What data do you have?

Understanding what data you actually collect is easier said than done. People often have more data than they think.

Step 1, whose data do you have?

- Employee data - this includes recruitment, current and past employees, contractors, interns, fellows, board members, investors, and volunteers.
- Partner data - this includes business contacts at companies you consider to be partners and collaborators.
- Customer data - this includes your businesses customers and/or end users. You might have separate lists for each of your products and/or services.

Action: make a list of whose data you have.

Step 2, what types of data do you have?

The Lean Data Practices approach is to evaluate what you have at a deep level to understand what data is sensitive. Group your data into categories (See Table 1) to better document and communicate what you have. This creates a useful framework for privacy and security. For example:

- your engineering policy might be that it's okay to collect certain types of customer data by default but that other data categories require extra reviews and approvals.
- your security policy might be to limit access or require more protections for certain types of data
- your privacy policy might use data categories to better explain to customers information they care about

Action: For each group in Step 1, add the data categories that you collect.

Step 3, what data is sensitive?

This answer will be different for every company. Sensitive information includes data that can identify people, reveal their private affairs or preferences, or which people would be surprised that you have.

Action: In your list, highlight the sensitive data that you collect. This is the information that you need to engage with your users about.

Table 1

Data Type	**Definition**	**Examples**	**What's the big deal? "Sensitive Data"**
Technical Data	Includes information about devices, client software, compatibility, logs and networks	Device identifier, OS, battery, any unique device attributes you are collecting (e.g. permissions to other apps); available memory, crashes and errors, outcome of automated processes like updates and version#s, language, timezone, preferences; websites, addons, and other 3rd-party software interactions with the client software; IP address, mobile carrier, device and browser information, dates and times of use	IP addresses combined with date and timestamps can be used to fingerprint a particular individual based on their activity on your service. Data minimization makes it harder to identify your customers to hackers and law enforcement.
Interaction Data	Includes information about direct engagement with the app, website, or service; session length; and content served	Examples: clicks and scrolls on icons, links, and buttons, interactions with features and actions taken; time, frequency and duration of activities; whether app is in foreground, background or active; content served (e.g. impressions) and actions taken by users in response	

Account Data	Includes information about required registration details and optional profile information	Email address, Name, phone number, address, username, photo, passwords	Account and profile information contains identifiable personal data. Strong security makes it harder for hackers to compromise this data. In most jurisdictions law enforcement can obtain this data without a court order.
Location Data	Includes information about the location of the device or individual	Country, metro, or precise location	Combined with Account and Technical data, this data can reveal a person's location history. This data may be interesting to advertisers, hackers, and law enforcement. In most jurisdictions law enforcement can obtain this data without a court order.
Marketing Data	Includes information about user journeys,analytics, campaign referrals, advertising IDs, advertising preferences	Analytics used to determine user journeys; Cookies; Advertising identifiers	This data can be used to target advertising to individuals based on their interests. Many people find this to be invasive of their privacy.

Content & Communications	Includes information about the websites, articles, music, videos, images that people watch, save, upload or share with others	Information about the websites, articles, music, videos, images that people watch, save, upload or share with others	This information is considered extremely private by most people.
Data about people received from other sources	Includes information about customers received by other services or people within customer's network	Data received by partners, advertisers, publishers, developers, public databases, social media platforms; customers' public posts, comments, reviews, or uploads to your service including contact	This contains identifiable personal data which people are unlikely to expect that you have. Strong security makes it harder for hackers to compromise this data.
Payment Data		Billing, shipping and contact details; data about transactions	This contains identifiable personal data that can be used to steal. Strong security makes it harder for hackers to compromise this data.