

# Reproducible Document Stack project

NAOMI PENFOLD  
Innovation Officer

**Slides available at**

<https://figshare.com/articles/Presentation-GA4GH-ContainersWorkflowsTaskTeam-Oct2017/5458882>



This work is licensed under a Creative Commons Attribution 4.0 International License.

# About eLife

@eLifeInnovation

innovation@elifesciences.org



eLIFE

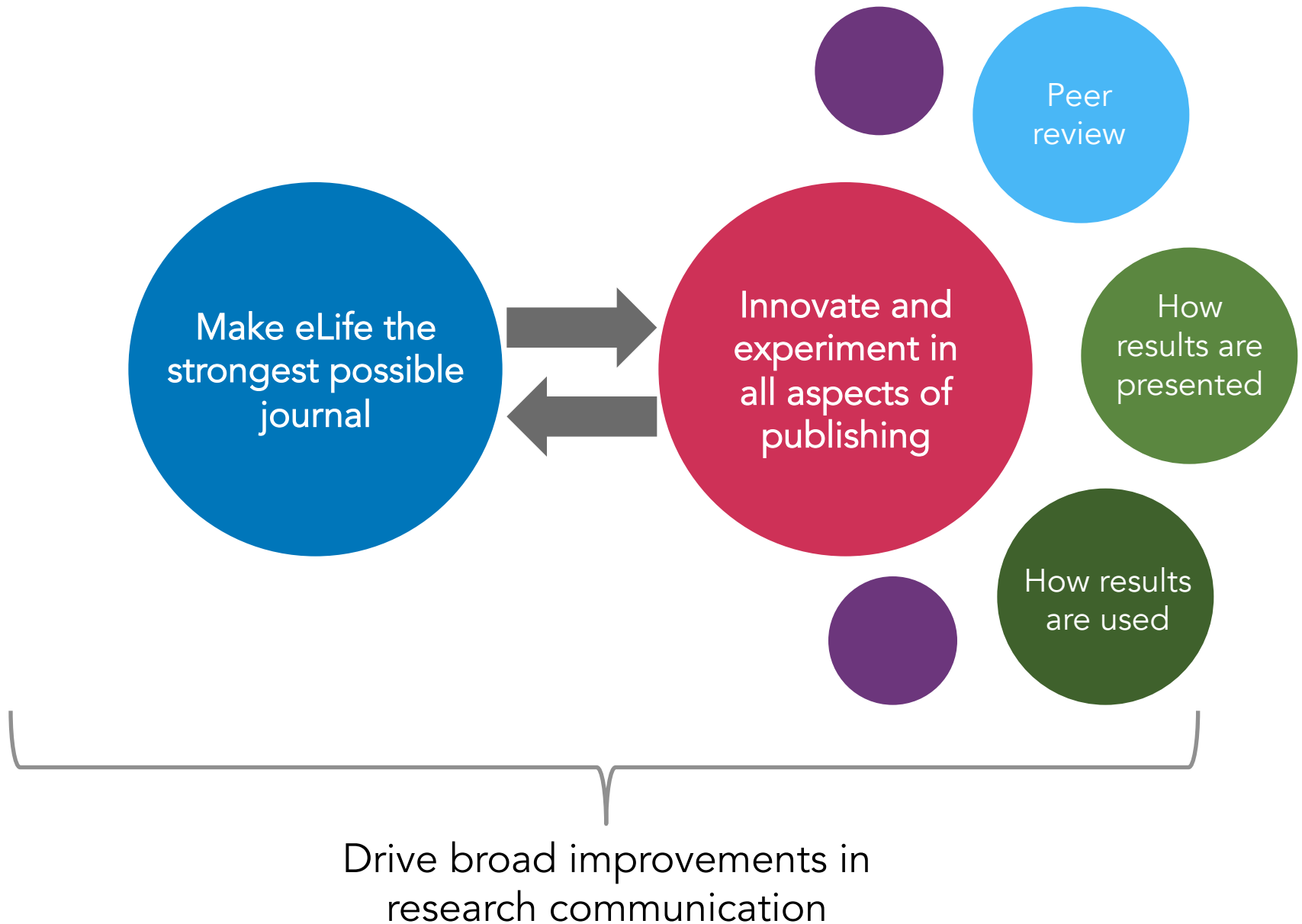
Helping scientists **accelerate discovery** by  
operating a platform for research **communication**  
that encourages and recognises **the most**  
**responsible behaviours in science**

## What do we mean by “responsible behaviours”?

- Sharing of data, tools, and resources
- Objective and comprehensive reporting
- Cooperation and collaboration
- Constructive feedback and encouragement

## eLife Innovation Initiative

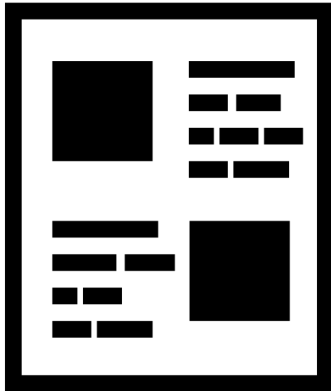
We invest in open source technologies, tools and processes that improve the way cutting-edge research is discovered, shared, consumed and evaluated



# **Publishing reproducible research documents**

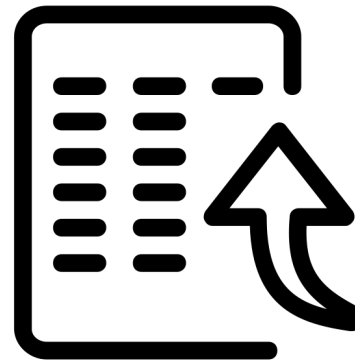


# Current situation



Created by Dmitry Podluzny  
from Noun Project  
[CC-BY 3.0](#)

+



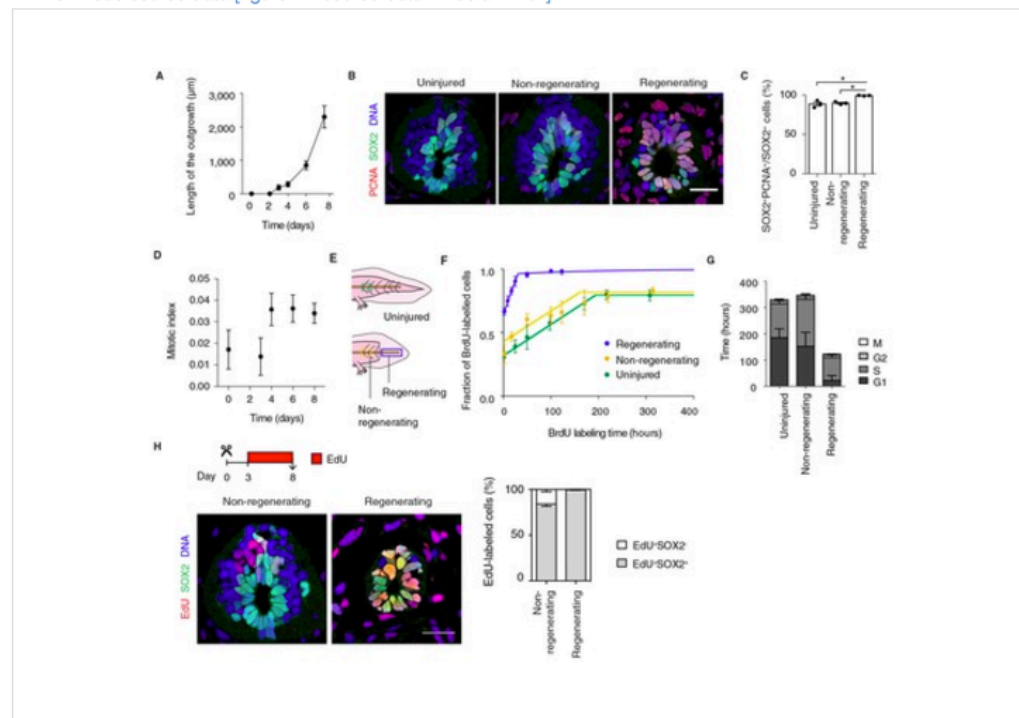
Created by IconfactoryTeam  
from Noun Project  
[CC-BY 3.0](#)

## Figure 1—source data 1

Gene expression changes in the regenerating axolotl spinal cord compared to the uninjured axolotl spinal cord.

DOI: <http://dx.doi.org/10.7554/eLife.10230.004>

[Download source data \[figure-1—source-data-1media-1.xlsx\]](#)



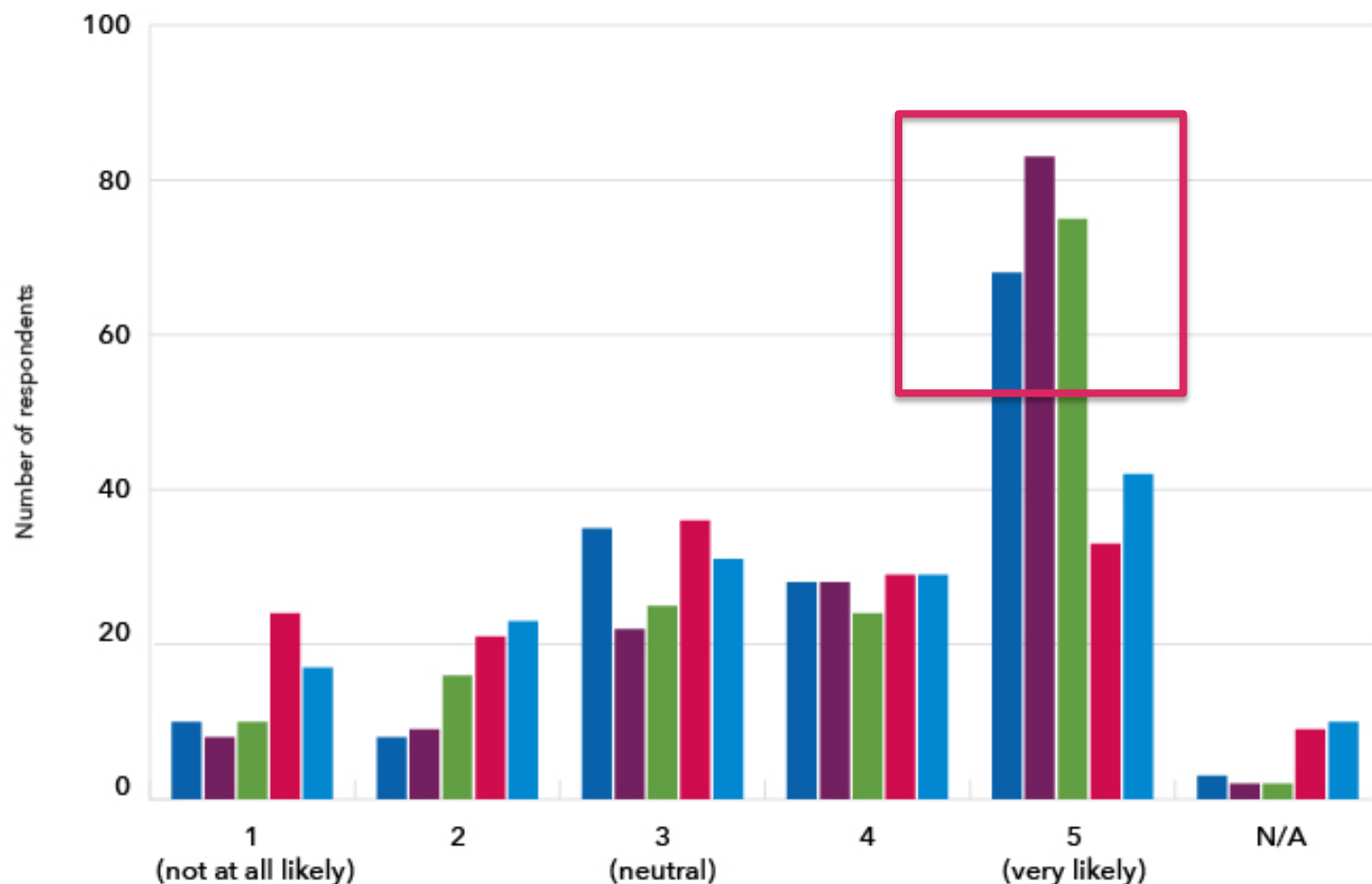
populations as in Figure 2E. Error bars, mean  $\pm$  s.d. of  $n \geq 5$  tails/time-point. (G) Cell cycle parameters of uninjured, non-regenerating, and regenerating neural stem cell populations, calculated from data in D and F. Error bars, mean  $\pm$  1 $\sigma$  confidence intervals. Data can be found in [Supplementary file 1](#) and <http://nbviewer.jupyter.org/gist/fabianrost84/3cc58a27b5688f4e2eba>. (H) EdU was injected daily from day 3 to day 8 of regeneration, when tails were collected for analysis. Representative images of non-regenerating and regenerating regions of the spinal cord. SOX2 labels neural stem cells, EdU labels cells that underwent DNA synthesis and their progeny, and Hoechst labels DNA. Quantification of the percentage of EdU+ cells that remain as neural stem cells (SOX2+) or differentiate (SOX2-) over total EdU+ cells after the six-day chase. Error bars, mean  $\pm$  s.e.m. of  $n=4$  tails/region. Scale bar, 50  $\mu$ m. A supplementary IPython notebook (Pérez and Granger, 2007) containing all the raw data and the code used for the estimations of mitotic index, cell cycle length and cell cycle parameters is available at <http://nbviewer.jupyter.org/gist/fabianrost84/3cc58a27b5688f4e2eba>.

DOI: <http://dx.doi.org/10.7554/eLife.10230.005>

eLife 2015;4:e10230

Survey of life scientists, June 2017 – Likelihood of using a feature if provided (n=152, all respondents)

<https://elifesciences.org/inside-elifesciences/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>

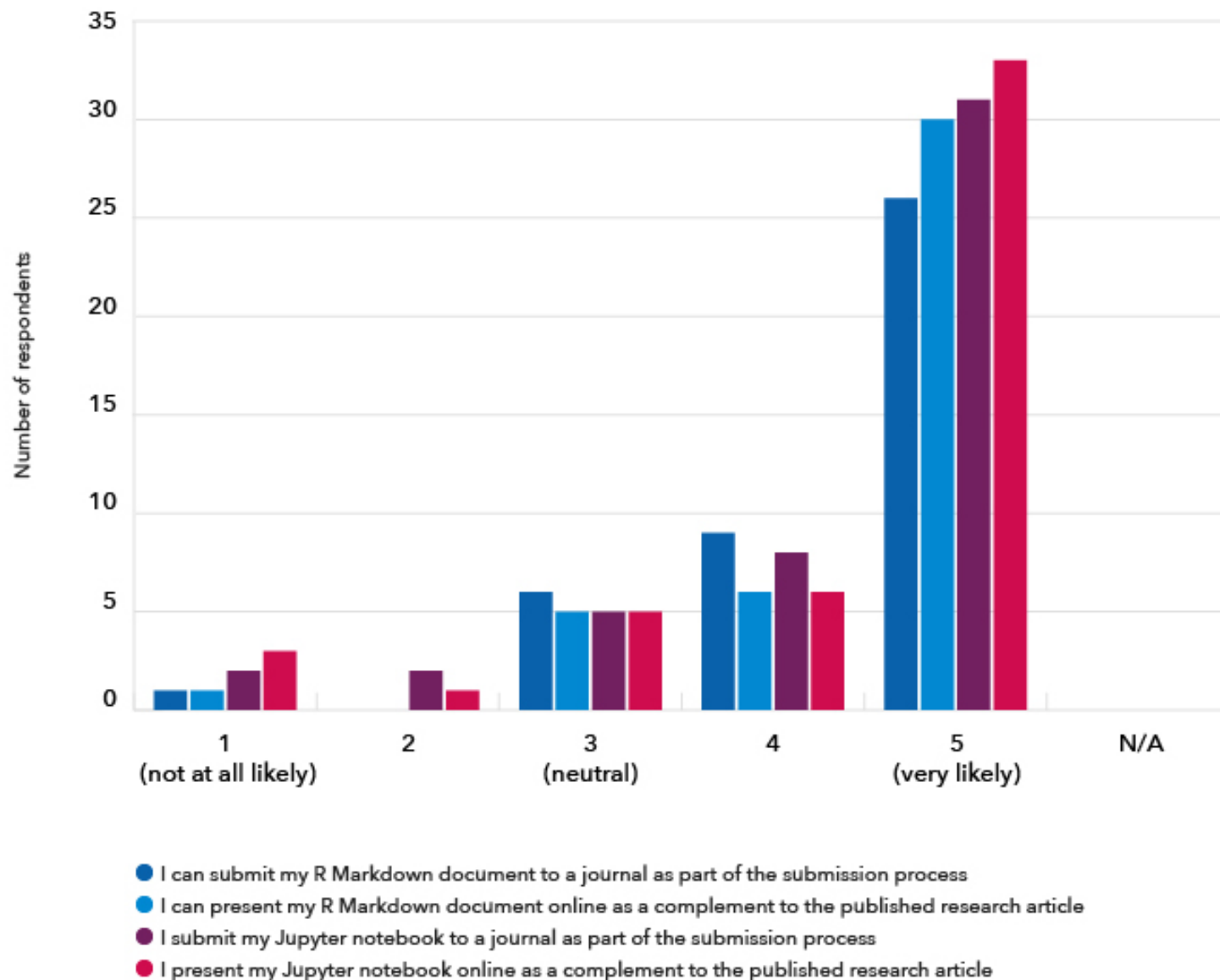


- I submit my research manuscript in a file format that preserves code and data alongside the text of a manuscript
- I view and download the data and code associated with a research article directly from the online publication
- I interact online with the data, code and/or plots in a published research article to gain additional insight
- I run the code directly from an online published research article to reproduce the author's results
- I modify the code directly from an online published research article and run my modified version to produce a new result



Survey of life scientists, June 2017 – Likelihood of using features specific to R Markdown (42 respondents) or Jupyter notebooks (48 respondents), were they to be made available

<https://elifesciences.org/inside-elifesciences/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>



# Opportunities

- Facilitating sharing **via the journal** means:
  - Persistence
  - Leverages incentive system
  - Streamline the process from initial discovery to replication
  - Engages community
- Doing it **open source** means:
  - Progress can be shared
  - Stimulates further innovation
  - Engages community

## The vision

- Progressive enhancement
  - From static research article, to full exploration and reuse of code and data
- Platform, tool, language agnostic
- Make it easy for everyone

# Reproducible Document Stack project, 2017/2018

Full stack from authoring to publication:

- Authoring platform
- Reproducible Document Exchange Format – to present online, and preserve as publisher
- An example published article, with enhanced reproducible version



Engagement between eLife, Substance and Stencila

## More info

- Announcement:  
<https://elifesciences.org/for-the-press/e6038800/elifesciences-supports-development-of-open-technology-stack-for-publishing-reproducible-manuscripts-online>
- About the project:  
<https://elifesciences.org/labs/7dbeb390/reproducible-document-stack-supporting-the-next-generation-research-article>
- June 2017 survey results:  
<https://elifesciences.org/inside-elifesciences/e832444e/innovation-understanding-the-demand-for-reproducible-research-articles>