

## **Why Data and Technology is Difficult The “Data and Technology Problem” in Marketing**

Many Chief Marketing Officers, Vice-Presidents of Marketing, and marketing managers share a common lament. The chorus is something along the lines of, “We supposedly have all this data, but we are not DOING anything with it!”

The reasons for the problem are many. In order to fully understand the problem one must examine various aspects of data and technology. The solution to the problem is not trivial. However, armed with the proper knowledge, resources and tools, marketers can better use both data and technology to improve their marketing.

This paper examines various aspects of the Data/Technology problem in an attempt to illuminate why these areas present difficulties. Understanding the problem is, of course, the first step in implementing a solution.

### **The “Data and Technology Problem” - Introduction**

Data and Technology are a problem because they both are complicated. They both often involve “new” aspects that have not yet been learned, and can be complicated and confusing even when understood. Fully understanding WHY there is a problem, or more accurately why there are actually many aspects of the problem will help clarify why the solution is not trivial.

The “Data and Technology Problem” lies in the complexity. The various aspects of the problem include:

- Data
- Data Integration
- Analysis of Data
- Using Data for Marketing
- People
- Process
- Software

To help understand the problem, and therefore the solution, each of these aspects is now examined.

### **Data**

The complexity of the data itself is part of the problem. It may seem like it should be simple, but examination of what exists can quickly reveal the complexity. Data can come from many sources. Data can be kept in many places. All of these things add to the complexity involved with the data, and to the problem.

Data sources can include the “customer house file”, coupon data, email data, web sales data, telemarketing data (inbound and outbound), retail Point of Sale (POS) data, click stream data, survey response data, third party data (e.g. Acxiom, Dun and Bradstreet) and others.

**Potential Data Sources**

House File	Surveys	Coupons	Email
Aggregate Sales	Retail Sales (POS)	eCommerce	Inbound Telemarketing
Outbound Telemarketing	Web Click Stream	Third Parties (e.g. D&B)	Etc.

As an example, a full fledged marketing campaign for a consumer packaged good (CPG) company can easily have ALL of these data sources, and each of the sources has its’ own inherent set of codes and complexity. The UPC code data and sales data from a grocery store’s transactional log can easily run into terra-byte file size. The same chocolate chip cookie can be sold in different forms (“Regular” and “Bits”) and different colors (specials for Halloween or “Holiday”) and of course different package sizes. Each of these items has a different code, which has to be assigned and tracked. Similar data source complexity arises in many other industries.

Data storage problems are also common. Often, if not always, a company’s data is kept in many different repositories. Each source mentioned above can actually have it’s own data repository or warehouse. A famous software company in Silicon Valley has at least seventeen (17) different data warehouses, all of which, of course, are completely discrete and independent. No data is shared between any of the warehouses.

**Data Integration**

With so many potential data sources available it is clear why data integration can be a problem. Combining data from different sources can be difficult and sometimes literally impossible.

Part of the complexity of data integration can lie in the ownership of the data. For example, if a company wants to combine e-commerce or web site data with catalog sales data it may require two different information technology (IT) groups to work together. In the case of a company with seventeen different warehouses, complete integration would require the cooperation of many different groups or factions.

Furthering the complication of integration is that each of the sources may not have a “link” between them. I.e., there is no way to methodically match the data from “Bob Smith” at “123 Elm Street” in the e-commerce database to the “B. Smith” at “123 Elm” in the catalog database. Often the databases will have a unique identification number for each customer, but since the systems were developed independently the numbers for the same customer in the two different databases are different.

Additional complexity in data integration is that the work often requires experts in certain software or database programs. Additionally, the data is often stored in a manner that makes it

easier for the storage of the data, but difficult for actual use of the data. This can result from data warehouse designers or IT experts not understanding how the data is going to be used, but knowing that they have to store “all of it”. The origin of this lack of understanding can lie in either the IT department or the marketing department.

Despite the data integration problems and complexities, it is imperative that serious marketing companies integrate data from different channels in order to maximize the return on their marketing efforts. In a clothing company, for example, the return on an additional catalog mailing cannot be fully measured unless the data from the e-commerce channel is combined with the catalog data. Some retailers have seen an additional 20% in sales from their e-commerce channel in response to a catalog mailing.

In order to optimize the allocation of resources across channels, it is necessary to be able to measure the results across channels to the degree possible and necessary. This requires integration of data from multiple sources and channels.

### **Analysis of Data**

To fully utilize the available data, marketers must employ advanced statistical techniques. Part of the “Data and Technology Problem” in analysis is that many traditional marketing analysts do not have the required training or knowledge that is needed to properly execute the more advanced analysis. While they can often analyze the data using basic (and very useful) reports (E.g., “Sales by Age Group”), more in-depth, simultaneous multiple-variable analysis will yield stronger results. Understanding when to use which technique is not trivial. And for each technique, knowing when to use which specific algorithm is also not trivial.

Furthering the “Data and Technology Problem” in the analysis area is that often those who do know how to execute the techniques do not fully understand the marketing problems they are trying to solve. Instead of focusing on the solutions to marketing problems, they focus on using the “sexiest” technique (that is a term that is *actually* used), whether or not it solves the marketing problem. Sometimes this results in “A canon being used to kill a mouse”, and other times “problems” are solved that are of no use to the marketer.

Some marketing areas in which the more advanced statistical methods should be used and the appropriate technique(s) are:

*Targeting:* There are a number of techniques that can be used when the marketing problem involves trying to target a specific type of customer. Common examples of this involve identifying most (or least) valuable customers and identifying customers likely to leave a program (aka “churn” or “attrition”). Other examples involve knowing which geographic area should be targeted. The statistical names for the techniques that could be employed include: Predictive Modeling, including Multiple Regression, Neural Net algorithms and Chi-Squared Automatic Interaction Detection (CHAID) models, and Factor Analysis.

*Segmentation:* Segmentation of customers can be based on many different variables, including amount spent, types of products purchased, geographic location, etc.. All of these possibilities

increase the complexity of the analysis. When the marketer is trying to develop a segmentation strategy questions can include: Which products are sold together? How many different types of customers exist? What natural groupings of customers exist? What products do they purchase? What services do they use? Who are my low, medium, and high value customers?

Appropriate statistical techniques for such analysis include: Cluster Analysis, CHAID, as well as predictive modeling techniques.

*Reporting:* While not really a statistical technique, incorporating sales data and data from multiple channels is an area that can be improved. The data can then be used to calculate improved total ROI measures. Reports involving tests (“Is Offer ‘A’ really better than Offer ‘B’?”) can be improved using statistical significance tests. “T-tests”, “significance tests” or “Chi-squared tests” are statistical terms used to describe these types of tests.

*Designing In-Market Tests:* If a marketer wants to test many offers, pieces of creative, colors, approaches, different text, etc., to many different customer segments and/or geographical regions, he will quickly find that he has many combinations of those variables to test. In order to maximize efficiency and effectiveness of the tests, the marketer should be using a test that is specifically designed to accomplish those goals.

While some companies do know how to answer the sample size questions that arise, most do not know about proper test design techniques. The statistical terms that are applicable include: Experimental Design, Design of Experiments, Taguchi Designs, as well as sample size requirements. All of these are techniques to maximize information gathered while minimizing the required sample size in the test phase of a marketing campaign. They can also be used as a part of ongoing testing.

*Product Recommendation:* Amazon’s “If you like this book you might consider...” is the most well known example of data based product recommendation. The “Beer and Diapers” correlation is another example that falls into this realm of analysis. Product recommendation techniques should be used for in-store and online purposes, as well as in customizing content for emails and catalogs. Statistical techniques that are appropriate in this area include: Collaborative Filtering, Cluster Analysis, and Regression

The marketing application and the appropriate statistical techniques are summarized in the following tables:

<b>Marketing Application</b>	Multiple Regression	Logistic Regression	Neural Net	Factor Analysis	Cluster Analysis	CHAID
Targeting	X	X	X	X		
Segmentation	X			X	X	X
Reporting						
Designing In-Market Tests						
Product Recommendation	X				X	

<b>Marketing Application</b>	Categorical Modeling	Reporting	Design of Experiments	Taguchi Designs	Sample Size	Collaborative Filtering
Targeting	X					
Segmentation	X	X				
Reporting		X				
Designing In-Market Tests			X	X	X	
Product Recommendation						X

All of the techniques require the use of appropriate software such as Statistical Analysis Software (SAS) or the Statistical Package for Social Sciences (SPSS). Some basic versions of the required algorithms have been incorporated into some campaign management/marketing software packages including those from E.piphany and Blue Martini.

The complexities of the areas of analysis as well as the complexity of the software are why analysis itself can be difficult for marketers.

### Using Data for Marketing

To take advantage of the available data marketers must have clear objectives and understand how to use the data and the analysis. Part of the complexity in this area involves knowing which analysis technique is appropriate, and what is and how to use the “output” of the analysis. Most marketing analysts have not been trained in or exposed to the full set of required tools.

The ramifications of not knowing how to use the data in marketing include improper planning, improper data acquisition, invalid marketing conclusions and incorrect marketing decisions, and ultimately sub-optimal performance.

Those who are conducting the advanced analysis often do not know the marketing objectives that apply to the data that is being analyzed. Those who are setting up the data collection process also do not always know what is trying to be accomplished. This “Profound Knowledge”, as quality guru/statistician Dr. Deming called it, can be the difference between a successful program and an unsuccessful one.

Marketers who want to optimize their campaigns must understand how data and analysis can be used for **precise targeting** and **economic analysis**. They must understand how to use “Media Mix Modeling” or Media Mix Analysis in the **planning** stages of campaigns. They need to optimize the **testing** of their new offers, products, etc., using Experimental Design. They need to use collaborative filtering, or something similar, to take advantage of data and technology to make **product recommendations** to customers. All of these are areas in which data can and should be used in marketing.

Unfortunately, a lack of knowledge or resources in these areas can cause a marketing program to be sub-optimal.

## **People**

Part of the “Data and Technology” problem can be attributed to management issues and employees. The “people” portion of the problem often involves lack of knowledge, “turf battles”, and lack of resources.

As mentioned earlier, if a data warehouse designer or engineer does not understand how certain data is going to be used, or why certain data is going to be collected, he may not adequately prepare the process of data collection or storage of the data. This is often a part of the “Data and Technology” problem. In essence the IT person is not a marketer, which may result in the “IT” problems for the marketer.

In another area, business analysts will not be able to choose or apply the appropriate analytical techniques if they have not been trained in or have not used the techniques. In essence, there is no way for them to know “What they do not know”, and they will not have the capability of recommending or implementing the correct technique. In addition the marketing analyst may not be trained extensively in IT methods or needs, which can cause problems in those areas.

The “turf battle” is also a common part of the problem. In the case of the software company that has seventeen different data warehouses, it should be self-evident that many of those warehouses have different “owners”. This can result in “office politics” situations where people do not want to cooperate. Even if it benefits the company, if the cooperation would diminish the importance of, or perhaps eliminate their specific warehouse, then the involved parties do not want to cooperate.

In addition there is sometimes a resource issue. Often the IT group may not have resources to handle the workload that is being requested of them. Sometimes the requests made of them are quite challenging, and they may not see the importance of the request. This can result in the request being ignored or the response being delayed, in hopes that the request might disappear.

In defense of IT and data warehouse groups (which are often a part of “IT”), sometimes the questions being asked of them are unclear (at best) or preposterous and impossible. This stems from a lack of knowledge on the marketer’s side regarding data capabilities or technological capabilities or requirements.

All of these issues contribute to the “people” portion of the “Data and Technology” problem.

## **Process**

The process portion of the “Data and Technology Problem” revolves around the need for precise, systemic processes to be put in place in order to facilitate the transfer of needed data, and to assure that current, applicable, and up-to-date data is used. Often these processes do not exist, and creating them requires time and expert resources.

The processes need to be engineered by a team of people that understand the available data sources, the software packages and database programs that are involved, and/or a programming

language, as well as the marketing objectives. The processes must be designed so that the result or “output” is acceptable for and useful to the marketing campaign. The output could be a report, data to be used for a report, a list of customers for a direct mail campaign, or many other things.

The reason process is a problem is that it takes planning, testing and proper implementation, as well as subject matter expertise to make everything work. None of these items are trivial matters. The proper resources must be available, knowledge of marketing objectives must be shared and understood, and the “heavy lifting” involved with data design, transfer, analysis, and delivery must all be properly executed in the process in order for the marketing objectives to be achieved.

The complexity of the data sources, combined with the engineering needed to construct required processes increases the size of the “Data and Technology” problem. However, the benefits of implementing properly designed processes exist in the form of greater success of marketing programs.

## **Software**

Another issue of the “Data and Technology Problem” is the competing claims of different software companies and application service providers (ASP’s) . The “magic bullet” software offering that will integrate all of the data and solve all of the problems does not exist. Yet, these types of claims are made in pitches and presentations every day.

While some software is certainly useful, the purchase of a “CRM package” alone will not solve a marketer’s problems. Software might provide a platform from which improvements can be made, but issues such as data integration (which is non-trivial, despite what claims are made) and others mentioned here will still exist. Unfortunately the claims of the ASP’s and software vendors do not mention this fact.

## **A Hint at the “Data and Technology Problem” Solution**

The above detailing of the “Data and Technology Problem” is given to help clarify why many marketers are saying, “We supposedly have all this data, but we are not DOING anything with it!” The problem lies in the complexity of all of the areas mentioned above.

Fortunately there are some definitive actions that marketers can take to improve their situation and their use of data for marketing. The information presented here sheds light on some of the areas that need improvement. While understanding the problem certainly does not solve the problem, it is a first step that illuminates the direction marketers must head.

Hopefully it is clear that the “Data and Technology Problem” in marketing can be improved through clarifications and sharing of marketing objectives, assessment and improvement of data and processes, improvement of skills, and the acquisition of appropriate resources (including specialists and software).