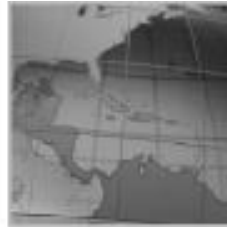


# Analysis



Business Development Services



February 2011

## Data 101: A Primer on Data Fundamentals

### Service Area

Business Development  
Strategies Production

[Comments or Questions?](#)

## Table of Contents

Introduction.....	3
What is Database Marketing?.....	3
Advantages of Database Marketing.....	3
What is Data and Where Does It Come From?.....	4
Types of Data .....	4
Anatomy of a Database .....	6
What to Do With Data.....	7
How to Use Data.....	7
Design .....	7
Logic.....	7
Putting It All Together .....	8
InfoTrends' Opinion.....	9
Drive and Deliver Better Data .....	9
Appendix: Glossary of Variable Data Printing Terminology .....	10
About the Author .....	12

## List of Figures

Figure 1: Does your company use the following types of customer data to create targeted messaging for a typical marketing campaign?.....	5
Figure 2: Basic Structure of a Database .....	6
Figure 3: A Data Driven Direct Mail Campaign for a Car Dealership .....	8

## Introduction

Data and database marketing has been the subject of conversation since the introduction of the digital press. With the recent economic recession and rise in marketing channels, data has taken a center stage position within many marketing campaigns. Despite all the talk about data driven marketing campaigns, there is still a lot of confusion surrounding data basics and fundamentals. Companies that do not have a solid grasp on data fundamentals will face more challenges presenting and instructing customers on the benefits of data driven marketing when compared to their counter parts who are already engaged in data driven print campaigns. Understanding the fundamentals of data and variable data terminology will help sales reps be more successful in selling digital print solutions while delivering a more valuable service to their clients. This document will address some of the core questions surrounding what qualifies as data, where it can be obtained, and how it can be used. We will also outline how entry-level projects are built and provide definitions for the key terminology used in database marketing.

## What is Database Marketing?

In short, database marketing is the collection of data to be used in a targeted marketing campaign. Database marketing is defined as a systematic approach to the gathering, consolidation, and processing of consumer data (for customers *and* potential customers) that is maintained in a company's database. Although databases have been used for customer data in traditional marketing for a long time (i.e., for direct mail), the database marketing approach is differentiated by the fact that much more consumer data is maintained, as well as that the data is processed and used in new and more sophisticated ways. Among other things, marketers use the data to learn more about customers, select target markets for specific campaigns (through customer segmentation), compare customers' value to the company, and provide more specialized offerings for customers.

## Advantages of Database Marketing

The advantages of data driven marketing are well documented with static forms of communication. According to the Direct Marketing Association's Response Rate Trend Report, the response rate for letter-sized direct mail sent to house lists in 2010 was on average 3.42%. By comparison, PODi.org highlights numerous case studies of simple data driven campaigns yielding response rates of 6%-8%, with highly-targeted campaigns yielding response rates in the double digits. Harnessing relevant information on recipients and using it to produce customized communication pieces that compel recipients to take action is the core of database marketing.

## What is Data and Where Does It Come From?

In short, data is any *information* (e.g., name, address, income level, age, life stage) that has been collected or can be bought that allows a marketer to better identify and communicate with their target audience.

There are two primary sources for data: internal data and external data. Internal data refers to data that a company may be collecting through a variety of direct customer interactions, such as credit card transactions, warranty inquiries, product registration, loyalty programs, and sweepstake giveaways. Internal data typically is gathered over a period of time and it contains personal, demographic, or purchasing information. It can be used to build a solid profile of a company's target audience.

External data is data that can be purchased through a variety of list providers, such as Accudata or USAData. External data is used to augment existing internal data. For example, a simple analysis of internal data may reveal that the typical buyer of a product is male, age 25-35, and lives in the zip code of 12345. By purchasing a list and augmenting the internal database to include recipients that match that a specific demographic, marketers are able to increase the number of recipients that a campaign will be targeted towards.

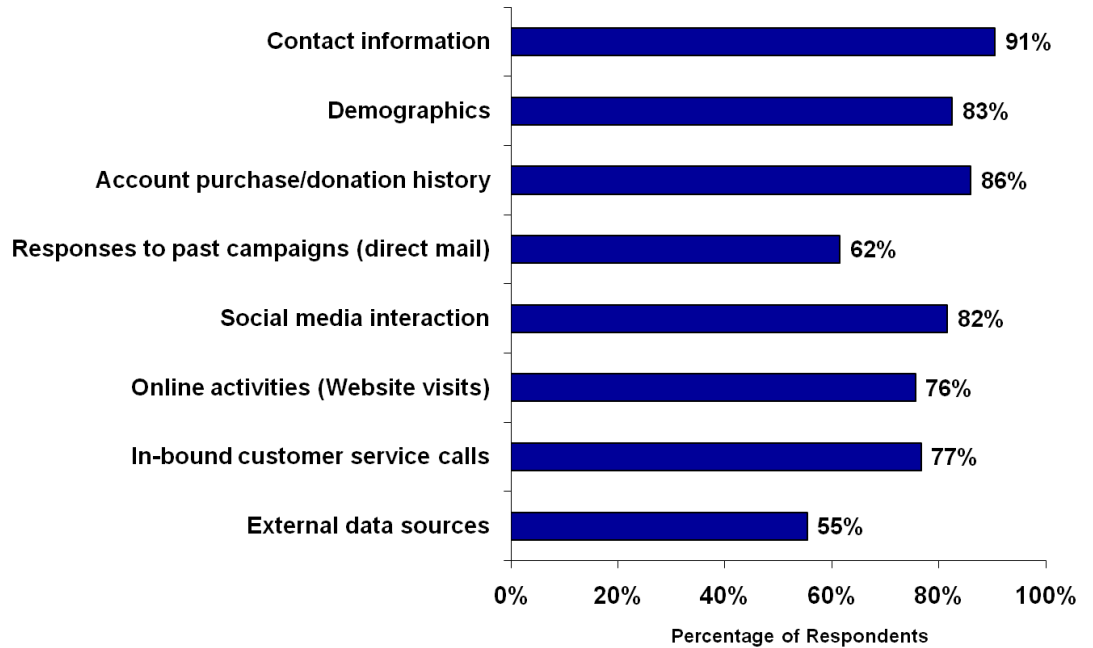
Regardless of the source, the data must contain correct information, such as the spelling of a person's name or e-mail address as well as current and correct mailing addresses. In an ideal situation, it should contain personal, demographic, or purchasing information that can be used to better target each recipient on a more relevant basis. Nevertheless, some very good campaigns have been created by using name and address only, which is typically what most people start out with.

## Types of Data

In a 2010 InfoTrends study titled *The Cross Media Direct Marketing Study*, respondents were asked which of the following types of data was used in targeted messaging campaigns. As expected, contact information was ranked as the most used type of data used in a typical marketing campaign.

Nearly 83% of all the marketers surveyed indicated that they store contact information within their databases. Meanwhile, 90% of these marketers are using this information to create targeted campaign messages. Basic contact information can be used in simple direct mail campaigns whereby the contact information is addressed onto a direct mail piece and may be used to personalize the message by including the recipient's name in the offer.

**Figure 1: Does your company use the following types of customer data to create targeted messaging for a typical marketing campaign?**



*InfoTrends: The Cross Media Direct Marketing Study 2010*

Traditionally, contact information is the only data used in a marketing campaign, data such as demographics (e.g., age, income, and marital status), account purchase/donation history, and social media interaction can be leveraged to create a more targeted marketing campaign.

## Anatomy of a Database

Figure 2: Basic Structure of a Database

	A	B	C	D	E	F	G	H
1	SALUTATION	FIRSTNAME	FULLNAME	ADDRESS	CITY	ST	ZIP	BARCODE
2	Mrs.	Helen	Helen Shirley Burgess	3364 Raes Creek Rd	Marietta	GA	30008-5702	!300085702648
3	Mr.	Lester	Lester Earl Smith	12541 Alcovy Rd	Covington	GA	30014-1031	!300141031413
4	Mrs.	Deborah	Deborah Jean Nelson	75 Five Oaks Dr	Covington	GA	30014-5715	!300145715755
5	Mrs.	Yeshika	Yeshika Marie Anderson	45 Dogwood Ter	Covington	GA	30016-5330	!300165330452
6	Mrs.	Mercedes	Mercedes Ileana Dixon	3391 Warbler Dr	Decatur	GA	30034-4230	!300344230917
7	Mrs.	Aisha	Aisha Natalie Smith Danzy	14207 Saint James Dr	Conyers	GA	30094-1219	!300941219071
8	Mrs.	Cynthia	Cynthia A. Nelson	2636 High St SW	Conyers	GA	30094-6843	!300946843366
9	Mrs.	Julie	Julie White Fancher	2212 Independence Dr SW	Conyers	GA	30094-6846	!300946846126
10	Mr.	Timothy	Timothy Kirk Driskell	200 S Main St	Alpharetta	GA	30004-1914	!300041914004
11	Mrs.	Rhonda	Rhonda Duffy	15275 Highgrove Rd	Alpharetta	GA	30004-3192	!300043192757
12	Mr.	Roger	Roger Layne Hopper	185 Dorris Rd	Alpharetta	GA	30004-3475	!300043475853
13	Mrs.	Connie	Connie Sue Abbott	3872 Avensoyng Village Cir	Alpharetta	GA	30004-4903	!300044903726
14	Mr.	Anthony	Anthony S. Fullam	333 Fernbank Ct	Alpharetta	GA	30004-5668	!300045668332
15	Mrs.	Mary	Mary Ann Brann	1635 Eversedge Dr	Alpharetta	GA	30004-7133	!300047133356
16	Mr.	Juan	Juan Carlos Castro	1625 Trellis King	Alpharetta	GA	30004-8447	!300048447252
17	Mr.	Samuel	Samuel W. Baird	2965 Manor Bridge Dr	Alpharetta	GA	30004-8813	!300048813654
18	Mrs.	Evelyn	Evelyn Gilder	7160 Harbour Lndg	Alpharetta	GA	30005-6990	!300056990605
19	Mrs.	Virginia	Virginia O. Diaz	772 Win West King	Auburn	GA	30011-2260	!300112260723
20	Mrs.	Betsy	Betsy Christian Gerald	2435 Staffordshire SE	Conyers	GA	30013-6478	!300136478357
21	Mrs.	Teresa	Teresa C. Jones	824 Fairview Club Cir	Decula	GA	30019-3119	!300193119243
22	Mrs.	Andreina	Andreina Dewendt	110 Caswyck Trce	Alpharetta	GA	30022-2689	!300222689102
23	Mr.	James	James Michael Settle	315 Ankonian Dr	Alpharetta	GA	30022-5117	!300225117155
24	Mr.	Christopher	Christopher Jon Everett	10570 Centennial Dr	Alpharetta	GA	30022-5767	!300225767701
25	Mrs.	Heidi	Heidi Ann Overton	10180 Groomsbridge Rd	Alpharetta	GA	30022-6098	!300226098801
26	Mr.	Robert	Robert L. McCormick	740 Mallory Manor Ct	Alpharetta	GA	30022-6439	!300226439406
27	Mr.	Michael	Michael Edward Bancroft	10520 Willow View Cir	Alpharetta	GA	30022-6550	!300226550202
28	Mr.	Carlos	Carlos Saldana	3065 Rivermont Pkwy	Alpharetta	GA	30022-6808	!300226808655

As shown in the Figure above, the most basic forms of data can come in the form of MS Excel (.xls), dBase (.dbf), FoxPro, and Delimited (.csv) file formats. Whatever the file format, it is important to know how the information is cataloged or organized. Basic organization of a database is done in the header of every database. The header is the first row or line of information within a database and it provides information about what is contained in the columns underneath. First Name, Last Name, Mailing Address 1, Mailing Address 2, City, State, and Zip Code are examples of basic header information in a simple mailing address database. More complex databases may have fields such as Gender, Age, or Income, which could be used in a more complex variable data campaign to trigger the change of images and create a greater level of personalization within a marketing piece.

## What to Do With Data

As discussed earlier, regardless of where the data is coming from, it must contain accurate data. There is little that can be done with inaccurate data. When accurate data has been gathered, it must then be “cleansed”.

Cleaning the data also involves going through the database file and making sure that upper and lower case letters are being used appropriately and eliminating any records that may be duplicated; this process is referred to as “de-duping.” De-duping the database file is an important component of the cleansing process since duplicate records can drive the cost of the campaign up due to extra postage, while also hindering its effectiveness. A prospective customer might pass on an offer if he/she receives multiple messages.

Once the data is de-duped, it needs to be optimized for postal distribution by sorting the records in zip code order. Optimization of a database is an important step prior to mailing. The inherent personalization of the piece requires that it reach the intended recipient to have the maximum impact. Optimization for postal distribution can occur through a coded accuracy support system (CASS) process performed by any mail service provider that is using a CASS-certified address matching software that standardizes the address according to USPS address guidelines. CASS processed direct mail also results in cost savings as it removes and updates mailing addresses as needed.

## How to Use Data

Once the data has been cleansed, it then has to be put to use. InfoTrends identifies two basic steps for developing a piece that incorporates a database, and they are Design and Logic.

### Design

The design of a variable data piece takes into consideration where certain components, such as personalized names and addresses, text, images, or signatures, will reside within a printed piece. These design fields are then populated with logic or conditional logic statements, which help determine what information is the populated into that field.

### Logic

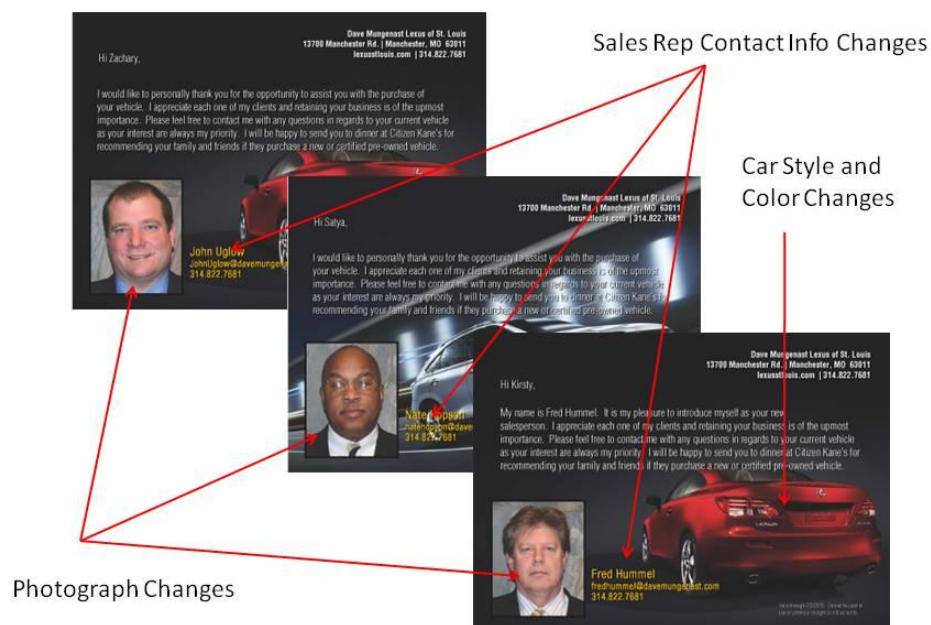
Logic is the business rules for a campaign and dictates how the images and text vary within the document. At its most basic level, a design field can have a simple logic statement that indicates which type of data, as indicated at the header level of the database, will be placed within that field. For example, a design field used to house address information could have a basic logic associated with it where a header titled FirstName will populate that field with the first name data housed in the column below the header.

Conditional logic is a series of If /Then statements that dictate which variable component will change and what it will become. In its most basic form, the data that resides in the

columns underneath the headers of the database is used to trigger a specific variable element within the design of the piece, which is based on the conditional logic associated with the field of design. For example, if a database header had a header titled “Color” with various colors populating the fields of the column below it, a simple conditional logic statement could be “IF RED Then insert ‘RedBallImage,’ ELSE insert ‘GreenBallImage.’” Most entry-level variable data solutions offer simple scripting languages or wizards for generating simple conditional logic.

## Putting It All Together

**Figure 3: A Data Driven Direct Mail Campaign for a Car Dealership**



The above Figure shows a simple direct mail postcard developed by Gateway Digital Press for an auto dealership located St. Louis, MO. By leveraging a database of cars sold over the past year, Gateway Digital Press created this personalized direct mail campaign for the dealership. The dealership’s database contained information that included their customer’s name, contact information, their assigned sales rep, and the details about the car the customer purchased (including make, model, color, and features). Data driven marketing campaigns such as this provide a level of sophistication that is not achievable through static campaigns, and these are a critical component to increasing the return on investment (ROI) of any marketing budget.

## InfoTrends' Opinion

### Drive and Deliver Better Data

The time to start engaging customers and working on developing campaigns that are measurable, targeted, and personal is *now*. Data driven marketing has proven itself as a means to deliver a higher ROI. Information opens up new ways to market more products and services for consumer and business-to-business marketers. Those who understand and master the fundamentals of where data comes from, how it can be used, and the basics of delivering a data driven marketing campaign can help their customers improve the efficiency and effectiveness of their marketing spend.

This material is prepared specifically for clients of InfoTrends, Inc. The opinions expressed represent our interpretation and analysis of information generally available to the public or released by responsible individuals in the subject companies. We believe that the sources of information on which our material is based are reliable and we have applied our best professional judgment to the data obtained

## Appendix: Glossary of Variable Data Printing Terminology

**Augment** - To increase, make larger, or supplement existing data with similar data through external data providers.

**Coded Accuracy Support System (CASS)** - A USPS approved process that improves address deliverability and standardization.

**Cleansing** - the act of detecting and correcting (or removing) corrupt or inaccurate records from a record set, Table, or database.

**Conditional Logic** – A series of IF, THEN, and ELSE statements that are used to dictate which assets are associated with a given variable.

**Customization** (See Personalization)

**Data** - Groups of information that represent the qualitative or quantitative attributes of a variable or set of variables. Within VDP, this could be name, address, favorite color, or purchase information.

**Database Marketing** - The use of customer profiles contained in a database to market to customers.

**De-duping/Purge** - The process of eliminating duplicates and/or unwanted names and addresses from one or more lists.

**External Data** - Data that is not in the current database and may be acquired through third-party sources, such as Accudata or USAData.

**Frequency Marketing** - A marketing program that recognizes and rewards customers based on their purchasing behavior (also known as Loyalty Programs).

**Internal Data** - Data generated by an organization's operations, such as sales and purchase orders or inventory transactions instead of being provided from a third-party study or database.

**Logic** - The rules associated with the data in a database that typically encode business policies. In VDP, logic controls where the data and assets are positioned within a printed piece.

**On-Demand** - Printing output only when it is needed.

**One-to-One Marketing** - A marketing process through which a business identifies its individual customers, differentiates among those individuals, interacts with customers and records responses, as well as customizes communications for individual consumers.

**Personalization / Personalized** - Customizing a document by varying the text, graphics, and layout to meet the needs of an individual consumer.

**Response Rate** - In a direct mail campaign, the percentage of recipients who responded. Response rate is a vital measurement of the success of any such campaign.

**Variable Data Printing (VDP)** - A digital print run where each printed page is somewhat different, with the variations usually determined by relating page content to customer information in a database (also known as Variable Information Printing or VIP)

**Versioned** - A form of short-run printing, where different versions of a document may go to different geographical areas or people with different income levels. Versioning can be done on a press that does not have VDP capability because hundreds or thousands of identical sheets are printed for each “version” of the job.

**Software Wizard** - A software wizard is a user interface element that presents a user with a sequence of dialog boxes that lead the user through a series of well-defined steps.

## About the Author



**Matt Rees**

Senior Consultant

[matt\\_rees@infotrends.com](mailto:matt_rees@infotrends.com)

+1 781 616 2100 ext. 145

Matt Rees is a Senior Consultant for InfoTrends' Business Development Services. In this role, Matt will support service initiatives by assisting customers in growing and developing new market opportunities. Mr. Rees has a broad range of experience within the photo and production printing/software space.

[Comments or Questions?](#)