Gina Helfrich, PhD

NumFOCUS Communications Director

What is the nature of 21st century science?

@numfocus | @ginahelfrich | gina@numfocus.org

Historically, scientific research has often looked something like this:

Source: U.S. Army RDECOM
https://www.flickr.com/photos/rdecom/15158050618

@numfocus | @ginahelfrich | gina@numfocus.org
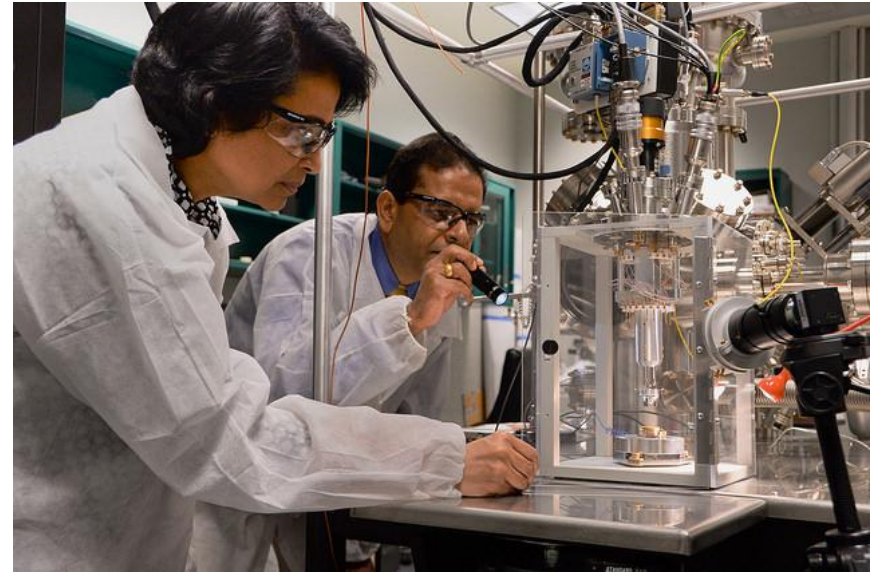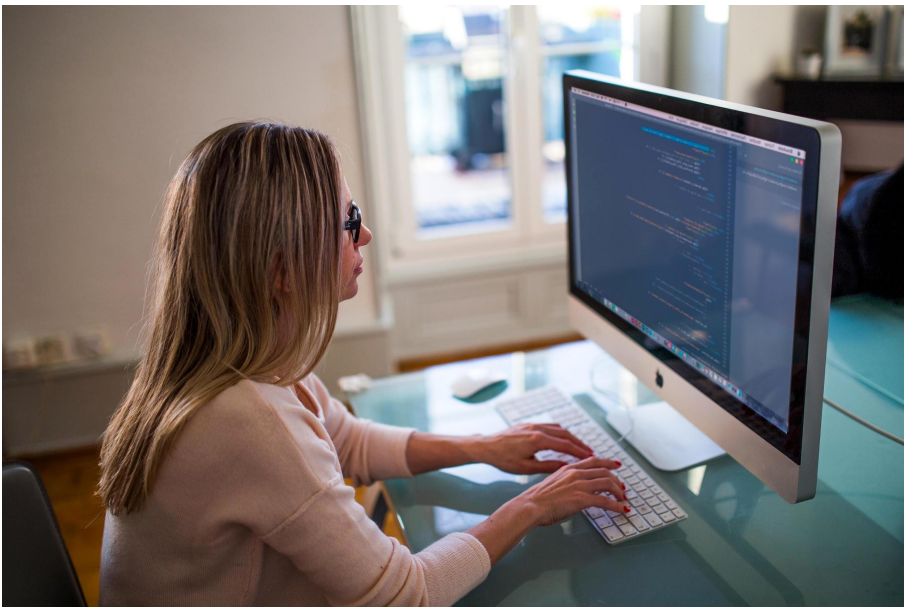
See: S. Hettrick, "2014 Software in Research Survey," Feb. 2018, doi:10.5281/zenodo.1183562

Increasingly, scientific research <u>requires</u> the use of software.

65% of researchers say they couldn't do their research without software.

@numfocus | @ginahelfrich | gina@numfocus.org

A quick definition of "Open Source"

In general, open source refers to any program whose source code is made available for use or modification as users or other developers see fit. Open source software is usually developed as a public collaboration and made freely available.

(https://whatis.techtarget.com/definition/open-source)

Different from "Free Software" — out of scope for this talk.

@numfocus | @ginahelfrich | gina@numfocus.org

Back to our modern scientist...

@numfocus | @ginahelfrich | gina@numfocus.org

A Very Brief History of Open Source Research Software

- Scientists need software for specialized research
  - They learn to **write it themselves** out of necessity

- Other researchers are interested in what their peers have built
  - They **want to extend it** for their own research purposes

But!

Most software is written to be *just good enough* to support **publication** of a research paper.

It probably doesn't have:

- User documentation
- Unit tests
- Flexibility (might be hard-coded)

A Very Brief History of Open Source Research Software

A **community begins to grow** around software packages that support research in various fields.

The community:

- Creates user documentation
- Writes unit tests
- Re-factors code
- Evangelizes the tools
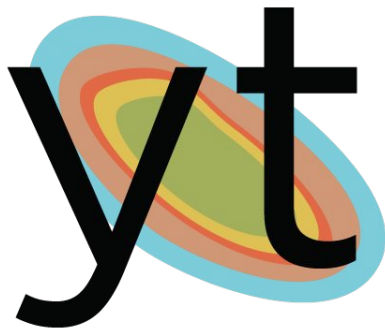- Develops governance structures

A Very Brief History of Open Source Research Software

- More mature packages start to find **applications in other disciplines**
  - Packages become more general, more flexible, less domain-specific

Astrophysics

Weather
Geophysics and Seismology
Molecular Dynamics
Observational Astronomy

OPEN CODE = BETTER SCIENCE

...but why?

将軍

sho   gun

[Reproducibility.
Accessibility.
Innovation.]

ROpenSci

@numfocus | @ginahelfrich | gina@numfocus.org

Modern scientific software practice follows cherished scientific tradition:

Building on the prior work of others.

@numfocus | @ginahelfrich | gina@numfocus.org
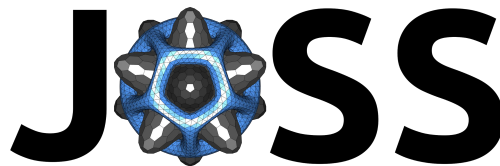
Major Challenges for OSS in Research

Difficulty receiving **career credit**/pathways

+

attractive **salaries in industry**
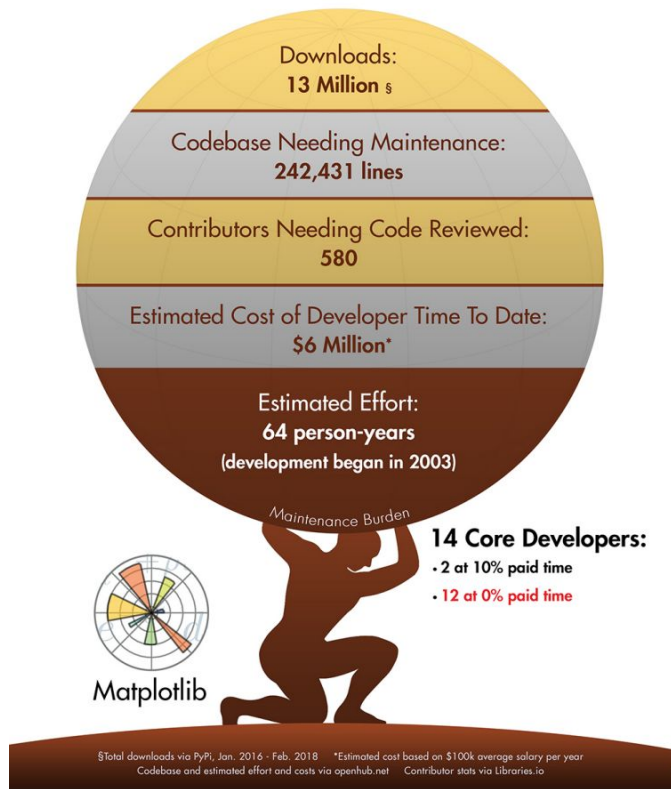
=

brain drain in academic research

@numfocus | @ginahelfrich | gina@numfocus.org

Major Challenges for OSS in General

"Bus Factor" is low.

Maintenance is generally unfunded work.

Burnout is a big risk.

Downloads:
13 Million §

Codebase Needing Maintenance:
242,431 lines

Contributors Needing Code Reviewed:
580

Estimated Cost of Developer Time To Date:
$6 Million*

Estimated Effort:
64 person-years
(development began in 2003)

Maintenance Burden

14 Core Developers:
• 2 at 10% paid time
• 12 at 0% paid time

Matplotlib

§Total downloads via PyPi, Jan. 2016 - Feb. 2018   *Estimated cost based on $100k average salary per year
Codebase and estimated effort and costs via openhub.net   Contributor stats via Libraries.io

The mission of NumFOCUS is to promote sustainable high-level programming languages, open code development, and reproducible scientific research.

@numfocus | @ginahelfrich | gina@numfocus.org

@numfocus | @ginahelfrich | gina@numfocus.org

NumFOCUS Fiscal Sponsorship Program

- Funding
- Stability
- Independence
- Logistical Support

- **LEGAL SERVICES**
  - Licensing
  - Trademark Registration & Ownership
  - Contracts & Agreements

- **FINANCIAL SERVICES**
  - Accounts Payable & Receivable
  - Grant Administration
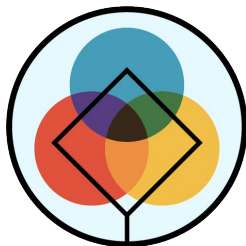 - Contractor Agreements
  - Taxes & Reporting

- **OPERATIONAL SERVICES**
  - Technical Infrastructure
  - Conference & Event Planning
  - Consulting on administrative requests

@numfocus | @ginahelfrich | gina@numfocus.org

Programs for the Community

DISC
Diversity and Inclusion in
Scientific Computing

Google
Summer of Code

John Hunter Matplotlib
Summer Fellowship

PyData
Educational Program

@numfocus | @ginahelfrich | gina@numfocus.org

NumFOCUS brings together key players to address the sustainability challenge and **connect projects with funding**.

- **Corporate Sponsors**
- **Grants**
- **Individual Donations**

@numfocus | @ginahelfrich | gina@numfocus.org

YOU can contribute to NumFOCUS!

- You can **contribute code** - projects always need help

- You can **contribute time** - organize hackathons & sprints

- You can **contribute money** - Supporting & Sustaining Members

- You can **contribute connections** - introductions to potential

  funders: corporations using the tools, NSF program officers, etc.

Get Involved + Give Back

[**numfocus.org**]

@numfocus | @ginahelfrich | gina@numfocus.org