Audience Database Development

Increasingly, media organizations are coming to realize that they control a typically underdeveloped asset, an asset that in fact may be essential to the ongoing success of the organization. That asset is the organization's audience data. Maximizing this asset typically involves development of an audience database.

A robust audience database contains deep knowledge and insights about the readership of your media properties. It is the keystone of your value proposition to advertisers. It's also a powerful tool for marketing paid products as well.

by **Nancy Ciliberti** | Senior Consultant and Analyst

In this White Paper:

▸ Discussion of the value and use of audience databases

▸ Explanation of the ICG "sources and uses" approach to audience database development

▸ Description of key process elements and considerations in building an audience database

Typically, media organizations know relatively little about their audiences. Indeed, until quite recently, they viewed audience data as a cost center – an area of the business where the goal was to do as little as possible in order to spend as little as possible. With this view, it is not surprising that much audience data is poorly keyed, unevenly maintained, and held in many information silos across the organization. The advent of popular software-as-a-service (SAAS) applications (e.g. conference registration software), while sometimes working to create better quality data, generally also create additional external data silos as well.

A good audience database will, at a minimum, provide what is often referred to as a "360 degree view" of the reader/subscriber which will enable more precise targeting and better return on efforts. This means that every type of relationship your organization has with a specific individual is reflected in a single database (it should be acknowledged there are many possible approaches to uniting your data from a systems perspective, but this is a technical conversation beyond the scope of this paper. For simplicity, this paper will assume all audience data will be physically maintained in a single database). The immediate benefit that results is a powerful means to identify and target your best subscribers, as well as your most engaged readers.

Knowing that a specific individual reads your online publication, participates in your webinars but has never attended one of your conferences has clear value for cross-selling and up-selling purposes. The next level of value is established by working to learn not only how these individuals engage with you, but more about who they are. For B2B media companies, this is the process of developing detail about the individual and the business with which that individual is associated. You will want to know, for example, business size, industry, number of employees, number of locations, etc. For the individual, you will want to know such things as job title, job function, whether the individual has purchasing authority, etc. You can get this information in some cases by simply asking for it, and in other cases from overlay data sourced from third-party data providers. Sometimes, you can even infer information. Suppose your audience consists of accountants, and you offer an e-newsletter on the topic of forensic accountants. It's reasonable to infer that subscribers to this e-newsletter are involved in forensic accounting.

Knowing your audience and the ways they engage with you is clearly powerful, but you can add an almost breathtaking additional layer of insight by setting yourself up to track what is sometimes called behavioral data. By tracking your audience and how they engage with your information products at a very granular level, you can actually infer not only what is of interest to them, but when they are in the market to buy something. There are several possible approaches to gathering behavioral data, and new software is emerging all the time, but for most publishers, it involves mining their server log files. For instance, let's say you own an online commercial real estate publication. If one of your readers was to read three articles in short succession about New York City real estate price trends, occupancy rates and recent transactions, it's a reasonable bet this reader is thinking about an imminent commercial real estate transaction. It's easy to see how this information creates an extremely valuable sales lead. And this is just a straightforward example for the purpose of illustrating the concept. With behavioral data, you are only limited by your creativity. Best of all, most of this valuable insight is already being collected by your web servers – for free. You need only set yourself up to tap into it. Privacy issues? Yes, they must be considered, but there are proven paths to minimize concerns in this area.

A Quick Review of Benefits

As you probably are already seeing, building an audience database positions your media company to compete more successfully, by offering the data, insights and sales leads that advertisers want. As advertisers increasingly move to performance-based advertising, a strong audience database allows you to generate and deliver quality sales leads with enough information for advertisers to make smart

choices about how to follow-up on these leads (in fact, you can even help advertisers here by assigning a score to each lead). And as we've just discussed, you can generate leads of people actively in the market and ready to buy.

It's also worth noting that this deep audience data also provides a powerful aggregate profile of your audience that can be useful in selling advertising by providing a compelling portrait of your audience. With so many media companies focused on selling largely anonymous traffic, this can provide a significant point of differentiation.

For those who sell information content, an audience database is equally valuable. It serves as a foundation to efficiently target. monetize and better serve an existing audience. A full understanding of what your customers do and don't buy from you is the essential basis of cross-sell and up-sell campaigns. Understanding who buys from you is fundamental to intelligent and cost-effective new subscriber prospecting efforts. And watching, through behavioral data, how frequently and deeply your subscribers engage with the content they have purchased from you is the basis for sophisticated subscriber retention strategies and even price optimization exercises.

From a product development and management perspective, deep audience data provides a sense of what you can do to improve your products as well as significant insight into which new products to develop.

And last but not least, there is the emerging issue of email governance. Simply put, almost every media company is over-promoting to its email lists, damaging its reputation, eroding response and putting a major business asset at risk. The only way to maintain and grow revenue from email promotions while minimizing list exhaustion is to carefully target these promotions. Fewer, but more targeted and compelling email offers is a business imperative for almost every media company, and an audience database is what provides the needed targeting capabilities.

So How Do I Get Started?

Building an audience database from scratch can seem like an overwhelming exercise. But step back a bit. Chances are you already have much if not all of the data, so the big work is assembling it in one place in a useful format. Similarly, you should have a reasonable sense of how you want to harness your audience database, so the job is to solicit input broadly and make sure you are building an audience database to address these needs.

That's why ICG advocates what it calls a "sources and uses" approach. Break the project into two parts:

▶ Sources – inventory your organization to determine what audience data you already have, where it comes from, how it is maintained, who has control of it, and in what format it is stored

▶ Uses – identify and interview stakeholders to determine their current audience data needs and also get them to think forward about what they'd like to do but can't do currently.

With these two inputs, you are in a strong position to start planning around the mechanics of building your audience database.

Audience Database Sources

Think of your effort to identify sources of audience data within your organization as akin to an audit. Chances are you'll be the first one in your organization to ever attempt this. You'll be doing a lot of poking around, asking lots of questions and ending up in unexpected places. You are likely to be surprised, if not shocked, by what you find.

Since organizations vary widely in scale and complexity, ICG suggests that as a general rule, you chase the biggest repositories of audience data first. For many media companies, this will be a third-party fulfillment company. These vendors can often provide a wealth of information about the many audience data collection points within your organization. Your journey begins. Chase down all these sources to determine what information is being collected, where and how.

When you've finished with this effort, consider your other audience-facing business activities. Many organizations, for example, use a third-party email service provider (ESP). Your key email lists may possibly be stored and maintained here. You also need to learn how names get added and purged from these lists as well as any other preferences your audience has provided like how often to contact them.

Move from there to your other content products. Do you have live conferences? Many media companies now use third-party conference registration companies (sometimes more than one!), which means off-site databases. Do you sell subscriptions? Expect that subscriber data to reside in its own database. Do you offer webinars? That means at least one webinar registration database exists. Do you have registration on your websites? More databases. You're starting to get the idea.

You may find your organization is maintaining dozens of audience databases. Some will be well-known across the organization (like your email database) and used heavily. It's also quite common for small groups or even individuals to be maintaining their own databases, generally out of convenience, but sometimes because they want exclusive use of them. Databases such as these typically are maintained as large Excel files. Your job, if you choose to accept it, is to bring together all these databases to build a modern, powerful, centrally accessible corporate resource.

The next step in the process is a filtering exercise. Some of the databases you identified simply won't be worth integrating. They may be too small (though small size does not automatically equate to low value). They may lack key data elements (a website registration file might contain only first name, last name, username and password – hard to do much with this). They may be hopelessly outdated (it's tough to let go of data, but introducing bad data into your new audience database won't help your cause). You see where this is going: whittle down the number of databases you identified to those that really matter to stay focused while keeping the project scope reasonable.

The next step is a necessary piece of drudge work: detailing what data is held in each database. You're after three main things:

- Detailing the fields in each database (this is useful for your work, but will also jump-start the programming effort when that time comes)

- Determining how often each field has data in it – this is called the fill rate. A database might have a field for email address, but contain few if any email addresses. You want to know this before you are unpleasantly surprised by this later. You can limit this effort to the most important and valuable fields

- Assessing data hygiene and uniformity. Just because the name of the field is "last name" doesn't mean it consistently contains last names. Most of the databases you have identified have been keyed by people who don't like data entry, aren't good at it, or have been incented to move as quickly as possible. The results typically are not pretty. You shouldn't expect perfection or anything close to it -- you are simply trying to identify real messes up-front so you don't get excited about data you really can't use.

This investigative and research activity completed, you should have a strong sense of what you've got in terms of data. Now it's time to find out what you need.

Audience Database Uses

Particularly in large organizations, you can't determine in a vacuum how your audience database will be used. This means the proper approach is to survey all the stakeholders in the audience database. At a minimum, this survey of stakeholders should include representatives of the groups that use the databases that will provide the input into the audience database. You'll also want to include a representative from your advertising and inside sales groups, your research group and your marketing services group. Your goal is to identify from the sales and marketing stakeholders what specific databases are being used, how frequently, and in what manner (e.g. are promotions and messages broad-based or targeted based on information in the databases?). You'll also want to understand how the stakehold-

ers contact the audience, for example is it by email, phone, postal mail or some other means? From the advertising sales stakeholders, you want to learn how audience data are used in the selling process and what information advertisers care about. Finally, ask the stakeholders what they like and don't like about data quality, current policies and processes. Ask them what's working and what isn't. All this input will help you get to an understanding of where to focus your energies and resources in terms of perfecting the data you've already got, and the type of system and capabilities you want to build around it.

Finally, ask the stakeholders what specific data that they don't now have that would be most useful to them. To avoid long laundry lists, you might want to ask the stakeholders to prioritize their lists, select the five most important, etc.

At this point, you how what data you currently have, where they are located and what condition they are in. You also know what data are most important to them, and what data they would most like to have. You are ready to start talking with your IT group.

Designing the Audience Database

A discussion of the technical considerations in building an audience database is beyond the scope of this paper. However, the following are useful considerations for the non-technical manager in framing and evaluating the technical conversations around development of the audience development database:

The Goal

The goal of the audience database is to unify disparate sources of internal audience and customer information, normalize it, make it readily accessible, and thus enable increased marketing efficiency, support of lead generation activities, generation of audience statistics and profiles for advertising sales purposes and to generally enable advanced audience analytics.

A Back-End System

Audience databases typically sit behind existing systems such as your email platform, your conference registration system and your subscription fulfillment system. Technically, an audience database

can be thought of as a data warehouse, but there is more involved than simply storing data. Audience databases depend on the integration and normalization of the data. In addition, real thought has to be put into how end users will access the audience database to get full value from it.

Not a Cost-Saving Exercise

In the course of this project, you've identified a number of internal databases, some small, some duplicative and some not properly being shared within the organization. It can be tempting to think that the audience database could replace some of these databases, with users keying directly into the audience database.

While this thinking is logical, it tends not to be practical, and we recommend against it. You want the audience database to be a pure back-end database, not a hybrid. Further, you will add surprising complexity and cost if you build an audience database that can support data entry. That's because you not only have to support data entry, but all the other miscellaneous reporting and other functions of that database. It's rarely worth it. Consolidating small databases yields small savings, and consolidating big databases quickly becomes a re-engineering of your entire organization. Remember that the purpose of the audience database is to generate new revenue, not shave your expenses.

Match/Clean/Standardize

The most important thing to understand about building an audience database is that the first and most important step is to match all these different input databases together. That means there has to be enough common data fields across these databases to match them. Sure, if all your databases have email addresses for every record, you're home free, and the matching process will be a breeze. It's rarely that easy in real life. If your IT people tell you matching is easy, they are either assuming there is a common field like email address in every input file, or they are exhibiting confidence not backed by experience. File matching is a hard, inherently imperfect process. Many media companies actually turn to outside specialist firms for assistance with file matching.

Once you've matched your files, you also have to deal with inconsistent data. If database "A" says the customer name is "John Smith" and database "B" says the name is "John Smyth," which one do you trust? Generally, this is handled with what are called "business rules," a series of hard decisions about which input database takes precedence in cases of conflict, and under what circumstance a conflict needs to be flagged for manual resolution. These business rules get written into the program that is bringing all the databases together.

You also need to thinking about standardizing and normalizing your data. You want to get the same data in the same fields (for example, one database might have a field for full name, and another might have two fields, one for first name and one for last name). You want your data consistent (e.g. you don't want some records saying "LA" and some records saying "Los Angeles" in the city field). And you may need to harmonize taxonomies. For example, one database might categorize a company as a "software company." Another might categorize it as an "applications provider," and a third might categorize it as a "vendor." You have got to standardize your categories to maximize the value of the database, but you have to do it in a way that doesn't compromise existing activities within the company. Don't under-estimate the work involved.

Frequent, But Not Real-Time

Since your audience database is a back-end database, you'll be continually drawing in new and updated data from the source databases you previously identified. But how frequently? There is a real temptation to go for a real-time updating. After all, sales leads are best when they are freshest. The reality is that real-time systems are expensive. Further, not all your input databases are being updated on a second-by-second basis. Finally, the most valuable sales leads will be developed inferentially, meaning you have to watch activity over time. Generally, no one single activity is by itself going to create a high value lead. Determine updating frequency based on a realistic review of your resources, the nature of the input databases and the needs of your market.

Plan for Change

Nobody can fully anticipate what data might become available in the future, what audience-based products you will be selling, or the evolving needs of your marketers. In the design of any audience database, you have to build in the flexibility to, at a minimum, add new fields and incorporate new data sources. Fortunately, today's database technology makes this fairly easy.

Third-Party Data

While this merits a detailed discussion of its own, keep in mind that there are many third-party data sources available to you. You can license data, scrape data, or even have it keyed for you. Third-party data can help you quickly and cost-efficiently deepen your insight into your audience. It can be used to backfill missing data. It can be used to validate the data you already have. You can even license taxonomy and company numbering systems to better organize and analyze your data.

It is important to note that licensed data isn't free, so you have to be judicious in its use. Scraped and keyed data may look like a less expensive route, but remember that data you collect yourself you must also maintain yourself.

About Us

Since 1995, InfoCommerce Group has been the thought leader in exploring, charting, analyzing —indeed, defining—the most important issues reshaping the business information industry.

The reason InfoCommerce Group is recognized as the leading expert organization in business information publishing: our nuanced understanding of the industry's dynamics coupled with unique access to the people who drive it. In addition to our long standing experience and commitment to the industry, our reputation is also based on our publications, Models of Excellence awards and conferences which we use to connect our customers to new and important players and identify industry-changing trends.

In turn, our customers provide a reality-based source of intelligence about what is and isn't working in the marketplace. This makes InfoCommerce Group a clearinghouse for industry information, and we draw on this privileged base of insight in all of our consulting assignments.

While we consult broadly across the entire spectrum of business information publishing, InfoCommerce Group is particularly recognized for its expertise in the areas of data publishing and health information content.

The information, data, analyses and opinions contained herein represent general views relative to various aspects of the marketplace and are provided solely for informational purposes. As specific business situations vary widely, InfoCommerce Group Inc. shall not be responsible for any decisions, damages or other losses resulting from, or related to, this information, data, analyses or opinions or their use.