

in

inspire
innovate
involve



FREE GUIDE

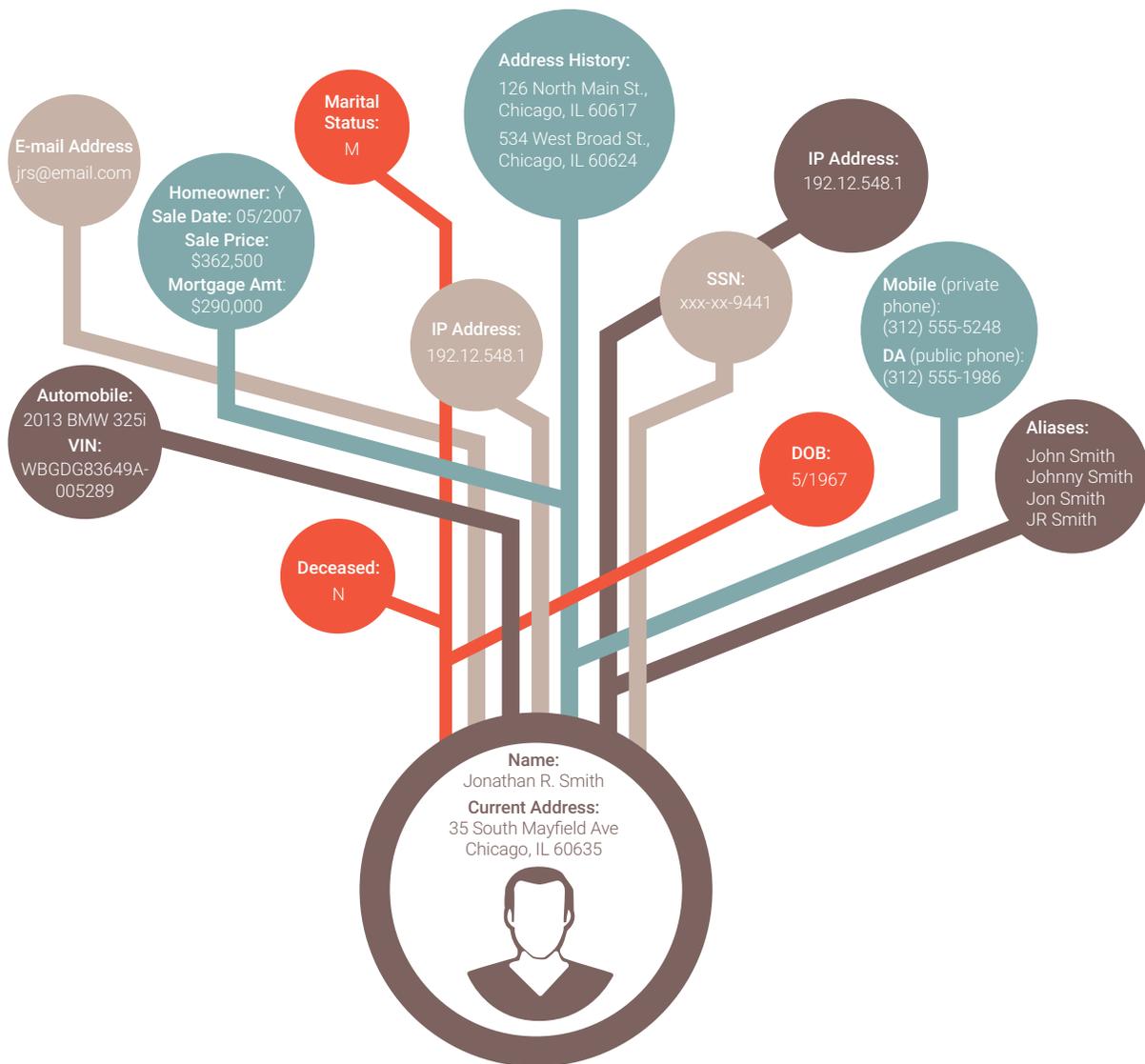
Data Appending 101:
16 Questions to Ask Your Data Provider

infutor

1 What is Data Appending?

What do the neighborhood pizza shop, the big box department store, Fortune 500 financial institution, and the national non-profit organization have in common? All four need to know as much as they can about their customers and prospects, and they need to know how to reach them.

The only way to gain this information is through data appending: taking the information you have and turning it into much more. A list of names and addresses can be transformed into a telemarketing list complete with demographics and purchase behavior, for example. In layman's terms, a data append is when a vendor receives input data, matches it to a master database, and returns a file containing both datasets.



2 Forward vs. Reverse Append

A **forward append** begins with a name and address, and adds a variety of additional information about an individual. This can include phone numbers, email addresses, vehicle information, property data, demographic information, and much more.

Forward appending is the most common form of append, but it is not the only one. Another common type of append is called **reverse appending**, whereby one field of data is used to return all other data tied to it. What if a call center has a list containing phone numbers of callers who expressed interest in a product, but lacks name or address? A reverse phone append will fill in the blanks. Using the phone number for exact matching to a master file, the call center can obtain the name, address, phone type, email address, and even complete demographic information associated with that particular number. Based on this additional information, they can segment and send postal mail or email to the prospects, trusting they are marketing to people who are interested in their product.

Reverse appending is available for virtually any other distinct dataset, from email address to Vehicle Identification Number. Typically a reverse append uses exact matching, whereas a forward append can use exact and/or fuzzy matching, which is explained in further detail below. Because of the nature of exact matching, reverse append is more accurate and takes less processing time than a forward append.

F

FORWARD APPEND

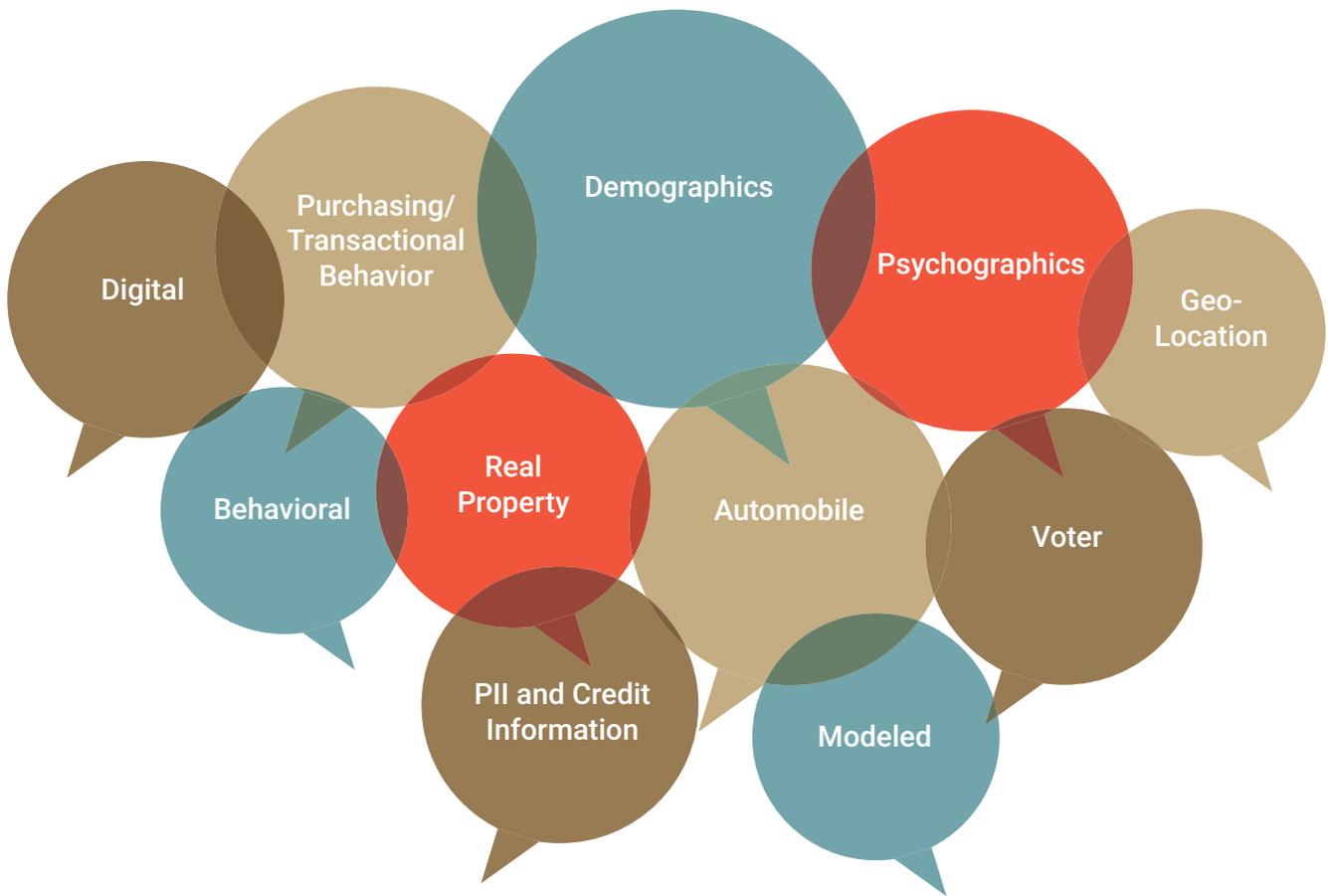
A forward append will begin with a name and address and match to a number of additional data elements.

Commonly appended elements include phone number, email address, demographics, automobile, lifestyle data and much more.

R

REVERSE APPEND

A reverse append will begin with a piece of information such as a phone number or email address and identify a consumer's name, address, as well as additional data elements about the consumer.



3 What Type of Data can be Appended?

Phone numbers, email addresses, and demographics are the most frequently appended elements. However, there are thousands of other elements that specialty companies can provide, from motorcycle ownership to social media propensity. If you can think of it, it likely exists as an element that is available for append.

Appending is not just a one-time process. Some of the country's largest companies run monthly or quarterly appends to follow their clients' behavior and demographics; as these items change, so do their needs as consumers.

For example, a bank benefits from knowing that their client has recently married, because direct marketing to a newly married 30-something looking to buy a home is different than to the single 20-something who lives at home with parents.

4 Real-time and Batch Delivery

There are two ways to obtain appended data: through **real-time lookup** or **batch processing**.

In **real-time appending**, a set of information is sent to a server, which completes the match and returns the additional data elements – all within a fraction of a second. This is perfect for companies that need information on the spot, for one record at a time. Examples of applications for real-time appending include a retail store signing people up for a customer loyalty program or a customer call center instantly routing inbound calls.

Batch processing refers to processing a file with a consistent layout, whether it contains 100 or 100 million records, for simultaneous matching.

5 Manual vs. Automated Batch Appending

Batch processing can be completed manually or automatically. In a manual append, a processor will load the data into a database program like Oracle, write code containing match instructions and logic, and then execute the process. The processor needs to be sure that the input file is ready for append, with all necessary fields containing quality data (no unknown characters, extra delimiters, or foreign addresses, for example).

Manual appending leaves room for human error and variance, but it can be advantageous for complex, custom match jobs. A reputable company will add an extra quality control step to ensure that both system and human processor have completed all steps correctly. At one time, all data appends were manual. Because of the potential for error or slight variances in methodology between processors, larger companies have created **automated systems** that complete tasks from basic matching to complex, multi-step fuzzy matching projects.

For some automated batch systems, it's as simple as 1-2-3.

- 1.) Users load a file to an FTP site
- 2.) The vendor's system receives the file and performs the append
- 3.) Users receive an email when the file is processed and back in the FTP folder for pick up

How does the system know what to do? Many automated systems offer complete flexibility in initial configuration including input layout, match levels, output layout, suppressions and more. Once the append project is configured to meet client specifications, the automated process runs itself. However, each data provider offers slightly different options and will have their own, unique way of handling automated batch appends.

6 How is Data Matched?

For a forward phone append, input data would need to include, at the most basic level, a street address and a zip code. City and state should not be required; a quality data provider will have tables that infer city and state from zip code.

First and last names are important for tighter matching. Without a name, you may input an address and get the intended person's roommate's phone number returned. Some applications need the tightest level of matching; in skip tracing, for example, the goal might be to find one specific person, not their spouse or roommate.

In a reverse append, the required information is a lot less complicated: one field for exact matching, whether it is phone number, VIN, IP address, email address, or other input data field.

SAMPLE PHONE APPEND



7 The Technical Process for Appending Data

The first step is standardization. If your file lists an address as **123 Main St**, but the master database lists **123 Main Street**, there won't be an exact match – even though the addresses are clearly the same to human eyes.

Standardization converts every input record into a standard format. A good vendor database will already be standardized, so the input data needs to look the same. If the master database refers to every “Street” as “St” and every “Place” as “Pl,” then standardization is vital to ensure proper matching.

After the input data is standardized, it is compared to what the vendor has on their master file using a given set of match logic. Typically, the vendor will have varying levels of matching available. Some allow the client to choose match levels, while others are rigid about using all of their match levels – and while these companies flag the output data with match levels, clients are not free to choose which levels are acceptable for their particular application.

Matching at the individual level refers to a match using first name, last name, and address; the household level refers to a match at the last name and address.

There are two ways vendors perform data matching. In exact matching, the input data matches the vendor's data exactly and to the letter; in fuzzy matching, there are variances between the two data sets, but the vendor believes the match to be correct. For example, Bob Smith is a fuzzy name match against Robert Smythe.

There are several ways to match names, but the most commonly known matching algorithms are Soundex and Jaro-Winkler. Soundex looks at the phonetics of a name (Smith vs. Smythe), where Jaro-Winkler compares two data sets (in this case, names) and assigns a score based on the similarity of the two pieces of information. Data compilers with advanced capabilities will often go beyond these algorithms and create their own proprietary match logic.

After the match process takes place, the standardized records are typically changed back to the exact same format as the input file; standardization is for the benefit of the append process, not to improve deliverability (address standardization is a product available from many vendors – but it is separate from data appending).

8 16 Questions to Ask Your Provider Before Appending Data

Are you looking for the most matches or the most accurate matches?

This illustrates the classic quality vs. quantity dilemma. If a phone append rate is higher than expected, you should ask how recently the appended data has been validated and how tight the match logic was. If the append rate was lower, you may be able to loosen up the match logic to see how many more matches you might receive. Understanding this is particularly important when evaluating two or more data vendors to ensure equal comparisons.

Do you have the ability to match at different match level criteria such as individual, household, or address level?

Your provider should be able to tell you exactly how the information was matched. Common matching levels include Individual, Household, or Address level matching. Some vendors will allow you to customize this and others will hold more rigid guidelines. It is important to know ahead of time if multiple match levels are available and if your project can be customized according to the various levels.

What will the Input and Output data look like?

A client-focused data provider will allow for flexibility on both input and output layouts, as far as format (comma delimited, fixed) and fields allowed. Understanding specific input and output information from the beginning will save you time and resources long-term.

What algorithms are available for fuzzy matching?

As mentioned in the article above, there are specific matching algorithms available for fuzzy matching. Make sure you understand how your provider matches data to ensure a quality append.

What are all the data elements available for my specific append?

While you may be looking for a specific data element, it is likely there are other elements available that could help you further segment your consumer audience. Check with your provider to gain a full understanding of all elements available to maximize your data append project.

Where does the appended data come from?

The source, quality, and quantity of the underlying data are critically important to understand prior to selecting a data provider. Ask whether the data is from an original compiler or aggregated from several compilers. Find out about whether the data is from double-verified, trusted sources. Understand how the database is built and maintained so you can be confident in the accuracy of the data you receive.

Does your data have verification dates for the sourcing of the data?

Receiving verification dates within the appended data gives you the power to filter out older records or use all records, depending on the goal of your project. The more information you have about your data, the more accurate decisions you can make.

Does your data include confidence scores regarding the quality of the data?

Confidence scores can indicate levels of perceived accuracy of the data. Many providers will include these scores on appended records to provide you with their degree of confidence in the data provided. Confidence scores are another way for you to filter through appended information and make the best decisions for your initiatives.

Are you able to flag deceased individuals and flag/scrub Do Not Call, wireless records, or opt-out emails?

Flags within data can be very useful and provide significant cost savings. Removing deceased individuals from calling or mailing campaigns will result in immediate savings. Flagging or removing wireless records, phone numbers on the Do-Not-Call Registry, and opt-out email addresses can save significant money in fines or potential lawsuits.

Am I able to run an append test before I complete a new project?

When you are evaluating new vendors, ask if they provide opportunities for testing and evaluating their data append service. Many vendors will allow you to test both the data and the append process prior to committing to a larger project.

How many records are you able to process per hour?

Particularly critical if you have large projects, it is important to understand how robust your data provider's system is. A strong platform will be able to process millions of records per hour, giving you the best chance to meet your project deadlines.

Is your platform available 24 hours/day, 7 days per week?

Many data vendors who offer an automated batch processing platform or real-time API make these systems available 24 hours per day, 7 days per week. Many even include off-hours support to help with potential system challenges. Others may have different hours of service. It is important to know the availability and support of the batch processing system.

Do you have architectural redundancy supporting your delivery platforms?

Simply put, ask whether your data provider has a back-up plan if the platform is experiencing problems. Some data providers offer redundancy, meaning they have two identical systems processing the data on different servers. If one is not working, the other immediately takes over until the problem is resolved. For you, this means no project delays due to unforeseen technical challenges of the provider.

What measures do you take to ensure the safety and integrity of my data?

With much buzz about data breaches among the largest US companies, you need to understand how your data provider is securing your data. Security among database vendors is a constantly changing science, with the top providers always staying ahead of the curve.

Do you have the training in place to handle Personally Identifiable Information (PII)?

Taking security a step further, specific training and data handling procedures are necessary when a provider is handling sensitive Personally Identifiable Information (PII). PII protection processes go above and beyond standard data security practices. Your provider should be able to provide detailed information in their handling of PII data.

Does your provider offer a dedicated Account Management team to manage your project?

This may not seem like a critical question, but a good Account Management support team can make or break a project. Having a dedicated support team will ensure your project criteria and deadlines are met.

9 Conclusion

Data appending is a practice that has been utilized for years as a way to further identify, verify, and segment consumers. As technology and data collection has advanced, so has data appending. From lengthy manual appending to receiving appended data in subsecond time, the pace of processing has been driven by the speed of consumer demand. Data appending will continue to evolve as marketers and fraud specialists refine and improve processes for consumer validation and segmentation.

10 About Infutor Data Solutions

Infutor Data Solutions, Inc. is a privately-held corporation whose primary focus is creating and maintaining the most accurate and complete data assets available. Compiled from public and proprietary information, these data assets are highly regarded in the information industry and are used by many of the industry's largest information resellers, advertising agencies, fraud and analytics firms, large non-profit organizations, and retailers.

Our clients include four of the five largest marketing service providers/agencies in the United States, two of the three major credit bureaus, one of the country's largest non-profit organizations, and Fortune 1000 retailers, skip tracing, fraud and analytics providers. Infutor's solutions are used as a foundation for applications in:

- Marketing segmentation and targeting
- Digital identity validation and verification
- Direct marketing
- Retail consumer acquisition/retention
- Database management
- Fraud detection and management
- Inbound call center identification and call routing
- Enhanced/proprietary change of address processing (PCOA)
- Skip tracing
- Corporate data integration (CDI)
- Law enforcement searches

Data is accessible through on-site licensing, automated batch processing, real-time XML data transactions, and online query and order systems, all of which have been developed and maintained by Infutor.

11 More Information



2017 South Route 59
Plainfield, IL 60586



(312) 348-7900



sales@infutor.com



www.infutor.com



www.linkedin.com/company/infutor-data-solutions



www.facebook.com/pages/Infutor-Data-Solutions/190927214263550



plus.google.com/+Infutor

