

# How to structure and centralize audience data for a strong, integrated data foundation

*By Joan DiMicco*

*In order to fully leverage your audience data for sustainability, you need to bring it all together in one place. Here's how.*

First-party data is information you collect directly from your audience. It's called "first-party" because you own it. No platform can take it away. No algorithm change can cut off your access. No third party sits between you and your audience.

First-party data is strategically valuable because it connects users' preferences and behaviors to specific user profiles, providing a picture of each audience member's relationship with your journalism: who they are (name, location), how they engage (which stories they read, which newsletters they open, which events they attend), how they support you (subscription levels, donation history, renewal patterns), and most importantly, a way to directly communicate with them (email, cell, address).

In this guide:

- The Core Data Systems Every Newsroom Has
- Database Fundamentals for News Organizations
- Designing Your Data Architecture
- Data Collection Best Practices
- Building Your Team's Capacity for Data Management

Newsrooms have spent decades building audiences on platforms they don't control. Every Facebook follower, every Apple Podcast subscriber, every YouTube viewer represents a relationship mediated by someone else's algorithm and business model. When those platforms change their rules, or worse, when a platform disappears entirely, that relationship disappears with them.

In this volatile platform environment, every newsroom needs to own and manage its own relationships with audience members.

But collecting the audience data is not enough. It needs to be brought together. When identifying data is connected to engagement and transaction data, you can not only understand what levers to pull to deepen your relationship with your audience, but also send personalized communications directly asking them to take the next step.

Integration also allows you to answer questions that span multiple contexts: Which event attendees are also donors? What content topics resonate with your most generous supporters? Who engages across multiple platforms and might be ready for a deeper relationship?

These insights are impossible when data remains scattered across disconnected tools.

Unfortunately, that's often the case. The task of bringing this data together can be so complicated and resource intensive that newsrooms never take the step to bring it all together (at WBUR, where I previously served as the executive director of product, [it took us two years to complete our data integration](#)).

That means newsrooms end up missing out on an array of opportunities for improvement, including:

**Personalized audience communication at scale:** You can send different messages to different audience segments without manual intervention. Through automations with triggering rules, emails and text can be sent immediately after specific actions take place. As you build up these rules over time, more and more of your communications can be tailored to each person's interests and relationship with your newsroom.

**Better resource allocation:** When you can easily analyze patterns in your data, you make better decisions about where to invest your staff's time. Which social channels bring the most readers to your stories? Which content topics drive the most newsletter signups? Which events attract people who later sign up for subscriptions? Structured data makes these questions answerable, and tells you where to focus your time.

**Reduced operational risk:** Manual processes create vulnerability, especially when it comes to staff absences and turnover. Automated processes with proper documentation means knowledge isn't trapped in one person's head. Processes become institutional rather than individual.

**Refined use of AI:** Newsrooms are navigating how—and whether—to incorporate AI tools into their work. Whatever your organization decides, artificial intelligence depends on well-structured data for processing. Having clean, standardized data with consistent fields and adequate volume will put you in a strong position to experiment with AI tools when they are available and they make sense for your newsroom.

This guide takes you through all the necessary steps to achieve those capabilities.

We'll start by explaining what data platform types exist and how to choose the right one for your needs. Then we will dive into more technical details of how a database needs to be structured to produce value, and finally we discuss data collection best practices and how to set your team up for managing a complex, business critical data environment.

# The Core Data Systems Every Newsroom Has

Newsroom operations hinge on two fundamental databases: your Content Management System (CMS) and your Customer Relationship Management System (CRM). Understanding the distinction between these systems is essential because they serve completely different purposes but are often confused or conflated.

Your **Content Management System** is organized around content and content production. The core unit is a piece of content: a story, headline, image, video, or newsletter. This system manages the creation, editing, and distribution of news content. Your reporters, editors, and producers likely spend their days in your CMS. A CMS can help you track what you've published, at what volume, and how it is performing.

Think of it as the newsroom's production line, where stories are made, edited, and pushed live. News organizations typically choose between buying a hosted CMS solution that offers built-in support and scalability like Arc XP, Newspack, or Piano, or building a custom CMS using open-source frameworks like WordPress, Drupal, Django, or Ghost, which provide more flexibility but require greater technical capacity. WordPress is a common CMS in newsrooms because it offers a customizable foundation supported by a wide ecosystem of tools and plugins.

Your **Customer Relationship Management System** is organized around people. The core unit is an individual, whether that's a reader, subscriber, event attendee, donor, or corporate sponsor. This system tracks relationships with your audience as people and groups. A CRM can help you track who is in your audience, what interactions you've had with them, and what their status is in terms of engagement, subscriptions, donations, etc. Common CRM platforms for newsrooms include Salesforce and HubSpot, though many smaller organizations use flexible tools like Airtable or email marketing platforms with CRM features. If you are looking for a basic solution for managing audience data, but not the complexity of a CRM, consider advanced spreadsheets and project management platforms that allow teams to coordinate and organize data in different worksheets.

Historically, newsrooms used CMS audience data to monitor engagement with content on their website. Then they began accessing audience data through third party platforms (e.g. social media and web analytics platforms) and downloading it into spreadsheets for one-off explorations.

This approach made sense when the business model was primarily advertising-driven and anonymous pageviews mattered more than individual relationships. That is no longer the case.

Financial sustainability now requires understanding your audience as people, not just traffic numbers. You need to know not just that 50,000 people visited your site, but which of those people might become subscribers, which ones care most about education coverage, and which ones attend your events. That requires a robust CRM that can store your audience data and provide dynamic reporting.

## The Supporting Players

Beyond these two core systems, newsrooms often have multiple supporting platforms handling specific functions:

**Email Marketing Platforms** sit at the intersection of content and people. They deliver your journalism (content) to specific individuals (people) based on their sign up preferences. These platforms might be standalone tools like Mailchimp or Constant Contact, or they might be integrated into your CRM as marketing automation features. They handle newsletter delivery, campaign management, and email engagement tracking, and many of them can easily integrate with a CRM.

**Donation and Payment Processors** like Stripe, PayPal, or specialized donation platforms handle financial transactions. These systems need to connect to your CRM so that when someone makes a donation or subscription payment, that transaction information updates their contact record automatically.

**Event Management Platforms** like Eventbrite or Cvent handle registration and ticketing for community forums, fundraisers, or other newsroom events. Event attendance is valuable first-party data because it signals strong engagement, so these systems should also connect to your CRM.

**Web and Mobile Analytics Tools** like Google Analytics measure user traffic on your website and digital platforms. They answer questions about content performance, traffic sources, and user behavior patterns for the overall audience. This data is rich with insights, but it is typically only tracking anonymous visitors and therefore cannot provide first-party data about specific individuals.

It's crucial to understand this limitation: while these analytics platforms measure audiences, they do not capture identifying data about specific people. Because of this, web and mobile analytics tools cannot directly flow data into your CRM to augment your information about individuals. For most newsrooms, web traffic will be anonymous and web analytics platforms will only be useful for trend analysis, not for sending data to your CRM.

The only time website activity connects to a known person is when someone takes an action on your website that identifies them: logging into an account, subscribing to a newsletter, registering for an event, or making a donation, typically through one of the other supporting platforms listed above. Until someone enters their identifying information, they remain anonymous.

This is why the email registration, payments, and events are so valuable: they are moments when anonymous visitors become known contacts you can store in your CRM and begin to communicate with directly.

**Product Analytics & First-Party Tracking Tools** like Mixpanel, Heap, or PostHog help solve that challenge by allowing you to track the behavior of logged-in or registered users across your

site or app. These tools bridge the gap between anonymous web analytics and your CRM by tying behavioral data—like article views, feature usage, or conversion actions—to identifiable users. This enables more personalized engagement strategies and deeper retention analysis. These platforms often work in tandem with your CRM or marketing tools to enrich user profiles with meaningful interaction data.

## The "Thousand Flowers" Problem

Here's a common pattern in newsrooms: the membership team finds a great tool for processing donations. The events team uses a platform perfect for ticket sales. The audience team starts using a survey tool to gather audience feedback. The marketing team adopts an email platform with features they need. Each decision solves a real problem for that specific team.

Fast forward a few years, and the organization has a dozen different tools, each containing one or more sources of data the organization needs. But no single system has the full picture, and worse, the systems often have conflicting information. Is Sarah Jones's email address sarahjones@email.com in the donation system but sjones@email.com in the event platform? Or are these different people?

This is what happens when you "let a thousand flowers bloom." Each individual tool makes sense in isolation, but they can't be brought together to give you a holistic picture of your audience members and their unique interests. Only centralization allows that.

Centralizing data into a single platform requires choosing a small set of flowers over all 1,000. It often means asking teams to give up tools they like and have customized to their workflows. It means change management, training, and accepting that the new central system might not do everything the old tool did, at least not in exactly the same way.

Newsrooms can survive for long periods of time without integrated databases – and many do, because of the cost and time required to integrate systems.

But if your databases aren't integrated, answering questions about your audience will require that you export data from multiple sources and manually match audience members based on similar information.

Before WBUR had a centralized CRM, teams were adept at downloading newsletter lists and cross referencing them with donor database exports to ensure the right appeals went to the right type of audience member. This was not just time-consuming, it also introduced opportunities for duplicates and other human errors.

Here's an example: Every February, WBUR runs a fundraising campaign in which donors can send Valentine's Day flowers through the station. A portion of the purchase is donated to WBUR.

Before we integrated our data, executing this campaign was an all-hands-on-deck, late-into-the-night affair. Tracking orders, addressing customer support issues, and keeping tabs

on how the drive was performing required staff to manually cross-reference multiple spreadsheets and databases, often late into the night.

There was a particularly painful part of the process that required everyone to exit the order database for several minutes while a member of the membership team executed a 45-minute process of exporting orders and sending them to the flower vendor.

After we centralized our data, the team could export orders for the vendor in 10 minutes and without asking colleagues to stop work. This meant orders could be sent in small, more frequent batches to the vendor, which meant orders could be processed faster. The platform also allowed for real-time tracking of revenue, alongside performance of digital marketing efforts, so the team could all see how to adjust promotion to increase revenue.

A drop from 45 minutes to 10 minutes is very measurable, but in addition the team gained confidence in how the process was running. Anxiety decreased. Staff could focus on customer service and campaign strategy rather than manually matching data records.

This is the fundamental value of structured, first-party data: it turns repetitive manual work into automated processes and it connects data together for complete views of audience interactions.

Integration will also help you spot opportunities for collaboration and revenue maximization. Before centralizing our data, WBUR had two separate CRMs for donor and sponsorship management, meaning there was no single place where a supporter's relationship with the station could be understood. As a result, the grants and sponsorship teams could be talking to the same potential supporter without realizing it. An informal office encounter was one of the only ways to spot any overlap.

But when a regional tourism organization approached WBUR in 2024, after we integrated our systems, it was immediately obvious in the CRM that two teams were simultaneously communicating with the same agency. The teams met and designed a dual-pronged pitch: the sponsorship team pitched a custom podcast and the grants team pitched a festival sponsorship that aligned with the podcast but took on a complementary format. The pitches served the client's promotion goals in complementary ways versus contradictory or confusing ways. Both deals closed, maximizing revenue for WBUR. This cross-departmental visibility opened doors to revenue opportunities that simply weren't possible when data lived in silos.

## Database Fundamentals for News Organizations

All of your data systems follow the same core concepts. Understanding the basics of how databases work will help you make better decisions about what data to collect, how to structure it, and how to ask questions that your systems can answer.

These fundamentals also matter when you're talking with technical staff and contractors. Understanding terms like "tables," "records," and "key-value pairs" means you can participate meaningfully in conversations about your data infrastructure.

## The Building Blocks: Tables, Records, Fields, and Relationships

The core components of a database are tables, records, fields, and relationships. Think of a database as a collection of interconnected spreadsheets, referred to as **tables**.

The rows of the spreadsheets are referred to as **records**. And the details in each column of a row are the database **fields**. What makes a database different from a set of spreadsheets is that each table is connected to other tables, and each row (record) of each spreadsheet has a link (relationship) to a row in another spreadsheet.

The fourth key component is these **relationships**. These are the links across tables making the connection between different contexts. For example, a newsletter subscription record can link to a specific donation record in another table. This match allows you to see both the email subscriptions and the donations of the same person. A database with links across tables is known as a **relational database**.

## What is a Table?

A **table** stores information about one type of entity, in this case users.

Users			
user_id	name	email	signup_date
USR_01	Sarah Jones	sjones@example.com	12/3/2024
USR_02	Bob Martinez	bob.martinez@example.com	27/4/2024
USR_03	Carla Nguyen	carla.nguyen@example.com	5/6/2024
USR_04	Daniel Okafor	daniel.okafor@example.com	15/7/2024

Each row is a **RECORD** (an individual user).

Each column is a **FIELD** (a property of the user).



## Key-Value Pairs: How Your Systems Recognize the Same Person

These relationships between database records are established through a concept called **key-value pairs**. A key value is one that is unique to each row (or person) in a table. A key-value pair is when these values are linked between two tables.

## What is a Key-Pair?

A Key-Pair is a pair of fields (primary key ↔ foreign key) that creates a relationship between two tables.

*Primary keys are unique identifiers for each record in a table.*

Users	
user_id	name
USR_01	Sarah Jones
USR_02	Bob Martinez

Articles	
article_id	headline
333	Local School Board Drama Hits Boiling Point
334	City Council's New Housing Plan Is Basically Vibes

*Foreign keys reference a primary key from another table.*

Comments			
comment_id	user_id	article_id	comment_text
10001	USR_01	333	I came for the school board tea. Stayed for the facts.
10002	USR_01	334	This plan is giving... confusion.
10003	USR_02	334	FAKE NEWS.



A common, unique identifier for newsroom contacts is an email address. And emails are a good starting point for establishing unique identifiers between platforms because each email address enables you to contact a unique person. But emails as identifiers are not perfect – people change email addresses, share email addresses, or use different addresses for different purposes. They are much preferred over using First Name Last Name, because you can have many Sarah Jones in your database, each with a different email, but email addresses should still be used with caution in database design.

More typically in relational databases, email is combined with an auto-generated ID for each row. This ID is a persistent, unique primary key that will remain constant even if a person changes their email address. These IDs get matched between tables to make the key-value pairs that link information about your audience across tables.

The most important concept to understand is that in order to integrate systems, you must have a consistent way to recognize when different records, spread across different tables, represent the same person.

Consider this practical example: Sarah Jones subscribes to your newsletter, attends a community event, and comments on education stories. In a relational database:

- Sarah's contact information (mailing address, phone number) is stored in the contacts table.
- The details of which newsletters she is subscribed to are stored in a subscription table, which links to her contact record through the key-value pair.
- Her event attendance is recorded in an event attendee table, which also links to her contact record.



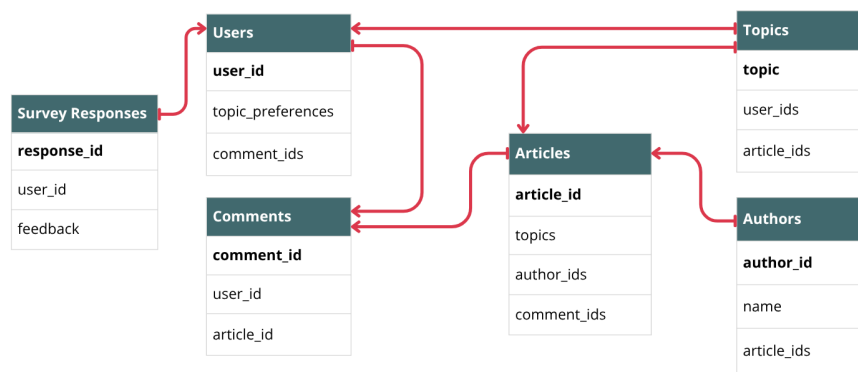
- Her comments on articles are stored and associated with both articles and her contact information, pointing to two different tables.

This structure of storing information in different tables eliminates the dangerous issue of duplicating information in multiple places. Across these four tables there is only one version of Sarah's interaction with your organization.

The way these tables connect to each other is defined by what's called a **database schema**. A schema is essentially the blueprint of your database: it defines what tables exist, what fields each table contains, and how those tables relate to one another. Think of it as the architectural plan that determines how information can flow through your system.

## What is a Relational Schema?

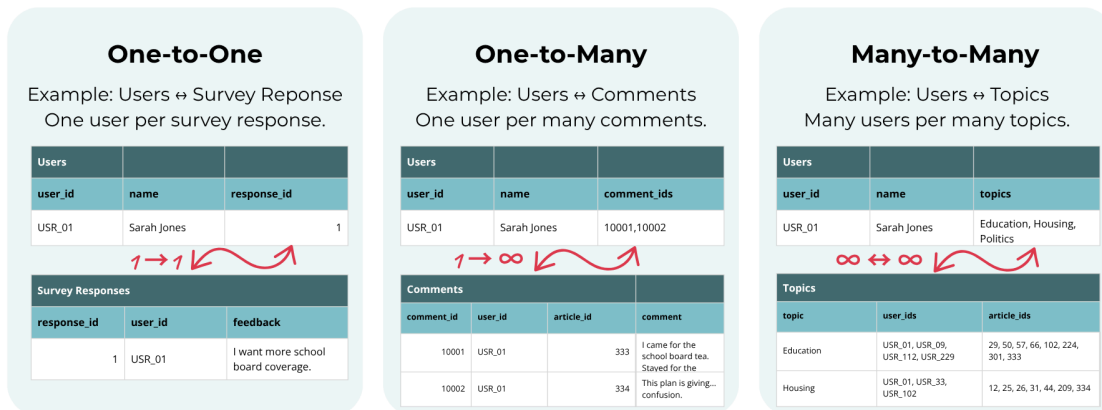
A relational schema is a map of the connections between tables that form a relational database.



Within this schema, relationships between tables follow specific patterns. A one-to-one relationship means that one record in a table corresponds to exactly one record in another table. For example, each contact might have one and only one primary email address record. A one-to-many relationship is more common and means that one record can relate to multiple records in another table. Sarah's example demonstrates this: she has one contact record, but that single record can link to many newsletter subscriptions, many event registrations, and many article comments. These relationship types determine what questions you can ask of your data and how efficiently your system can answer them.

# Types of Data Relationships

Key-Pairs work together to define relationships and structural rules of engagement between tables.



## Data Normalization: Why "Boston" and "Boston, MA" Will Break Your System

Data normalization is the process of organizing information to reduce redundancy and improve consistency so that it is clean and usable. For newsrooms, this means establishing standards for how information gets entered and stored.

Sounds straightforward, but in many newsrooms, each department has its own standards for data collection. That leads to data inconsistencies like these:

- **Contact names:** Is it "Sarah Jones," "Sarah P. Jones," or "Jones, Sarah"?
- **Geographic data:** "Boston," "Boston, MA," "Boston, Massachusetts," or "02101"?
- **Content categories:** "Politics," "Government," "City Hall," or "Local Politics"?

Each of these inconsistencies seems minor in isolation. But when you're trying to segment your audience by location so that you can send a personalized appeal, having the same city entered four different ways means your reports will be inaccurate. You need one consistent way to find all the right people.

Establishing data entry standards before you start collecting information saves enormous time and effort later. Your organization needs to decide on a style guide for common data fields and train everyone who enters information or creates forms to ensure everyone follows consistent formats.

Data normalization isn't as straightforward as it sounds. Some transformations can be handled when a person first shares their data by enforcing a structure in your forms. For example, you can use dropdown menus instead of free text fields for location, forcing people to select from standardized options.

Some normalization can be programmed into formula fields in tools like Notion or Airtable that automatically format data as it's entered. For instance, you can create rules that automatically capitalize names or format phone numbers consistently.

Teams can also create automations using helpers like Zapier or Make, or code snippets, to detect discrepancies and transform data as it flows between systems. These "middleware" solutions clean data in transit.

But sometimes, you have to manually clean data using tools like OpenRefine, especially for historical data that was collected before you established standards. This is tedious work, but it's essential for making your database useful.

Without it, you'll run into issues like duplicate donor records that make it almost impossible to know how long someone has been giving and how much that adds up to.

My WBUR team learned this lesson the hard way when we migrated data records into our new CRM in 2024. We had duplicate and inaccurate records throughout the system, which led to a multi-month data cleanup project. We could have avoided that with proper preparation. Duplicate records also consumed valuable storage space.

### Why these concepts matter for your newsroom

Understanding these database fundamentals helps you in several practical ways:

- When you're evaluating CRM options or working with developers on integrations, you'll be able to ask informed questions about how data is structured, what fields are available, and how systems will connect.
- You'll establish consistent data entry standards from the start, saving you from the harder task of cleaning up messy data later.
- You'll be able to develop more realistic timelines for data management projects.
- You'll be able to create more sophisticated audience segments, generate more useful reports, and ask more strategic questions of your data because you understand how the information connects.

You don't need to master these concepts overnight. But as you build your data infrastructure, these fundamentals will help you make better decisions and communicate more effectively with everyone involved in the process.

# Designing Your Data Architecture

Once you've chosen your tools and understand the structure of your data, you can tackle the challenge of connecting your data systems so information flows automatically between them. This integration work determines whether your CRM becomes a powerful tool for understanding your audience or just an expensive contact list that requires manual syncing.

## Audit of your current data sources

Most newsrooms already collect more data than they realize. Before you can integrate systems, you need to understand what you have.

Start by documenting every place your organization collects audience information, with a focus on audience data that includes uniquely identifiable information about each person.

This inventory will reveal both opportunities and gaps in your current data collection. If you haven't done this before, head to the News Product Alliance's [introduction to first-party data](#), where author Kevin Anderson offers advice on how to conduct a data census in your organization.

## Choose how you will integrate your data sources

Integration – connecting different systems so information flows between them automatically – can be complicated, but most CRMs already integrate with at least some common audience data tools.

Understanding your integration options helps you choose the right approach for each connection. There are three primary ways in which platforms can be integrated:

**Direct integrations** are pre-built connections between platforms. These are the easiest options because the platforms already know how to talk to each other. You typically just need to authorize the connection and configure what data should sync. Many CRM and email marketing platforms offer dozens of direct integrations with common tools. When direct integrations are available, you should choose this option.

**Automation tools** function as a bridge when direct integrations don't exist. These tools are referred to as automation platforms because they automate data transfer between systems. These platforms let you create "recipes" or "workflows" that say "when X happens in System A, do Y in System B." For example: "When someone registers for an event in Eventbrite, create or update their contact record in our CRM." Zapier, Make and IFTTT are popular options.

Automation tools are powerful and relatively affordable, but they add a layer of complexity. Each connection requires setup and ongoing monitoring to ensure it continues working correctly. They also typically charge based on the number of "tasks" or data transfers per month, so costs will increase as your audience grows.

**Custom code integration** is the third option. If you have in-house engineering resources, this might be the right approach for more complex integrations, such as connecting a modern CRM with a legacy financial system you cannot upgrade. To build a custom integration, engineers utilize a platform's available APIs (application programming interfaces). An API is a documented way to send, receive, and update data in a database, which enables an engineer to write code that directly transfers data between systems as needed.

The benefit of a custom solution is that it works exactly as you want it to. You have complete control over what data transfers, when it transfers, and how it's formatted.

The downside is that custom code needs to be monitored and maintained by your organization. When a vendor makes an update to one of your platforms, integrations can break. If your custom integrations are not well documented, it is difficult to repair them when they break. This level of risk may be too high to absorb by your organization.

### Prioritize your integration work

Integrations can be complex and time consuming. If you have limited resources to devote to this, start with your highest-volume data flows. If you send a weekly newsletter to 500 people, automating that connection provides immediate value. Event attendance that happens monthly might start as a manual process that you automate later once the more critical integrations are stable.

Focus on connections that will:

- Save the most staff time (look for high-frequency manual tasks)
- Reduce the most errors (look for places where manual data entry often goes wrong)
- Unlock the most valuable insights (look for ways that connecting data can reveal audience patterns)
- Support your most critical revenue activities (for example, donation processing and subscription management)

### Integration Best Practices

**Map data flows.** Understand how information moves between systems. Create a diagram showing which systems connect to which, what triggers data transfers, and what information flows in each direction. This documentation becomes essential when troubleshooting issues or onboarding new team members.

**Test thoroughly.** Ensure integrations work correctly before relying on them. Test with real data in small batches first. Create a checklist of scenarios to verify: Does the integration handle new contacts? Existing contacts? Updated information? What happens when someone unsubscribes or their email bounces?

**Monitor continuously.** Test integrations periodically to make sure they are still working as intended. APIs change, platforms update, and connections can break. Schedule regular checks of your most critical integrations, especially after any platform updates.

**Assume you will still have some manual processes.** Some integration might require periodic manual work, especially in the beginning. That's okay. It's better to have a documented manual process that works consistently than to rush into a complex automation that breaks frequently.

**Document everything.** Write down how each integration works, what data it transfers, and how to troubleshoot common issues. Future you (and your colleagues) will be grateful.

## Data Collection Best Practices

You've chosen your CRM, mapped your integrations, and you understand how databases work.

Now you need to collect data in ways that serve your newsroom's goals. That doesn't necessarily mean collecting *more* data. It means collecting the right data, in consistent formats, with clear purpose, and with respect for your audience's trust.

Poor data collection practices undermine even the most sophisticated systems, but thoughtful approaches can make simple tools remarkably powerful and support your work of serving and engaging audiences.

Start with your goals

More data isn't always better data. Focus on collecting data that serves a purpose you can easily identify and that you have capacity to use effectively. At the very least you should collect the data that lets you contact people effectively:

- Email address (a way to directly communicate with someone)
- First name (so you can personalize your communication)
- Subscription/engagement preferences
- Location (whether it is city, zip or full address, so you can provide relevant information)

After that, identify your priorities.

- If your organizational priority is audience growth, your data priority should be developing strong data input and data hygiene processes at the point of ingesting new contacts
- If your organizational priority is developing a loyal audience, you're likely focused on audience engagement tactics such as events, surveys, and comments and your data priority should be integrations between these input sources with your CRM's known contacts so that you ensure you have only one record for each person
- If your organizational priority is growing audience revenue, your data priority should be real-time syncing between your financial systems and your CRM, so you can track each

person's status and appropriately communicate with them based on their financial relationship with your organization.

### Build contact profiles gradually over time

When collecting information from your audience, it can be tempting to ask for as much data as possible immediately. But you risk frustrating your new contacts if you bombard them with lengthy or frequent forms. Research consistently shows that form abandonment increases with every additional required field on a form.

Instead, collect information progressively. Start with the essential contact data mentioned above and gather additional details through ongoing interactions.

For example, after someone subscribes to their first newsletter, you can follow up with a short series of emails that both introduce your newsroom and ask 1-2 questions per message – "What topics interest you most?" in email one, "How did you hear about us?" in email two. This feels like conversation rather than interrogation and allows you to start an interest profile.

Here are a few other examples of ways you can progressively learn about your audience:

- **Preference centers:** Give people a self-service way to update their interests and communication preferences. Link to this from every email you send. Some people will proactively tell you what they care about if you make it easy.
- **Content gates:** Requiring registration to access special reports, in-depth investigations, or premium newsletters creates a natural moment to collect information. People expect to provide details in exchange for substantial value.
- **Periodic brief surveys:** Once or twice a year, send a short survey (three to five questions maximum) to your most engaged audience members. Frame it as helping you serve them better.

The key is spacing these requests out and making each one feel purposeful, rather than extractive. Every time you ask for information, make sure the request is aligned to your audience goals and not just for the sake of data collection.

### Maintain data quality as an ongoing practice

Poor data quality—duplicate records, inconsistent formatting, outdated information—undermines even the best-designed systems. At WBUR, one of the biggest lessons learned is that data cleaning isn't a one-time migration task. It's ongoing work that requires dedicated attention.

Data governance standards and documentation of these standards are essential. How are you labeling things? What fields do you input dates and amounts into? You can't analyze data if you're not tracking things consistently over time.

Follow these strategies for ensuring your data remains clean and usable:

- **Create and enforce a style guide.** Document exactly how to enter common data: name format, location format, phone numbers, source attribution (a set list of options, not free text). Make this guide accessible to everyone who enters data.
- **Train everyone who touches data.** This includes not just your membership team, but anyone designing forms, entering information manually, or importing data from other systems. One person entering data inconsistently can create hundreds of records that need cleanup.
- **Use system validation tools.** Most CRMs offer ways to enforce data formats automatically. Required fields prevent blank records. Format validation ensures email addresses include @ symbols and phone numbers contain the right number of digits. Dropdown menus eliminate free-text variations.
- **Establish clear data ownership and protocol for fixing errors.** When data records have errors or are duplicative, someone needs to assess the source of the problem. Who is that person? And is it human error, process error, or automation error? Think of it like fixing a leak: first stop the water from leaking and then clean up the water. This requires someone who understands both the technology and the business processes well enough to diagnose root causes and identify how teams can both clean up the water and prevent further leaks.
- **Schedule regular data hygiene reviews.** WBUR's teams have "house cleaning" on the third Friday of every month, when teams review prospect lists and update contact information, opportunity status, and pending amounts. Making it a recurring calendar event to ensure it actually happens.

## Don't Collect Data You Won't Use

Finally, resist the temptation to collect data "just in case." Every field you add to a form or profile increases complexity. Every piece of data you collect requires storage, maintenance, and governance. And under privacy regulations like GDPR and CCPA, you may need to justify why you're collecting certain information and provide ways for people to access or delete it.

Before adding a new data field, ask:

- What specific decision will this data inform?
- Who will use this information and how often?
- Do we have the capacity to keep this data current?
- Can we collect this through behavior instead of asking directly?

If you can't answer these questions clearly, you probably don't need that field. Start lean. You can always collect additional data later, when you have a proven use case for it.



# Building Your Team's Capacity for Data Management

You don't need a massive budget or a large team to improve how you manage audience data. But you do need specific skills, clear processes, and realistic expectations about what's possible at your organization's current scale.

This section addresses two critical questions: what capabilities does your team need to manage data effectively, and what can you accomplish right now, even if you're not ready to invest in new platforms?

## The essential skills

There were three data skills that were essential to our successful CRM migration and ongoing digital transformation. It doesn't matter whether you find these skills in one person or three different people – you just need to have them.

- **CRM Administrator:** This is someone who knows the nuts-and-bolts of your audience platforms. For WBUR, this meant both our CRM and email platform, but for smaller organizations it might just mean deeply understanding all your audience systems. This person needs to maintain the system, troubleshoot issues when they arise, create reports, and train others.
- **Data Analyst:** This is someone who can extract strategic insights from your data, not just generate reports. This role requires an understanding of the goals and metrics of success well enough to draw the right conclusions from the data. At WBUR, our data analyst doesn't just pull numbers—they interpret patterns, identify opportunities, and help teams understand what the data means for strategy. [Read more about how Chicago Public Media developed this capability across the organization in our case study.](#)
- **Business Analyst/Process Champion:** This is someone who can work with teams to design better workflows. This person acts as a bridge between newsroom needs and the capabilities of your tools. They ask clarifying questions, document requirements, and help everyone understand the implications of different choices.

## Write your own documentation

WBUR quickly realized that the generic training offered by its CRM vendor wasn't enough – it didn't cover how WBUR was using the platform to accomplish its goals.

The membership team wrote a 30-page manual documenting not just how to use the CRM more broadly, but how WBUR uses it. The manual included details such as the purpose of each field on a donor's profile, how to fill it in, and how to utilize it for strategic purposes.

This documentation was so valuable that it spread beyond the membership department to people across the organization who needed to establish their own rules for using the CRM. Your version of this might only be a couple pages, but it's still valuable to have something like this.

## When and how to use consultants

WBUR brought in consultants to oversee their data migration and system integrations. These experienced professionals were invaluable. But at the end of their project, the conclusion was consultants are great, but consultants leave – and take with them valuable expertise and knowledge. For example, the business processes of the old systems were replicated by consultants in the new systems, which was intentional, but when we wanted to adjust and improve processes, we didn't have the expertise to make the changes.

We realized in hindsight that we needed internal team members to make more of the implementation decisions because the decisions needed to factor in not just the technological capabilities but our specific audience goals.

This doesn't mean avoiding consultants entirely. They provide specialized expertise that it's unrealistic for most newsrooms to have. But you should focus on building internal capacity as well, so that you don't become dependent on external resources.

Do this by insisting on thorough documentation and knowledge transfer. Make sure internal staff members shadow the consultant and understand not just what was built, but why decisions were made. Budget time for your team to learn, not just for the consultant to execute.

## What to do when your budget is limited

Not every newsroom is able to invest in digital integration and technology infrastructure projects right now. Budget constraints, limited staff capacity, or uncertainty about needs might mean the better path is optimizing what you already have.

Here's what you can do to improve your systems without any new technology investment.

**Maximize your current tools.** Make sure you're actually utilizing your existing systems to their fullest extent. Are there features or capabilities you're not using? Are there updates or extensions available? Many organizations use only a fraction of their current platform's capabilities.

**Excel and Google Sheets are powerful.** Don't overlook how useful spreadsheets can be when managed well. With proper structure, consistent formatting, and documented processes, spreadsheets can serve small organizations remarkably well. The key is treating them like a database: one spreadsheet for contacts, one for donations, one for events, with clear ways to connect them (like using email as a consistent identifier across sheets).

**Focus on data quality over new systems.** Clean your data. Spend the time on this because garbage in equals garbage out. Clean data in simple systems is more valuable than messy data in sophisticated platforms.

**Start with one improvement.** Don't try to fix everything at once. Pick one pain point—maybe it's your newsletter segmentation process, or your event registration workflow, or your donor thank-you system. Document the current process, identify where it breaks down, and fix that one thing. Then move to the next problem.

**Small improvements compound over time.** The organization that spends a year optimizing its spreadsheets and processes might be in a much better position to evaluate new CRM options than the organization that jumps immediately to new technology without understanding what they actually need.

**Invest in people.** If you have limited resources to invest in your data infrastructure, consider putting that money toward people before technology. Hiring a part-time data analyst who can make strategic sense of your existing data might provide more value than a new CRM you don't have capacity to implement well.

**Or invest in training for existing staff.** Many CRM platforms offer certification programs. Online courses can build essential skills in data analysis, database management, and marketing automation. A staff member who develops deep expertise in your current systems becomes more valuable than a sophisticated system nobody knows how to use. As your organization grows and your needs become more complex, you'll be better positioned to evaluate what technology you actually need because you'll have people who understand both the business requirements and the technical possibilities.

## The Path Forward

Building a first-party data strategy isn't about achieving perfection. It's about incrementally building your knowledge of your audience in ways you can leverage more and more over time. Start where you are. Use what you have. Build the skills and processes that will serve you regardless of which specific platforms you're using.

Whether you're managing audience data in spreadsheets or implementing an enterprise CRM, the fundamentals remain the same: clean data, documented processes, trained staff, and a clear purpose for why you're collecting information in the first place. Creating a robust audience data foundation requires patience, persistence, and a clear vision of how data serves your newsroom's mission. The goal isn't to become a technology company — it's to use technology to strengthen the relationships that sustain local journalism.

Remember that behind every data point is a real person who chose to engage with your journalism. Honor that choice by using their information responsibly, transparently, and in service of the community you serve together.

- 
- [Digital Transformation Must Be Led by Business Transformation: Lessons from WBUR - Greater Public](#)

- [The True ROI Timeline of Digital Transformation: When Technology Investment Finally Pays Off - Greater Public](#)
- [How WBUR Restructured Teams and Roles to Support Digital Transformation - Greater Public](#)
- [Setting the Stage for Organizational Transformation: The Importance of Strategic Positioning - Greater Public](#)
- [The Payoff from Centralizing Your Data - Greater Public](#)