LEARNINGS *from*

OPEN-SOURCE TOOL TEAMS

about

PRIVACY PRESERVING MEASUREMENT

APRIL 2020

Interviews conducted by Okthanks on behalf of Clean Insights pioneered by Guardian Project, supported by Internews as part of the BASICS project.

OKTHANKS
okthanks.com

GUARDIAN PROJECT
https://guardianproject.info

Internews

INTRODUCTION

Clean Insights is both a methodology and an open-source SDK for privacy-preserving measurement. It aims to help answer key questions about app usage patterns, while not enabling invasive surveillance of all user habits. Clean Insights gives developers a way to plug into a secure, private measurement platform and provides user interactions that are empowering rather than alienating.

The Clean Insights work grew out of the need for open-source teams to have a better understanding of how their apps are used and the impact they have. Its immediate focus is to provide a solution for small-medium sized open source teams. This report represents the user research that was done to understand the needs of various open-source teams.

We held interviews with the following 10 tool teams who are part of the BASICS project led by Internews. They include TunnelBear, OONI, Least Authority, Qubes OS, KeePassXC, MISP, CalyxOS, Umbrella App, Psiphon, and Tella.

Our aim was to understand:

- What teams are interested in measuring
- What is meaningful for their mission
- What level of tolerance their users would have for being measured

In addition, we collected questions and concerns pertaining to the Clean Insights work.



KEY TAKEAWAYS

People have questions. With no answers.

There was an overwhelming positive response to the Clean Insights work. Some teams have used analytics tools or implemented their own methods for collecting data. Others have not. Both are eager to learn more about the project. Many are faced with questions that they do not have answers to.

- How are people using my app?
- How many users do we actually have?
- How can we even do this (measure usage) while respecting user privacy?
- Are we doing everything right? Apple adds noise. Should we be doing that?

For more on measurement questions, see 'Measurement Interests'.
For more on tech-related questions, see 'Questions & Concerns'.

Low capacity. A prominent theme.

We don't have a dedicated analyst.

What this means: Dashboard results need to be easy to digest for any member of a team. In addition, one participant expressed the desire for non-developers to be able to access and interpret the dashboard.

We don't have someone on our team thinking about data privacy all the time.

What this means: Clean Insights can bring a lot of value to teams by filling this gap and representing 'the person that always thinks about data privacy.' It can do this by setting a standard for open-source, privacy-centric tools that takes into account the modern practices of good actors and the work of academics.

Criteria for analytics tools.

Below is a cumulative list of criteria shared by interviewees.

- Open-source.
- Well-maintained: Has a big existing user community or is backed by a company with a sustainable business model.
- Not locked in to a vendor.
- Modular enough that we can turn on the additional tracking features we need.
- Offers the option to self-host. Though we would prefer that it's hosted by someone else, if that person is trustworthy and has established principles of how they are handling data.
- Completely stand alone.
- No connection with third parties.
- Supports a dashboard that could be in-house.
- Completely disconnected from the internet.

SYMPOSIUM CONSIDERATIONS

Below are challenges to discuss and work on as we gather great minds during the month of May!

How can we understand ‘how people use apps’ in a noninvasive way?

So many teams are interested in knowing how people use their apps. Within this question, there are various potential methods of measuring. How does Clean Insights learn about user behavior differently? And how does it do it in a way where users are comfortable and eager to give consent? This seems to be one of the biggest human challenges.

How can we combine data in meaningful ways?

If we’re thinking about the minimal viable data, we should consider which data is needed to allow teams to draw meaningful conclusions about their users’ experience.

What’s the best way to ask? And when?

Participation in measurement is key to having an accurate data set to guide decisions. The experience a user has when asked for consent will drive or stifle participation.

> *I assume we have a small user base. Having 80% [of people opting in to measurement] is ideal. We want as much [input] as possible.*

How can Clean Insights analytics cross-pollinate with user research?

Analytics is a helpful supplementary tool for user research. It’s an opportunity to get quantitative (observation-like) data and qualitative input from a large pool of users.

Are users at risk in any way? Which unintended consequences can we anticipate?

Threat modeling is an important piece of Clean Insights. Along with this, we should consider if there are any unintended side effects of implementing the SDK.

TOPICS FOR DISCUSSION

- Overall approaches to ethical data collection for organizations
- Censorship resistance
- Dealing with concerns of corroborated data & subpoenaed data — Threat modeling

> *We wouldn't want to collect any data that, together, would reveal personal info.*
>
> *We don't want any data that would be subpoenaed by the government. If we had data that was useful in a court case, that wouldn't be good.*

Note: Metrics is a broad topic. It would be helpful to clarify the 'Types of measurements' we're talking about.

POSSIBLE OUTCOMES

- Guidelines and best practices. A checklist of things you need to be thinking about when implementing analytics.
- Risk-Reward model for measurements
- Data storytelling examples
- Sample consent design wording
- An outline of privacy-preserving methods we should consider
- An outline of technical considerations

MEASUREMENT INTERESTS

APPLICATION METRICS

- How many users
- Number of installs
- Number of uninstalls
- How many installers are using the app
- Monthly active users

> *Funders want to know the number of users. It would be useful to show that people are using it on a regular basis.*

- Version numbers

> *It would be helpful to see the number of devices on a particular version.*
>
> *It would be helpful to connect users who haven't updated with their country to better understand why.*

- How many users on each version in use
- Countries and languages

> *We put a lot toward localization. It would be nice to know its impact and to know where localization may be needed.*

PERFORMANCE

- Crash reports

> *Right now we get a user report. We're left with haphazard [reports].*

- Internal debugging

> *If it could capture that [a log of debugging activity], that would be incredible.*

- Diagnostics
- Contextual data points that may contribute to errors
- Device health check

BEHAVIOR

- What are people using and doing within the app?

> *I'm a big proponent of having a streamlined application. [...] "Can we trim it in here?" [...] I'd like to say we made that decision with data.*

- Which features are people using?
- How much time is spent using different features?
- How many people are using a new feature after we roll it out?

> *It's hard to know if people use it. Real raw data would be way more useful.*

- How are people using a feature?

> *I'm not sure if the way we put it in would be the way they'd want to use it.*

- How many people spend how much time?

> *How do you know if they are spending a bunch of time because they really enjoy the feature or because they have stumbled?*

DESIRED CAPABILITIES

- Survey

> *There's a divorce between us and the user. Unless they literally go on Github, there's no way for us to know [...].*

- AB testing

A user survey or probing the user is not as effective. A data driven approach would be more useful.

OTHER

- How is the experience? A satisfaction rating from users.
- Does using the app put users at risk or in danger?
- Which content is useful?

Refer to the 'Use Cases' for additional insight.

USE CASES

The use cases below represent 3 of the 10 teams we interviewed.

Umbrella App

A training app for digital and physical security.
https://secfirst.org/
Android and iPhone and Web (Beta)

WANT TO MEASURE

- Crash data (what the stack for the crash is/chain of functions)
- Language use to get insight into localization needs and impact
- Which parts of the app are used the most
- Where people spend their time
- Which content is useful
- If the design of a feature is intuitive

> *I'm not sure if the way we put the feature in would be the way they'd want to use it.*

DEPLOYMENT

- Opt-in

Tella

A documentation app for human rights defenders. Their objective is to replace other camera apps with Tella, because it's a more privacy-conscious app.
https://hzontal.org/tella/
Android

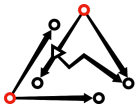
WANT TO MEASURE

- How much time do they spend in the app
- What are they doing
- How long is it taking them
- Are they using feature A, B or C
- Language
- Crash reports
- Our users are very often on no or poor internet. How they use it when they have a poor connection and if they use offline mode is helpful to know.
- It would be interesting to know urban vs rural

> *It would be good to know how they're using the app with more precision. For example, If there's a big crackdown in Egypt — seeing during that week that people are using those features in that time period. And when people are using the feature. We would have a sense of how people are using it.*
>
> *Funders want to know the number of users. It would be useful to show that people are using it on a regular basis.*

DEPLOYMENT

- Opt-out



Least Authority
PRIVACY MATTERS

TAHOE-LAFS

A secure decentralized cloud storage system. Least Authority is actively researching how to adapt Tahoe-LAFS to make it more usable for human rights groups.
https://tahoe-lafs.org/trac/tahoe-lafs

WANT TO MEASURE

- How people are using it
  - Quantity of storage
  - How often do they upload things
  - How often they open the application
- Where are its users
- How many users and how many deployments
- Crash reports

Even if it's a crash report, you're sending a lot more personal data.

QUESTIONS & CONCERNS

The following list captures the various questions and concerns expressed by teams.

- How are we pulling this off? (tracking data but not people)
- What are we doing with IP addresses?
- Is there any risk for my users? Could the data collected on their device put them in danger in any way?
- Are you collecting data points that could be corroborated to reveal personally identifying information on a user?
- Is any data valuable in a court case if subpoenaed?
- Deanonymization - How are we handling this potential?
- My app doesn't ask for trust. We demonstrate it. How does Clean Insights align with this value?
- Because we're an anti-censorship tool, some of the data we're collecting is likely censored. How can we handle this?
- If session time is 1 hour, it's useless statistics. If it's 1 year, it may be too much. What's the minimal session length that gets you useful statistics to report usage but not to de-anonymize?

Responses related to certain measurements

DEVICE HEALTH CHECK

- How would it work? Would it check the phone periodically?
- OONI probe tests could be very helpful here - detecting how much bandwidth, how good their network. https://ooni.org/nettest/ndt/

DIAGNOSTICS

- We use https://sentry.io/welcome/. It helps triage and prioritize errors.

SENSORS

- It could require extra permissions.

- It could be an open door for exposing certain things.

CRASH REPORTS

- Even if it's a crash report, you're sending a lot more personal data.
- Because they're dealing with files, personal information may be in crash reports.

CONCLUSION

These insights will inform the development of the Clean Insights methodology, symposium, SDK and website content. Thank you to each participant who contributed to this portion of the project!