Newsrooms Must Prepare for AI by Getting Their First-Party Data Right

Written By Ariel Zirulnick

AI can't help your newsroom if your first-party data is a mess.

When conversations began last spring that led to the creation of the NPAI Co-Lab on how we could collaborate across the news industry to build open-source AI technology to serve unmet newsroom needs, I pitched my moonshot — a unified CRM.

It was a solution I'd been thinking about since 2016, when I first encountered a sales CRM while running a local media startup in Miami called The New Tropic. I wondered: What if we had a similar system to log and better serve the hundreds of community members who contributed time, ideas, and energy to our journalism? Journalists often hold valuable information about their audiences — the issues they care about, their expertise, the events they attend, the neighborhood they call home, and even their connections to each other. But it's typically scattered across myriad teams and places — inboxes, spreadsheets, conversations, comment threads, and external platforms like Facebook and Mailchimp — making it nearly impossible to act on strategically across the organization.

At the Membership Puzzle Project, I worked with dozens of newsroom leaders from around the globe who shared the understanding that the information we collect from our audience is key to aligning our audiences' needs with a diversified, sustainable revenue strategy, but

nearly all were held back by the lack of technical infrastructure to build a unified database of first-party data. At SRCCON: Product in 2020 (the conference that led to the formation of the News Product Alliance), I [led a packed session](#) with Texas Public Radio CEO Ashley Alvarado, then Director of Community Engagement at LAist, on how to tackle the technical challenge of managing engaged newsrooms' disconnected data sources.

While the Co-Lab group quickly agreed a unified CRM was too ambitious to tackle head-on, the need at the core of that idea was undeniable. For newsrooms to adapt to rapidly changing customer behaviors and evolve their failing revenue models, they must be able to collect, understand, and act on first-party data to connect with audience members *directly*. It's no longer just a technical advantage—it's a strategic necessity.

That belief became the foundation of the News Product AI Co-Lab. Product professionals, journalists, technologists, and researchers from across news organizations, universities, and support organizations have come together around a shared goal: help small and local news organizations strategically use first-party data to ship quality products that respond to changing audience behavior – and to ensure they aren't left behind as AI capabilities accelerate across the industry.

Validating our idea

Before building solutions, the Co-Lab needed to test our biggest assumption — that a lack of tools and technical knowledge was the main obstacle holding small and local newsrooms back from making better use of first-party data — and learn more about what open-source AI tools would actually be helpful.

As we talked to more newsrooms and partners — via a newsroom survey, interviews with vendors and audience consultants, and participants at Story Discovery at Scale in April — it became clear that even if we solved the technical challenges, structural and cultural challenges would continue to hold newsrooms back. Parallel research by The Rebooting has further validated the need for this cultural work across the news industry.

So we shifted the question: what would it take to get newsrooms ready?

In order for open-source AI technology to make the greatest impact, particularly for small and local organizations with limited capacity, we must address three foundational challenges:

- First, news organizations need a **shared understanding** across departments of what first-party data is and how to collect it.
- Next, they need to achieve **system integration**,connecting first-party data sources into a cohesive structure.
- Finally, **strategic leadership** is required to apply insights extracted with AI to achieve editorial, audience and revenue goals.

All three are needed to strategically utilize AI technology to derive and apply insights from first-party data to develop more resonant editorial choices, stronger products, and smarter revenue and engagement strategies.

Skipping steps leads to tangled efforts and missed potential. If you don't have a shared understanding of first-party data across departments, you'll never fully connect your systems. If you build the integrated system without a plan to apply the insights, you won't get a return on your technical investment. If you try to apply first-party data without connecting your systems, you're unlikely to get beyond demonstrating the opportunity because you're just not working with enough data.

We'll unpack these three challenges below.

There is a lack of shared language or structure for first-party data.

Before newsrooms can utilize tools built to help them use first-party data strategically, they need to understand what first-party data is. That's where many organizations are stuck. It's where the industry is stuck.

First-party data has a simple definition: it's any information that an organization collects about its audience directly. But our research so far has made clear that there isn't a shared understanding of first-party data. Almost every newsroom team – editorial, engagement, audience revenue, product, marketing – collects some form of first-party data. But they're often working toward different goals, using different tools and even the terminology used to talk about it is inconsistent.

Engagement teams might think of "audience data" as survey responses or community feedback. Meanwhile, editorial colleagues are referencing "newsletter metrics" or "reader signals," analytics teams are sharing "site traffic" or "behavioral data," membership is discussing "donor lists" and the product team is dissecting "subscriber journeys." Part of the challenge is most people tend to think about first-party data from the narrow lens of their role — and they rarely call it first-party data. Even within the Co-Lab team, we discovered early on that we were using the term "first-party data" to describe different things, depending on the newsroom context we came from. Without a shared definition, you can't identify or teach best practices.

This confusion reminds me of the early days of membership models in digital journalism. One of the first things that my Membership Puzzle Project colleagues Jay Rosen and Emily Goligoski did was draw a distinction between membership and subscription so that we could be sure everyone was operating with a shared understanding of what membership is – and what it isn't. It also mirrors the early days of product management in newsrooms: the work was happening, but without a clear discipline or shared understanding, it was inconsistent and hard to scale.

That's why one of the Co-Lab's first deliverables will be a shared guide to first-party data—what it is, how to use it, and how to talk about it across teams.

Data collection is fragmented.

In our research we heard about disconnected systems, redundant records, and little capacity to integrate or maintain data infrastructure.

Just like the siloed teams within news organizations, the tools used to capture valuable first-party data often don't communicate. Without shared systems or intentional integration, audience data becomes fragmented and hard to use beyond its original context and share across teams.

I saw this challenge up close at *LAist*, where I served as Director of News Experimentation. As a newsroom committed to listening across many platforms, we had no shortage of audience signals. But none of it lived in a central place.

In 2023, we experienced what was possible if we could centralize that data. [A citywide survey run by the engagement department](#) collected enough responses and was structured such that we were able to turn it into a database.

We were able to promote new editorial products and conduct targeted surveys to inform our reporting based on meaningful characteristics, such as personal identities, life stages, and zip codes. But the database couldn't be integrated with our other systems, so it was static. The data became dated and we ran the risk of exhausting people with our outreach because the same profiles were showing up each time we drew a segment.

This challenge is shared by many newsrooms, especially small and local ones, who are often solving for significant technical debt that prevents them from tackling complex data challenges like this one.

Before open-source tools can clean or centralize data, newsrooms need guidance on what data should be collected, why it matters, and how to structure it to integrate systems. Although each newsroom needs to decide this for itself, we think the News Product Alliance can help by taking a first pass at establishing standards that newsrooms can adapt.

This is why the Co-Lab is investing in open-source infrastructure to help small and local newsrooms integrate their existing tools and build toward a centralized data layer—one that can eventually support more advanced AI-driven insights.

Newsroom leaders need to own their role in this work.

Clean data and connected systems aren't enough. Most organizations are still missing something critical: strategic leadership that helps teams identify what insights are most valuable and how to apply them across editorial, product, and revenue strategies.

We heard over and over that no one owns the full data picture and leadership is too removed from data decisions to coordinate across silos or guide prioritization when there are tradeoffs.

As we've conducted research, we've found that the newsrooms using first-party data most strategically share a critical characteristic: their revenue model prioritizes reaching the right people, not the most people. These organizations succeed by deeply understanding and serving specific audiences rather than chasing maximum reach. They

have rich audience segments based on personal characteristics, not just their depth of habit with the news organization.

That kind of clarity doesn't emerge on its own. It requires leadership. As The Rebooting put it in *The Audience Data Opportunity*, enabling smarter audience strategy with AI is a key goal of publishers, but organizational readiness is not keeping pace with AI's quickly advancing technical capabilities — "culture, not tooling, is the biggest blocker."

Mark Zohar of Viafoura calls it the "last mile" challenge. Newsrooms often collect and analyze data, but few enrich it by combining sources and even fewer activate it by applying it to strategic decisions. As a result, the complete strategic value of first-party data is rarely realized.

To take full advantage of insights derived from AI tools, newsrooms need a clear vision for how audience data flows across departments, how it supports editorial and business goals, and how to turn those insights into action.

That's where the Co-Lab sees its greatest opportunity. Yes, we're building tools—but we're also working to elevate and support the leadership structures required to use those tools well as part of the News Product Alliance. We believe the NPA community as product leaders are uniquely positioned to help news organizations strategically leverage AI tools and first-party data . They understand the technology, the editorial mission, and the need for alignment across teams. They're the ones who can turn raw data into actionable intelligence by setting priorities, building shared processes, and ensuring the right insights reach the right decision-makers. AI won't transform a newsroom that lacks the strategic clarity and cultural readiness to act on what the data reveals

Where we go from here

Since launching this work, we've interviewed more than a dozen newsrooms, from legacy outlets to startups, and spoken with consultants and vendors who support them. Our research is ongoing, but one thing is already clear: tools alone won't close the gap. The barriers to strategic first-party data use are as much cultural and structural as they are technical.

That's why the AI Co-Lab is taking a holistic approach. Yes, we're building tools, but we're also mapping the landscape, identifying the barriers, and sharing what's working. Our goal is to help small and local newsrooms build the capacity they need to leverage audience data with confidence and clarity amid tectonic shifts in news discovery and consumption. That's why we've taken these findings and started another round of research to learn how newsrooms collect audience data, segment their audiences, and make decisions about how to apply audience data.

If you would like to participate in the Co-Lab's work or share your experience with applying first-party data strategically, <u>fill out this form</u>.