

# The Role of a Customer Data Platform

**Seth Earley**  
Earley Information Science

**Editor:**  
Seth Earley, Earley  
Information Science;  
seth@earley.com

At the end of the day, everything boils down to providing value to our customers. The more we know about our customers, the better we can provide that value and stand out from our competitors. Although this objective sounds simple, many challenges arise along the way. The entire customer experience relies on data, so creating a positive and compelling experience depends on the quality and types of available data to understand the needs of the customer. The challenges to getting the right data fall into four broad categories related to systems that produce and

consume customer data: the ways that data is modeled, the data itself, gaining insights from the data, and acting on the data.

Customer data platforms (CDPs) aggregate data from many different sources to provide a 360-degree view of the customer. These platforms are designed to be managed and used directly by marketers, and they eliminate the need to access multiple systems to create customer profiles, develop marketing campaigns, test the effectiveness of marketing strategies, and predict customer behavior.

## WHAT IS A CDP?

According to the Customer Data Platform Institute (CDPI), a CDP is “a marketer-managed system that builds a unified, persistent customer database that is accessible to other systems.” The CDP acts as a centralized clearinghouse and repository for all sorts of data from various internal and external systems. Consider any place where a customer interaction is recorded, tracked, or managed. Past purchases constitute a big category of customer behavior, of course. But so do social media interactions and website visits, even when nothing is actually purchased. Collectively, this data produces signals that can be thought of as “electronic body language.”

Some data is reasonably straightforward (such as name, address, and demographic details). However, some information requires processing and interpretation. Clickstream data, for example, tracks part of the customer journey and can be very informative, but understanding what it means requires effort and human intervention. Data about website behavior can be stored in a CDP, but the dataset is large and has numerous components that are time- and context-dependent. Before the data can be meaningfully acted upon, it must be analyzed and interpreted.

## Ability to Summarize Data or Surface Trends

A CDP can be used to summarize certain types of data to present a trend or characteristic that can then be exploited. Rather than require a marketer to sift through hundreds of log lines from a

website visit, certain events or behaviors can be segmented and acted upon. These could include offers, promotions, content that similar users find valuable, or other relevant signals derived from large datasets.

## Integration of Varying Data Formats and Structures

Another valuable function of CDPs is their ability to accommodate different data types and formats that might have varying structures and naming conventions. Data might come into the CDP through a live feed via an API or web service layer, or might be input on a batch basis through a file transfer. Formats can be structured by transaction such as accounting and purchase data or unstructured content such as chat logs or Facebook conversations, tweets, and even images from Instagram. This varied data can then be put into a consistent format (through rules retained in the CDP) that can be more easily interpreted by a marketer or acted upon by other systems without repeated manual running of extraction, translation, and load scripts.

## Ability to Cleanse and Process Data

Data might have redundant records and content; for example, it might be missing details (fields or values) or contain incorrect data that has to be reconciled with another system. When new data is available, only certain values can be updated. Other data sources can be used to enrich the data or append missing information. Rules for cleansing, enriching, appending, and correcting data can reduce the cost and complexity of data hygiene by automating remediation.

## Exposing Data for Use by Other Systems

A major function of the CDP is to act as a centralized location so other systems can access and act upon customer data. The CDP becomes a broker or orchestration layer that can take the outputs from one customer-facing application, process it, convert the format, and export it or make it available via an API for a personalization engine. The personalization engine might then send data back to the CDP, which can use the results to inform or drive another process.

A major function of the CDP is to act as a centralized location so other systems can access and act upon customer data.

## Doesn't My Marketing Automation System Do This?

A CDP can provide some of the functionality of other marketing systems and customer engagement platforms, but it is fundamentally different in design and function. Marketing automation systems can integrate with other tools but usually in a limited fashion to suit a narrow set of use cases. CDP tools are designed from the ground up to talk to other systems. They also retain details from other systems that the engagement or automation tool does not. This is valuable for trend analysis, predictive analytics, and recommendations that can leverage large amounts of historical data.

The various functions of a CDP, from data ingestion and normalization to identification of relevant signals and output to downstream systems, are illustrated in Figure 1.

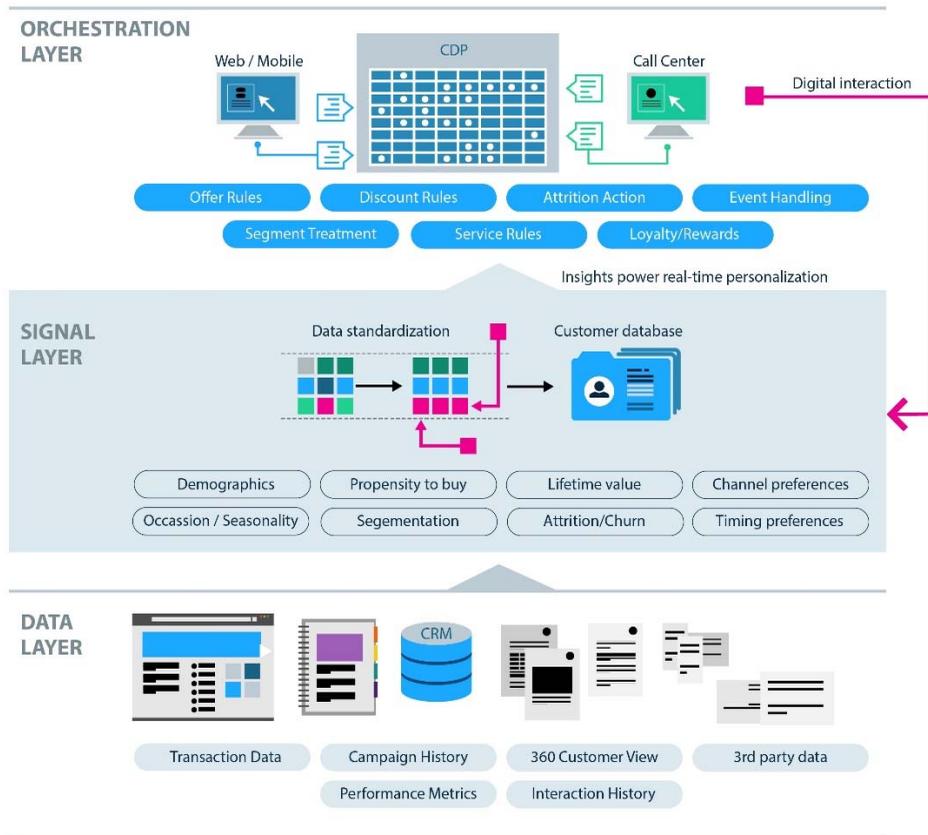


Figure 1. Functions of a customer data platform (CDP).

## CUSTOMER DATA MODEL CHALLENGES

A customer data model identifies the factors that a marketer believes to be important in understanding and predicting a customer’s behavior. Without a model, there is no way to systematically segment customers into groups and test the effectiveness of different marketing strategies. A data model captures a variety of data, from unambiguous, clearly defined attributes like name, address, and demographic details to data that can be derived and inferred through interactions and by processing data produced by other systems.

The data model contains attributes that might be created when an event occurs—for example, when a customer’s purchases exceed a certain threshold. Or an attribute can be based on who the customer is and where they live. Segmentation models can be based on a combination of explicitly defined data points, such as purchase history, user-declared values (for example, an expressed preference), externally referenced information (subscription or membership information), and attributes and values that are inferred by comparison to large numbers of customers with similar characteristics. Some techniques find hidden or latent attributes, or create relationships based on numerous subtle data signals.

Data models can be mathematical, rules-based, visual, or based on a list of the relevant factors that are believed to produce certain behavior; for example, all people under the age of 40 will prefer slim-cut jeans and those over the age of 40 will prefer a looser fit. (This is not a valid general statement, but it could form the basis of a hypothesis that could then be tested.)

A data model represents the customer and the collective insights and understanding of their real-world needs.

Creating a data model is a valuable exercise for non-technical specialists because it allows marketers to use language to describe what they know, believe, or can infer about their customer. These characteristics are converted to a structure that the system leverages (either capturing the details or defining rules and algorithms for inferring them). Data models can also inform marketers about the types of metrics they should track, and how well their strategy and specific campaigns are performing relative to a particular customer segment or characteristic.

Customer data is usually collected from a large variety of systems that come from different vendors or, if homegrown, are created by different groups. Therefore, they will have varying formats, architectures, and naming conventions. As a result, customer data models could be inconsistent, which makes it challenging to create a unified model that incorporates meaningful data in actionable formats. For example, one system might define one customer at the individual level and another at the household level. If one system totals all of the purchases for multiple members and another tracks individual purchases, the analysis of sales per customer will produce different results.

Despite variations in the data, the model must contain enough detail and the correct attributes to support advanced functionality such as effectively predicting purchase patterns or recommending an appropriate product that meets the precise needs of the customer. A customer data model is analogous to a content data model (typically called a content model). Customer profiles contain attributes that are used by other systems to improve their outputs. For example, for personalization to work correctly, the model needs to provide signals to customer engagement systems that tell those systems how to differentiate the customer’s experience—what content to present, what products or solutions to offer, and the overall experience that will move them forward in their journey. What is it about the customer that can be captured as metadata (or attributes) in the customer data model and represented in the details of their profile that will drive a unique interaction? It might be the customer’s age, or whether they were active on social media, or whether they had children. The CDP stores data about the customer that can be leveraged by various downstream systems to predict and influence the customer’s behavior.

These signals can come from many sources. Some are based on explicit attributes such as demographics, content preferences, and account information (see Figure 2). Others come from subjective or behavioral attributes (see Figure 3). These might be interest profiles, past purchase behaviors, social media patterns, loyalty scores, and real-time website behaviors.

## Explicit, Objective or Applied Metadata

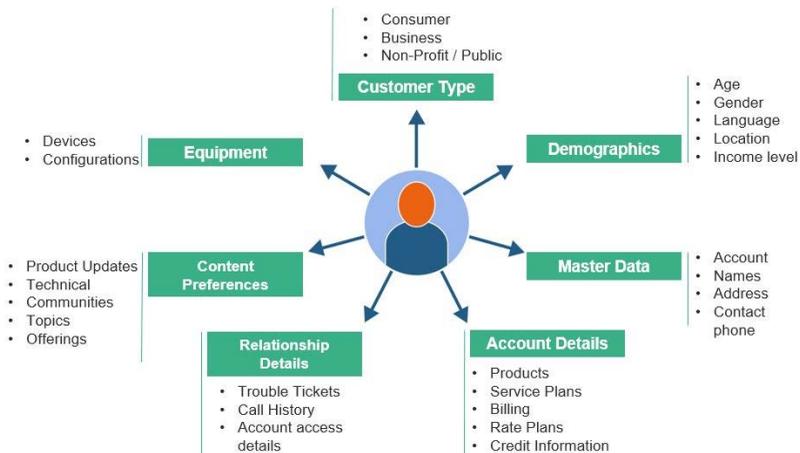


Figure 2. Customers can be described with explicit metadata from a variety of source systems.

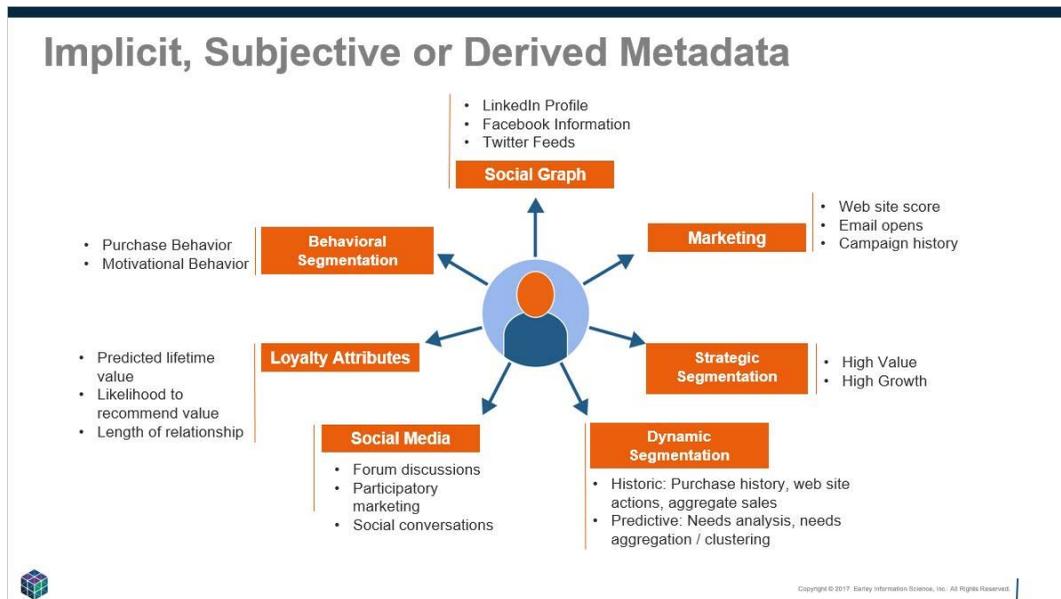


Figure 3. Implicit metadata about a customer is based on judgment and/or derived from other data sources.

## ATTRIBUTES, METADATA, AND SIGNALS

Attributes are descriptors that define all the things we know about the customer and describe their characteristics, needs, and behaviors. They determine how systems and tools provide an experience that helps customers meet their objectives and solve their problems. For systems and tools to meet user needs, these attributes have to be represented as metadata in the customer experience applications and in back-end systems that contain transaction and interaction histories.

Metadata supports the machinery of back-end operational and front-end customer-facing technology. Metadata describes the “is-ness” and “about-ness” of a customer. Are they consumers or are they businesses? The metadata field of “customer type” contains the attribute of consumer.

A signal is the data produced by an action of the customer when interacting with an application. The applications can fall into multiple categories of functionality including those that generate explicit, objective, or applied metadata (see Figure 2). This metadata and associated data is typically sourced from customer relationship management systems, enterprise resource planning, e-commerce, sales automation, order management, and external data sources like credit bureaus, data aggregators, and credit card processors (some of which can also fall into the derived metadata category). Examples of this type of metadata include:

- customer type (consumer, business, or nonprofit),
- demographic (age, gender, language, location, income level),
- master customer data (account, name, address, contact phone, email),
- account details (products, service plans, billing, rate plans, credit information),
- relationship details (trouble tickets, call history, account access details),
- content preferences (product updates, technical, communities, topics, offerings), and
- equipment (devices and configurations).

Applications that generate implicit, subjective, or derived metadata (see Figure 3) are sourced by analyzing outputs from diverse customer experiences and operational and social media applications, and by applying conventional analytics and machine-learning algorithms to create

new ways of understanding and describing the customer. Examples of this type of metadata include:

- social graph (LinkedIn, Facebook, Twitter, Instagram),
- marketing applications (website score, email opens, campaign history),
- strategic segmentation (high value, high growth),
- social media (forum discussions, participatory marketing, social conversations),
- loyalty attributes (predicted lifetime value, likelihood to recommend value, length of relationship), and
- behavioral segmentation (purchase and motivational behavior).

The overall information architecture (for customer data models and content models) needs to be aligned so that specific pieces of information can be surfaced to the user depending on the real-time signals from customers' digital body language. This step requires human understanding—which pieces of information most contribute to a solid model for behavior. While some of these features can be defined in advance, many are based on upstream and downstream system architecture and algorithms.

## MINING THE DATASTREAM

What does customer data reveal about what the customer needs? Many different data sources are available. For instance, social media data might contain information about a preference. How is that captured? Clickstream tells us something about how customers are consuming content and traversing the website: whether they click through an offer, whether they respond to a promotion, or whether they are able to complete their purchase. The data tells a story—the question is how to understand that story.

Every tool and technology in the digital engagement ecosystem produces datastreams that need to be interpreted to be acted upon. The challenge lies in identifying what data is important, understanding what it is saying, and determining what to do with it. These questions must be asked over and over again to keep the focus on the purpose of collecting and analyzing data, which might change over time.

Metrics fall into several classes, including search, behavior, utilization, content, and response. Each category of metric can have dozens of details and reports. The goal of understanding these metrics is to drive an action to optimize or improve an outcome. When users browse to a certain point and then leave the site, they were unable to complete their task. What can be changed to impact this behavior? Experiments need to be designed to find the best combination of user outcomes. These are the insights that lead an organization to change its strategies for campaigns and offers. The CDP not only allows ready access to all the metrics associated with customer behavior, but also provides the ability to execute an appropriate response based on the data, sometimes by accessing and triggering actions in external systems.

## DERIVING CUSTOMER INSIGHTS

A good sales associate knows both their customer and their products and solutions, and makes recommendations that meet the customer's needs. In the digital world, organizations need a way to capture and act on the insights that come from a digital representation of this knowledge. This means first interpreting the signals, which arrive in a different form from those a sales associate receives. Marketers and merchandisers have been dealing with the issue of signals for years, but in a different context—the signals were in the form of customer feedback, market intelligence, and sales trends captured over a period of weeks or months and across broad segments of customers rather than at the individual level.

The new pieces of the puzzle are the scale of the challenge, the velocity of commerce, and the number of variables that need to be interpreted as well as the granularity of responsiveness to customer needs. It is now possible to act on detailed understanding of customer needs based on signals from data; however, it is not possible to manually design and create the combinations of

products, services, and solutions that meet each customer’s unique needs. In the physical world, this is what a great salesperson does—they know the customer and offer solutions based on that knowledge. Digital technology is the stand-in for the best salesperson in an organization.

The process begins with a hypothesis about what offers the customer will respond to and what components of that offer can be recombined. It could be as simple as shopping basket analysis, where purchase history combinations are mined and presented to customers exhibiting similar signal patterns. Other sources of variables can be mined from human experts, product engineering documentation, maintenance manuals, and support call chat logs.

Personalizing search results can be based on customer data models that are related to product data and content models mapped through an ontology. The ontology tells the search engine how to interpret a search query based on interests, industry, or other derived profile data. For example, “mold stripping” means one thing to a manufacturing engineer and another thing to a construction manager. In the first case, the task is related to an injection mold maintenance procedure. In the second, it is related to an approach for renovating a flooded building. The ontology can disambiguate a search query by associating the context of the term with the industry or task of the user. This requires a customer data model that captures information about the user’s world at a more granular level.

Marketing organizations need to understand users and their tasks and journeys at a level of specificity that allows them to anticipate needs and determine what offer, product combination, solution, advice, or content will move the customer closer to their objective. Marketers have to ask, “What does it mean to personalize my customer’s experience? How can I differentiate between one type of customer and another, and what does that mean for how they interact with us? What content and information can be served that will be different? Why is it important?”

Digital technology is the stand-in for the best salesperson in an organization.

An analysis might reveal that for a specific segment of customers, a particular combination of products leads to increased conversions under certain circumstances. The hypothesis is that extending the product breadth with new audiences will also lead to increased revenue. To test this, the CDP needs to integrate with downstream systems that orchestrate the user experience. These systems include content management systems, e-commerce applications, product information systems, and the CDP.

Once the insights are gained, the challenge becomes one of converting knowledge into action.

## ACTING ON INSIGHTS FROM CUSTOMER DATA

The goal of understanding the customer is to take action based on that understanding. How does the organization interpret the interactions, preferences, experiences, and all the signals that stream from every customer relationship? What do you do with the data? This is the “what next” question that has to be continually asked and answered. What do we do when we know our customers’ needs? The obvious answer is that we try to meet those needs as cost-effectively as we can. The question is always, “How do we act on the information?”

Human intervention is critical at this stage. Acting on insights requires a human to interpret the data and recognize a pattern, the ability to test hypotheses about actions that will create the desired behavior, and the machinery to operationalize the confirmed (or fine-tuned) hypothesis. The end result is dynamic functionality derived from a combination of human judgment, expertise, and creativity. However, this outcome cannot be achieved unless the foundational information, data, and content architecture is in place for the digital machinery.

Because data comes from different processes and applications, there are likely varying constraints on the data—from usage and permissions to downstream enrichment. In many cases, owners of data in one application might not be aware of the downstream impact of their decisions about customer data. They also might not be impacted by quality issues that would be

critical to another use case at the downstream orchestration layer. In that case, there would be no incentive to fix the data and no business case to invest in remediation. These complexities make CDPs an enterprise priority that requires executive support, sponsorship, and funding that considers the entire customer data ecosystem.

## CONCLUSION

CDPs are increasingly essential to integrated digital marketing programs. Deploying these technologies reveals gaps and challenges throughout the entire enterprise and the digital supply chain serving the customer. If the customer cannot be fully understood from every point of view of the enterprise, it is not possible to serve them optimally. These gaps and challenges can only be remediated with board- and C-level resources and attention. If you are not serving your customers optimally, they will go to your competitor.

Successful use of customer data requires the development of a robust model, judicious selection of data, careful interpretation of analytics, and the ability to act on the results. Each of these steps poses its own challenges. By providing access to data from numerous systems in one database and supporting the systems that can produce an appropriate customer experience, the CDP overcomes the limitations imposed by fragmented point solutions and presents a holistic approach to customer interactions.

## ABOUT THE AUTHOR

**Seth Earley** is CEO of Earley Information Science ([www.earley.com](http://www.earley.com)). Contact him at [seth@earley.com](mailto:seth@earley.com).