building capacity for open science

kaitlin thaney
@kaythaney ; @mozillascience
COASP / 15 sep 2015
getting open practice to stick

kaitlin thaney
@kaythaney ; @mozillascience
COASP / 15 sep 2015
doing good is part of our code
we empower researchers to do more open, collaborative research on the web.
OBSERVATIONES SIDEREAE
ex parte feliicit Orientali duo aderant Stellae, una vero Occidum verus. Orientalior atque Occidua, reliqua paulo minusque appararente, de distantia inter ipsas & Iouem minime feliicitus fuit; fixe enim vti dimissus primo creditur fuerunt; cum autem die octava, necio quo Fato ductus, ad inscriptionem candeur reseruit, longè aliam constitutionem reperiri, erant enim tres Stellae occidentales omnes à Ioue, atque inter se quam superiori nocte viciniores, paribusque interstitiibus mutuo disperaratur, veluti appositae praefert delineatio. Hic licet ad mutuum Stellarum appropinquationem minime cogitationem appuisset,

Ori. * * * Occ.

exitate tamen carpe, quomam pauci Jupiter ab omnibus predictionis axis potuit orientali repetiri, cum autem ex ilis pridie occidentalis fuerit; ac prorsum verius sum ne forte, fatus à computo astronomico, directionem notarit; ac propriis moris proprie Stellae illas anteoconsiderat; quapropter maximo cum defidero sequentem expectauerit noctem; verum, ut fructus fuit, nubes primariae obscurum die celerat.

At die decima apparet Stellae in eiuem modi ad Iouem posita: duae enim tantum, & orientales ambæ

Ori. * * Occ.

aderant, tertia, ut opinatur fuit, sub Ioue latitante. Erant pariter veluti ante in cæcum recta cum Ioue, ac iuxta Zodiaci longitudinem adamasfim locatae. Hae cum vidissent, cumque mutantes confusiles in Ioue nulla

RECENS HABITAE

nulla ratione reponi possit intellegere, atque infuper sepectatas Stellae semper caelestem fuisset cognoscerem, nullæ enim aliae, aut præcedentes, aut conseqüentes intra magnum inferiorum in Iuxta longitudinem Zo-diaci aderant suam ambiguitatem in admirationem perturabant, apparentem commutacionem non in Ioue, sed in Stellis adnotatis reposti, et aca collegi, et scrupulos magis deinceps obseruantum fore lumen ratum.

Iuxta, et decima eiuem modi constitutionem vidi:

Ori. * * Occ.

Stellae feliicit tantum duas orientales; quorum media triplo distabat a Ioue, quam ab orientalis; erat orientalior duplo fecerat major reliqua, cum tamen antecedenti nocte aequalis semper apparuerit. Statuta idem, omnique procul dubio aedecem tuit, tres in cculis adesse Stellæ vagantes circa Iouem, initaverunt, atque Mercuriens circa Sol qui quoniam luce meridiana clarior in alis passum commodum pluribus inspicientibus obseruati est; ac non tantum tres, verum quatuor esse vaga Sydera circa Iouem suas circumvolutiones obseruavit; quorum permutationes exactae conseqüenter obseruatas subseqüentis narratio ministratis, interinitialia quoque inter ipsa per Peripileum, superius explicata ratione, dimittitum sum: horas infuper obseruationum praefere cum pluribus in eadem nocte habere fuerunt appauji, adeo enim celebres horum Planetarum extant revolutiones, ut horas quoque differentias plerunque liceat accipere.

Die igitur duodecim, hora sequentis noctis prima hæ ratione disposita Sydera vidi. Erat orientalis

E 2

Stella

http://bit.ly/1eZZC0f
current state of science
some have a firehose
quality versus quantity
measured systems

I didn't have any accurate numbers so I just made up this one.

Studies have shown that accurate numbers aren't any more useful than the ones you make up.

How many studies showed that?

Eighty-seven.
The diagram illustrates the decline of information content over time. Key events include:

- **Time of publication**: An initial burst of specific details.
- **Specific details**: Gradual decrease post-publication.
- **General details**: Further decline, providing context.
- **Retirement or career change**: Period of plateau, followed by a rise in accident details.
- **Accident**: Sharp increase in information content.
- **Death**: Final stage with minimal information.

Source: Michener, 2006 Ecoinformatics.
Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Affiliations  |  Corresponding author

Nature 483, 531–533 (29 March 2012)  |  doi:10.1038/483531a
Published online 28 March 2012
Clarification (May, 2012)

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.
a few fallacies of the research system
the published record is the only useful record
if it’s published, it’s usable.

“I think you should be more explicit here in step two.”
An example of a false-positive in the field of Solar—Terrestrial research: A Purported relationship between Neutron monitor changes and extremes of the Indian Monsoon

Dr. Benjamin A. Laken, Department of Geosciences, University of Oslo, Norway

In [1]:
```
# Some required modules and settings (compliant with future Python)
from __future__ import print_function, division, generators
%pylab inline
from scipy.stats.stats import pearsonr
import pandas as pd
import datetime as dt
```

Populating the interactive namespace from numpy and matplotlib

1. Read data

Fetch the required data from CSV files held on a public server. Then automatically create date object column (leaving it only as an index).

In [2]:
```
monsoon = pd.read_csv('http://www.files.benlaken.com/documents/Monsoon_data.csv', parse_dates=['Date'])
# nb. you may also download the file from the below link (although this will not work for fetching
# https://github.com/benlaken/Comment_BadruddinAslam2014/blob/master/Data/Monsoon_data.csv
monsoon.index = monsoon.Date
monsoon = monsoon.drop('Date', 1)

# Uncomment this and you will get the PD dataframe preview of Precipitation data
```

In [3]:
```
olou = pd.read_csv('http://www.files.benlaken.com/documents/Olou_counts.csv', parse_dates=['Date'])
# nb. you may also download the file from the below link (although this will not work for fetching
#https://github.com/benlaken/Comment_BadruddinAslam2014/blob/master/Data/Olou_counts.csv
olou.index = olou.Date
olou = olou.drop('Date', 1)

# Uncomment this and you will get the PD dataframe preview of NM data
```

In [4]:
```
# Produce a monthly climatology of Precipitation data
```

http://bit.ly/1MEcOmd
... that this is good job advice:

“What Des-Cartes did was a good step. You have added much several ways, & especially in taking ye colours of thin plates into philosophical consideration.

If I have seen further it is by standing on ye shoulders of Giants.”

- Isaac Newton, 1676
that there are other ways to job security
1. examples of what’s possible.
2. making the case for open (and teaching it).
3. sustaining momentum.
learning from (+ through) open source development examples of “science like the web” in practice
An example of a false-positive in the field of Solar—Terrestrial research: A Purported relationship between Neutron monitor changes and extremes of the Indian Monsoon

Dr. Benjamin A. Laken, Department of Geosciences, University of Oslo, Norway

In [1]:
```python
# Some required modules and settings (compliant with future Python)
from __future__ import print_function, division, generators
%pylab inline
from scipy.stats.stats import pearsonr
import pandas as pd
import datetime as dt
```

Populating the interactive namespace from numpy and matplotlib

1. Read data

Fetch the required data from CSV files held on a public server. Then automatically create date object column (leaving it only as an index).

In [2]:
```python
monsoon = pd.read_csv('http://www.files.benlaken.com/documents/Monsoon_data.csv', parse_dates=['Date'])
# nb. you may also download the file from the below link (although this will not work for fetching
# https://github.com/benlaken/Comment_BadruddinAslam2014/blob/master/Data/Monsoon_data.csv
monsoon.index = monsoon.Date
monsoon = monsoon.drop('Date', 1)
#monsoon # Uncomment this and you will get the PD dataframe preview of Precipitation data
```

In [3]:
```python
olou = pd.read_csv('http://www.files.benlaken.com/documents/Olou_counts.csv', parse_dates=['Date'])
# nb. you may also download the file from the below link (although this will not work for fetching
#https://github.com/benlaken/Comment_BadruddinAslam2014/blob/master/Data/Olou_counts.csv
olou.index = olou.Date
olou = olou.drop('Date', 1)
#olou # Uncomment this and you will get the PD dataframe preview of NM data
```

In [4]:
```python
# Produce a monthly climatology of Precipitation data
```

http://bit.ly/1MEcOmD
code as a research object

what’s needed to reuse?

# Data inspection

The codes for variable site are PC=Potter Cove AB=Admiral Bay

```r
dataInspection, echo=FALSE, message=FALSE, warning=FALSE
```

```r
chla <- read.table("Meteo_stat.txt", header=T)
pairs(na.omit(chla[,2:6]), panel=function(x,y) { points(x,y); lines(lowess(x,y)) } )
```

#source("panelutils.R")
#pairs(chlaR, panel=panel.smooth, diag.panel=panel.hist, main="")
#hist(chlaR$ClorMAX, col="bisque", right=FALSE)
#hist(sqrt(chlaR$ClorMAX), col="bisque", right=FALSE)

#hist(chlaR$DegDay_NM, col="bisque", right=FALSE)
#hist(chlaR$ENSO_NM, col="bisque", right=FALSE)
#hist(chlaR$SAM_NM, col="bisque", right=FALSE)

An ungodly union of GitHub and Figshare http://fidgit.arfon.org — Edit

- 10 commits
- 1 branch
- 3 releases
- 1 contributor

branch: master

This branch is 0 commits ahead and 0 commits behind master

Update README.md

- arfon authored on Oct 22, 2013

screens
- Screenshots
  - 9 months ago

tmp
- Initial commit
  - 9 months ago

views
- Initial commit
  - 9 months ago

.gitignore
- Initial commit
  - 9 months ago

Gemfile
- Initial commit
  - 9 months ago

Gemfile.lock
- Initial commit
  - 9 months ago

Procfile
- Initial commit
  - 9 months ago

README.md
- Update README.md
  - 8 months ago

HTTPS clone URL

https://github.com/
Project Metadata

To help people find your project, it's useful to provide as much of this information as possible:

Name
mozillacience/fidgit

Home Page
http://

Maintainer
First Last <first.last@example.com>

The current maintainer of the project

Contributors
First Last, First Last, First Last

People who contributed to the project

Summary
An ungodly union of GitHub and Figshare

Summarise this project in a single sentence

License
Select a license...
syndication and storage (via APIs)

[code repos>  figshare  zenodo

[institutional archives>  [national archives>
This objectives of this “Software Discovery Index” would be:

1. to assign standard and unambiguous identifiers to reference all software,
2. to track specific metadata features that describe that software, and
3. to enable robust querying of all relevant information for users.

If broadly used, this Software Discovery Index will form a cornerstone in a software ecosystem that benefits software developers, software users, journal publishers, and funding agencies.

http://softwarediscoveryindex.org/report/
from campaigning to discovery

Sex Work: A Comparative Study

Bill McCarthy, Cecilia Benoit, Mikael Jansson

Abstract

Explanations of adult involvement in sex work typically adopt one of two approaches. One perspective highlights a variety of negative experiences in childhood and adolescence, including physical and sexual abuse, family instability, poverty, associations with “pimps” and other exploiters, homelessness, and drug use. An alternative account recognizes that some of these factors may be involved, but underscores the contribution of more immediate circumstances, such as current economic needs, human capital, and employment opportunities. Prior research offers a limited assessment of these contrasting claims: most studies have focused exclusively on people working in the sex industry and they have not assessed the independent effects of life course traumas.
Sex Work: A Comparative Study

Bill McCarthy, Cecilia Benoit, Mikael Jansson

Abstract

Explanations of adult involvement in sex work typically adopt one of two approaches. One perspective highlights a variety of negative experiences in childhood and adolescence, including physical and sexual abuse, family instability, poverty, associations with “pimps” and other exploiters, homelessness, and drug use. An alternative account recognizes that some of these factors may be involved, but underscores the contribution of more immediate circumstances, such as current economic needs, human capital, and employment opportunities. Prior research offers a limited assessment of these contrasting claims: most studies have focused exclusively on people involved in sex work, and the literature is replete with a confusing overlap of definitions, classifications, and methodologies.

* Final gross prices may vary according to local VAT.
Sex Work: A Comparative Study

Bill McCarthy, Cecilia Benoit, Mikael Jansson

Abstract

Explanations of adult involvement in sex work typically adopt one of two approaches. One perspective highlights a variety of negative experiences in childhood and adolescence, including physical and sexual abuse, family instability, poverty, associations with "pimps" and other exploiters, homelessness, and drug use. An alternative account recognizes that some of these factors may be involved, but underscores the contribution of more immediate circumstances, such as current economic needs, human capital, and employment opportunities. Prior research offers a limited assessment of these contrasting claims; most studies have focused exclusively on people with a particular characteristic, and the work on peer experiences has been too small or too narrow to argue for direct causation. To address this gap, a random sample of individuals was surveyed, who were enrolling in college and who had a family member, friend, or partner who was engaged in sex work. Data analysis confirms negative experiences in childhood and adolescence, but also shows that current circumstances play a significant role in women's decision to involve themselves in sex work.
Sex Work: A Comparative Study

Bill McCarthy, Cecilia Benoit, Mikael Jansson

Abstract

Explanations of adult involvement in sex work typically adopt one of two approaches. One perspective highlights a variety of negative experiences in childhood and adolescence, including physical and sexual abuse, family instability, poverty, associations with "pimps" and other exploiters, homelessness, and drug use. An alternative account recognizes that some of these factors may be involved, but underscores the centrality of personal circumstances, such...
open, iterative development

the “work in progress” effect
Digital Credentials
Conceptualization Data Curation Data Visualization Writing - Review Formal Analysis
Methodology Project Administration Resources Supervision Testing
OPEN BADGES ANATOMY

Image CC BY Kyle Bowen

- Badge name
- Description
- Criteria
- Issuer
- Evidence
- Date issued
- Standards
- Tags

Badge image
Contributorship Badges - Proposed Implementation / Workflow

Option 1: ORCID implements ID lookup/creation via email address

http://mozillascience.org/contributorship-badges-a-new-project/
what does “web literacy” mean for research?
the web as a platform
communication
access, reuse, scale
distributed environment
power, performance, scale
“web-enabled research”

- access to content, data, code, materials.
- emergence of “web-native” tools.
- rewards for openness, interoperability, collaboration, sharing.
- push for ROI, reuse, recomputability, transparency.
putting open ideals into practice
(+ paying it forward)

https://commonspace.wordpress.com/2015/07/15/web-literacy-and-leadership/
service learning: n. hands-on, experiential learning where people develop skills by working on a project in service of a bigger goal.

http://bit.ly/1JTMBSb
how do we amplify within research?
(... and beyond software development?)
community-driven contributorship

1. Mozilla 101
   1.1. Working Open
   1.2. Who are Contributors?

2. Project Setup
   2.1. Lowering Barriers
   2.2. Roadmapping
   2.3. Legal
   2.4. Project Checklist

3. Working with Contributors
   3.1. Mechanics of Contributing
   3.2. Engaging Contributors
   3.3. Developing Leaders

4. Getting Unstuck
   4.1. Where to Get Help
   4.2. Toxic Members

'Working Open' Project Guide

Working openly with contributors enables your community to learn how to build and collaborate together. This document is a guideline on how to work openly and involve others in your projects with Mozilla. We want to help you engage your community in a way that encourages contributors and builds other leaders.

Help Us Finish This Book

To add content to this book, follow these steps:

1. On the page you want to edit, click on the 'Edit' button at the top of the page

2. Make your edits in the GitHub interface. When you're done, scroll down and add a commit message summarizing your changes

3. Click on the 'propose file change' button

http://mozillascience.github.io/leadership-training/
mozillascience.org/collaborate
2 days, 30+ cities, 53+ hours

http://bit.ly/1N331JV
100+ pull requests

(code, content, learning resources)

http://bit.ly/1N331JV
(3)
how do we build capacity?

furthering adoption of open, web-enabled research
rewards, incentives, reputation
Data Scientist: The Sexiest Job of the 21st Century
by Thomas H. Davenport and D.J. Patil

Computational Thinking: A Digital Age Skill for Everyone

The 21st Century WORKFORCE: Skills Gap & The STEM DILEMMA

The gap between the skills available in the current workforce and those needed for many 21st century jobs creates a serious challenge.

ONLY 42% of employers believe new graduates in the workforce are adequately prepared by their colleges or other pre-employment training programs.

A workforce prepared to tackle science, technology, engineering and math (STEM) is critical to driving future growth and innovation.

WHILE ONLY 4% OF THE NATION’S WORKFORCE IS COMPOSED OF SCIENTISTS AND ENGINEERS, THIS GROUP DISPROPORTIONATELY CREATES JOBS FOR.

THE OTHER 96%
supports needed for "professional development"
“Reliance on ad-hoc, self-education about what’s possible doesn’t scale.”

- Selena Decklemann
lowering barriers to entry (+ leveling the playing field)
Mozilla Study Group Handbook

Mozilla Study Groups are a fun and informal way to meet up with friends and colleagues to share code, work on bug fixes, and share knowledge. This handbook describes everything you'll need to start your own study group.

Made by the Mozilla Science Lab

Introduction

Writing and working with code is becoming a necessary part of daily life for scientists and researchers in almost every field - and yet, the skills to do so are still not included in most formal curricula, few places exist to support and encourage the discussion of scientific code and coding, and opportunities to collaborate on this crucial element of research are rare and hard to find. As a result, students and new researchers struggle to build the skills they need to pursue the science they aspire to, leading to an epidemic of imposter syndrome in the lab, and risking the waste of a tremendous amount of time and money as we fail to identify places to share and reuse research code, and instead reinvent the wheel, over and over.

https://mozillascience.github.io/studyGroupHandbook/
Pathogens & Disease Immunity
Contributorship Badges
WormBase Website
Cytoscape.js
Matplotlibdash

mozillascience.org/collaborate
100+ pull requests

(code, content, learning resources)

http://bit.ly/1N331JV
Roadmap

We're building a prototype to test ways to use open badges to assign digital credentials to contributions on academic papers.

Using Mozilla's Badgekit-api to implement our badges, we can issue and fetch badges from badgekit via the badgkit-api-client. By authenticating against ORCID, a user can reliably issue badges to a valid ORCID, the standard unique researcher identifier.

Want to jump in?

- **Good first bugs:** This is a good place to start if you're new to the project. These issues are all mentored.
- **Design:** We need some help designing interfaces and deciding how these badges will look on a paper page / on an ORCID page.
- **Writing code:** Ready to write some JavaScript with us? These issues are waiting for you!
- **Discuss:** There are still some open ended questions for this prototype. Join in on the discussion.
starting soon!

The Mozilla Open Science Fellowship

The Mozilla Open Science Fellowship enables early-career researchers to spend 10 months working on open, web-enabled research and further open science within the community. Mozilla Open Science Fellows will learn open research skills, write code and help build community during their fellowship year.

https://www.mozillascience.org/fellows
we need your help.
rethink beyond just access to usability.

don’t be afraid of open source collaborations.

remember the non-technical challenges.
we’re here to help.

http://mozillascience.org
sciencelab@mozillafoundation.org
kaitlin@mozillafoundation.org
@kaythaney ; @mozillascience

special thanks:

ALFRED P. SLOAN FOUNDATION

THE LEONA M. AND HARRY B. HELMSLEY CHARITABLE TRUST