WHAT ABOUT THE DATA?
ORIGINALLY WRITTEN IN 2004 BY JOHN TAYLOR & JONATHAN LINDSEY (DECEASED)
UPDATED BY JOHN TAYLOR IN JANUARY, 2017

It seems that it was only 20 years ago that we felt that the only data we needed in order to effectively solicit donors and recruit volunteers was where they lived, where the worked, and (maybe) what they were involved in when they were as student at your institution. Gosh, that was just 20 years ago!

Competition has changed all that. We are not only competing with other institutions (one school's high-schooler is another schools undergraduate alumnus is another schools graduate degree-holder is another schools doctorate holder is another schools professor!), but we are competing with nearly 1,000,000 additional worthwhile nonprofit organizations. And that's just in the United States, and does not fully address the temporary shift in giving we see in response to natural disasters – or worse. September 11th changed many donors outlook toward charitable giving. In fact, we saw the first decline in giving to education the year following the attack on America. But what many people may have missed when reading about this decline in the press is that giving to religious and disaster relief organizations actually increased.

We are also competing with life in general. There are many more two-income families today – necessary in many cases just to make ends meet. And when we are not working, we are driving our children to soccer matches & ballet classes and the like. And, if we can find a spare moment, many of us are volunteering our time – in increasing numbers – to local nonprofit organizations that serve our children, families, and communities.

The reality is that all of these endeavors are good ones, certainly in the minds and eyes of those of us engaged in them. So with all of these activities going on around us it has become imperative for the conduct of a successful fund-raising campaign to understand as much about our friends, alumni, and donors as possible so that we can identify that which may encourage them to place a higher priority on our organizations when they are considering how and where to dispense with what is seen to be a decrease in discretionary assets, but an increase in funding opportunities.

With thanks to our friends in the software industry we store vastly more data on our constituents than ever before. So much so, in fact, that many organizations are facing data-saturation problems and are contemplating what data they should purge when just a few years ago they were asking what data they should obtain!

So, what about the data? Or, more precisely, what is data? There are many definitions:

- Factual information, especially information organized for analysis or used to reason or make decisions.
- Numerical or other information represented in a form suitable for processing by computer.

- Something given or admitted; a fact or principle granted; that upon which an inference or an argument is based; -- used chiefly in the plural.
- A collection of facts from which conclusions may be drawn; "statistical data".

Our problem, however, is that while we may gather data it's not data that we need – it is **information** we seek with which we can make informed decisions:

> "Data on its own has no meaning, only when interpreted by some kind of data processing system does it take on meaning and become information.

> "People or computers can find patterns in data to perceive information, and information can be used to enhance knowledge. Since knowledge is prerequisite to wisdom, we always want more data and information. But, as modern societies verge on information overload, we especially need better ways to find patterns."[1]

To paraphrase a familiar quote, the devil is in the data. We have saturated our databases with data. After all, that's what a database is for – storing data. Yet by gathering endless amounts of data we have created for ourselves the age-old forest and tree problem. It is almost impossible for us to identify the golden nugget that might result in our realizing a major contribution. Thus, we tend to internally revert to using our databases for mass-mailings, conducting telethons, and generating fund-raising reports. All worthwhile tasks, to be sure. But the return on investment many not be that great. Where the real money is, literally and figuratively, is in mining the data that you have.

An individual at your institution can certainly do data mining. But rarely is that process thorough enough to find patterns across the database. More often than not the 1:1 data-mining activity is just that. One person evaluating one person. We call this prospect research. But effective prospect research relies, first, on the identification of a likely prospect. Unless they self-identify themselves to you, you are going to have to find them on your own. However, without a robust analytic tool, strong IT staff, and talented analysts on staff to perform these functions you are going to have to use "outsiders."

Some of these "outsiders," will come in to your organization and look at the data you already have pertaining to your donors. They will conduct data-informed forecasts for gift receipts and investment levels to achieve continued growth over a period of time. However, this requires looking back at an institution's fundraising performance over up to 30 years. This process absolutely works! But it works best when you know you have good data – and lots of it, and have a fairly robust budget!

So, given the problems of too much data, uncertainty regarding the quality of data you have, and limited budgets, we must turn to vendors who specialize in this arena for help. But because so many of us have very limited budgets we must take care with how we

---

[1] *The Free On-line Dictionary of Computing, © 1993-2003 Denis Howe*

spend those precious dollars. We can ill-afford to ship off thousands of data-elements that may or may not be necessary for accurate analysis knowing that every element we require the vendor to sort through is going to cost us. So before taking this path we must evaluate the answers to 3 critical questions:

1. What data is essential for analysis?

    The simple answer to this question is, "Only that data which is necessary for analysis!" If you were going to take a cross-country trip and wanted to find the most direct, efficient, and quickest route, would it makes sense to review Delorme Topo Maps©, showing all of the back roads and side streets for every state in a 60-page book? While the data may be useful **at some point** (like when I don't listen to my wife's suggestion and get us lost), all you really need at the outset is a map of the interstate system (and not even that – a cell phone with a GPS app will work fine!). The same is true when working with outside vendors. What your overall objective is will dictate the amount of data you will need to send and they will need to analyze. The key here is to:

    -   Document your objective
    -   Seek input from the vendor as to what **minimum** data is required by them to achieve that objective
    -   Confirm that you have the minimum data necessary for the analysis
    -   Negotiate with the vendor regarding additional data needs for achieving other objectives not originally contemplated (remember, the vendor does this for a living – they may know something you do not!) – see the next question

2. What is the optimum number of elements to ship off for analysis?

    Much depends on the objectives you and the vendor determined are appropriate to achieve the desired outcome (and within budget!). But there is more to it than that. Remember that planned cross-country trip we discussed? Would you really need to pack all 50 Topo Maps© in your car in case you got lost? Not hardly. Just as you know there are states you will not remotely get close to on your trip, there are many individuals you will not want to include in the data you ship off for mining – lost individual, international prospects, current students, etc. Additionally, depending on the vendor and their reputation for protection of data (or lack thereof), you may feel that sending some sensitive information is not appropriate. And remember, the vendor is going to charge (typically) on the volume of data given and reviewed. So they have an interest in your giving them all that there is! But that may not be necessary. Check with peer institutions to see what they have done. It may be possible to do more with less. See question #3.

3. How can you avoid choosing the least valuable variables?

Shakespeare said, "Know thyself and to thine own self be true." This statement could not be more to the point in the world of data mining. **You** know your institutional culture and history. **You** know the context in which the data is going through analysis. Your vendor does not. Vendors tend to have one model that works best in most situations. And that is perfectly acceptable – the same can be said for your organization. That it why it is imperative that you must help the vendor know you – and you know the vendor. A pat formula will not always work except, perhaps, in the most rudimentary of analyses. Understand the power of emotional-based responses pre-existing on your database and link those with the new information provided to you. And most importantly, weight the value of various variables in order to properly contextualize the information now in hand.

Some refer to data mining as a science, especially when you hear terms like "predictive modeling" and "forensic analysis." But the truth is that data mining is an art. It is the art of knowing what is important and what is not. It is the art of negotiating with information providers for precisely what it is you which to achieve. But most importantly it is the art of painting an accurate picture as a result of your understanding how the information you have makes the most sense in the context in which it is observed.

Enjoy your road trip. And remember that some of the most wonderful journeys are those you had never anticipated taking.