Helping Mobile Application Developers Create Accurate Privacy Labels

Jack Gardner,* Yuanyuang Feng, Kayla Rieman, Zhi Lin, Akshath Jain, and Norman Sadeh

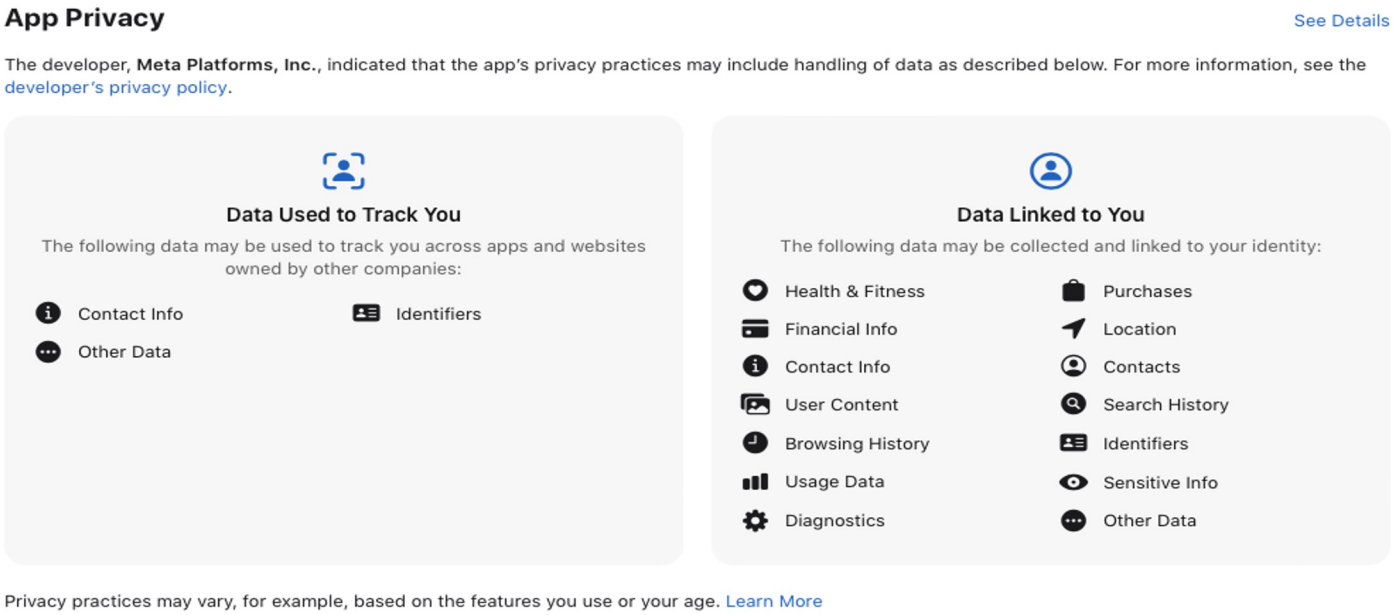
usableprivacy.org privacyassistant.org

*presenting today

Outline

- What are mobile application privacy labels
- Challenges for developers
- Relevant prior work
- Development and evaluation of the initial *Privacy Label Wiz*
- Refining *Privacy Label Wiz*
- Remaining challenges and future work

An Apple privacy label



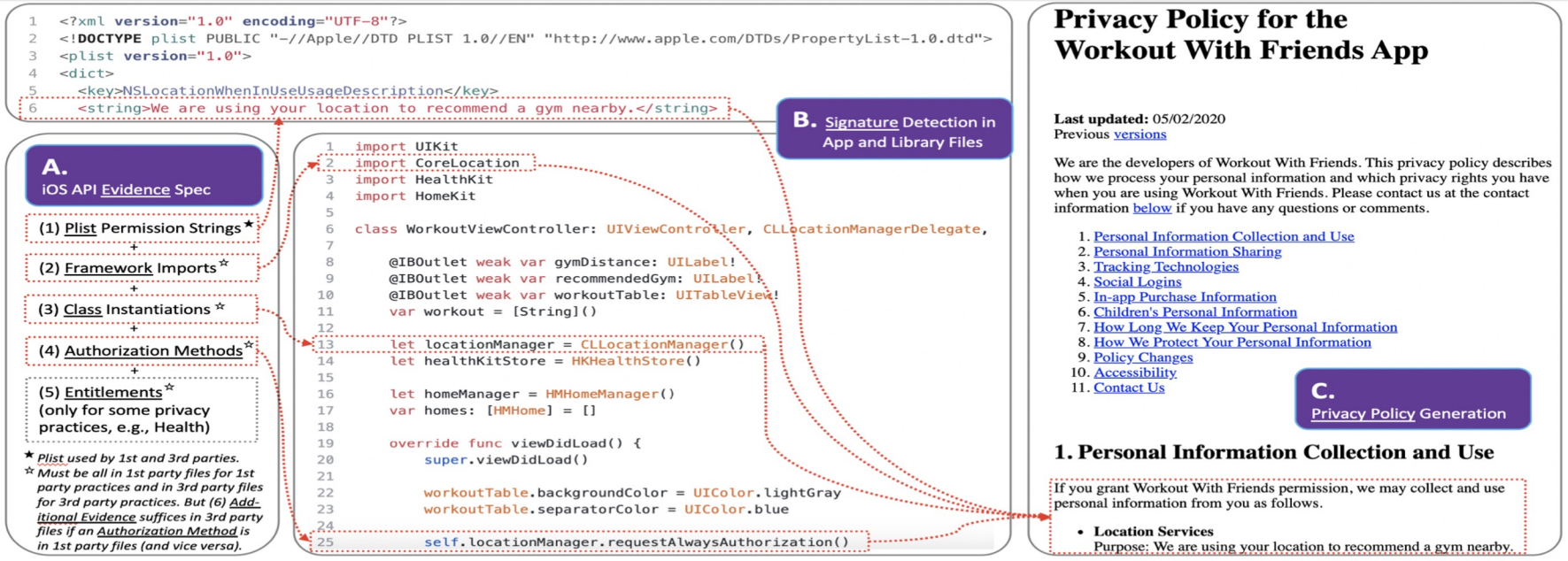
Facebook's privacy label: https://apps.apple.com/us/app/facebook/id284882215

Challenges in creating privacy labels

- Time intensive for applications that collect a lot of data
- Misconceptions about data linked to users
- Under or overreporting data types
- Underreporting of third-party data collection and use
- Underreporting implies lack of awareness, overreporting implies lack of understanding

Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 588, 1–24. https://doi.org/10.1145/3491102.3502012

Building on the analysis framework of Privacy Flash Pro

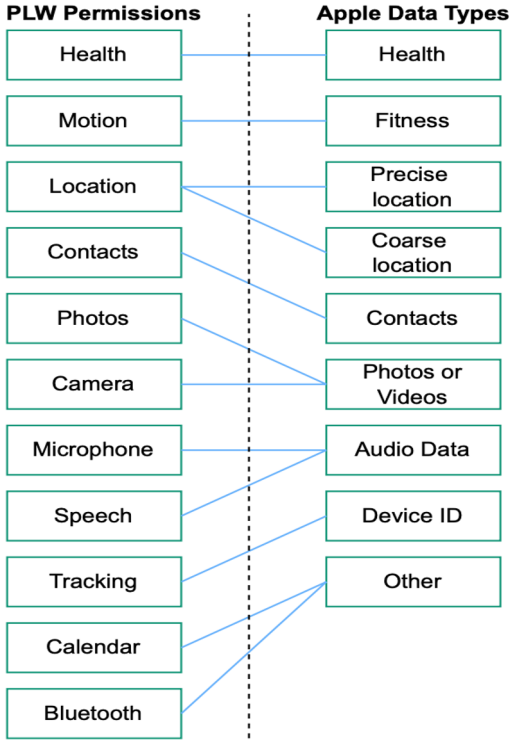


S. Zimmeck, R. Goldstein, and D. Baraka, "PrivacyFlash Pro: Automating privacy policy generation for mobile apps," in 28th Network and Distributed System Security Symposium (NDSS 2021). NDSS, 2021.

S. Zimmeck, P. Story, R. Goldstein, D. Baraka, S. Li, Y. Feng, and N. Sadeh, "Compliance traceability: Privacy policies as software development artifacts," in Open Day for Privacy, Usability, and Transparency (PUT), Stockholm, Sweden, 2019.

Initial version of Privacy Label Wiz

- New UI with specific focus on privacy labels
- Use static analysis to start dialogue with developer
  - Mapped detected iOS permission to Apple's data types
  - Included full list of Apple's data types for developers to review
  - Provided a list of reflection questions for developers to more deeply consider their application's data practices



Evaluating the initial version with developers

- Conducted in depth think-aloud interviews with 4 developers
  - Developers bring their own application to the interview
  - Working with the tool led some developers to update their labels
  - Helped us identify fundamental shortcomings:
    - Lacked introduction to the purpose of the tool and privacy labels
    - Didn't tell developers where PLW depends on their knowledge
    - Needed to help developers understand third-party data flows
    - Developers also struggled with workflow of the initial UI, felt overloaded with information – highlighted the need to breakup the UI and support iterative workflows

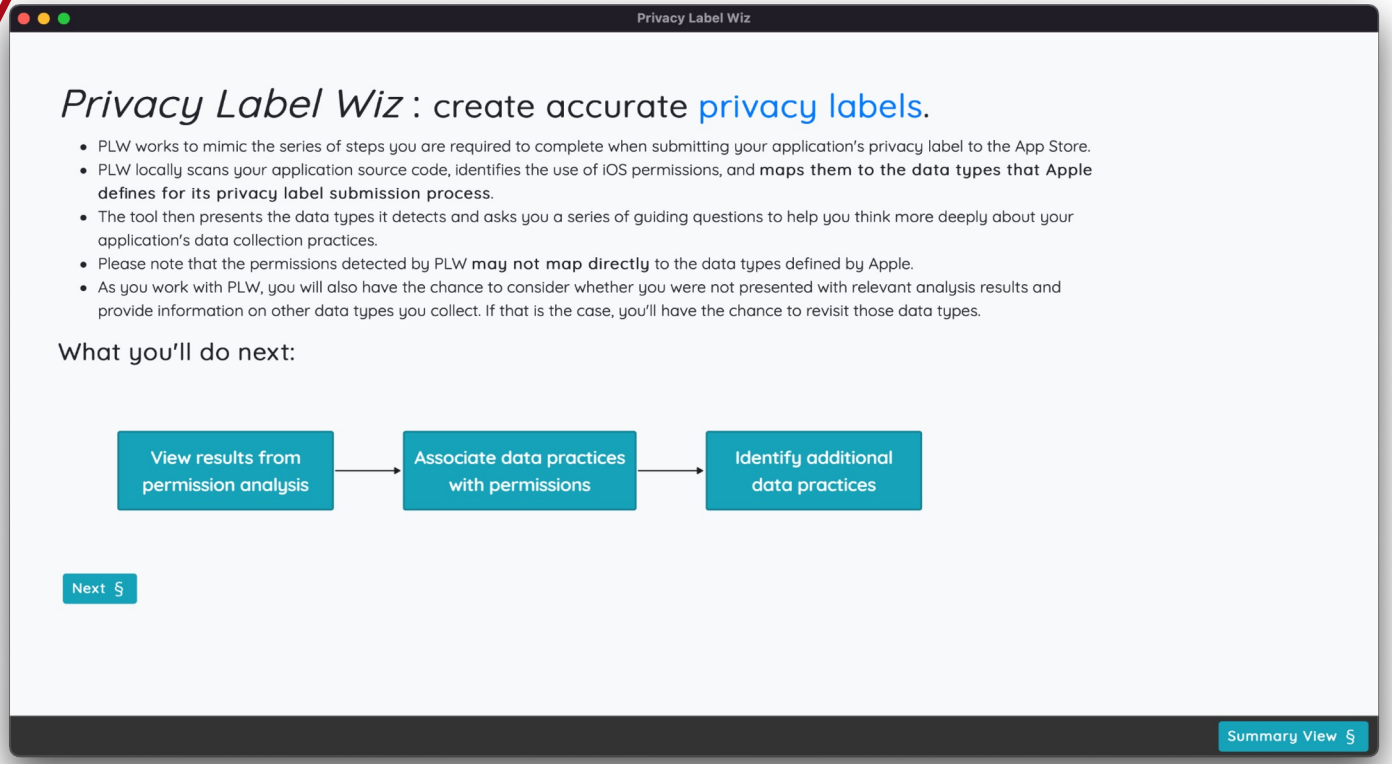
Refining Privacy Label Wiz

- We redesigned the initial version of Privacy Label Wiz
- Feedback from our initial interviews led us to:
  - Provide instructions to introduce the tool to developers
  - Modify the UI structure to resemble that of Apple's web form
  - Add room for developers to mark and revisit their decisions
  - Improve reflection questions presented in the UI

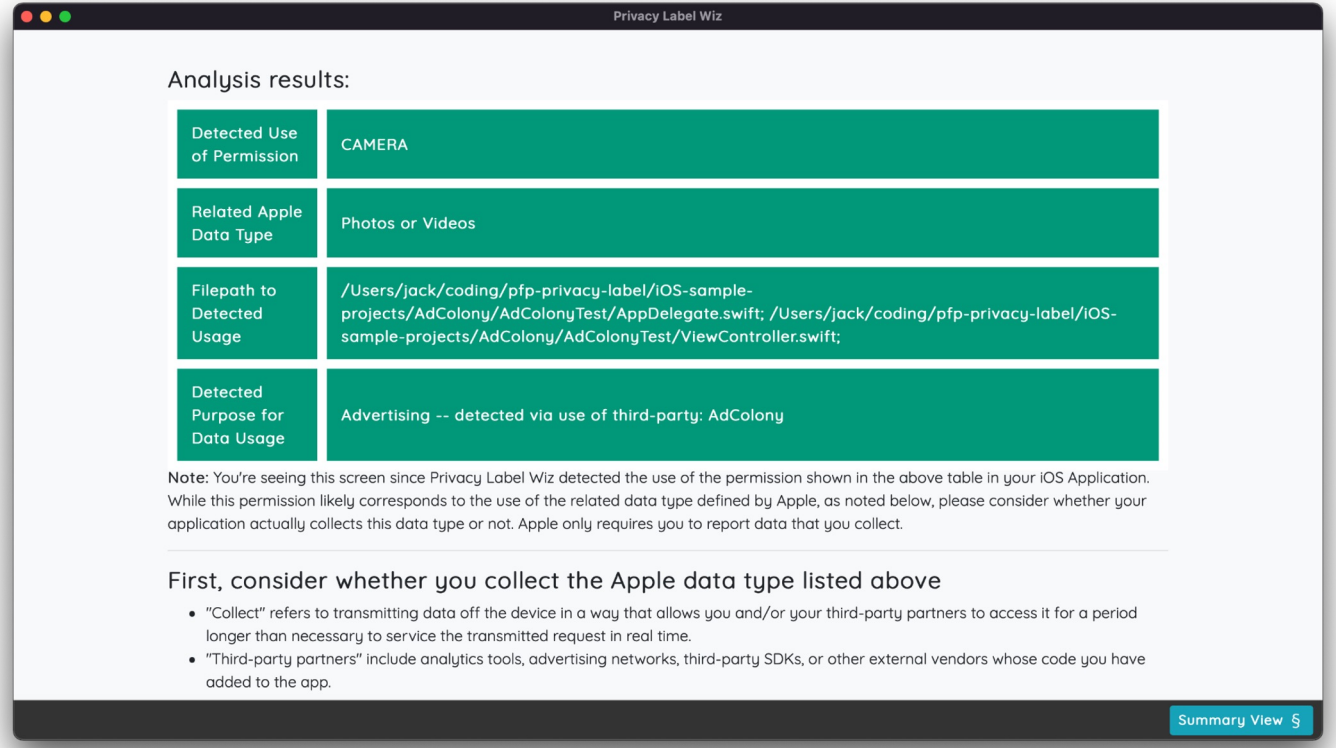
Privacy Label Wiz Workflow

- Download and run the software tool
- Load iOS application (.xcodeproj folder)
- PLW runs analysis (nearly instantaneous)
- Initiate dialogue with developer driven by analysis results
- Iterative workflow designed to allow the developer to revisit and refine data practice disclosures

Overview the purpose and structure of *PLW*



One data type at a time



One data type at a time

Select whether the above data type is linked to the user's identity:

Indicate if the data collected from this app is linked to the user's identity (via their account, device, or details).

Note: If you do any of the following to de-identify or anonymize user data, you may not need to report your data as linked.

- Stripping data of any direct identifiers, such as e-mail address or name, before collection.

Select whether the above data type is used to track the user:

Indicate if the data type will be used for tracking purposes

Do you or third-party partners use the data type for tracking purposes?

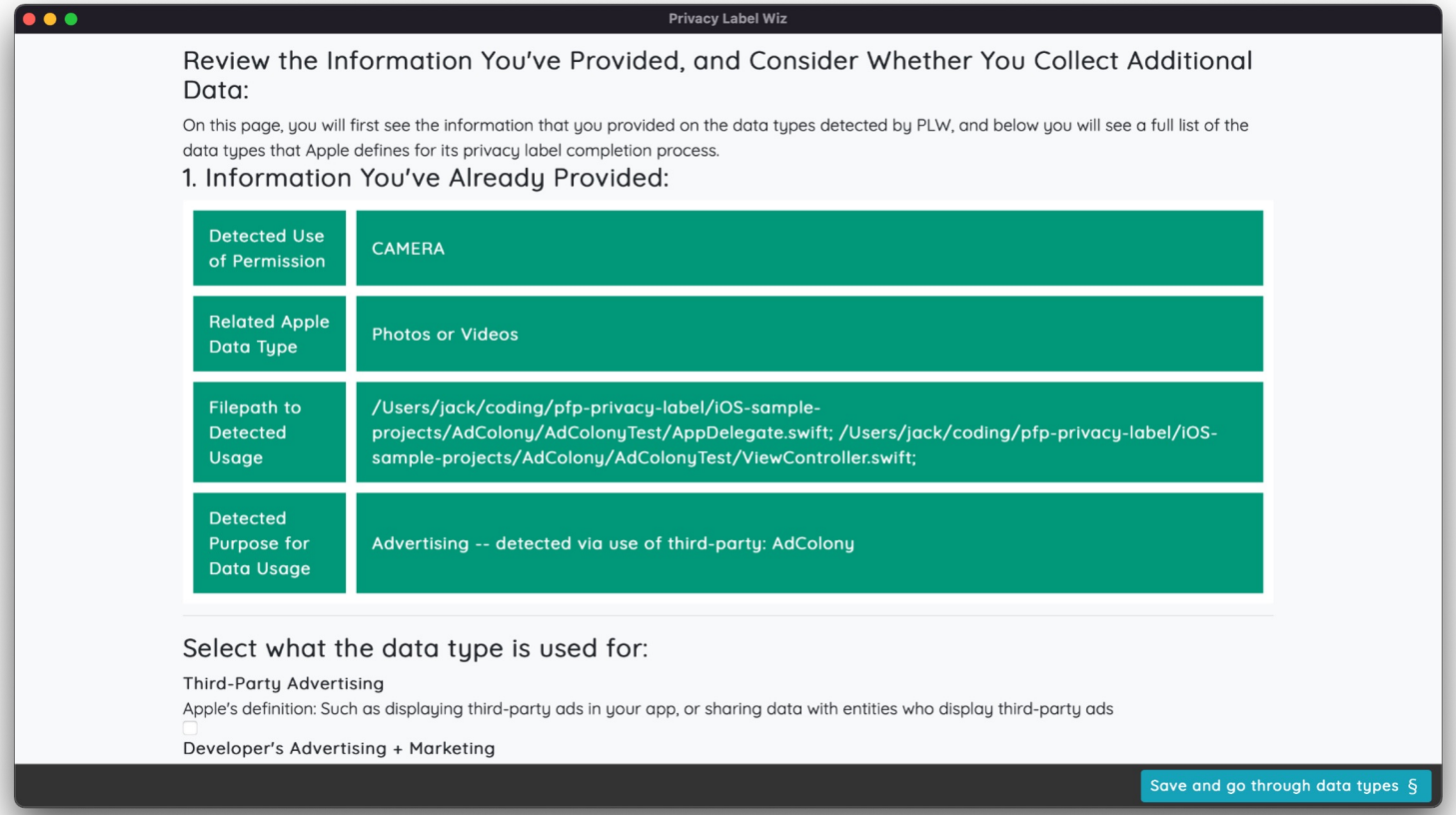
Yes, we use these data for tracking purposes

No, we do not use these data for tracking purposes

Not sure (mark this to revisit this choice later)

Save and continue §

Reviewing answers on the summary page



Reviewing answers on the summary page

Select whether Photo or Video data is linked to the user's identity:

! (remember, you indicated you weren't sure about this answer)

Indicate if the data collected from this app is linked to the user's identity (via their account, device, or details).

Note: If you do any of the following to de-identify or anonymize user data, you may not need to report your data as linked.

- Stripping data of any direct identifiers, such as e-mail address or name, before collection.
- Manipulating data to break the linkage and prevent re-linkage to real-world identities.

Note: additional requirements for data to be considered **not linked** are below.

- You must not attempt to link the data back to the user's identity.
- You must not tie the data to other datasets that enable it to be linked to the user's identity.

Are the Photos or Videos collected from this app linked to the user's identity?

○ **Yes, Photos or Videos** collected from this app are linked to the user's identity.

○ **No, Photos or Videos** collected from this app are not linked to the user's identity.

◉ **Not sure** (mark this to revisit this choice later)

Selecting additional data types on the summary page

Full list of data types:

Consider whether you collect any of the additional data types listed below. As mentioned earlier, you should only select a data type if you collect it.

- **"Collect"** refers to transmitting data off the device in a way that allows you and/or your third-party partners to access it for a period longer than necessary to service the transmitted request in real time.
- **"Third-party partners"** include analytics tools, advertising networks, third-party SDKs, or other external vendors whose code you have added to the app.

Reflection Questions:

Please review each of the comments and questions below and consider whether they apply to your application's data practices.

Questions on third-party data practices:

Did you report all data practices of the SDKs, frameworks, or other third-party services you used?

- You may find it helpful to consult the documentation of these services. They may provide specific information on how to complete privacy labels.

Do you advertise (e.g., do you display ads to your users, send them marketing content, or share data with entities that will display your ads?)

- If yes, under the usage section for relevant data types, please consider whether you should check "developer's advertising or marketing".

Do you include third-party ads in your application, or, do you share data with entities that display third-party ads?

- If yes, under the usage section for relevant data types, please consider whether you should check "third-party advertising".

Revisiting uncertain answers and generally reconsidering the data you collect:

You marked "not sure" for your answer on Photos or Videos above.

- We encourage you to err on the side of caution and report a data practice if you are not sure whether it is performed. You may always update your privacy label at a later time.

Do you really need all the data you are collecting?

- We encourage you to always ask yourself this question. Collecting too much data can make your app non-compliant and is in general a liability.

Questions on diagnostics, data linkage, data used to track users, and relevant laws:

How do you run diagnostics?

- If you are using Apple's default App Analytics, you do **not** need to report this.
- If you are using any other service (e.g., Crashlytics, Yahoo's Flurry, or other analytics like those used to log events from a WebView), under data usage, please consider whether you should check "usage data" and "diagnostics".

Save and go through data types §

Reflection questions at the bottom of the summary page

Reflection Questions:

Please review each of the comments and questions below and consider whether they apply to your application's data practices.

Questions on diagnostics, data linkage, data used to track users, and relevant laws:

How do you run diagnostics?

- If you are using Apple's default App Analytics, you do **not** need to report this.
- If you are using any other service (e.g., Crashlytics, Yahoo's Flurry, or other analytics like those used to log events from a WebView), under data usage, please consider whether you should check "usage data" and "diagnostics".

Do you have any way to analyze individual app users?

- If yes, please consider whether the data types you collect are correctly marked as "linked to users".
- If you also share your analysis of app users with advertisers, please make sure you have correctly marked "data used to track users".

Where do your users live?

- If your users live in the EU, California, or other places with privacy laws that define personal data, please consider your data practices with these laws in mind and think about whether you should mark relevant data types as "linked to users".

Developers still need more help

Future work:

- Provide ability to add notes / documentation as developers use *PLW*
- Provide easy access to third-party documentation
- Evaluate the redesigned tool with developers

Generally:

- Need to integrate these tools with developer workflows
  - Must understand developer time constraints, tools must be responsive
  - Record developer knowledge
  - Provide specific warning explanations

Thank you!

Q & A?