OPENPRISE

White paper

The complete data enrichment survival guide

for RevOps

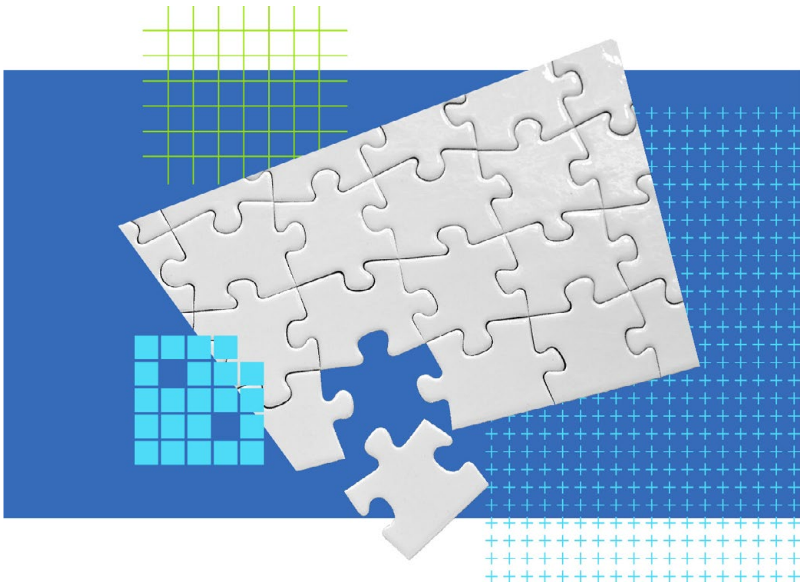

Table of contents

Table of contents, cont'd.

Who should read this guide?

So, you've come to the conclusion that you need help to enrich your marketing and sales data. Data enrichment can be a confusing project. How do you pick the right vendor? Do you need more than one vendor? What technologies do you need? What data do you need? How do you maximize the return on your investment?

This guide will cover the key topics for ensuring success with your data enrichment endeavor so you can maximize the impact of your budget. If you're doing enrichment for the first time, this guide will save you from making some common mistakes. If you've purchased enrichment services before, you probably already have war stories, but we hope you'll still pick up some useful tips.

Before you contact any data providers

Before you start Googling data providers or looking for comparison reports, you first must understand what your needs are so you can determine the criteria to select and evaluate third-party providers.

Determine your goals and purpose for data enrichment

How do you expect to use the third-party data? Here are the common use cases we see, in order of frequency:

1. Identifying new prospects

By far the most common reason Openprise customers procure third-party data is to acquire new prospects, whether they're accounts, contacts, or "total addressable market" (the latest fad). With the elevated interest in account-based marketing, there's an even stronger need to acquire new prospects based on some definition of an ideal prospect, or ideal customer profile, developed either through human analysis or predictive analysis.

2. Enriching existing prospect data

The second major use of a third-party data service is to improve your data in terms of completeness, accuracy, and recency. Enrichment is no longer just about contact information, as it can include these major categories of data:

- Contacts (people data)
- Accounts (firmographic)
- Technology used (technographic)
- Device type
- Intent (activity)
- Sentiment

3. Validating first- and third-party data

Even if your data is complete and accurate, or is supposed to be complete and accurate, there's often the need to verify the data with another third party. For example, the data in your CRM system may be a little long in the tooth, as no one's updated it in the past twelve months. So it's a good idea to validate whether it's still accurate.

If you've just acquired a list of prospects but don't have the utmost confidence in the data source, what can you do? It may be a good idea to validate some aspects of the data, like email deliverability, before loading it into your systems.

When it comes to validation, you want the most accurate data, as with enrichment, and you also want to know if the data isn't working and why not. This information lets you make disposition decisions as well as assess the quality of the data source.

Your goals and purposes may very well be a combination of more than one of the above, but it's important that you rank the three in terms of usage volume and importance. As you'll see later, your goals and purposes for the data are the most important factor when evaluating a data provider.

Data enrichment is primarily the concern of marketing, especially when doing it at scale. Sales' enrichment usage tends to be manually done one at a time by the individual salesperson.

Determining which processes the data will support

It's tempting to just say, "The data must be perfect." The truth is that the perfect database doesn't exist. Depending on how they develop and maintain their database, each data provider has strengths and weaknesses. By understanding precisely what aspects of the data are most important to you, you can maximize your success in finding the right mix of data providers that best fit your needs. Here are a few common examples:

1. Sales prospecting efficiency

Not having quality contact data can drastically reduce the efficiency of your sales team. It can take a salesperson 30 minutes to try to find a phone number, navigate the phone tree, and talk his way past the gatekeeper before he even has a chance to leave a 30-second voicemail. If your goal is to improve the efficiency of your sales team, then direct dial phone numbers and verified deliverable emails are very important to you.

2. Tactile campaigns

Are you adding some "old-fashioned" touches to your marketing campaigns by going tactile? Sending a bottle of nice wine or having a gift basket delivered can be terribly expensive if your address data isn't accurate and the gifts are returned. If you're doing high-volume mailing, then just having an accurate physical address may not be good enough, you may also need postal service optimized address data. For example, not just city, but USPS postal city.

3. Segmentation, scoring, and attribution

Anything marketing wants to do that involves buyer personas and ideal customer profiles will require data to support database segmentation. Segmenting at the contact level typically involves job function, job sub-function, and job level. Some data providers only provide job titles, while others provide job function and job level as well. If the third-party job function and job level data is accurate and matches your segmentation logic precisely, then that data is important. However, that's rarely the case. The providers' data is generic and isn't likely to be a perfect fit to your go-to-market strategy. Thus, in most cases, job title is the raw data you need for deriving job function, job level, and persona.

Are you segmenting accounts by company size? Do you do this by the usual review of annual revenue and number of employees, or by more custom dimensions like number of office locations, number of countries operating in, size of vehicle fleet, number of open job requisitions, or number of rooms/beds?

How about segmenting by industry or technology used? Knowing your segmentation needs can help you prioritize what aspects of the third-party data is most important to you. Data that is specific to certain industries, like the number of beds in a hospital or the average cost of a hotel room, typically won't be found with generic data providers, and is usually sourced through a vertical-specific vendor.

Once you have clarity and priorities set for your high-level goals, next determine which specific business processes the data will support.

Determining which processes the data will support

4. List loading/ list enhancement

Have your found that you don't always get all the data you need from your business partners that send you leads? Are you too often missing job title, phone number, or address? Or do you get many personal email addresses when what you actually need are business email addresses? Do you get a lot of social media data like Twitter handles and Facebook names that need to be mapped to business contact data? If the process involves filling in missing sets of data fields, then you need to focus on the quality of those specific data fields.

5. Form append

Do you want to keep your sign-up forms as short as possible and minimize user fatigue or input errors? The best approach is to determine the minimum set of data fields needed from your prospect and partners filling out a form. Then, consider which additional data fields can reliably be appended from either open data or third-party data.

Determining your target market

At the risk of sounding like a broken record, we'll keep on repeating that there's no single perfect database for every company. This is perhaps the most important thing to understand, as it's the most critical factor in determining the success or failure of your data enrichment project. A good first step is identifying the data providers whose databases offer good coverage for your customer profiles. But before you can evaluate data providers, you must first have a deep understanding of your customers and ideal prospects.

Here are the key dimensions you must precisely define:

1. Business or consumer

This one is the most obvious, but also the most important. Depending on whether you're a B2B or B2C company, there's a completely different set of data providers who serve your needs. B2B data includes firmographic, technographic, and contact data. B2C data is often referred to as consumer data or audience data.

2. Company size

Do you sell to large enterprise, mid-size enterprise, or small businesses? Some providers specialize in large enterprise and others focus on smaller companies. It's not sufficient to just use a generic label like those mentioned earlier. Be very specific in your definition of company size and quantify it. For example, a mid-size enterprise is 200 to 500 employees; annual revenue is $100 million to $1 billion, or properties with over 1,000 beds. We haven't seen a data provider that has equally good data for both Fortune 500 companies like Apple and Exxon and small mom and pop businesses like the cafe and nail salon in your local strip mall.

3. Industry

If you have a strategic focus selling into specific sectors, you must include this in your data provider selection criteria. Some industry data is only available from specialized data providers, and not from general-purpose data providers. For example, data about the number of beds in a hospital or the average occupancy rate of a hotel property isn't available in most general-purpose databases.

4. Job function/department

What departments do you sell into: IT, HR, Finance, or Marketing? Many database providers who focus on large enterprises also focus heavily on IT buyers, so coverage in other departments may be thin.

Determining your target market

5. Job level/seniority

Determine the prospect's job level within the company. Are they C-level, executive, director, manager, or an individual contributor? Many data providers focus on decision makers and budget owners, so the data quality and coverage for senior people may be much better than other job levels. However, if you sell a SaaS product to developers and heavily rely on freemium and free trials for lead generation, even the large enterprise IT-focused database providers may not have the data you need to reach programmers and system administrators.

6. Geography

Most data providers have very good coverage for the United States and Canada. The coverage is often not as good in Western Europe, and can be even worse in Asia, Australia, South America, Eastern Europe, and the Middle East. To solve this issue, there are regional data providers that excel in specific regions outside of North America and Western Europe. This is one of the most common reasons why you may need multiple data providers. Since the GDPR (General Data Protection Regulation) went in effect in May, 2018, European data is harder to obtain and you have to engage with country-specific providers in the European Union. More on the impact of GDPR below.

7. Source of data

- Data providers use a multitude of methods, channels, and technologies to acquire their data. Some of the most common means are:
- Purchased or swapped from other data providers.
- Crawling the web, specifically social media and forum sites.
- Acquiring subscriber databases from magazine, newspaper, and other media publishers.
- Crawling government open-data sites, specifically for secretary of state filings.
- Crowdsourcing using manual entry, address book scraping, email scraping, or business card scanning.
- Manual research.
- Depending on your ideal customer profile, some of these methods would yield better results than others.

Selecting data providers

Many factors can influence how you buy and whom you buy from. All the internal requirements we discussed above play an important part in your sourcing strategy, but there are other tactical considerations as well.

Sourcing strategy

1. Single vendor or best-in-class

Whether you're looking for only contact data or comprehensive data that includes firmographic, contact, social, technographic, and intent, you'll be faced with the choice of using only a single vendor that gives you the widest coverage, or assembling a collection of best-in-class vendors.

The advantages of a single vendor include:

- Lower cost
- Fewer integration and automation requirements
- Less confusion for users
- Fewer data standardization issues

The benefits of the best-of-breed approach are:

- Highest possible data quality, depth, and coverage
- Cross-validation

Now that you're clear on your goals and needs, you must decide on your preferred consumption model. Here are the factors you need to consider.

2. Single source or multi-source

Match rate for data providers can range from under 10% to over 70%. We've seen most match rates tend to fall between 30% and 50%. So, if your preferred provider doesn't find a match or returns a low confidence match, you may want to try another provider. Another common reason for multi-sourcing the same data is to get the necessary geography, vertical, or company size coverage.

3. Cost model

Data providers use different pricing schemes. All else being equal, the best-fit cost model can help you maximize the ROI of your data investment. Here are the most common models and when they make the most sense:

Credits/a-la-carte

You buy a basket of credits and those credits can be used for different levels of services. For example, email validation usually costs the fewest credits, appending existing leads costs more, acquiring a new lead costs even more, and acquiring additional data like direct-dial phone numbers and social data may cost the most. The benefit of this model is that you can maximize your budget given your desired mix of data needs. However, tracking usage and budgeting can get tricky.

Selecting data providers

Per record

You pay once for each record you receive. You don't pay again if you send in the same record again. The benefits here are simplicity and that you will never pay more than once for the same record. This makes it simple when it comes to implementing the process and technology that will use this service, because you don't have to worry about coordination.

Per request/API call

You pay every time you make an inquiry, so if you send in an enrichment request for the same person three months in a row, you pay for that record three times. This makes the most sense when you are doing batch bulk enrichment. It makes the cost tracking very simple if you don't have multiple request channels.

All-you-can-eat

You rent the entire database or a subset of the database and can get as many leads and make as many inquiries as you want. This is the best option if you want to keep your records absolutely up to date using automation tools like Openprise.

Understanding the data supply chain

It's a very common practice within the data business to source data from other data providers. In fact, the majority of data providers don't have original data they produce themselves—they aggregate other providers' data instead. Even providers who do have original data often source third-party data to produce a more comprehensive offering. Here are some examples:

- If the data contains a DUNS Number, it's guaranteed that the provider has sourced firmographic data from Dun & Bradstreet (D&B).
- Salesforce's Data.com firmographic data is D&B data.
- Many providers' technographic data is sourced from the firm HG Data.
- Social media data are often sourced from SYNTHio and FullContact.
- Email validation is often sourced from BriteVerify and FreshAddress.

Whether you're going with a single provider or best-in-class approach, you should always ask providers which third-party data they've sourced. It's important to understand what the original data sources are because:

- You shouldn't unknowingly pay for the same data multiple times.
- If you're using a multi-sourcing strategy to achieve a more comprehensive coverage, you may not get the coverage you were targeting.
- This has great implications when it comes to GDPR compliance, which we'll discuss shortly.

Each one of these sourcing relationships is likely unique. The provider can source anywhere from just one data field like the DUNS number, or the entire data set. So, ask not only which third parties they source data from, but which data fields are sourced from whom.

GDPR compliance

General Data Protection Regulation (GDPR) from the European Union, went into effect May 25, 2018.

If you have European Union citizen data in your marketing and sales database, you're subject to GDPR compliance, and GDPR can severely limit your choice of data enrichment providers. In GDPR terminology, any data provider you use is a "Data Processor." To send any EU citizen data to any Data Processor for any purpose, including enrichment, you must have a Data Processing Agreement/Addendum (DPA) signed with the vendor. Here's an example from Salesforce.

A compliant DPA must contain the EU's Model Contract/ Clause, which requires the Data Processor to follow a set of standard security and privacy protocols. What it boils down to is you can't send any EU citizen data to any partner and vendor unless you have a compliant DPA in place, and not all data providers will sign a DPA. So, if you have a non-trivial amount of EU data and you'd like to include it in your enrichment and prospecting effort, you must pick a data provider that's willing to sign a DPA.

It's also worth clarifying what constitutes a transfer of data to a data provider. It includes all these consumption channels:

- Sending a spreadsheet or flat file to a data provider for bulk match.
- Getting per-record enrichment using any vendor-provided plugin for your sales automation or marketing automation platform.
- An "inquiry" that requires sending any personal information to the vendor for match purpose, including any API call, is a data transfer event.
- This also includes any "smart form" technology that does dynamic lookup as the prospect fills out a form.

Since many data providers source data from other third-party providers, GDPR mandates the entire data supply chain to be secured and compliant. Once a DPA is signed with a Data Processor, the liability is further shared down the supply chain. In other words, if any data providers are not GDPR compliant and are not willing to sign a DPA, then the data provider who is source data from them also cannot be GDPR compliant. This domino effect within an industry that has complicated sourcing relationships can make GDPR compliance extremely challenging.

It's worth clarifying that GDPR doesn't just apply to EU companies or companies with EU business units and physical presence, but it applies to any company worldwide that holds EU citizen data.

Alternatives to paid services

More often than not, the issue with data quality is not the complete absence of data, but just incomplete, non-uniform, and unstructured data. For example:

Incomplete data

- Address contains ZIP code but no city and state
- Address contains state but no country
- Phone number is missing country code

Non-uniform data

- State contains: California, CA, and Calif
- Country contains: United States, US, USA, and U.S.A.
- Puerto Rico is sometimes a country and sometimes a state

Unstructured data

- Has job title, but no job function or job level
- Too many variations on industry data
- Annual revenue and employee count are sometimes numbers and sometimes ranges and the ranges are all different depending on the data source

Most of these data quality problems can be resolved without paying a third-party provider. You should reserve your precious enrichment budget to acquire missing data or validate data. Here are some methods you can use to improve your data quality without spending money with a data provider. We'll discuss later that there's great benefit to "pre-clean" your data before sending them to an enrichment provider.

1. Take advantage of open data

Many data fields are related, especially address, phone number, email, and website data. The missing data fields can often be filled in easily if you have the right reference data. Here are some examples:

- If state = California, country can be inferred as US
- If country = US and ZIP code = 94403, city and state can be inferred as San Mateo, CA
- If country = Ireland, then the country code can be inferred as +353
- If area code = 510, then the metro area can be inferred as San Francisco Bay Area

This type of inference and filling in the blanks can be done with a combination of automation technology and open data. Automation technology can be as simple as Excel's VLOOKUP formula, or something more advanced like Openprise RevOps Automation Platform. For open data, check out the Marketing Open Data Project. It's a great source for the various lists you'll need, like city, state, county, country, area code, postal codes, stock ticker symbols, domain suffixes, free and disposable email providers, and urban and metro areas.

Alternatives to paid services

2. Use freemium services

There are plenty of free resources on the web that can be leveraged to improve your data quality. Google Maps is a great example. We probably all have used Google Maps, and know how powerful the service is. It can:

- Provide geolocation data, given an address.
- Return complete addresses, given partial addresses.
- Return business names, business types, and phone numbers.
- Standardize any address text to a standard format and break it down into component parts.

Many of the Google Services, like Google Maps/Places, Google Search, and Google Translate are available as APIs and come with a free usage tier. So, if you have the right automation solution, you can use these free services and take advantage of your free daily quota.

3. Leverage your own database

In many organizations, account and contact data is of much higher quality than lead data. Account and contact data have been scrutinized by the sales team and often contains valuable data from manual information gathering and research processes. You can easily improve your lead data by finding the right match to your account and contact data using any number of matching schemes. For example, if a lead record only contains the company name and state, if that can be matched to an account or contact record that has the full address, then the lead's address data can be filled out.

Match test

Once you've defined your requirements and assembled a shortlist of data providers to evaluate, it's time to run a match test. Here are some recommendations and tips on how to run an efficient and effective multi-vendor match test.

Sample size

The match test should be run on a representative sample of your database that is substantial enough to yield meaningful results. We recommend the sample size to be at least 10% of your database, or 10,000 records.

Representative sample

A representative sample isn't just a random sample of your entire database, it should be a random sample of the part of the database that you most desperately want to improve according to the process you want to support, as discussed previously. This usually means it's the worst part of your database. For example, if you're looking to enrich leads during the list loading process, and you would like to spend money to enrich only leads that are missing phone number and job title, then that scope should be the baseline for your representative sample. You should pull a sample of leads that have no phone number and no job title. This is important as it correctly evaluates the vendor's performance for the part of the database that you want to enrich.

Specify your critical data fields

Make sure you tell the vendor which data fields are critical to your evaluation. Specify whether it's direct dial phone number, email, job title, company size, industry, or DUNS number. Ask the vendor to provide match rate statistics on these fields explicitly.

Ask vendors about preparation work

Each vendor's matching algorithm works differently and its performance may vary drastically based on how the input data is formatted and structured. There's nothing wrong with a vendor doing manual preparation work on your data before they attempt the match, but you should ask the vendor what preparation work they have performed to achieve the match result. This is important—to achieve a match rate comparable to what they quoted you during the test, you'll need to do the same preparation work in your production environment. Here are some examples we have seen from the match tests we have run for our customers:

- Match rate is much better using separate first and last name data fields compared to using a single, full name field.
- Website data must begin with "www" or it won't match.
- Email syntax error and invalid suffix must be corrected before matching, for example, "jdoe@acme.con" must be corrected to "jdoe@acme.com."
- Phone number data cannot contain any letters. For example, "(800) Flowers," "650.555.1212 (FAX)," "4155551212ext1234" won't match.
- Multiple values in one data field won't match, like "(415) 555-1212, (650) 555-1212."
- Use fewer fields for matching can produce better results.
- Certain data fields are absolutely required, like a business email address.

Match test

Get samples of full data set

It's neither realistic nor fair to expect the vendor to return the full data set for all your matched records. After all, when you go to Costco and taste the samples, you don't expect them to give you the whole bag or bottle, do you? Expect to get back a report on:

- The number of records successfully matched.
- The number of matched records containing the critical data fields you specified.
- When the matched records were last updated.
- Match confidence score.

In addition to the report, it's fair to request enrichment for a small sample of records so you can see what the data looks like and do your own quality validation. What each vendor will agree to depends highly on the size of your potential contract.

Validate the results

So, even if a data provider tells you a match is found, it simply means they have a record. It doesn't mean the record is any good. If you have spent any time talking to data providers, you have no doubt heard the quality vs. quantity argument. Vendors whose match rate is lower than the others almost always claim the others have lower quality data. The only way to gauge a data provider's quality is to validate it yourself. Include a list of contacts from your own company, your customers, and your friends—data that you know is accurate. This doesn't have to be a large set of data; 100 records would do just fine. Use this sample data set to evaluate the quality of the matched records.

Implementation tips

Now that you've picked the best data providers for your needs, you're ready to put the services into production. Here are some tips and things to consider when deploying an enrichment data service into production.

Batch or continuous process

You can implement enrichment either as a batch process or a continuous process. Batch process is quicker to implement since it's usually a manual process. This typically involves simply extracting a file of data to send to your data provider for enrichment. The hard part is the manual process of incorporating the results back into your database. Most people just dump the data into custom data fields, which we do not recommend.

A continuous process means either an automated, standalone enrichment, or having the enrichment steps incorporated into other automated processes like list loading, lead routing, or segmentation. This means using RevOps automation technologies like Openprise. A combination of these may be the most practical approach.

Integrating enrichment steps into other automated processes makes sure the best data is available to support those processes when needed, thus optimizing the performance of those processes. Having a standalone process to maintain otherwise inactive records may be a good practice to ensure that no records go stale for a long period of time.

If you choose to do a batch process, make sure you run the batch process in short enough intervals to balance data quality vs. available budget vs. resource availability. Remember that the longer you go between batches, the bigger the reconciliation job gets.

Pre-clean the data

We highly recommended that you pre-clean the data before sending it out for enrichment. Pre-cleaning the data can drastically improve the match rate, thus maximizing the effectiveness and ROI of your data enrichment investment. Pre-cleaning involves:

- Correcting obvious mistakes like email syntax errors and invalid domain suffixes.
- Filling in the blanks that can be inferred from other pieces of data. For example, filling in city and state if zip code data is available.
- Transforming to a data standard that your data provider prefers in order to maximize the match rate.
- Cleaning up formatting in fields like phone number, zip code, and address.

Some data providers offer professional services to clean the data first before enrichment. This is a viable option for batch processing. For continuous processing, the pre-cleaning needs to be automated with technology.

A data provider's ability to use dirty data as-is varies, so it's best to not leave this to chance and do your own pre-cleaning.

Reconcile new data immediately

Perhaps the biggest mistake when it comes to data enrichment is not reconciling the new data into your primary data fields immediately after data acquisition. Most people simply keep the new data in custom data fields. This is a waste of money because:

- Your sales team won't be looking at this data unless the primary data fields are not working.

Implementation tips

- Your marketing automation platform or other marketing technology won't be able to leverage these custom data fields.
- You'll end up with multiple sets of information that may be different, and it'll become impossible to reconcile later once the context is lost and old and new is mixed together.

Come up with the appropriate reconciliation logic about which data fields to keep, which ones to throw away, and when to overwrite. Reconcile the new data into your primary data fields. No reconciliation logic will ever be right 100% of the time, but having a set that's 85% right is much better than not reconciling at all.

Once again, for batch processing, this reconciliation task can be done manually. A continuous process needs to be automated.

Simply put, if you don't reconcile your new data immediately, you're just throwing your money away.

Normalize and segment new data

Depending on your data provider's flexibility in delivery options, the data you get back may require further processing to meet your data standards. Common examples of these tasks include:

- The data you get back may not be in the right format, so you'll have to transform it. For example, new data says a lead's country is "USA," but your standard is "United States."
- Instead of receiving the data in your preferred standard, you may just get all the different versions, like "US," "USA," and "United States," so you need to pick the version you want.
- Data fields like industry, number of employees, and annual revenue are generically set by the data provider, which may not be the same way you segment your data. For example, the number of employee data from the provider may be "50-100," but your segments are "1-25," "26-200," "201-1000." You need to map the generic segments to your specific segments.
- Data fields like job function and job level are also generic segments that may not match your needs. You may need to do your own segmentation based on other data fields like job title.

The role of technology

Automation technology like the Openprise RevOps Automation Platform should be part of your data enrichment strategy, especially if your database is large or if you enrich frequently or continuously. RevOps automation technology like Openprise improves the outcome of your enrichment process by:

- Pre-cleaning data before it's sent for enrichment.
- Processing the new data, post enrichment, to ensure compliance to your data standards and segmentation.
- Reconciling new data into primary data fields.
- Integrating new data into your sales automation, marketing automation, support, or other databases.
- Integrating data enrichment inline of processes like list loading and lead routing.

Separate your data provider from your automation technology. Changing data providers is easy, but changing automation technologies is hard. Don't end up being stuck with a data provider you no longer feel is the best fit, but can't swap out, because it's the only one available with your automation technology.

OPENPRISE

Parting thoughts

We hope you found this white paper helpful. Data enrichment, when done right, can add tremendous value to your account and contact database.

Here are some of the key points worth reiterating:

- It's rare that one data provider can meet all your data needs. Consider a multi-source enrichment strategy.
- Understanding your own needs and requirements is the key to selecting the right data providers for you.
- Enrichment should be a continuous process and can be automated. The extra time and effort involved in automating the process will easily pay for itself within a year.
- Reconcile new data immediately. Data isn't like red wine. It doesn't get better with age. It's more like a ripe avocado that starts to go bad the moment you acquire it.
- There are a wide variety of data providers out there and affordable automation technologies are readily available, so there's really no need to compromise, whatever your budget is.

Happy enriching!

If you liked this white paper, you may also like our survival guide for deduplication.

About Openprise

The Openprise RevOps Automation Platform fuels company growth by automating hundreds of sales and marketing processes, helping RevOps teams realize the value promised from their RevTech investments. Openprise is a single, no-code platform that can help to simplify even the most complex RevTech stack. For more information, please visit www.openprisetech.com.

Contact Openprise:
info@Openprisetech.com
(888) 810-7774
www.openprisetech.com