

Participants

Sean Goggins <s@goggins.com>,
Andrea Wiggins <akwiggins@gmail.com>,
Jonathan Brier <Brierjon@umd.edu>,
Aaron Halfaker <ahalfaker@wikimedia.org>,
Becky Rother <becky@zooniverse.org>,
Carsten Oesterlund <costerlu@syr.edu>,
Greg Newman <gregory.newman@colostate.edu>,
James Howison <jhowison@ischool.utexas.edu>,
Hannah Cohoon <jlcohoon@utexas.edu>,
Raymond Paik <rpaik@linuxfoundation.org>,
Aaron Schechter <aschechter@uga.edu>,
Don Marti <dmarti@mozilla.com>,
Georg Link <glink@unomaha.edu>,
Kevin Lumbard <klumbard@unomaha.edu>,
Vinod Ahuja <vahuja@unomaha.edu>,
Holly Rosser <hrosser@unomaha.edu>,
Nicole Damen <ndamen@unomaha.edu>,
Matt Germonprez <germonprez@gmail.edu>

Abbreviations:

CS : Citizen Science

OS : Open Source

April 26th

9:30am - Community Health and Sustainability - Existing and Near-neighbor Efforts

[

Big difference in community health concept between oss and citz sci (AW)

Communities of users? Developers?

Think about it in different use cases. (MG) Answer can be many different things depending on the use cases

Financial sustainability?

Software Remains scientifically useful (JH);

Project being constituted by the software and the community (MG/DM)

Goal? Somewhat common language where we are using some of the same terminology.

Best practices to achieve sustainability

Evaluation - the degree to which you are achieving sustainability

Existing efforts:

Mozilla

CHAOSS

GMD need to be contextualized based on community

Risk is kind of emerging--software licensing; pony factor/bus factor; Concentration and distribution of effort/skew.

Value: impact of the project in a larger ecosystem. Metrics=software impact/contributor award.
(DM) Doing research to create new metrics if the metrics aren't working.

How does value account for specialization? (HC) Open source deadspots--we don't see that cycle of contributor award to build up that sector. (DM) Software isn't having enough impact to be meaningful.

Complementary goods.

Trace data heavily relied upon, but there is need to look beyond.

Map value metrics to other metrics? Hard! Not seeing it...

A lot of efforts out there with localization, but at a meta level are there things we can share?
Achieving sustainability and evaluating sustainability. From Concept to more specific (AW)

Bus factor: if someone gets hit by a bus, would your project go down?
If 1, then that's bad

Pony factor: nr of org. that contribute > 50% on the project

CS talks about skew while OS talks about risk (?)

H-index: How to understand the project in a larger eco-system

How many projects are dependent on my project, and how many are dependent on those dependencies.

Metrics that matter:

Does this activity metric predict something on how much value this project is going to provide for other dependent projects, and is it something that contributors get enough value on that they want to stick around?

> Problem : we can put many metrics up there, but we can't yet justify why those metrics need to be there and what their impact is.

How does value relate to specialization?

> discussion of the day

Value propositions?

Pain and gain

Metrics might be disconnected from value ((difficult to judge; trash for one and gold for another?))

Organizations differ in their ability to separate proprietary and open source parts of their project. E.g. accounting might be a 'closed'/hidden field, so maybe OS would be less suitable. However, Facebook is also very secretive yet does manage to open source some parts of their projects.

Value and risk may not be the right term. Impact?

CS: Churn and turnover vs. risk

Value for who? - citizen science;

How to measure change in people's capabilities; how are they learning, simple vs advanced work, and how to do that over time. what are good trace measures?

CS has the same discussions but uses different words, and the effort seems more dispersed. Somehow need to define a space for inclusion and collaboration.

Scistarter

Example of a similar effort within CS

Dynamics within cross-projects

Impacts of motivations; different levels of participation;

> pull from the website

Recruitment, retention and skew

> are people leveling up or are they sticking to the same thing?

Karen cooper?

Directing participants to projects of interest...no true tracking of participation-- progression and lifespan of participation.

OS has some software that is aimed at tackling these metrics (CHAOSS, Augur, etc > see slide)

> What are the things that can help us improve / take the first steps to improving diversity and inclusion.

What are the critical metrics that will address diversity and inclusion; also work with the software committee or community? Not just thoughts but implementations--what is valuable/not? How can we get things done in application and in practise.

By doing the field work and workgroups we can see what people are doing with the data and what actually adds value and what doesn't.

Understanding our communities is going to require some footwork--conferences, surveys, interviews, etc. Heavy lifting will be required.

Also must make an effort to go beyond conferences, that's those who have already succeeded. > Go talk to your customer > real people, community members who actually use and work with it. Go talk to the 'real people' who are using the software.

Local hackathons and events ipv large conferences which are expensive to go to, but some have bad reputation for newcomers, and since they're so big its hard to interact with new people and expand your network.

How do you get new audiences to investigate your work?

There's enough open source conferences to fill your calendar but enough of them that are not welcoming to newcomers.

There is an art to getting newbies involved.

Efforts that the attendees are aware of;

- Mozilla Diversity and Inclusion
- DC community : NSF funded: sustainable software institute (SSI)
 - Cyber infrastructure, software infrastructure,
 - Research-oriented/funding program
-
- WSSSPE1
 - Summary of the first workshop on sustainable software for science: Practice and Experiences - Katz Daniel et al.
 - Whisky?
 - Workshop on sustainable scientific software production
- - More about community exploring these questions.
 - Preceding set of literature in OSS, but never a concerted effort to put this into active practice.

OS: There are quite some papers on metrics, but there is less effort on combining those. Mozilla people were looking at those metrics and models but didn't find a use case for them (yet).

> There is a research component, but the ground testing > actually making an impact - component is a bit lacking

Back to **Scistarter**

- Broader than just the platform
- How much people move between project
- Social side of people's involvement
- Life-span of individual engagement

CS:

- Maturity and decline
- After a while forums get cluttered and its difficult to find things,.
- Tension between:
 - Personomies > individual tags to organize information
 - Folksonomies > collectivite classification schemes
 - > increases onboarding barrier for newbies

Tension between personomies--barrier to entry because there is more and more to learn. Easier to get started in a new project than to continue on with an older project.

Personamies → folksonomies?

Zooniverse

- Employs moderators
- Onboarding/Talk Boards is where we make our discoveries
- Helps with scalability
- Sense of ownership for participants; but no real sense of community is necessary.
- Onboarding approach needs to address participant personalities and willingness to socially engage.
- Welcome new members
- Struggle: scaling while keeping personal connections with all the community
 - The social component is not necessary, you don't have to interact to be a highly contributor. Different personalities and experiences: some people need lots of socialness vs less,

and some can get started right away while others need to be hand-held. Also lurkers; some people read/listen to social interaction going on,

- You can be a lurker in one community, but then be an active contributor in another community > translating and moving information to another community
- People often start lurking before they start becoming more active

- Participation without socialization is possible in citizen science
- Individuals want different levels of engagement/ hand-holding

Lurking? Listening to engagement qualify of social engagement?

Value of lurking?

Start as lurkers prior to contribution

Transfer that knowledge from lurker to other members of the community is very difficult.

Not trying to solve the goals, but looking on how to move them forward and identify next steps

]

1:00pm - Summaries and Discussion from Breakout Sessions

[

Openness Index

Issue of maturity?

It's hard to imagine health without a boundary.

Notion of interoperability--it might be more healthy if others can come and use it.

Summaries

OSS:

Apprenticeship model of communities?

Little focus on people in OS compared to Citz Sci.

Unified artifact in OS vs. Citz Sci

Citizen Science:

More ease in separating people from artifacts.

Because its not a reliable proxy; there are many different thing going on in each project, so there is big difficulty in comparing between projects. Also many different type of tasks and workflows (e.g. hiking vs reading vs spotting)

A genre notion would be helpful within CS.

Sustainability of the software, or what about new blood? Influx while others fade-out.

Barriers to engagement and community health observed OSS as well as CitzSci

Spectrum between finite project vs something that is meant to be continued.

CS is more constrained by funding, so if the science question has been solved then the project kinda ends. However in OS there is not really an ending of a project.

Shepherding that community from one community to another in CS (what is the community?)

In OS there are people who do float between projects, especially when they are an expert of a certain topic.

As a community manager I look for projects that are inactive and let them know it is okay to terminate project -- Go work on something else.

The repo gets locked down

Looks like there is no generic closing a project template. > Lack of project management skills / knowledge or is it a focus thing or attitude? (is it okay to close a project or not)

Who would maintain the code after the project terminated? Repo locked down. Different projects go through different cycles.

How do you convert a drive by contributor to become an active contributor. Goal of conversion in CS, but there is a consensus that the contribution of drive by is beneficial.

> Conversion rate is low

>

Lack of persistent identity - wikipedia drive by contributors

OSS - need skill to contrib

Corporate interest may drive skill development and requirements.

There might be a difference between OS and CS, in the sense that CS has a lower bar, since the idea is that everyone can just walk in and do it, while the OS requires some level of skill.

Additionally, in OS there are some inhibitions/barriers related to the style and topic, e.g. the banking api which is done mostly by banking, so that makes it hard to join even if you technically could.

CLA in OSS (contributing license agreement) barrier to OSS contribution.

OS also has CLA (contributor license agreement or something?) > In CS its part of the terms of agreement, but the volunteer is less legally restricted (?) > barrier that goes beyond the technical barrier

OS is a way of doing interorganizational R&S

OSS=New form of R&D; but there are community-based projects as well; but there is different forms of organizations.

Very little crowdsourcing in OSS; bug reporting maybe the closest example.

Assigned task CS vs. choice of task assignment in OSS

Co-created vs crowdsourced

Distinction between type of CS : some are created by regular civilians with a question or problem but go through the scientific method, while others are more like Zooniverse.

Other similarities;

Other differences;

]

3:00pm - Summaries from Consensus-building from Breakout Sessions

[

CS (zooniverse): Volunteers and moderators and science team

OS : Core team

Most [CS] only have two levels but may be variable

Setting up the forms and people filling out the forms. [cs]

There is no launch moment in OSS ?

Can a citizen science worker become part of the science team - unusual...

Challenge of new users/motivations

Yes or No metrics vs deeper or hard define metrics

Qualified developers?

Emergent skill levels - CS

All welcome and contribs sort themselves

Take all the data and sort it out at a later date [cs]

Is the documentation any good?

Who is welcome and who is encouraged?

Is diversity and inclusion valued within CS as well? It's valued but not mission critical.

However, no projects have been where this was something that was clearly value-adding

In OS its known to be important, but .. ?

For Wiki there are coverage gaps; problem and opportunity. Wiki does have barriers, but in addition there is also the thing where if everyone had equal access, then you might still attract certain people over others.

CS: Geographic diversity may have value but not demographic

Biggest D&I is in projects that are related to eco... in the local area. Personal relevance for a wider group of people than rich educated white people.

Both the rich and the poor can sleep under bridges

Outreach on diversity -- Example wikipedia female scientists pages

Diversity separate from inclusion?

A group can be a numeric diversity, but behavioral not inclusive.

Inclusion typically means people with disabilities.

Address issues in a timely manner - bugs, errors, syntax errors, and observation (lat and long flopped) - wiki bugs are like vandalism. Different from software - accidental. CS: science team vs contributors.

Use of the data phase in citizen science - could be a collective activity

- Maybe not so in OSS

]

3 reasons for people leave OSS engagement

- Project conflict
- Family responsibility
- Startup job

Startups do a lot of value extraction, but do little work to put value back in. They focus on immediate returns, product marketing and development. They are under pressure to perform and deliver. In a corporate environment your incentives are quarter/yearly, while start up incentives are daily/weekly.

Tragedy of the commons?

Do not feed back into the project - parasitic

Service industry connection to wikipedia > people work on the open source project when they are bored or in downtime of work.

Open source at work???

Short-term view vs long-term view

Users and devs want to release at a high level but some orgs may have interest in releasing early

Small group participation - Google Summer of Code - retention?

Location based activities for small groups

Project with in a project

Sub projects and local user groups - meetups

What causes the need to meet face to face?

People like being identified in a group - identity

Time thinking about swag

Meaningful for people

Face to face and meetings are very common

Hackathons common within OS > usually without an agenda, but mostly people use it to show up (and also people online) and get together > community aspect to recruit people (work on my project) and produce some output.

Financial health of Citizen Science endeavors is a concern

Best practices...

Shared language - what is a bug in OSS, Wikimedia and Citizen Science. How are they similar and how are they different?

April 27th

9am - Sharing work around Health and Sustainability -- Laying a Foundation

[

Similarities;

- Shared interest between communities
- Survey methods > great interest from everyone : these are challenges but not insurmountable problems.

Differences;

- Nature of the work is inherently different in CS vs OS. > however the people component is similar so that should be moved forward.

Take home's;

- It's all about the people
- Need to connect with Sociologists, Marketeers,
- Erin: Target the conferences where sociologists are and are actually looking to interact with tech minded/based people
- Quick read of papers - service. Not a peer review, but make suggestions on literature or on how to improve the paper
- Tap into HR and business field, org psych (I/O psych/organizational sociology)
 - Forming effective teams
 - Big 12 guy ? there were some people on virtual teams
-

Hope is to move forward; engaging sociologists--target mutual spaces for workshops
What can I do--there's a lot of work to be done.

SURVEY DISCUSSION

Insights into community roles and leadership opportunities (see slides from 4/27/18)

Tools are available to develop community outreach.

- International issues can be a show-stopper for CS (laws, regulations, etc)
- People need the information, so could be unethical not to share?
- In research on volunteer/user demographic gets slumped by privacy and international laws. Also the question on if the data will be outwards or inwards facing.
- Paper on deception and joke answers on demo questions

- > outliers on other fields are correlated to height
- work in this field is getting complicated in their efforts to be inclusive and not exclusive
- also no sample rate, because no sample frame is available (X people responded, but no pool size --? Thus can't do any of the traditional survey methods ; Pew foundation as exemplar, internal validation,)

Response rate issues because sample frame is not addressed.

- Once there are answers then it is taken as representative of the population.
 - >It's a lot easier when the person distributing the survey is part of the community.
- Being a member of a community vs an outsider makes a big difference: If you have a good idea of the size of the community, then you can make an educated guess on the size. Not quantitative but more subjective indication.
- Confidence in the data; are we trying to understand the ecosystem or something smaller? Define your community;
- Try to come up with new ways of asking questions so that integrates more with what people are actually doing in the community.

WikiMedia: big response rates because they look at a funnel within reader-contributor path > they have control over the software

Can't turn money into engineering. Independent but trusted organization that could take care of the heavy technical lifting.

Conversations within WikiMedia on setting up a committee that would be tasked with surveys and information gathering.

Pro-publica - lots of news media don't have resources but they can tap into this (?)

- Browser add-on available
- These infrastructure type

Within Zooniverse / CS there are some native survey actions by project leaders, however those are not very effective (cuz they're new at it and dont know how to do it). However a model for this would be greatly value adding because getting project leaders to buy in would yield interesting cross-project data.

Model needed; but none of the issues are hugely insurmountable.

Next topic

Who needs to be part of this conversation beyond this group? Who do we know are missing?

Next steps to advance

Taxonomy and mapping of shared aspect of people. Lower ground to start at.
Should have hypotheses that come out of this immediately.

Resources

- Visibility
- Dollars/fundraising strategies
- People
- Etc.

Mapping taxonomy between communities

- Shared aspect between people. The technology might be trickier, so start at a lower ground for people

There should be some hypotheses that can inform the other group

]

11am - Summaries and Discussion from Breakout Sessions

[...]

Section proposal for CSA Conference (May 2019)

Open Collaboration (conference deadline 4/29/18) Open Source Summit?

SEAN/Dean's Conference Room:

- Went through three stages leading to conceptual mapping between OSS - CS
- Social and technical support for Diversity and Inclusion collaborative measures

Place to share things with this group other than e-mail;

- Google Drive
- GitHub
-

SLOAN (Relevant people):

- Michael Brennan
- Josh Greenberg
- Digital Infrastructure :

Outcomes from the CS/PKI250 group;

- Reading List
- Mapping Taxonomy list on Github or Wiki > enables you to do things that *Are* publishable > does need more work to be well publishable
 - What's the method to get the mapping published?
 - Mappings between papers; in WikiData (example)
 - Not possible to do this in one publishable paper, since it's so big and complex. So need to chunk it in separate papers and works. a) to distribute the workload and b) to adequately represent the information
 - Chunk it by field would not be helpful, because that's what we've already been doing.

Paper?-

Who wants to what?

Concepts

Ways to instantiate

mapping

Review of the data gap--gap analysis staging future papers

- Need another layer or two to develop abstraction and applicability.