/ Whitepaper

REQUIREMENTS FOR ENTERPRISE CDP

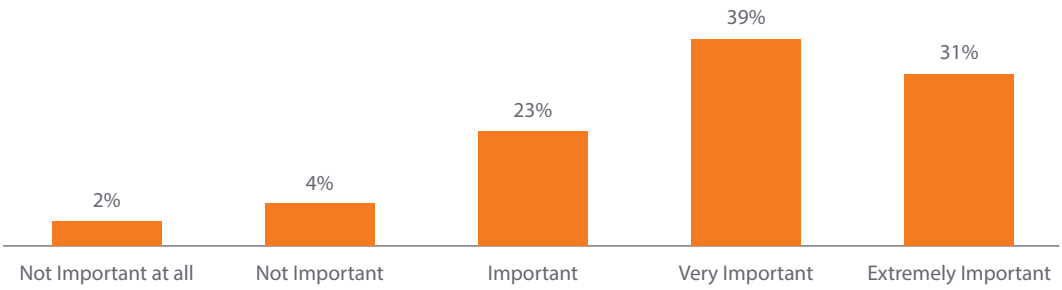
ACQUIA
EXPERIENCE DIGITAL FREEDOM

CUSTOMER DATA PLATFORM INSTITUTE

INTRODUCTION

The vast majority of enterprise marketers rate unified customer data – often called a Single Customer View – as important for success. Many of those marketers are turning to Customer Data Platforms to build that unified view.* But CDPs differ greatly how they create the unified view and in what other features they provide. So the question for many marketers isn't whether they need a CDP, but which CDP to pick. That choice ultimately comes down to finding a CDP with features that meet the marketer's own requirements.

Importance of Single Customer View



Those requirements are determined by the nature of their business and the specific uses they have in mind for their unified customer data. Business factors include the channels used, state of existing systems, and technical resources available. Uses range from customer insights to personalized messages to marketing performance measurement.

This paper will look at several CDP functions that are important for enterprise marketing. For each function, we'll define what it is, who needs it, and what to look for. You won't necessarily need every feature but it's good to consider them all before deciding whether they apply to your situation.

Integrated Features vs Partner Systems
CDP functions are sometimes delivered by partner systems rather than the CDP itself. It's generally more convenient to use built-in functions. But a partner system may better fit your needs or let you use a CDP with other strengths. Rather than rejecting CDP systems that lack a particular function, marketers should assess the CDP-plus-partner package as a unit, considering both the costs of integration and the capabilities of the combined system.

REPORT AUTHORED BY THE CDP INSTITUTE

- Data Quality & Identity Management
- Online and Offline Data
- Reporting
- Analytics & Machine Learning
- Data Actionability

* Customer Data Platform is defined by the CDP Institute as "a marketer-managed system that creates a persistent, unified customer database that is accessible to other systems". Visit www.cdpinstitute.org for more information.

DATA QUALITY & IDENTITY MANAGEMENT

What It Is

Processes to transform raw customer data from scattered sources into a usable unified view.

Who Needs It

Everyone to some degree. How much and exactly which features you need depend on the state of your source data. Enterprises getting data from many source systems, from sources with many errors, and from different communication channels will need the most help converting those inputs.

What To Look For

Data Quality Features
General data quality features ensure that data entering the system is complete, accurate, and consistent. Initial quality checks should scan new inputs to ensure the contents are readable, in the correct formats, include key elements, and contain reasonable values. For example, purchase transactions could be scanned to ensure they have a valid date, reasonable price, populated currency indicator, and valid product ID. Other items, such as email address, might be optional but can still be checked to ensure they have a valid format it they're present. The system needs rules for dealing with quality problems when it finds them: users should be able to Importance of Single Customer View decide which errors are acceptable and which will cause the data to be rejected altogether.

Cleaning & Standardization
Cleaning and standardization make corrections. Cleaning might apply rules to identify clearly incorrect values, such as impossible birthdates. Typically the original value is retained and the cleaned values are stored separately and used in additional processing. Standardization converts input values to common formats, such as a standard date or telephone number format. It may also introduce new values by converting variations to standard forms (Bob, Rob, Robert all become Robert; product names in different languages become a standard name and ID).

Postal addresses are also placed in country-specific standard formats and then may be verified or corrected by matching against a reference file of valid postal addresses. Cleaning and standardization allow data from different systems to be used without reference to the original source.

Transformation & Enhancement
Transformation and enhancement further convert data into usable forms. Transformations might convert product names into ID codes or add tags that identify an item as part of a group. More advanced transformations might add derived or calculated values such as cumulative lifetime purchases, time since last response, or consumption trend over time. Enhancement applies data from external sources, such as personal or business demographics, interests, or location history. While data cleaning and standardization are generally applied when data is loaded, transformation and enhancements may run continuously as new information becomes available from external sources. For example, a system might alert a company when any of their customers has moved to a new address or switched to a new mobile phone.

Identity Management
Identity management links records that belong to the same customer. Identity resolution methods include "fuzzy" matching of similar names and addresses (which is made a little easier by standardization); "deterministic" matching of identifiers that are definitely linked, such as an email address and phone number on the same account; and "probabilistic" matching based on estimated likelihood that two identifiers should be linked, such as different mobile devices that are frequently used at the same location. Identity resolution may draw on external services that assemble relevant data, such as old and current postal addresses for people who have moved or mobile devices that have been associated with the same person. Persistent identities allow the CDP to recognize a former customer when they open a new account or to link an old and new email address to the same individual.

DATA QUALITY CHECKLIST

Input Assessment

- [ ] Scan new inputs for issues
- [ ] Rules to handle issues

Identity Management

- [ ] Similarity match ('fuzzy')
- [ ] Associate known IDs ('deterministic')
- [ ] Associate likely IDs ('probabilistic')
- [ ] Apply external data
- [ ] Retain persistent ID

Clean & Standardize

- [ ] Change to correct values
- [ ] Convert to standard formats
- [ ] Standardize postal address
- [ ] Verify postal address

Transform & Enhance

- [ ] Add ID codes
- [ ] Apply tags
- [ ] Calculate derived values
- [ ] Append external data

ONLINE & OFFLINE DATA

What It Is

Ability to combine online and offline data in the same database, and in particular to build customer profiles that unify both data sets. Online channels include Web site interactions, ecommerce, Web advertising, email, mobile apps, and social media. Offline channels include retail purchases, in-store activities, community events, call center, product usage, and service calls.

Who Needs It

Retailers, banks, airlines, and anyone else whose business spans online and offline channels, especially if they want to build a complete customer view for marketing performance measurement and to coordinate customer treatments across those channels.

What To Look For

Ingest Data From All Sources
Online data is usually acquired through Application Program Interface (API) or Software Development Kit (SDK) connectors. Web sites, email, and mobile apps are currently the main sources, although new sources such as smart TVs and Internet of Things devices are on the rise. Online data tends to be very large, ingested from continuous streams, and loosely structured. Offline data is more likely to be structured and come through batch file transfers. Sources can include point of sale systems, loyalty programs, physical events, mailing lists, and third party data appends. Your CDP will need to support both online and offline inputs. Look for prebuilt connectors to your current source systems and explore their capabilities in detail: there are often limits on which items are exposed, volumes, access speed, and other factors that could mean a connector doesn't meet your needs. Also look at the time, skills, and cost to build custom connectors to additional systems or to extend the standard connectors if needed. Ideally, new connections can be configured within the system interface rather than built from scratch. Assess the types of data the CDP can ingest: every system accepts customer attributes and transaction records but not all support other types such as Web logs, texts, videos, graphics, audio, and streaming feeds. Streaming data in particular may need specialized processing.

For everything other than structured data, look for special features to examine the input and extract features such as keywords, locations, topics, entities, entity relationships, and sentiments. These become attributes that make the data usable for conventional databases, query tools, rules engines, and analytical systems. Check that the system can accept real time inputs if you're going to need them. Be sure to understand how quickly those real time inputs become available for use.

Scale To Your Requirements
The system must be able to ingest the amount of data you expect to feed it each and to store the total volume of data you will feed it over time. Other dimensions of scalability include number of users and data requests, response time when handling multiple real-time interactions, and extraction volumes. Be sure to understand your options to expand the system's capacity if you outgrow the initial installation: the possibilities range from automatically and instantly reconfiguring itself to meet new demands, to spending weeks or months moving all data and programs onto new hardware.

Easily Accommodate Input Changes
The system should easily handle new attributes, such as adding size or color to a product record. In some systems, each attribute must be identified in advance and any new attribute requires a redesign of the database. Other systems can automatically examine inputs to find new attributes and load them without database changes. The automated approach is a big advantage if you frequently add new data sources or new attributes in existing sources. Also understand the process to add a new source system, to change the load frequency or method, and to map input attributes to common data elements. Is it done by the CDP vendor or by the client? What can marketers do for themselves and what requires technical skills? Is the work done by filling out forms in a structured interface or by writing in a scripting language? Some CDPs give marketers or business analysts an interface to do simpler jobs, such as labeling a new attribute, while giving technical experts a different interface for complicated tasks such as connecting to a new system.

Manage Anonymous & Identified Profiles
Online data often includes some information that is tied to an entity but not an identified person. Web browser cookies for anonymous visitors are a common example; profiles linked to an IP address or mobile device are others. Some CDPs can connect anonymous entities with personal identifiers, either based on customer behavior or through external services. This reduces reliance on browser cookies, which have become increasingly unreliable as more users block them, delete them, and interact through mobile apps. Connecting anonymous profiles to personal identifiers lets the system add individual attributes such as age, location, interests, customer status, or purchase history. These can then be used to target online advertisements. The identifiers themselves can then be removed to preserve privacy when the enhanced profiles are loaded into external ad buying systems or shared with partners.

Persist the Data Over Time
Creating a persistent database (that is, one which stores data over time) is part of the CDP definition. But you still need to understand the details: will the CDP store all inputs as provided or will it create summaries? If an attribute such as address or customer status changes, will the CDP keep the old value or simply replace it? Does the system periodically snapshot derived values such as lifetime purchases? Can the system easily recreate a record as it looked at a specified point in time, including derived values and changed attributes? How hard it is to access raw details? Is there a limit on how long old data is retained?

Control Data Access
The CDP should let users control how data is combined, accessed, and retained, to ensure compliance with privacy policies, government regulations, or contractual agreements. Access policies may be based on customer location, permissions they've granted, specific data elements, how and where the data was acquired, which system or user is accessing it, or the purpose of the access. Policies should be easy to create, understand, and change. You may need to document changes over time in case questions arise over the policies in force at a particular moment. Large enterprises are especially likely to need rigorous access controls and to be able apply different controls on data from different geographic regions.

Access To External Data
Some relevant data may not be stored in the CDP but still be needed for specified purposes. This might include sensitive transactions or health information, data streams such as current location or weather, or masses of detail such as Web logs. If you might need these, look for a CDP that can support them by connecting to the external source to gather them when needed. This might be a real-time connection if the data is needed during an interaction or it might be a batch query to capture the data for historical analysis. These will usually be API connections. Check for existing to connectors if you know the systems you want to include. Assess whether the response time is fast enough to support your intended application and how the CDP informs the external system which data it needs.

ONLINE & OFFLINE DATA CHECKLIST

Ingest Data

- ☐ Online sources
- ☐ Offline sources
- ☐ Structured data
- ☐ Unstructured data
- ☐ Prebuilt connectors

Feature Extraction

- ☐ Real time inputs

Scalability

- ☐ Daily update time
- ☐ Total storage volume
- ☐ Simultaneous users and requests
- ☐ Expansion process

Access Policies

- ☐ Specify access rules
- ☐ Comply with permissions
- ☐ Document rule changes

Anonymous & Identified Profiles

- ☐ Connect anonymous and identified
- ☐ Re-anonymize enhanced data

Accommodate Changes

- ☐ Effort to add new sources
- ☐ Effort to add new objects/tables
- ☐ Effort to add new attributes/fields
- ☐ Vendor vs. user tasks

Persistant Data

- ☐ Retain raw detail
- ☐ Retain changed attributes
- ☐ Recreate old records
- ☐ Retention periods

External Data

- ☐ Real-time connections
- ☐ Existing connectors

REPORTING

What It Is

Presenting CDP data in ad hoc analysis, standard reports, and custom reports. Specific goals include profiling customers, finding segments, showing campaign results, and tracking operations such as data loads and access.

Who Needs It

Reporting on CDP operations is essential for the people who manage the CDP. Other CDP reporting is needed when operational systems don't provide adequate reporting on their own data and when users don't have other tools available to report against the CDP contents. Even when users do have other reporting tools, the CDP needs to make the data available in formats those tools can use.

What To Look For

Standard Reporting Features
Any reporting system should provide out-of-the-box features such as: selecting data elements to include; tabular and cross-tab formats; grouping and subtotals; user-defined labels; time-series and trends; highlighting of outliers; value frequencies; graphs such as line, bar, and pie charts; export to flat files and spreadsheets; and interactive drill downs. There are many others. Most users today will have experience with third party reporting tools that provide a frame of reference.

Customer Profiles
More specific to CDPs, the system should present customer profiles that combine data from all CDP sources. These profiles should include summaries, such as total number of Web site visits or cumulative purchase value, as well as details such as a timeline showing all interactions. Visualization should allow users to more easily understand profiles and compare different profiles to each other.

Segmentation
Users should be able to define customer segments based on complex expressions that include multiple attributes and operators, such as “customers who live in California and purchased from any two of these five product categories in the past six months”. They should be able to run reports that give aggregated results for each segment, including profiles, comparisons of different segments, comparison of the same segment over time, and comparisons against benchmarks or averages. Users should be able to highlight a segment on a tabular or graphical report and automatically save it as an audience available for further analysis or marketing execution. When users save a segment in this fashion, they should be able to specify whether the list is static (the specific customers selected when the segment was created) or dynamic (whichever customers meet the segment definition each time the segment is used).

Usability
Reporting features should be available to non-technical users. The selection interface should let users create groups or segments without writing code in SQL, PHP, or other languages. Metadata that describes CDP contents should be available to the reporting system. The system should provide a library of pre-built reports to answer common questions, with options for users to modify these as needed. Users should be able to save their queries and reports and to automatically run and distribute saved reports on a specified schedule. The system should be able to automatically scan reports for exceptions or trigger conditions and issue alerts when it finds them.

REPORTING CHECKLIST

Standard Reporting

- ☐ Access all data
- ☐ Multiple formats
- ☐ Visualization
- ☐ Interactive drill down

Segmentation

- ☐ Define with complex expressions
- ☐ Compare over time, vs other segments
- ☐ Save segment as promotion audience
- ☐ Static and dynamic segments

Customer Profiles

- ☐ Summary statistics
- ☐ Timeline detail
- ☐ Visualization

Usability

- ☐ Non-technical user interface
- ☐ Prebuilt report library
- ☐ Reuse queries and reports
- ☐ Scheduled reports and distribution
- ☐ Trigger conditions and alerts

ANALYTICS & MACHINE LEARNING

What It Is

Use of advanced statistical methods to explore CDP data, predict customer behaviors, recommend actions, and optimize choices over time. Machine learning performs these tasks with little human intervention and automatically adapts its models to new information over time. It allows marketers to tailor treatments to very large numbers of individual customers or micro-segments.

Who Needs It

Enterprises with multiple products, segments, messages, campaigns, channels, or other dimensions that are too complicated to optimize through manual methods alone.

What To Look For

Prepare Data for Analysis
CDP users need tools to extract subsets of the full CDP database for specific analytical tasks. The extracts will often reformat the data by combining information from separate sources into a single table and sometimes into a single row for each customers. Analytics may also require specialized cleaning, standardization, transformation, and enhancement beyond the steps taken when the data was loaded. The extract will often be placed into a flat file, database table, or specialized format different from the primary CDP storage format. The CDP should provide options to save a user-defined extract process and rerun it automatically on a regular schedule. The CDP should be able to alert analytical processes when the new extract is ready and to distribute the extracts to external analytical systems.

Analytical Functions
CDP data is used for a wide range of analytical tasks. Descriptive tasks include data exploration, segmentation, and cluster analysis. Predictive models calculate the likelihood of customers taking a defined action. Recommendation engines suggest the best choice among multiple options; in addition to predictions, this requires a way to specify the value of different outcomes so the system can compare them.

Optimization programs find the best approach to take when making many individual choices; in addition to predictions and values, this requires considering constraints such as advertising budgets, number of messages per person, calling capacity, or available inventory. Determine which of these tasks you need your CDP to perform.

External Systems
Analytical functions may be built into the CDP or provided by external systems. Many enterprises will have an existing analytical staff that prefers to use its current tools instead of or in addition to tools built into the CDP. If that's your situation, look for a CDP that has prebuilt connectors for the tools of choice, can prepare its data in a format compatible with those tools, allows those tools to connect to raw CDP data for ad hoc exploration, and can import products of those tools such as scores or scoring formulas. Understand what work, if any, Is needed to make new data objects or attributes accessible to the external analytical tools.

Automation
Analytical tasks can be handled by statistical specialists, data scientists, marketers, business analysts, or others with different skill sets. Each type of user needs different tools: in general, more skilled users want more technical tools that let them take precise control over each step, while less skilled users want to give guidance and let the system handle more of the details. Today's most advanced systems can almost fully automate the model building process, handling much of the data preparation as well as the actual model creation and scoring. Machine learning systems go further by automatically adjusting their models as new information is received Other forms of automation can watch user behavior and make relevant changes such as adding indexes or precalculated variables to support common queries.

User Control
The degree of automation will determine how much effort users need to put into analytical projects. Even fully automated systems will need users to define the audience, outcome to predict, value calculations, and constraint parameters. In less automated systems, uses also do data preparation and variable selection. Assess the effort and skills needed to do these tasks, taking into account your current staff, ability to hire new staff, and how many analytical projects you expect to need. If you own resources are limited, look for a CDP vendor who can supplement them with its own staff or partners.
Whatever the degree of automation, look for a system that lets users assess its work through features such as showing the weights assigned to different inputs, displaying the range and frequency of outputs, comparing predicted results with actual results, and showing the improvement gained by using predictions.

Access to Analytical Results
Some analytical outputs are reports or visualizations that simply present the findings. But predictions, recommendations, and optimizations are usually intended to directly drive business decisions. The CDP should make these available to use in segmentations, business rules, personalized messages, media plans, and elsewhere. How results are accessed will depend on the situation but it might involve reading analytical outputs from a database table, calling for them through an API, or loading them into an external system. Real-time processes, such as feeding recommendations onto a Web page, impose specific requirements for quick response and immediate use of information gathered during the current interaction.

ANALYTICS & MACHINE LEARNING CHECKLIST

Prepare Data

- [ ] Clean, standardize, transform, enhance
- [ ] Reformat to match target system
- [ ] Save and schedule standard extracts

Analytical Functions

- [ ] Descriptive analytics
- [ ] Predictive models
- [ ] Recommendations
- [ ] Optimization

External Systems

- [ ] Existing connectors
- [ ] Reformat data for external access
- [ ] Direct external access to raw data
- [ ] Import scores and formulas

Automation

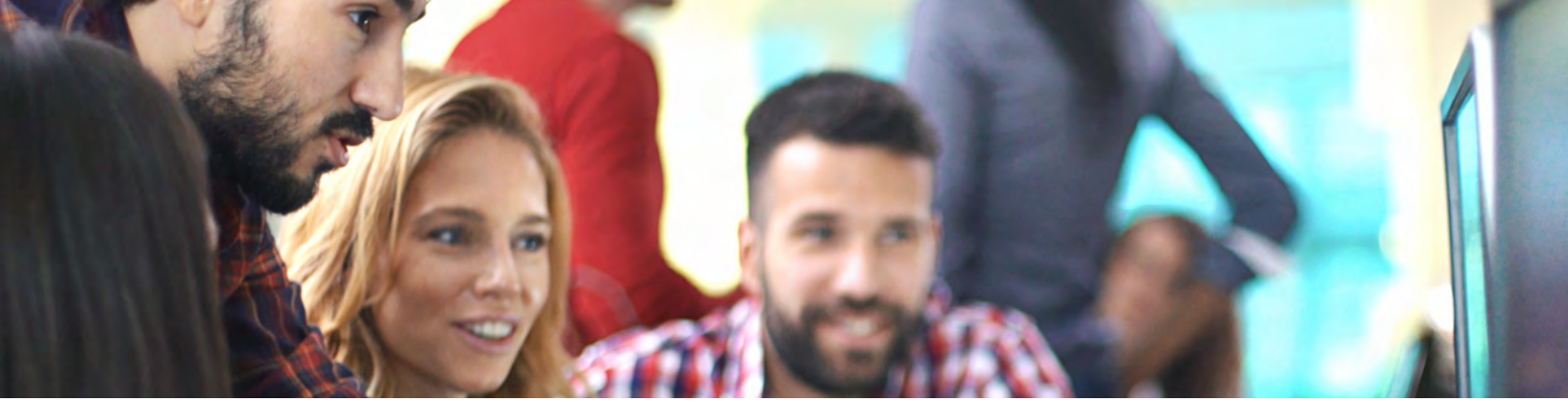
- [ ] Automated processes
- [ ] Technical user intervention
- [ ] Self-adjusting models
- [ ] Automated indexes and summaries

User Control

- [ ] User skills needed
- [ ] Vendor resources
- [ ] User review of results

Access to Results

- [ ] Output to table
- [ ] Call through API
- [ ] Load to external system
- [ ] Real-time response



DATA ACTIONABILITY

What It Is

Making CDP data available for marketing execution, especially in real time.

Who Needs It

Marketers who want to use CDP data to drive marketing programs, as well as analysis. The CDP lets systems in different channels draw on the same set of unified customer data and can support central decision systems that coordinate customer messaging across channels.

What To Look For

Accessible Formats
Marketing execution systems include campaign managers, email engines, Web personalization tools, call centers, ad buying platforms, retail kiosks, and others. These may be part of the CDP or external systems connected to the CDP. Either way, they typically need data in specific formats that are different from the main CDP repository. The CDP needs to prepare data for those systems by extracting selected elements and placing it standard formats such as database tables or in Data Actionability Checklist Accessible Formats Specify elements to include Scores and calculated values Indexes and summary tables Update Speed Load and prep time Support required volume Real Time Access Specialized formats Interaction system integration Background processing flat files. It may add calculated attributes such as cumulative purchases, last purchase date, predictive model scores, recommended next offer, segment codes, interests, or affinities. It may also need indexes or other features to speed data access.

Update Speed
Data enters the CDP from many sources and goes through several processes before it is added to the main CDP data store and reformatted for access by execution systems. The time to complete this process can vary from a few seconds to days. Even greater delays may be due to source systems that send their data to the CDP at longer intervals.

Users need to understand how often data will be fed into the system and how quickly it will become available. Update speed may vary based on data type, volume, and source. Users should compare their expected volumes with CDP capabilities to ensure the system can deliver updates as quickly as they're needed.

Real-Time Access
Some execution processes require real-time access to the CDP to use its data during live interactions such as a Web site visit or retail checkout. Depending on the application, this may require response time as low as 30 milliseconds. Meeting these standards may require specialized data storage and processing methods. In some cases, only a small amount of data is made available in this way. Real-time access doesn't necessarily imply real-time updates of the underlying CDP data, although up-to-the-second information is often important in real-time situations. In some cases the CDP will temporarily make some new data available immediately, for example to recalculate model scores or recommendations, while simultaneously running the same or other data through slower preparation processes before adding it to the permanent record.

DATA ACTIONABILITY CHECKLIST

Accessible Formats

- ☐ Specify elements to include
- ☐ Scores and calculated values
- ☐ Indexes and summary tables

Updated Speed

- ☐ Load and prep time
- ☐ Support required volume
- ☐ Real Time Access

Specialized Formats

- ☐ Interaction system integration
- ☐ Background processing

FINAL THOUGHTS

The functions listed above are important to nearly all CDP users but especially relevant to large enterprises that have the most demanding requirements. But even enterprise requirements vary substantially based on each company's needs and resources. The capabilities of CDPs also differ greatly, with the additional complication that features not found in a particular CDP can often be provided through integration with external systems. In all cases, marketers need to start by setting their goals for the CDP, then define the functions needed to support those goals, and then define the CDP features needed to provide those functions. Only then will they really know what to look for in a CDP system.