What savvy open scholars know and do

@LorenaABarba
Main messages

- Open scholarship has more impact
- Use preprints! Green vs. Gold
- Deposit data packages for reproducibility
- Take part in new publication models
- Check your biases
About me

› Reproducibility PI Manifesto
  figshare, 2012

› “The hard road to reproducibility”

› “Repro Packs”
  Nature blogs, Apr. 2017

http://lorenabarba.com
Open Access

Top 10%

98% of your research is free to read online. This level of availability puts you in the top 2% of researchers.
Reproducibility PI Manifesto

- I teach my graduate students about reproducibility
- All our research code (and writing) is under version control
- We always carry out verification & validation (and make them public)
- For main results, we share data, plotting script & figure under CC-BY
- We upload preprint to arXiv at the time of submission to a journal
- We release code at the time of submission of a paper to a journal
- We add a “Reproducibility” declaration at the end of each paper
- I develop a consistent open-science policy & keep an up-to-date web presence
The hard road to reproducibility

Early in my Ph.D. studies, my supervisor assigned me the task of running computer code written by a previous student who was graduated and gone. It was hell.

“My students and I continuously discuss and perfect our standards.”
ReproPacks

- For main results in a paper, we share data, plotting script & figure under CC-BY.

- File bundle with input data, running scripts, plotting scripts, and figure.

- We cite our own figure in the caption!
Grid convergence of PyGBe with a spherical molecule near spherical surface

Reproducibility package containing data, running script, plotting script and final plot of convergence.
Preprints

- In physics, math, CS... arXiv is a way of life
- Preprints growing by all metrics
- Explosion of ’Xiv sites
Paul Ginsparg (arXiv) is giving today's UW Data Science Center: Adventures in Little Data.
### List of academic journals by preprint policy, Wikipedia

<table>
<thead>
<tr>
<th>Journal</th>
<th>Preprint Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature Publishing Group</td>
<td>Compatible</td>
</tr>
<tr>
<td>IOP Publishing</td>
<td>Compatible</td>
</tr>
<tr>
<td>Oxford Journals</td>
<td>Compatible</td>
</tr>
<tr>
<td>Elsevier</td>
<td>Compatible</td>
</tr>
<tr>
<td>Springer, incl. SpringerOpen</td>
<td>Compatible</td>
</tr>
<tr>
<td>Journals and BioMed Central (BMC)</td>
<td></td>
</tr>
<tr>
<td>Taylor &amp; Francis</td>
<td>Compatible</td>
</tr>
</tbody>
</table>

Communication between researchers includes not only conferences but also preprint servers. The ArXiv preprint server is the medium of choice for (mainly) physicists and astronomers who wish to share drafts of their papers with their colleagues, and with anyone else with sufficient time and knowledge to navigate it. [...] If scientists wish to display drafts of their research papers, they can upload them to preprint servers. This adds another layer of dissemination to the traditional publication process. As a result, preprint servers have become integral to the scholarly communication landscape.

You may submit the revised article to another journal. The authors are strongly encouraged to submit their article to the journal they would have submitted to if it was not for the preprint server. It is also possible to post the article by P’s allowing original, derived from wish to article available in their as follows:

Prior on the subject in their this a not post your article.

This is your original manuscript (often called a “preprint”), and you can share this as much as you like. If you do decide to post it anywhere, including onto an academic networking site, we would recommend you use an amended version of the wording below to encourage usage and citation of your final, published article.
“The rising tide of preprint servers”

<table>
<thead>
<tr>
<th>Name</th>
<th>Fields</th>
<th>Start date</th>
<th># 2016 submissions (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprint servers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arXiv</td>
<td>Physics, mathematics, computing, quantitative biology, quantitative finance, statistics</td>
<td>1991</td>
<td>113,308</td>
</tr>
<tr>
<td>bioRxiv</td>
<td>Life sciences</td>
<td>2013</td>
<td>4712</td>
</tr>
<tr>
<td>PeerJ Preprints</td>
<td>General</td>
<td>2013</td>
<td>~1000</td>
</tr>
<tr>
<td>Preprints (MDPI)</td>
<td>General</td>
<td>2016</td>
<td>~1000</td>
</tr>
<tr>
<td>SocArxiv</td>
<td>Social sciences</td>
<td>2016</td>
<td>633</td>
</tr>
<tr>
<td>PsyArXiv</td>
<td>Psychology</td>
<td>2016</td>
<td>191</td>
</tr>
<tr>
<td>EngrXiv</td>
<td>Engineering</td>
<td>2016</td>
<td>35</td>
</tr>
<tr>
<td>ChemRxiv</td>
<td>Chemistry</td>
<td>2017</td>
<td>Not launched</td>
</tr>
<tr>
<td>AgriXiv</td>
<td>Agriculture</td>
<td>Upcoming</td>
<td>Not launched</td>
</tr>
</tbody>
</table>
Open access

- Know your OA: Green vs. Gold
- You don’t have to pay $$$
- IR, preprint servers, free repositories
Free data repositories:

- figshare
- zenodo
- DRYAD
- Open Science Framework
- The Dataverse Project
Our workflow:

1. Write paper
2. Zip figure files
3. Upload
4. DOI

Icons from Icons8.
New-wave journals

- To slash the hurdles for researchers to receive academic credit for all their output, including software and data.
The Journal of Open Source Software

A developer friendly journal for research software packages.

http://joss.theoj.org
JOSS infrastructure

- GitHub—open-source software hosting & collaboration
- Zenodo—data repository by CERN
- ORCID—author identification
- CrossRef—DOI minting
- custom web app and Ruby bot
JOSS sustainability model & licensing

- CrossRef membership — $275 /year
- CrossRef DOIs — $1 /article
- Heroku web hosting — $19 /month
- 1st year, ~$6 / article
JOSS peer-review process

- Write article as `paper.md` file in code repo
- Submit via web app and submission tool: code repo URL, author ORCID, metadata
- Pre-Review issue created; editor assigned.
- Editor finds reviewer, assigns, starts review.
JOSS year 1: May 2016–May 2017

- 111 articles published, 41 under review
- median 32 days under review
- 93 unique reviewers
Submit software for review

Before you submit
Please make sure you've read the submission instructions before submitting. In particular please make sure there is a README present in your repository that is structured like this. We promise this will make things go much more quickly during the review process.

Title
What's the title of this paper?

Repository address
What's the URL of your software?

Software version
e.g. v1.0.0

Suggested editor. View editors here »
Suggested editor

Description
Please give short (1-2 line) description of your software.

☐ I certify that I am submitting software for which I am a primary author

☐ I confirm that I read and will adhere to the JOSS code of conduct
[REVIEW]: PlateReaderDataAnalyzer: a SAAS service to compute growth rates based on plate reader collected data [paused] pending-major-enhancements review
#381 opened on Aug 27 by whedon

[REVIEW]: Histogram-weighted Networks for Feature Extraction, Connectivity and Advanced Analysis in Neuroscience review
#380 opened on Aug 28 by whedon

[PRE REVIEW]: Finch: MinHashing for Sequencing Data with Abundance Calculation and Adaptive Filtering pre-review
#378 opened on Aug 24 by whedon

[PRE REVIEW]: bikedata pre-review
#375 opened on Aug 24 by whedon

[REVIEW]: dms2dfe: Comprehensive Workflow for Analysis of Deep Mutational Scanning Data review
#362 opened on Aug 12 by whedon

[REVIEW]: prisonbrief: An R package that returns tidy data from the World Prison Brief website review
#361 opened on Aug 11 by whedon

[REVIEW]: ClimDown: Climate Downscaling in R review
#360 opened on Aug 11 by whedon

[REVIEW]: Spectrum: Spectral Analysis in Python review
#348 opened on Aug 2 by whedon

[PRE REVIEW]: ELIXIR TeSS: The Training eSupport System pre-review
#345 opened on Aug 1 by whedon
JOSS was born to answer the needs of research software developers to thrive in the current merit traditions of science...but we may have come upon a generalizable formula for digital science.

— article about JOSS, submitted

arXiv:1707.02264
Free & Open-source Software (FOSS)

People can coordinate their work freely, within the confines of copyright law, while making access and wide distribution a priority.
Open-source licenses:

It's not sufficient to make the source public to read. We must attach a license that allows others to modify and distribute the code.
▶ 92% of academics use research software
▶ 69% say that their research would not be practical without it
▶ 56% develop their own software
▶ 21% of those have no training in software development

Open-source licenses:

Everyone developing software in an academic setting should have working knowledge of software licenses.
A Quick Guide to Software Licensing for the Scientist-Programmer

Andrew Morin¹, Jennifer Urban², Piotr Sliz¹*

¹ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, United States of America, ² Samuelson Law, Technology & Public Policy Clinic, School of Law, University of California Berkeley, Berkeley, California, United States of America
Permissive vs. copy-left?
Permissive licenses

- Fewest restrictions
- Allow use, distribution, modification
- Only require giving credit to code authors
- Best choice for academic use!
- e.g., Berkeley Software Distribution (BSD), MIT License, Apache License
MIT License

Copyright (c) [year] [fullname]

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

http://choosealicense.com/licenses/mit/
Copy-left licenses

- Guarantees perpetual access to the source code
- Requires any derivative work be under the same license
- a.k.a. “share-alike” licenses
- Are considered restrictive
- e.g., GPL license
GNU GENERAL PUBLIC LICENSE
Version 3, 29 June 2007

Copyright (C) 2007 Free Software Foundation, Inc. <http://fsf.org/>
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The GNU General Public License is a free, copyleft license for
software and other kinds of works.

The licenses for most software and other practical works are designed
to take away your freedom to share and change the works. By contrast,
the GNU General Public License is intended to guarantee your freedom to
share and change all versions of a program--to make sure it remains free
software for all its users. We, the Free Software Foundation, use the
GNU General Public License for most of our software; it applies also to
any other work released this way by its authors. You can apply it to
your programs, too.
License compatibility

Compatible licenses allow source code from different works to be combined to make new software. Not all licenses are compatible!
How to choose?

For academic work: simple & permissive is best.

http://choosealicense.com/
Subtlety: MIT vs. BSD 3-clause

MIT License:

“The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.”

BSD 3-clause:

“Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.”
**Bonus advice:**

Write into your grant proposals that your research software will be released under an OSI-approved license.
Check your bias

—known weakness of open review: it can disadvantage women and minorities, and researchers from institutions of lesser prestige, due to implicit bias.
Threats to objectivity in peer review: the case of gender

Anna Kaatz¹, Belinda Gutierrez¹,², and Molly Carnes¹,²,³,⁴

SciPy 2017 Chair and Reviewer Guidelines

The Program Committee's job is to select the talks that maximize the benefit to the entire conference audience, within the space and time constraints given.

SciPy 2017 will implement a double-open review: authors and reviewers are known to each other. Area chairs will make acceptance decisions, informed by reviewer comments and conference schedule limits. The intention is to offer transparency in the process, building trust in the community. (There is also research attesting to higher quality in signed reviews, compared to anonymous reviews.)

But, research shows that peer review can suffer from implicit bias—just like the hiring processes—that disadvantages women, minorities and people from less-prestigious institutions. For this reason, we want to educate our reviewers about bias in the review process, and equip them with tools to interrupt this bias.

Bias in the review process

Human beings are consistently, routinely, and profoundly biased. We not only are profoundly biased, but we also almost never know we are being biased. (Interested?[1],[2])
What savvy open scholars know and do

@LorenaABarba