in how we communicated problems

Before 'Open' Mindset

- **Confusing** screenshots
- Sent over Slack so **not accessible** to others
- Only providing useful information when **prompted**
- Difficult to identify problem and the solution
- Does not provide **long term documentation** of error and solution

After 'Open' Mindset

- Improved **readability**
- On GitHub - **accessible**
- Provide **useful information** making it easier to solve the problem
- **Long term record** of how problem was overcome
- New mindset, **same mistakes, better solutions**

3mma-mack commented on 25 Jun 2021 — Author

I have tried running the config files with both versions of adapters, however both fail at the step BamToH5. I have looked into the .command.log file for both runs and both contain the following:

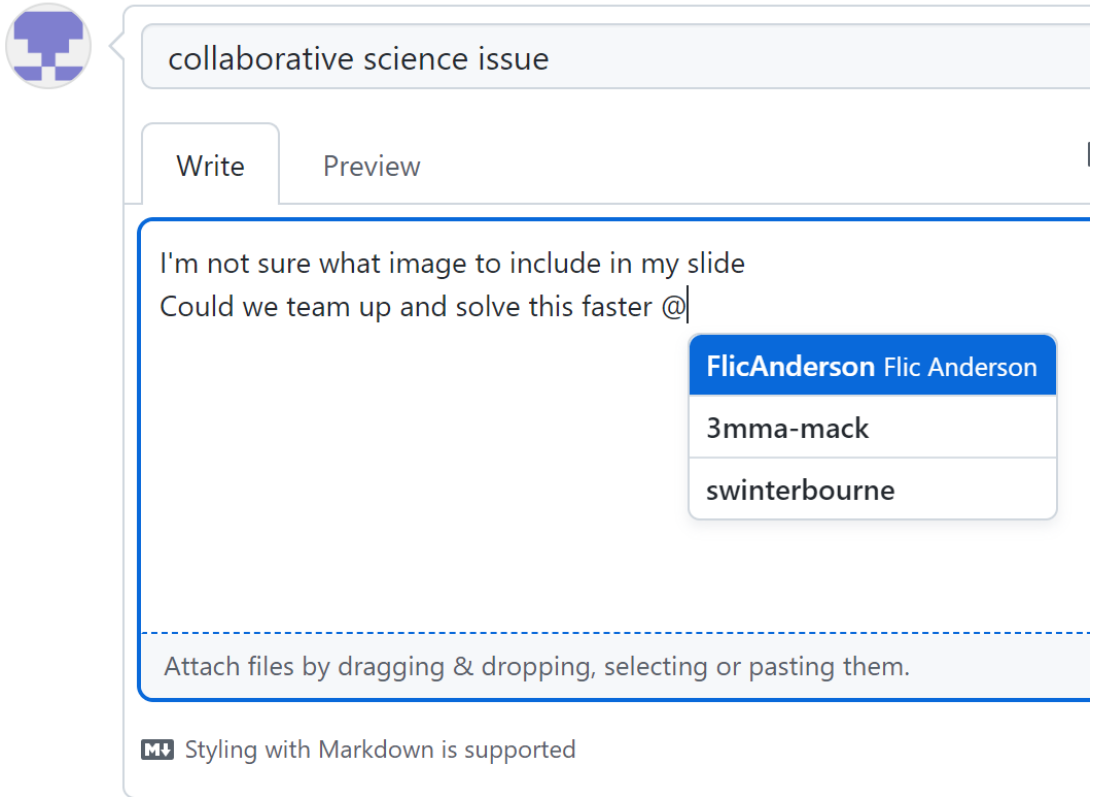
```
Error in library(getopt, quietly = T) :
  there is no package called 'getopt'
Calls: suppressMessages -> withCallingHandlers -> library
Execution halted
```

Having a look, the line library(getopt, quietly = T) appears on line 28 of Bam_to_H5.R, and the package is also present in the wallace_rna/Rlibrary folder on eddie, so I don't understand why it is failing here. I have started a run of the Duncan et al Schizosaccharomyces dataset to see if the error occurs for that dataset as well, as I successfully ran the Duncan et al dataset on the 23rd .

This is being run in the develop branch of riboviz, with the latest commit being 106a204bc64525db8de7d18e99d5421be1246394, on the 21st of June.

Collaborative Learning

- Documenting mistakes **helps future users** if they run into the same error
- Using issue tickets can **solve problems faster** – more knowledgeable people can jump in and help
- Learn from and adapt existing code
- **If there is a mistake to be made, a student will make it**



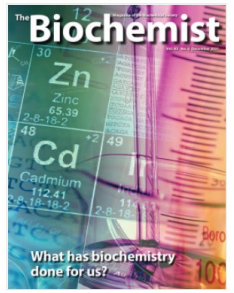
Gain Life and Professional Skills

Before student project	After student project
Poor **communication**, struggling to work collaboratively	Experience working collaboratively, communicating issues effectively
Inconsistent **documentation** of tasks and solutions, need encouragement to provide relevant information	Improved approach to documenting tasks in future projects. Write with knowledge that notes may be used by others in future
Need **guidance** on how to approach tasks	Understanding of processes and able to provide guidance to new users
Uncomfortable making **mistakes**, assume unexpected results are due to own error	Comfortable communicating mistakes and working towards collaborative solutions

Open Research Leads To Positive Outcomes

- Evidence of work
- Good projects
- Data re-analyzed
- Publications
- Conferences

A bioinformatics big-data approach to quantify protein synthesis in fungi: Translational regulation by use of the 5'UTR in Schizosaccharomyces pombe.



riboviz 2: a flexible and robust ribosome profiling data analysis and visualization workflow

Alexander L Cope, Felicity Anderson, John Favate, Michael Jackson, Amanda Mok, Anna Kurowska, Junchen Liu, Emma MacKenzie, Vikram Shivakumar, Peter Tilton, Sophie M Winterbourne, Siyin Xue, Kostas Kavoussanakis, Liana F Lareau ✉, Premal Shah ✉, Edward W J Wallace ✉

Bioinformatics, Volume 38, Issue 8, 15 April 2022, Pages 2358–2360, https://doi.org/10.1093/bioinformatics/btac093
Published: 14 February 2022 **Article history** ▾

Edinburgh Open Research Initiative

FEATURE | DECEMBER 13 2021

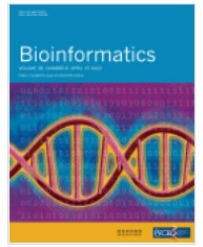
The benefits of an open-science approach in student research projects FREE

Emma MacKenzie ; Sophie Winterbourne ; Felicity Anderson ; Edward Wallace

Check for updates

Biochem (Lond) (2021) 43 (6): 66–73.
https://doi.org/10.1042/bio_2021_198

20



26

A bioinformatics big-data approach to quantify protein synthesis in fungi: investigating translational control by inhibitory codon pairs

Edinburgh ReproducibiliTea
@Edinburgh_Tea

An excellent read detailing the experience of applying #OpenResearch principles to student projects in The Wallace Lab https://t.co/JOe7UTr5E5! https://t.co/BAR8f4Qca6 With insights from students Emma MacKenzie, Sophie Winterbourne, supervisor @Flic_Ande

21 Dec 2021

Thanks for Listening!

Any Questions?

Student Vs Lab?

	Student Project	**Lab Research Project**
Question Scope:	Something **achievable**	Something **novel**
They Aim To:	Learn new skills, get good grades, build research experience	Discover new insights, progress lab research outcomes, adopt new methods
They Need This To:	**Achieve their degree** (=> career goals)	Contribute high-quality **publishable results** (=> secure future funding)
Timescale:	SHORT! 1 x Semester	LONG... Multi-year

Very different goals – how can Open Science approaches help resolve this?

… Think of students as instant collaborators!