

# Understanding First-Party Data: A Guide for Newsrooms

Written By Kevin Anderson

On its face, understanding first-party data is deceptively simple – it's the information publishers collect directly from their audiences and about their behaviour.

As news organizations shift away from platform dependence and toward deeper, more direct relationships with audiences, collecting and utilizing first-party data has become a major focus.

Yet this task is complex, and doing it well requires news publishers to convert anonymous users to known ones through newsletter sign-ups, site registration, membership, subscriptions, and events.

First-party data is critical to every aspect of news organizations' operations today.

- It improves content strategy by providing insights into the content and formats that meet user needs.
- It provides invaluable insights for product operations.
- It supports marketing efforts to engage newly known users and drive conversion.
- It increases the return on advertising by improving its relevance.

While almost every department across a news organization collects some form of first-party data, they often do so in isolation – using different systems, speaking different jargon, and aiming toward different goals. This lack of shared infrastructure and understanding of audience

data makes it challenging for news organizations to coordinate efforts, align on strategy, and fully realize the potential of first-party data.

This guide was created to help navigate these challenges. Especially in small and local newsrooms, where resources are tight and people wear many hats, it's crucial to build shared understanding and more intentional collaboration around audience data.

We'll walk through the different types of audience data—**first-party**, **second-party**, and **third-party**—explaining where the data comes from, what it's good for, what its limitations are, and how it's typically used in a news organization. We'll also highlight why the *source* of data matters: it influences its quality, ethical implications, and legal constraints. And lastly, we'll provide a framework to assess your audience data and identify gaps.

## **From first-party to third-party: The source of audience data impacts its value**

Most publishers understand the value of audience data, but the volume can be overwhelming. It is helpful to categorize audience data by its source and the level of consent involved, as this shapes both its value and its limitations.

The terms first-party, second-party, third-party, and zero-party data are categories that describe where the data comes from and how directly it was collected. As the numbers get smaller, the publisher gets closer to the audience—and the data tends to be more relevant, accurate, and ethically usable. Each has a role to play in helping news organizations achieve their goals, and the data sources work best when used in concert. Understanding the strengths and roles of each helps product managers orchestrate their integration and application.

Data type	Zero-party	First-party	Second-party	Third-party
<b>How it's collected</b>	Voluntarily shared by the user (e.g., in forms or settings)	Captured through use of your site, app, or services	Purchased from or provided by a trusted partner (e.g. platform or co-hosted initiative)	Purchased or acquired from data brokers and aggregators
<b>User relationship and consent</b>	Direct. Data shared intentionally and explicitly by the user.	Direct. Data collected through direct interaction with implicit or explicit consent.	Indirect (shared by a partner organization). Consent given to the original collector.	Indirect (shared by external data aggregators). Consent varies, and may have been collected without the user's knowledge.
<b>Examples of sources</b>	Account, site and app preferences; User surveys; User research	Newsletter sign-ups; Subscriptions; Membership; Registration; Site, app and product analytics	Data brokers; Social media platforms; Marketing partners; Platform-shared analytics (e.g., from Facebook or Substack); Partner event registrations; Shared CRM data	Commercial data providers (e.g., Google, Acxiom, Nielsen); Third-party cookie tracking; Website tracking data
<b>Data fidelity</b>	Very high – user-reported and specific	High – directly collected, behavior-based, including registration and subscription data	Moderate to high – can vary based on source	Low – often inferred, outdated, or misaligned
<b>What it's good for</b>	Personalization; Direct engagement; Audience segmentation	Measuring behavior and trends; Content/product optimization; Conversion tracking, audience segmentation	Benchmarking; Cross-channel insights; Lookalike audience targeting	Ad targeting; Market research; Broad audience profiling
<b>Limitations</b>	Limited in scale; users must volunteer info	Requires infrastructure to collect, store, and apply ethically	May not fully align with your audience or consent standards	Privacy concerns; often restricted by regulations (e.g., GDPR, CCPA)

## First-Party (and Zero-Party) Data

First-party data is the most foundational and strategic category of audience data available to news organizations.

### What it is

**First-party data** is information a publisher collects directly from people interacting with its owned platforms—such as websites, mobile apps, email newsletters, live events, and registration or subscription flows. It includes both anonymous and known users, depending on whether access is open, metered, or paywalled. As publishers' digital sophistication has grown, so has the value and richness of this data.

Examples include:

- **Attention data** — how users interact with content, such as pageviews, time on page, scroll depth and user journeys.
- **Product usage and engagement data** — actions like clicks on buttons or links, commenting, saving content, setting user preferences, opting into push notifications, and results from A/B testing.
- **Transaction and access data** — collected via newsletter sign-ups, event RSVPs, membership and subscription purchases. For publishers with e-commerce services, it can include purchase history, which can be mined for interest and intent data and tied to individual users.

**Zero-party data** is an important subset of first-party data used to distinguish information shared with a publisher intentionally and voluntarily by the user, rather than collected passively through interactions on a publisher's owned platforms. You might also hear zero-party data referred to as declarative data. In a news context, zero-party data includes responses to surveys and content preferences selected that shape a personalized experience. With zero-party data, publishers do not need to infer or guess what the individual needs, because they've declared their preferences.

### **Why it's valuable:**

In recent years, the rise of digital audience revenue models—and the decline in referral traffic from search and social media—has made converting users from unknown to known a strategic imperative for many publishers. In other words, publishers are increasingly focused on acquiring identity data, typically in the form of an email address, to unlock the full value of first-party data.

This shift allows publishers to move from generalized audience metrics to user-level insights: not just what percentage of traffic goes to housing coverage, but which specific users return to read housing stories. This deeper understanding enables more personalized content, marketing,

and user experiences—ultimately improving engagement, conversion, and retention.

Many publishers now routinely segment their audience analytics based on identity tiers—such as anonymous, registered, and subscribed users—and build product and marketing strategies to move users through that funnel.

Because it comes directly from user interactions with your products, first-party data is highly relevant, trustworthy, and actionable. It enables strategic decision-making across teams and unlocks more personalized, targeted, and sustainable approaches to growth.

Publishers use it to:

- Build strategies around converting unknown users to known users to support long-term engagement, loyalty, and monetization.
- Understand which formats and topics engage their most valuable audiences.
- Design and iterate features or products based on real user behavior.
- Support direct-sold advertising and reduce reliance on third-party ad networks.

As publishers have come to understand the outsized value of known users, first-party data is increasingly leveraged across the organization to improve editorial strategy, distribution and revenue. For example, zero-party data can deliver what FT Strategies calls [active personalization](#), where users customize their own experience. It is in contrast to passive, often algorithmically delivered personalization which is inferred based on first-party behavioral data and third-party demographic or interest data. Again, the value is in that news publishers do not need to infer users' preferences. Active personalization led to an 86% increase in engagement at the Financial Times, a 60% increase at Gannett and a 23% increase at Mediahuis.

Here are more examples of how publishers are deriving value from first-party data:

- The Financial Times developed a framework to measure how engaging content is for their members and subscribers.
- Bloomberg leverages first-party audience segments to serve targeted advertising without relying on third-party ad exchanges. It has improved user experience by reducing the site load time introduced by third-party tags and improved CPMs by 20%.
- Cityside analyzed donor behavior with AI to inform its fundraising strategies.
- The Boston Globe tripled push notification open rates—from 2% to over 6%—by letting users choose which alerts they receive through a preference center.

### **Where you'll find it:**

First-party data is collected across nearly every touchpoint in your digital ecosystem. It often lives in disconnected systems used by different departments.

Common tools include:

- Product analytics (e.g. Mixpanel or event-based data in Google Analytics)
- Email/newsletter platforms (e.g. Mailchimp, Sailthru)
- Registration/membership platforms (e.g. Piano, Zuora)
- Push notification services (e.g. Pushly, Airship, OneSignal)
- A/B testing tools (e.g. Optimizely, VWO)
- Survey tools and preference centers

Publishers are also increasingly implementing intentional strategies to convert unknown users into known ones, typically by prompting users to provide an email address or create an account. These tactics support identity collection and expand the pool of usable first-party data.

Common conversion tactics include:

- Registration walls – gating access to content behind a free registration form
- [Cookie walls](#) – requiring users to accept cookies to continue browsing
- Engagement features – like [commenting](#), saving content, or app personalization that require an account
- Limited-run newsletters and “register-to-receive” resources – used to collect email addresses without requiring a long-term newsletter commitment

### **Considerations and limitations:**

Building an effective first-party data strategy requires a technical and organizational infrastructure to collect, manage, and apply insights in alignment with both business and audience goals. It also requires ongoing maintenance and governance. Preferences change, contact information expires, and behavior evolves. Regular audits, deduplication, and thoughtful data expiration practices ensure your insights remain relevant and actionable.

Nearly every newsroom department collects first-party data, but often for different purposes and using different tools. As a result, a news organization’s first-party data often suffers from fragmentation, duplication, and inconsistent definitions. Email and physical addresses can help unify records, but must be handled carefully.

Trust is critical to the collection of first-party data, as nothing destroys confidence more than leaks of personally identifiable information. Robust privacy policies and secure data infrastructure are essential to sustaining long-term audience relationships.

When collecting first-party data, consent and transparency are critical. Some data (such as zero-party) is proactively provided with clear intent,

while other data (like behavioral tracking) may require explicit user consent under privacy regulations like GDPR and CCPA. Publishers must establish ethical and compliant practices for collecting, storing, and using this data.

Tactics for collecting data — such as registration walls — must be carefully calibrated. If the value exchange isn't clear, users may opt out—or never convert at all. Consent mechanisms, frictionless UX, and a transparent purpose for collecting data are essential to maintaining user trust and meeting compliance requirements.

Zero-party data, while especially high-quality, is typically offered by only a small fraction of users. It can take time to collect enough to inform meaningful strategy. Publishers must make the value of providing this information clear, whether through more relevant content, time-saving features, or personalized offerings.

## **Second-party data:**

Second-party data is simply someone else's first-party data that has been shared or sold to you.

Acquiring second-party data allows a publisher to increase their store of first-party data from a trusted source.

### **What it is:**

Second-party data is audience data collected directly from users by another organization, then shared with or sold to your organization. This includes, for example, event registration lists from a co-hosted event, or data purchased from a research firm or community organization

An easy way to think of it: it's someone else's first-party data that you obtained through an agreement.

The line between first-party and second-party data can get blurry. For example, if you co-host an event and gain access to attendee data, is that

first-party or second-party? It depends on consent language, data handling protocols, and who the user believed they were sharing data with.

Social media and platform-shared data add to the complexity—Instagram Insights or Facebook Page Analytics feel like first-party data because they're tied to your account, but are technically third-party platform data about your presence, not your users. These gray areas highlight the need for clear definitions and governance when integrating data sources.

### **Why it's valuable:**

Second-party data is useful for reaching and understanding an audience outside of your owned platforms. It can help publishers extend their reach and enrich internal datasets by adding context or filling in gaps.

Publishers use it to:

- Reach local communities or demographics they don't yet serve
- Conduct outreach based on shared lists or aligned engagement goals
- Seed targeted messaging or cold outreach, such as SMS campaigns tied to a civic issue
- Generate revenue by selling their own data under transparent, privacy-compliant terms

For example, Outlier founder Sarah Alvarez [purchased marketing lists in order to cold text thousands of Detroiters](#) to offer personalized information about housing issues tied to their address.

Selling your first-party data to advertisers or marketers as second-party data can also be a source of income for publishers. For example, publishers who host live events may share RSVP lists with event sponsors as part of their advertising agreement. In these cases, publishers need to be careful about how consent is obtained and communicated—attendees should be clearly informed that their information may be shared with

sponsors, ideally through an opt-in mechanism or transparent privacy notice at the point of registration.

### **Where you'll find it:**

Second-party data is often obtained through in-kind or paid sharing of data, such as through direct collaboration, event partnerships, shared initiatives, or paid acquisition.

Common sources include:

- Co-hosted or sponsored events
- Cross-promotional campaigns with aligned organizations
- Voter files or contact lists purchased for outreach
- An [exchange of first-party data](#), or [data collaboration platforms](#)
- Platform-shared analytics for your social media accounts (e.g. Instagram Insights, YouTube Studio, Facebook Page Analytics).

### **Considerations and limitations:**

Integrating second-party data requires careful planning. It may come in a different format or structure than your first-party data and will likely require cleaning and transformation.

Because the user did not directly engage with your organization, outreach based on second-party data should be handled with care. Clear messaging and value exchange are essential. Also, second-party data may need to be flagged separately for auditing, segmentation, and compliance purposes.

Publishers should always validate that proper consent was collected by the original organization and determine how (or whether) the data can be reused.

## **Third-party data**

Third-party data offers a broad but generalized view of audience demographics and behaviors collected across the web and beyond your organization's control.

### **What it is:**

Third-party data refers to anonymous, aggregated data collected by ad networks, social platforms, and data brokers. It is typically inferred from browsing behavior, demographic models, and platform interactions—not volunteered by users—and can include demographic, psychographic (interests, opinions and values), purchasing, social media, and high-level behavioural data.

Examples include:

- Demographic estimates (age, gender) in Google Analytics
- Interest segments from Meta Ads or X/Twitter
- Behavioral profiles from ad tech vendors like Nielsen or Experian

Platforms can incorporate multiple types of data. For instance, Google Analytics collects first-party data through your site or app—such as pageviews, session duration, and click behavior. If Google Signals is enabled, it can also add aggregated, anonymized demographic and interest data from signed-in users who have opted into ad personalization. The integration of third-party data on the platform enables publishers to segment audiences not only by on-site engagement, but also by inferred attributes such as age, gender and interests.

### **Why it's valuable:**

Although less precise, third-party data can help publishers understand broad trends and reach users outside their known audience—especially for marketing and audience development. Since publishers have less first-party data about individuals who interact with their products, but don't register or subscribe, third-party or aggregated, anonymized data can help them infer more information about these audiences.

Publishers use third-party data to:

- Target ads or promoted content to interest-based segments using third-party data in analytics platforms.
- Run paid social campaigns to get people to sign up for their newsletters, download their app or reach special coverage.
- Target advertising using ad networks.

### **Where you'll find it:**

This data comes from third parties, hence the name, and can be acquired via ad platforms, analytics tools, and data vendors. For example, analytics platforms like Google Analytics and [Parse.ly](#) will provide views of aggregated demographic data. Digital advertising networks use third-party data for targeting by [interests and habits, their search behavior and their online activity](#), and social media platforms like Facebook provide breakdowns by geography and interests for paid social campaigns.

### **Considerations and limitations:**

Third-party data should be treated as directional, not definitive. Third-party data provides broad audience interest and profile data that is often modeled or inferred, and publishers usually cannot fully assess the accuracy or source methodology of that data.

Additionally, publishers have no control over the structure of third-party data. When relying on third-party data, publishers must reconcile discrepancies across platforms and clearly differentiate it from more precise first-party insights.

It also cannot be used for direct outreach or personalization. Many sources of third-party data are being phased out due to regulatory pressure and browser privacy changes (e.g., cookie deprecation). Ethical use demands transparency and compliance with evolving privacy standards, and as audiences become aware of digital privacy issues, some are less willing to share their data with third parties.

# Getting Started: Conduct a Data Census

Once a foundational understanding of audience data types is in place, one of the most effective ways to apply this knowledge is by conducting a **data census**. This exercise helps map where audience data lives within your organization, how it's collected, and how it connects—or doesn't—across teams and tools.

A data census provides the baseline needed to begin building a coherent, cross-functional data infrastructure aligned with organizational goals. It helps uncover gaps, reduce duplication, and break down silos that often prevent teams from using data effectively.

## Why a Data Census matters

Conducting a data census is the first step to becoming a data-driven organization. Nearly every department in a news organization collects or relies on some form of audience data—from digital analytics and membership systems to survey tools and marketing platforms. However, these sources are often fragmented, underutilized, or disconnected.

By cataloging all data sources, systems, and data owners across the organization, publishers can:

- Identify gaps or duplication across tools
- Evaluate the interoperability of platforms and data formats
- Understand where high-value first- and zero-party data exists and how it's maintained
- Surface opportunities to unify or share data across departments
- Assess whether the organization is collecting the right data to support strategic goals

A data census has the added benefit of fostering cross-functional relationships that can help break down silos in your organization, preparing you to make the most out of the rich stores of first-party data you have. That makes a data census a good opportunity to also revisit

organizational goals and your ability to measure them. If you do a data census without clarity on organizational goals, you're unlikely to reap the benefits connecting your systems or be able to prioritize effectively.

## **Step-by-step, how to conduct a data census:**

### **1. Identify a cross-functional team and appoint a lead**

The team should include representatives from every department that manages or uses audience data — product, editorial, marketing, development, operations and more. The team lead should be someone with experience working across all the departments, likely someone on the product team or leadership team who is empowered to coordinate collaboration.

### **2. Make sure you understand the organizational goals**

The data census team should work with senior decision-makers to deepen their understanding of organizational goals, including why those goals were chosen and how senior leaders want to measure progress toward them. This sets the census team up to ensure the organization is actually collecting the data it needs to measure progress towards these goals.

### **3. List all current data sources**

Each departmental representative should make an initial list of all the sources of data collected in their department. Include structured and unstructured data, digital and analog formats, and operational or legacy systems.

### **4. Review these sources with each department**

Review the list with each department, asking the data owners to identify any gaps, additional sources or clarify how existing tools are used. Consider using the overview of data types (first-, second- and third-party) above to jog their memories.

## **5. Categorize data by type**

For each source, classify whether the data is first-, zero-, second-, or third-party. Keep in mind that platforms often contain multiple data types. . For instance, Google Analytics has both first-party behavioral attention data as well as third-party age, gender, location and interest data.

## **6. Review data for currency and recency**

User surveys and preferences are rich sources of zero-party data, but they have a shelf life. Consider identifying a best-used-by date for the data. Are you still referencing the results of a survey from six years ago? Maybe it's time to re-run the survey. It's also worth checking how frequently a user sets their preferences. Consider establishing a cadence for reminding audiences to review their preferences.

## **7. Conduct a gap analysis**

Zoom out and ask: Are there goals you can't track progress toward with your existing data? Is there a question you have that you aren't getting from your data? Is that a gap in the data, a missed opportunity to leverage data from one of your sources better, or does the data need to be integrated differently?

I was first introduced to the idea of a data census when I worked at Ideastream Public Media. Working across departments and all our major outlets, we eventually identified more than 33 sources of data, including third-party TV and radio audience ratings, first-party membership data and a full range of first-party digital platform data from our website, streaming services and podcasts. Conducting a data census was also the first step to developing a product practice at Ideastream. It helped create cross-functional relationships and allowed for data to become a common language across disciplines and departments.