

# Exploring Reproducible Science

coding practices & analysis sharing

March 9<sup>th</sup>, 2017

Analise Hofmann

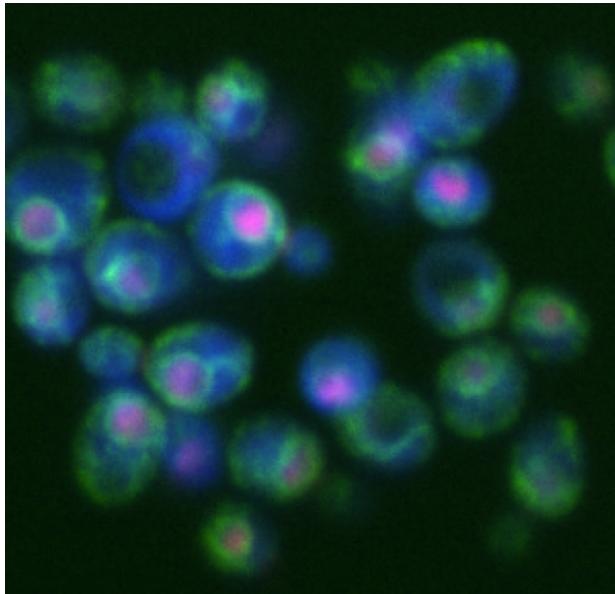
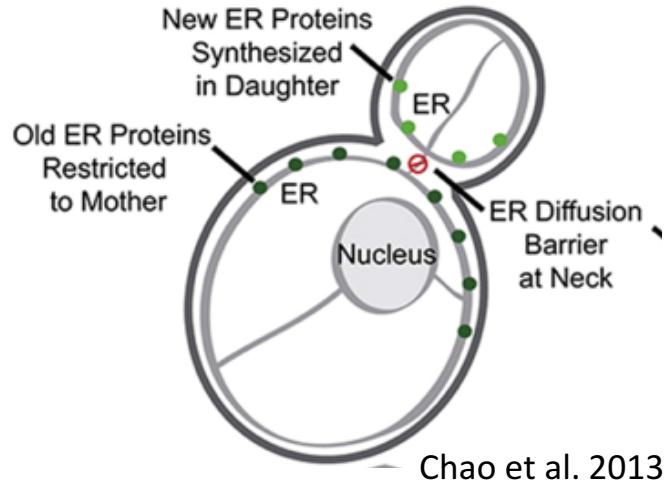
PhD Student, Loewen Lab

Genome Science & Technology

# About Me:



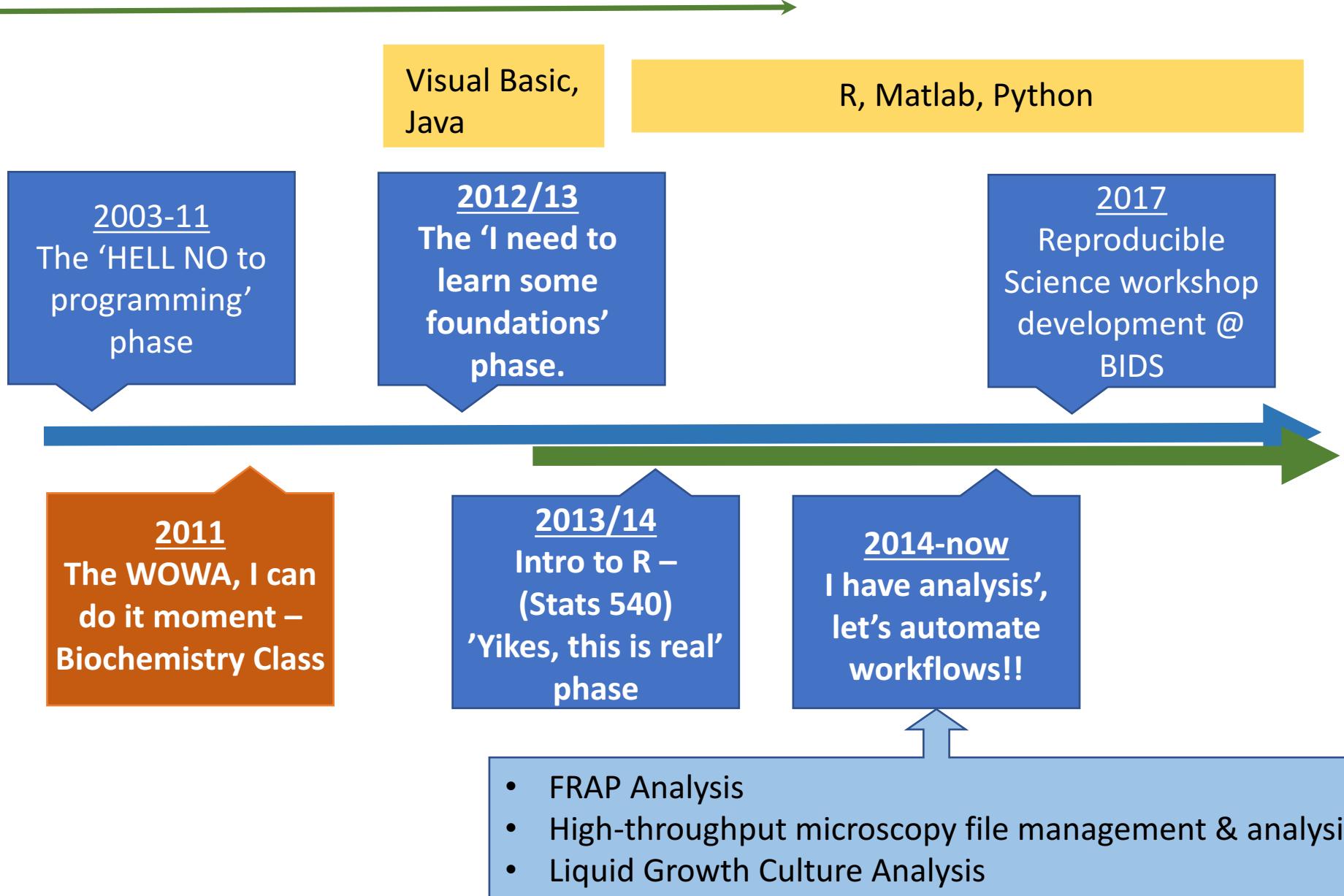
Main fields of research: Cell Biology, Genetics



GitHub: [ahofmann4](https://github.com/ahofmann4)

Twitter: [@yst4Sci](https://twitter.com/yst4Sci)

# My Coding timeline



# Goals of this talk

---

- 
- 1) Provide an overview of on going efforts to make science more reproducible
  - 2) Provide an overview of key coding practices and tools to help with reproducibility efforts
  - 3) Provide examples of resources available at UBC, and online to help you share your work publicly
  - 4) Provide further reading, and upcoming events



Marc Pervès @Marcusinbrux · Mar 6

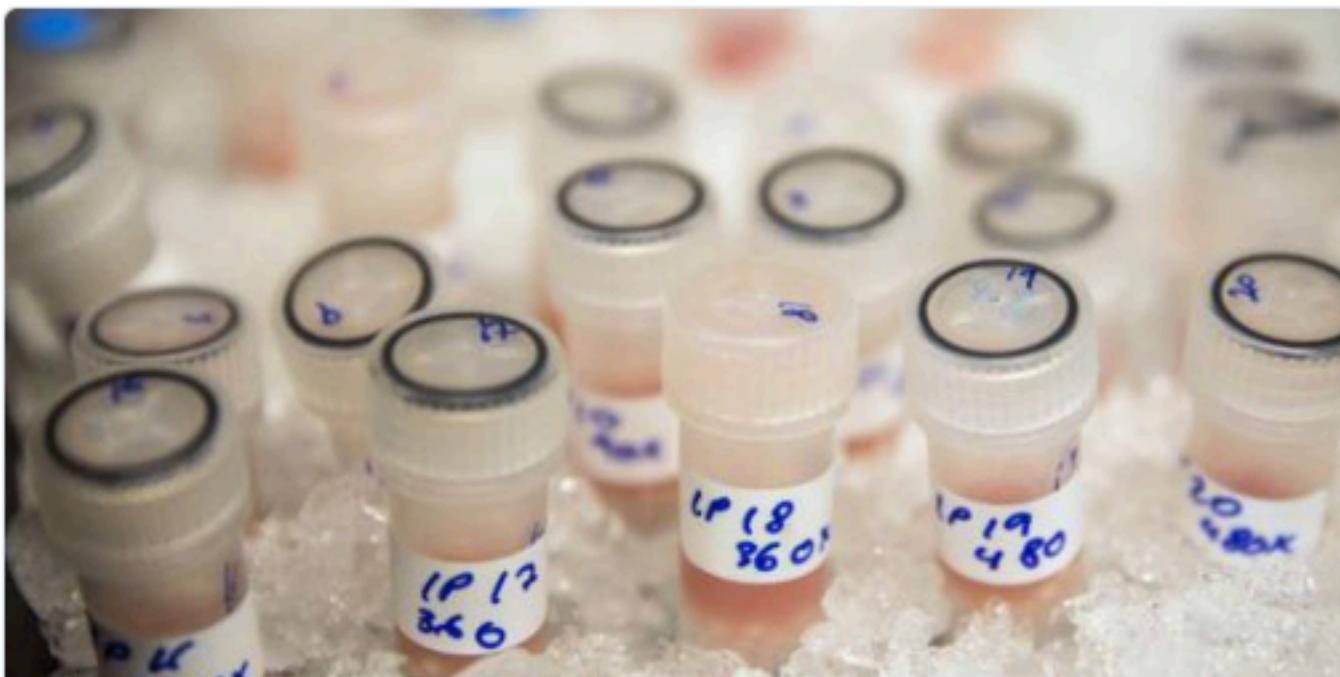
Markram of @FrontiersIn making a strong case for #openscience; picture speaks for itself. Scientific results must be shared #tedxbrussels





Rishabh Jain @rishabhmjain · Feb 26

'2/3 of researchers have... failed to reproduce another scientist's experiments'  
we need #openscience and #scicomm

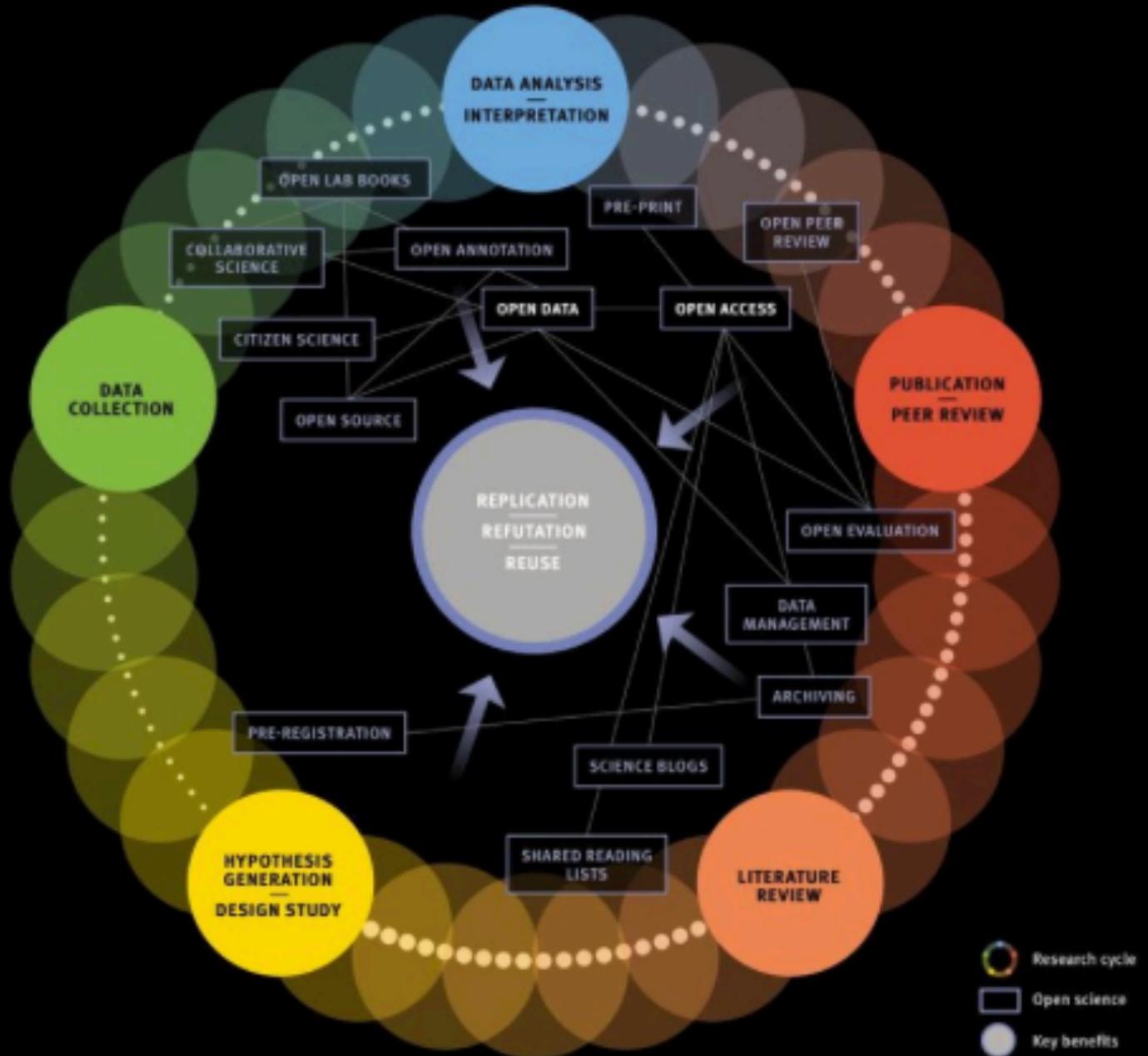


**Most scientists 'can't replicate studies by their peers' - BBC News**

Science is facing a "reproducibility crisis" as scientists fail to reproduce others' work, it is claimed.

[bbc.com](http://bbc.com)







**Julie van der Hoop** @jvanderhoop · 15h

We need to encourage this more -great example to set and model to follow.  
#openscience #opendata @ESIPfed @yolandagil

**Dragonfly** @dflydsci

How to science: recent paper from @sean\_anderson in PNAS, with data, code, and paper all openly available [github.com/seananderson/h...](https://github.com/seananderson/h...)



1



4



Open Science Retweeted



**Ra Ramana** @rravi · Feb 27

Open Design 3D-Printable Micropipette meeting ISO Standard accuracy  
[biorxiv.org/content/early/...](https://biorxiv.org/content/early/...) #oshw #openscience Now this is gd #diybio



### Open Design 3D-Printable Adjustable Micropipette ...

Scientific communities are drawn to the open source model as an increasingly utilitarian method to produce and share work. Initially used as a means to develop ...

[biorxiv.org](https://biorxiv.org)



5



5



Open Science follows



**Andrew Piper @\_akpiper** · 9h

How \*everyone\* should report more about their research process and not just what worked. **#openscience**

the paper. What is not included is how many incorrect tests were made that found nothing or that found results that were the opposite of the final set of conclusions. We need a track record—a log of statistical exploration. How many equations did you estimate? What proportion of them found something different than what you reported? Can you name them as well as the ones that you now feel are the “right” ones? And please explain why the ones you think are the right ones are not wrong. The ideal would be an econometric Do-It-That would allow

**Russell Roberts @EconTalker**

How economists should report their empirical findings:  
[medium.com/@russroberts/w...](https://medium.com/@russroberts/w...)



1



7



18



OpenCon and 2 others follow



**Dario Taraborelli @ReaderMeter** · 15h

**#openscience** needs unrestricted access to citations. Many people have been hard at work to get us here, stay tuned for more announcements.

**Alex@nder Grossmann @SciPubLab**

+1 @SpringerNature becomes largest publisher to #open up all reference lists to make mining of databases easier: [stm-publishing.com/springer-natur... #OA](https://stm-publishing.com/springer-natur...)



11

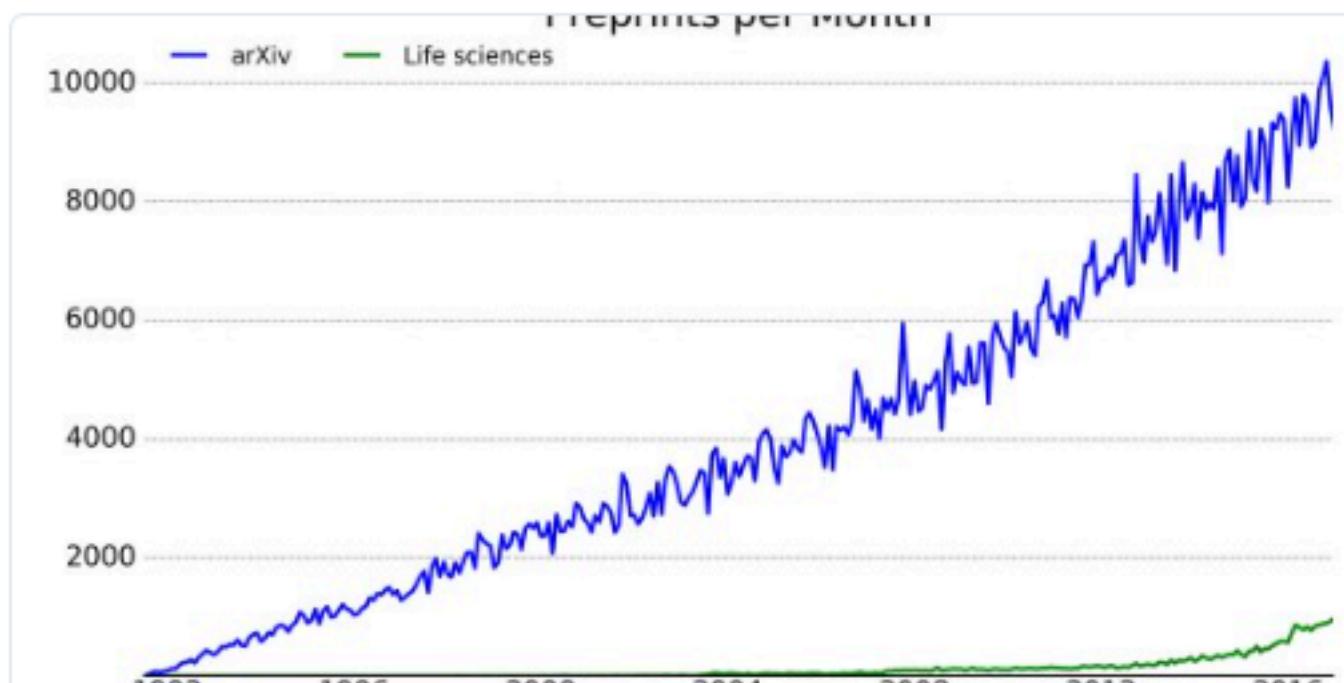


15

 OpenCon and 3 others follow

 Authorea  @authorea · Feb 26

Want to #preprint in another language besides English? Do so at Authorea.  
#openscience



### Introducing the 21st-century preprint: HTML, versioned, citable, data...

Authorea was founded with the mission of improving how researchers write and publish their findings. We created a platform that allows researchers t...

[authorea.com](http://authorea.com)



4



5

# Defining Reproducible Science

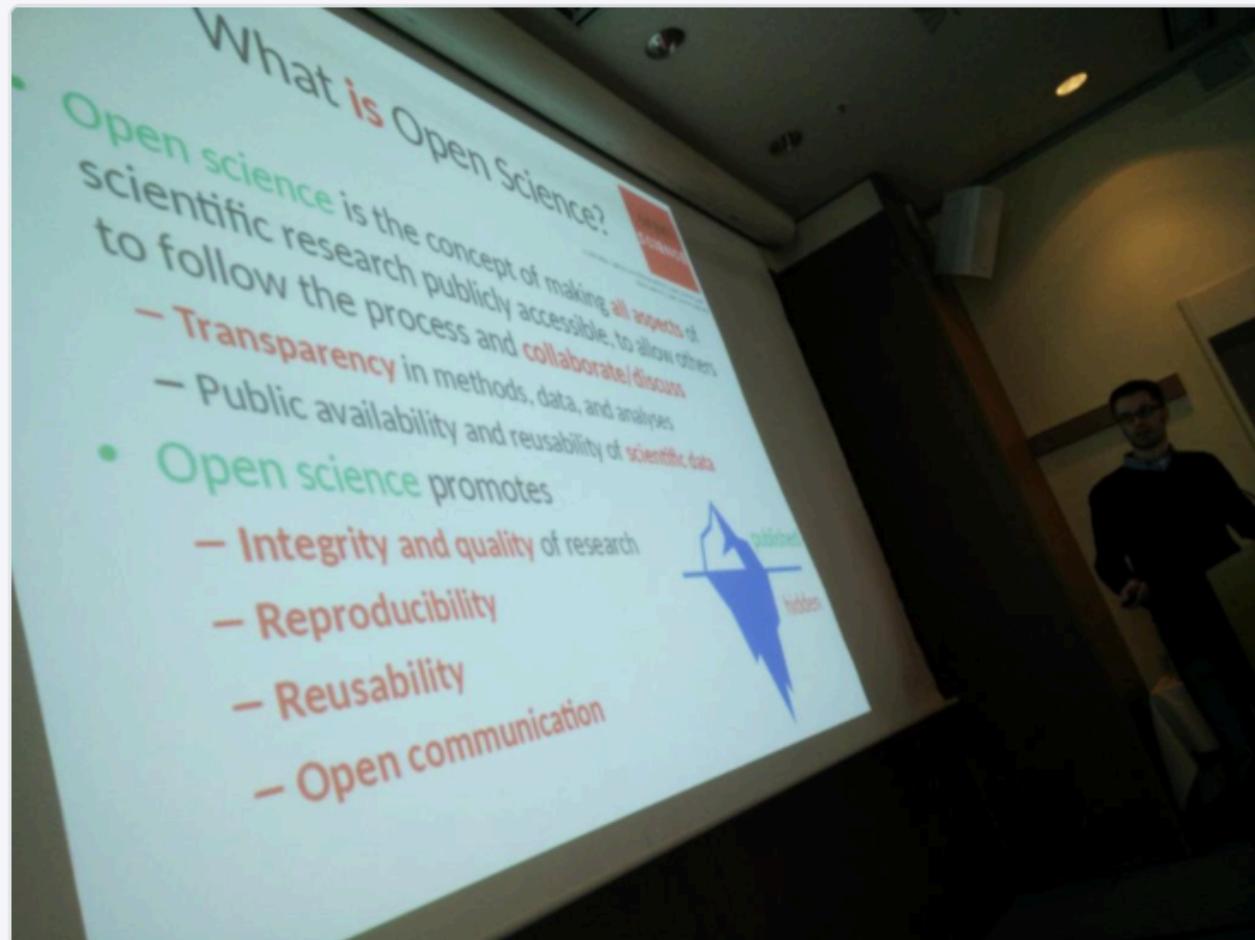


Open Science Retweeted



**Konrad Förstner** @konradfoerstner · 22h

@aleimba in action during the #OpenScience workshop #VAAM/#DGHM



# Basic Programming Syntax – follow it!



- File naming conventions
- Variable naming conventions
- Code spacing conventions
- Code documentation
- DRY programming – use loops, functions, etc
- Consider saving your functions into a R package or equivalent to pass on to other lab members!

# R Markdown Files – an easy step!

---

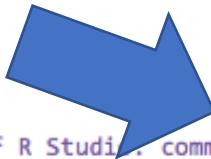
```
---
```

```
title: "Day_1_Recap"
author: "Analise Hofmann"
date: "January 26, 2017"
output: html_document
---
```

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

# Intro to R Studio
- demonstrated the different parts of R Studio: command on the command line, Git - how to commit and push changes up, Packages - where you can find the packages you have documentation on commands, Viewer - use to look at knit - made a new R Project with a github repository

## R Markdown Review of template when you open a new .Rmd file
**We also looked at a Markdown [cheat sheet](https://g
```



| title       | author          | date             | output        |
|-------------|-----------------|------------------|---------------|
| Day_1_Recap | Analise Hofmann | January 26, 2017 | html_document |

```
knitr::opts_chunk$set(echo = TRUE)
```

## Intro to R Studio

- demonstrated the different parts of R Studio: command line, Environment (where to see your variables record of code run on the command line, Git - how to commit and push changes and work, Files - use to present, Plots - where graphs will pop up, Packages - where you can find the packages you have installed ones loaded and 'active'), Help - use to search help documentation on commands, Viewer - use to look at files (from .Rmd files) )
- made a new R Project with a github repository

## R Markdown Review of template when you open a new .Rmd file

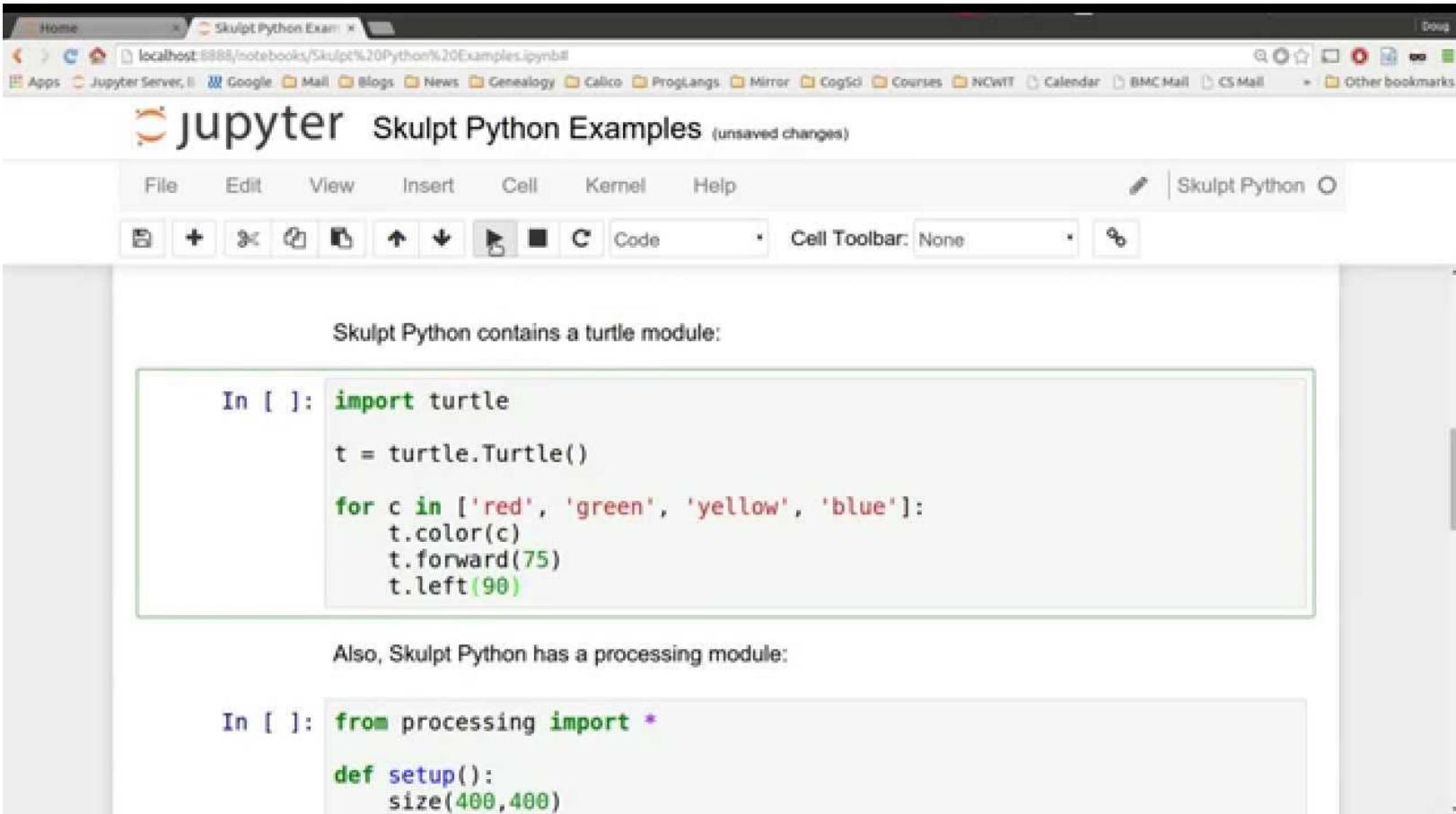
We also looked at a Markdown [cheat sheet](#)

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the Knit button a document will be generated that includes both content as well as the embedded R code chunks within the document. You can embed an R code chunk like this:

# Jupyter Notebooks - an easy step!

Easiest way to start → download Anaconda



The screenshot shows a Jupyter Notebook interface with the title "jupyter Skulpt Python Examples (unsaved changes)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, and Help. The toolbar below has icons for New, Open, Save, Run, Cell Type, Cell Toolbar, and Cell Kernel. A code cell is selected, containing the following Python code:

```
In [ ]: import turtle  
  
t = turtle.Turtle()  
  
for c in ['red', 'green', 'yellow', 'blue']:  
    t.color(c)  
    t.forward(75)  
    t.left(90)
```

Below the code cell, a text cell contains the following text:

Also, Skulpt Python has a processing module:

```
In [ ]: from processing import *  
  
def setup():  
    size(400,400)
```

Reproducible Data Analysis in

Jupyter: <http://www.youtube.com/playlist?list=PLYCpMb24GpOC704uO9svUrihl-HY1tTJJ>

UBC's online Jupyter Notebooks  
Server: <https://ubc.syzygy.ca>

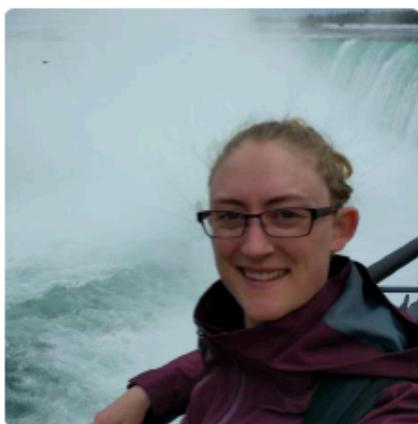
# Version Control: GitHub – do not avoid!



Features Explore Pricing

Search GitHub

Sign in or Sign up



Overview

Repositories 4

Stars 0

Followers 3

Following 21

## Pinned repositories

[XX\\_CodingClub](#)

working towards creating efficient coding workflows

HTML ★ 1

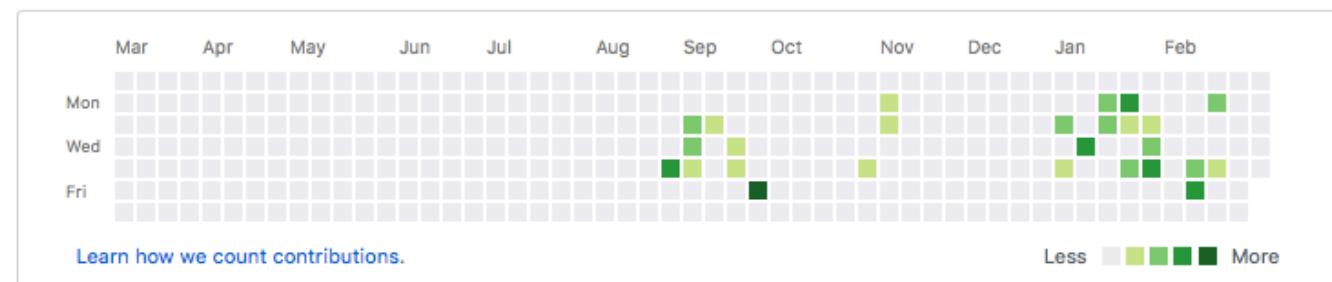
## Analise Hofmann

ahofmann4

PhD Candidate - Genome Science & Technology, Loewen Lab, The University of British Columbia

[Block or report user](#)

93 contributions in the last year



## Contribution activity

March 2017

Jump to ▾

2017

2016

# GitHub



ahofmann4 / XX\_CodingClub

Watch 4 Star 1 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Pulse Graphs

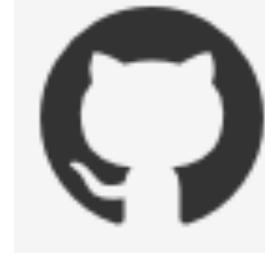
working towards creating efficient coding workflows

15 commits 1 branch 0 releases 3 contributors

Branch: master New pull request Find file Clone or download

|                     | ChiaraToselli Merge branch 'master' of https://github.com/ahofmann4/XX_CodingClub | Latest commit ee8bc6c 14 days ago |
|---------------------|---|-----------------------------------|
| Day_03_WorkAlong    | Merge branch 'master' of https://github.com/ahofmann4/XX_CodingClub               | 14 days ago                       |
| .gitignore          | minor cleaning  | a month ago                       |
| Day_01_Recap.Rmd    | recap of day 1  | a month ago                       |
| Day_01_Recap.html   | recap of day 1  | a month ago                       |
| Day_02_Recap.Rmd    | create new recap day 2  | a month ago                       |
| ReadMe.md           | day & time added  | 17 days ago                       |
| XX_CodingClub.Rproj | recap of day 1  | a month ago                       |

# Introducing Zenodo!



(All) Research.  
Shared.

## **— your one stop research shop!**

all research outputs from across all fields of research are welcome! Zenodo accepts any file format as well as both positive and negative results. We choose to promote peer-reviewed openly accessible research, and we curate the uploads posted on the front-page.

Citeable.  
Discoverable.

## **— be found!**

Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. Zenodo further supports harvesting of all content via the OAI-PMH protocol.

Community  
Collections

## **— create your own repository**

Zenodo allows you to create your own collection and accept or reject uploads submitted to it. Creating a space for your next workshop or project has never been easier. Plus, everything is citeable and discoverable!

## Intrc

(All) Research.  
Shared.

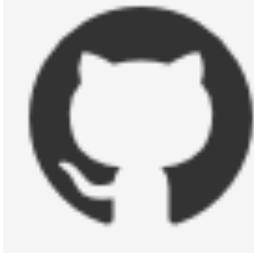
— your one stop research shop!

all research outputs from across all fields of research are welcome! Zenodo accepts any file format as well as both positive and negative results. We choose to promote peer-reviewed openly accessible research, and we curate the uploads posted on the front-page.

### Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** — because not everything is under Creative Commons.
- **Safe** — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

odo!



### Community Collections

— create your own repository

Zenodo allows you to create your own collection and accept or reject uploads submitted to it. Creating a space for your next workshop or project has never been easier. Plus, everything is citeable and discoverable!



British Columbia Research Libraries'  
Data Services



Hosted at the University of  
British Columbia

**UBC Library offers a robust data management software to host your data:**

### UBC Abacus Dataverse

An open source application to publish, share, reference, extract and analyze research data. UBC Library can create a Dataverse setup for your research group where you can actively manage your data, share it securely with colleagues, provide version control, establish a permanent data citation, get a DOI for your dataset, and more. All Dataverse files will be digitally preserved (coming soon).

Wondering about the difference between Abacus Dataverse and cIRcle? Try [cIRcle Dataverse Data Guidelines](#).

### **Learn more:**

- [Publish your Data](#)
- [License your Data](#)
- [Support for the Tri-Agency Statement of Principles on Digital Data Management in UBC](#)
- [Cite and get credit for your Data](#)

<http://dvn.library.ubc.ca/dvn/>

<http://researchdata.library.ubc.ca/share/>

# NSERC & CIHR Requirements



As publicly funded organizations, the agencies are strong advocates for making the results of the research they fund as accessible as possible. In promoting access to research results, they aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. These aspirations align with the Government of Canada's commitment to open science, as described in *Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation*(2014).

Governments and research funders across the globe are becoming increasingly aware of the value of digital research data, the importance of fostering reuse of digital research data and the need for policies to enable excellence in data stewardship. Canada has joined many other countries at the forefront of this movement, as shown in its support for the Organisation for Economic Co-operation and Development's *Declaration on Access to Research Data from Public Funding* (2004); its commitment to the *Open Government Declaration* (2011); and its approval of the *G8 Science Ministers Statement* (2013).

# NSERC & CIHR Requirements



Responsibilities of **researchers** include:

- incorporating data management best practices into their research;
- developing data management plans to guide the responsible collection, formatting, preservation and sharing of their data throughout the entire lifecycle of a research project and beyond;
- following the requirements of applicable institutional and/or funding agency policies and professional or disciplinary standards;
- acknowledging and citing datasets that contribute to their research; and
- staying abreast of standards and expectations of their disciplinary community.

### **3.1 Peer-reviewed Journal Publications**

Grant recipients are required to ensure that any peer-reviewed journal publications arising from Agency-supported research are freely accessible within 12 months of publication. Recipients can do this through one of the following routes:

**a. Online Repositories**

Grant recipients can deposit their final, peer-reviewed manuscript into an institutional or disciplinary repository that will make the manuscript freely accessible within 12 months of publication. It is the responsibility of the grant recipient to determine which publishers allow authors to retain copyright and/or allow authors to archive journal publications in accordance with funding agency policies.

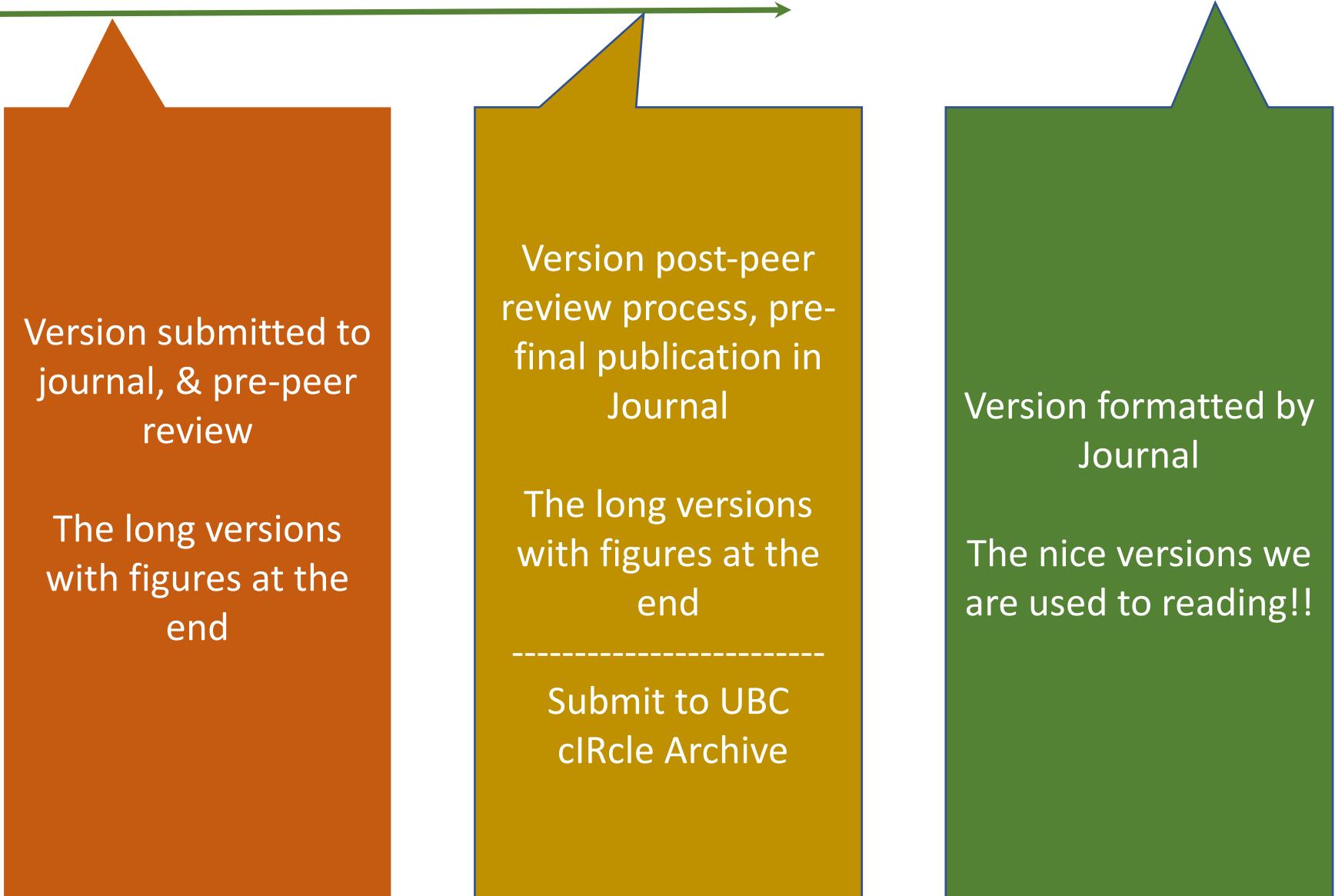
**b. Journals**

Grant recipients can publish in a journal that offers immediate open access or that offers open access on its website within 12 months. Some journals require authors to pay article processing charges (APCs) to make manuscripts freely available upon publication. The cost of publishing in open access journals is an eligible expense under the [Use of Grant Funds](#).

These routes to open access are not mutually exclusive. Researchers are strongly encouraged to deposit a copy of the final, peer-reviewed manuscript into an accessible online repository immediately upon publication, even if the article is freely available on the journal's website.

Grant recipients must acknowledge Agency contributions in all peer-reviewed publications, quoting the funding reference number (e.g. FRN, Application ID).

# Preprints vs. Postprints vs. Published



# Research Data Management

## Learn

All about research data management.

## Plan

Meet grant requirements and preserve your data.

## Share

Share your data with the world.

## Find

Search and browse existing data sets.

**UBC Library is here to help our faculty, students and staff manage their research data.**

Select a topic above to learn more, jump directly to a Data Management tool below, or [contact us](#) for assistance

DMP Assistant

Get DOIs

UBC Abacus  
Dataverse

# Support for the Tri-Agency Statement of Principles on Digital Data Management in UBC

- **Data Management Repository** — UBC Library has implemented robust data management software – [Abacus Dataverse](#) – collaborating with Harvard and supporting other BC schools (UNBC, UVic and SFU). The system is designed to manage and preserve data and it is opened to UBC researchers, labs and institutes.
- **Data Management Guidance** — [Research Data Management Website](#) – is a valuable tool for researchers to learn about data management, specifically at UBC. Open access materials were developed to provide training for UBC researchers, faculty and students. Please see our [DataGuide](#) to get started and data privacy and security best practices [document](#) that outlines key considerations for researchers when working with sensitive data and Personal Information. Workshops are offered monthly at the library [HERE](#).
- **Data Management Plans** — Our national [DMP Assistant](#) software – is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management. Hosted by the Canadian national [Portage](#) initiative, DMP Assistant is designed to meet the anticipated Data Management Plan requirements (in English or French) of most major Canadian funders.

# Support for the Tri-Agency Statement of Principles on Digital Data Management in UBC

- **Discoverability of Data** — The Library is working to increase the visibility of research datasets already added to the UBC's data repository. These datasets can be discovered through many search interfaces, including Google and Summon. We have embarked on a project to assign DOI's to all UBC Library digital assets including datasets, via our new Open Collections portal – <https://open.library.ubc.ca/>. DOIs will increase the further citability and discoverability of UBC research data.

# Impact Story



Ethan White

University of Florida Associate Professor

open access 90%

OVERVIEW

ACHIEVEMENTS

TIMELINE

PUBLICATIONS

## ACHIEVEMENTS

[view all](#)



### Wikitastic

Your research is mentioned in 6 Wikipedia articles! Only 6% of researchers are this highly cited in Wikipedia.



### Open Access

89% of your research is free to read online. This level of availability puts you in the top 8% of researchers.



### Hot Streak

People keep talking about your research. Someone has shared your research online every month for the last 58 months. That's a sharing

## TIMELINE

[view all](#)

4060

Online mentions  
over 1 years

3.7k 126 120 36 8 7 6 6 4

## PUBLICATIONS

[view all](#)

### Best Practices for Scientific Computing

2014 PLoS Biology

2165 W

### The Case for Open Preprints in Biology

2013

527 W

### Elevating The Status of Code in Ecology

2016 Trends in Ecology & Evolution

166

# Upcoming events

---

- ECOSCOPE Workshop: Reproducible Science: March 20, 22 + 24
- UBC Library Creative Commons Workshops – on going
- CSV Conference: May 2-3, Portland Oregon
- Jupyter Notebooks Conference – JupyterCon August 22-25, 2017 New York

## On-going Meet-ups at UBC

- UBC R Study Group (<https://github.com/minisciencegirl/studyGroup>)
- Python Machine learning (Patrick Walls, [pwalls@math.ubc.ca](mailto:pwalls@math.ubc.ca))
- Matlab (Jeff LeDue, [jledue@gmail.com](mailto:jledue@gmail.com))
- XX\_Coding Club ([github.com/ahofmann4/XX\\_CodingClub](https://github.com/ahofmann4/XX_CodingClub))

# Suggested Resources

---

- <http://ropensci.github.io/reproducibility-guide/>
- Twitter: @OpenScience @ ROpenSci  
@RLadiesGlobal @DataCarpentry  
#OpenScience #OpenData #OpenResearch ...
- UBC Librarians
  - Subject Librarians
  - Data Management Librarians

## Community Page

# Best Practices for Scientific Computing

**Greg Wilson<sup>1\*</sup>, D. A. Aruliah<sup>2</sup>, C. Titus Brown<sup>3</sup>, Neil P. Chue Hong<sup>4</sup>, Matt Davis<sup>5</sup>, Richard T. Guy<sup>6✉</sup>, Steven H. D. Haddock<sup>7</sup>, Kathryn D. Huff<sup>8</sup>, Ian M. Mitchell<sup>9</sup>, Mark D. Plumbley<sup>10</sup>, Ben Waugh<sup>11</sup>, Ethan P. White<sup>12</sup>, Paul Wilson<sup>13</sup>**

**1** Mozilla Foundation, Toronto, Ontario, Canada, **2** University of Ontario Institute of Technology, Oshawa, Ontario, Canada, **3** Michigan State University, East Lansing, Michigan, United States of America, **4** Software Sustainability Institute, Edinburgh, United Kingdom, **5** Space Telescope Science Institute, Baltimore, Maryland, United States of America, **6** University of Toronto, Toronto, Ontario, Canada, **7** Monterey Bay Aquarium Research Institute, Moss Landing, California, United States of America, **8** University of California Berkeley, Berkeley, California, United States of America, **9** University of British Columbia, Vancouver, British Columbia, Canada, **10** Queen Mary University of London, London, United Kingdom, **11** University College London, London, United Kingdom, **12** Utah State University, Logan, Utah, United States of America, **13** University of Wisconsin, Madison, Wisconsin, United States of America

## Elevating The Status of Code in Ecology

K.A.S. Mislan  , Jeffrey M. Heer, Ethan P. White

Published Online: December 15, 2015

Open Access DOI: <http://dx.doi.org/10.1016/j.tree.2015.11.006>

 Article Info



Summary

Full Text

Images

References

Code is increasingly central to ecological research but often remains unpublished and insufficiently recognized. Making code available allows analyses to be more easily reproduced and can facilitate research by other scientists. We evaluate journal handling of code, discuss barriers to its publication, and suggest approaches for promoting and archiving code.

# Questions / Discussion

---

Together, we can make science more accessible and  
reproducible!