

Paper 321-2011

## **Taxonomies, Tone and Tweets – Where text analytics and social marketing collide**

John Bastone, SAS Institute, Tampa, FL, USA

### **ABSTRACT**

Text analytics is an established science. Social marketing has exploded in popularity. The latter phenomenon has generated an enormous amount of interest in the former. As the software landscape is quickly adapting to the needs of next year's business, there are some core issues that companies must wrestle with in embarking on any social media analytics engagement:

- Listening – By growing bigger ears to hear beyond the usual sites, to pull in relevant conversations about your business.
- Processing – To ensure your organization understands what it is hearing.
- Responding – Both literally and procedurally, to ensure the organization adapts and learns from direct customer feedback.

This paper will explore many different aspects of the social media analytics discipline, spanning the evolution of listening techniques, the fundamentals of word taxonomies and sentiment analysis that get discern positive and negative tone, and recommendations on how to handle tweets and any other form of real-time communications.

## REMEMBER MOSIAC? DON'T WORRY, NEITHER DO A BILLION OTHER'S CURRENTLY ON FACEBOOK

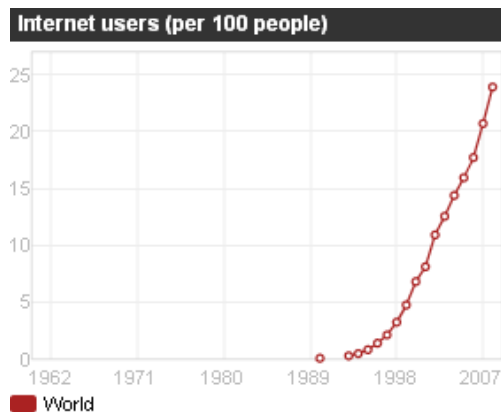
Look up the word Ubiquitous on dictionary.com, and you'll find the following definition:

*"existing or being everywhere, especially at the same time".*

Under that definition, it could be argued that the internet is still not that. According to stats compiled by <http://www.internetworldstats.com>, close to 3 out of every 10 human beings on this planet is an internet user, or just shy of 2 Billion people globally. But to put things in perspective, Greenpeace estimates that only 75% of the globe even has access to electricity. So while not ubiquitous in the purest sense of the word, its impact is profound, particular among us who have enough voltage coursing through our communities to be able to read this text on our iPhones, laptops and tablets.

Consider life back in 1993. Seinfeld was in its 5<sup>th</sup> season as one of the highest rated television shows in the United States. Bill Clinton was starting his first of 2 terms as president of the United States. And Mosaic was developed by Marc Andreessen at the National Center for Supercomputing Applications (NCSA). Mosaic became the basis of our modern internet browser, and it is sometimes shocking to realize that in seemingly modern 1993, less than 1% of the world's population had access to the internet.

Consider the growth curve since that time, every year, more than 2% of the world's population is getting a virtual spigot to this information firehouse. Just as AOL was for millions of consumers the gateway to the internet in the mid-1990's, many coming of age in this millennium think about online primarily through the lens of social media sites like Facebook.



Data from World Bank

**Figure 1. Internet World Statistics.**

The website internetworldstats.com estimates that 26% of people with access to internet spend some amount of that time on Facebook. What is critical about this statistic is not the number of eyeballs that Facebook controls, no, what matters here is that consumers are fundamentally changing their expectations of what the internet is all about. Where the 1990's and 2000's were a period defined by ever greater access to information, we've now moved into a world where relationships, opinions, and clout are central to the digital experience.

Businesses, of course, have taken notice.

## MARKETERS MOBILIZE

Ask a room full of marketers what they see the benefits of social media being, and not surprisingly, you get a lot talk about awareness and brand building. Consider the top 3 reasons approximately 2000 business leaders gave in an October 2010 Harvard Business Review report “The New Conversation: Taking Social Media from Talk to Action”:

What have the three primary benefits that use of social media has brought to your organization?



Figure 2. Primary benefits of social media.

I would suggest that if you asked these same people 10 years ago the benefits of having a website, some of the same answers would appear. So why haven't companies moved beyond branding and into more revenue-centric and strategic goals for social media? Take a look at the challenges cited in that same report for some clues:

Which of the following are the three most pressing challenges that your organization currently faces (or anticipate you'll face) with regard to social media?



Figure 3. Primary challenges in social media.

As it turns out, social media, is so much harder to manage, measure and understand, because what's most important to track, isn't necessarily easily quantifiable clicks or page hits. As it turns out, you have to listen.

## THE SCIENCE OF LISTENING

As complex as social media may seem, the foundation process of all good social media analytics efforts are rooted in the same fundamentals that toddlers are exposed to when first learning to communicate:

- **Listening**
- **Processing**
- **Responding**

In essence, good social media analytics requires developing the ears, brain and voice necessary to do the job. What follows is a primer on what's involved in each step.

## TO BETTER LISTEN, YOU FIRST NEED TO GROW A BIGGER SET OF EARS

As I write this paper in 2011, if I ask 10 people what sites they think of when they visualize the term social media, the most likely response, at least in the United States, will be Facebook, followed by Twitter. The 2 sites dominate all media chatter around the phenomenon, and the influence of that chatter can be seen on every website that has scrambled to set up a Facebook page or twitter handle linked to their website. And while I am a practitioner of direct marketing, and I understand the Pareto principle of 80 percent of gains being mined in 20% of the customers, focusing exclusively on 2 sites for all listening efforts, is a flawed concept.

Brian Solis shows us this better than I can write it, in his visual rendering of all things social called “The Conversation Prism”. The reality is, social media listening has to extend to wherever consumers are talking with other consumers about your brand, your products, your services, and influencing real purchase decisions. If you are a hotel chain, you might want to check out the reviews being written about your properties every day on the sites that other consumers are making reservations, such as Orbitz, Expedia, Travelocity and Priceline. And if you are a major global airline with loads of flights into Chicago, you may want to check in on YouTube to ensure no parodies of your customer service are being viewed and commented on by millions (type “United Breaks Guitars” if you don’t get the joke).

But you may be looking at all of this and saying, I’d need an army to monitor all of this. And yes, if you have money to burn, that is one way, but not the only way.

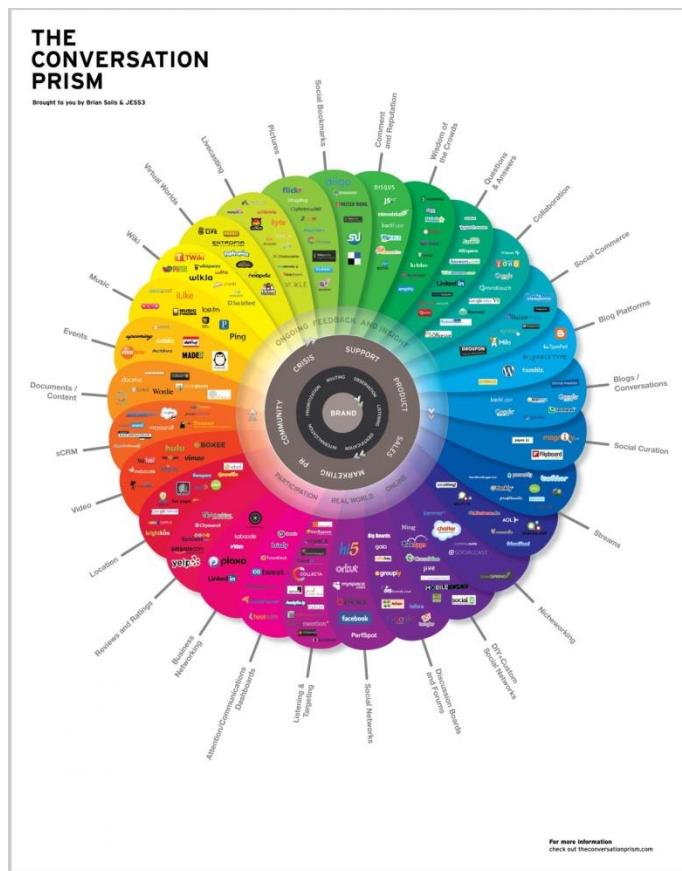


Figure 4. The conversation prism.

## EVOLUTION OF LISTENING AT SCALE

When I work with clients, I tend to see 1 of 3 mindsets to the capture of online conversations. The first is:

**Reactive** — If you set up Google alerts for every permutation of your brand name, you might be reactive. If you hire interns to spend their time checking on twitter and Facebook, you might be reactive. If your only real involvement in social media is in responding to a crisis, you might be reactive. And that's OK, frankly. Social Media involves a learning curve, and almost all organizations start with many of these tactics while they get up to speed. What's key is not stopping here, but progressing to...

**Proactive** — This is where technology is your friend, and allows you to start to scale beyond just a few sites. Use APIs to pull data directly from Facebook. Use online aggregators of social media content to pull down thousands of other sites in one shot. Use web crawlers to scrape whatever you can, when no other options exist. These technologies in place allow you to ultimately think...

**Strategic** — Ever hear of email? Yes, despite rumors to the contrary, it is alive and well, and millions of your customers are still capable of firing over a flaming email now and then. Does your business have a call center? Do those calls get transcribed? The organizations that really understand this space, understand that social media, while an important gauge on what's being talked about, is not the only channel through which relevant conversations occur. Viewing conversations across many different channels, is hugely important. Treating all conversations, regardless of channel, as an information asset for the enterprise, is what separates the leaders from the laggards, and it is this investment in social media analytics that ultimately will drive success.

## WITHOUT A BRAIN, CONVERSATIONS FLOW IN ONE EAR, AND OUT THE OTHER

For those of you reading this in relationships, either personally or professionally, have you ever had an instance where you are talking to your wife, husband, significant other, or boss, and they are nodding, but you know they aren't really listening to what you are saying? Oh, they hear you, but it isn't processing? Well, that's a good analogy for what's going on today. Type in listening software on Google, and there will be no shortage of solutions that will come up. There are a lot of different ways to identify key words that occur across a variety of social media sites. But so what? What good does it do Coca-Cola to know that their brand is mentioned in approximately 52,000 blog posts in January – March 2011 (<http://boardreader.com/s/coca%20cola.html>). Hearing doesn't become listening until the mind processes the words by determining their meaning, the relationships to one another, the tone, the context and all of the other nuances that make communications the essential social construct that it is. What separates social media analytics from social media monitoring, is a brain to process these conversations. In this same way, the grey matter of the social media analytics brain, is formed through taxonomies.

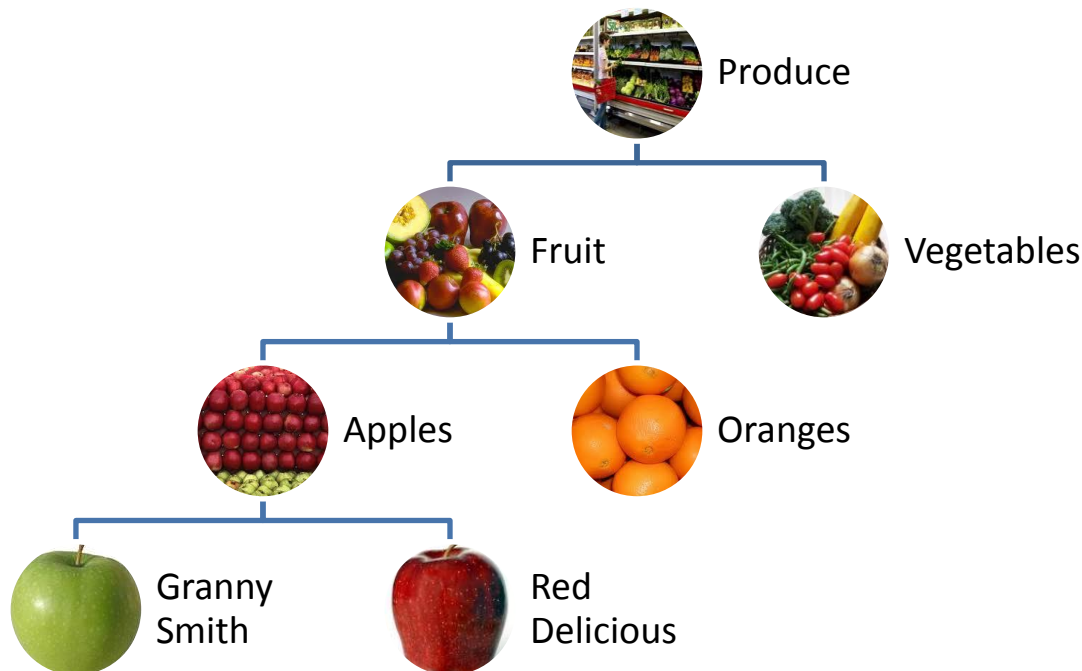
## TAXONOMIES

As a conversation is analyzed, what does it pertain to? Is it a complaint about customer service? Is it a random comment about your brand? Is it someone thrilled about a new product or service? How would you go about understanding these things? You do it, by listening for the words any elementary student is familiar with. You are listening for nouns.

Nouns of course, are the people, places and things that all sentences are built around. Nouns are among the first indications in a sentence to provide context and meaning. If someone talking about a hotel chain is making reference the "price", you'll be expecting someone's opinion on the overall value of their stay. If someone makes reference to the CEO's "ethics", that's likely a conversation about corporate reputation. These word-to-topic mappings are the basis of word taxonomies that are the heart and soul of making good text analytics work for organizations.

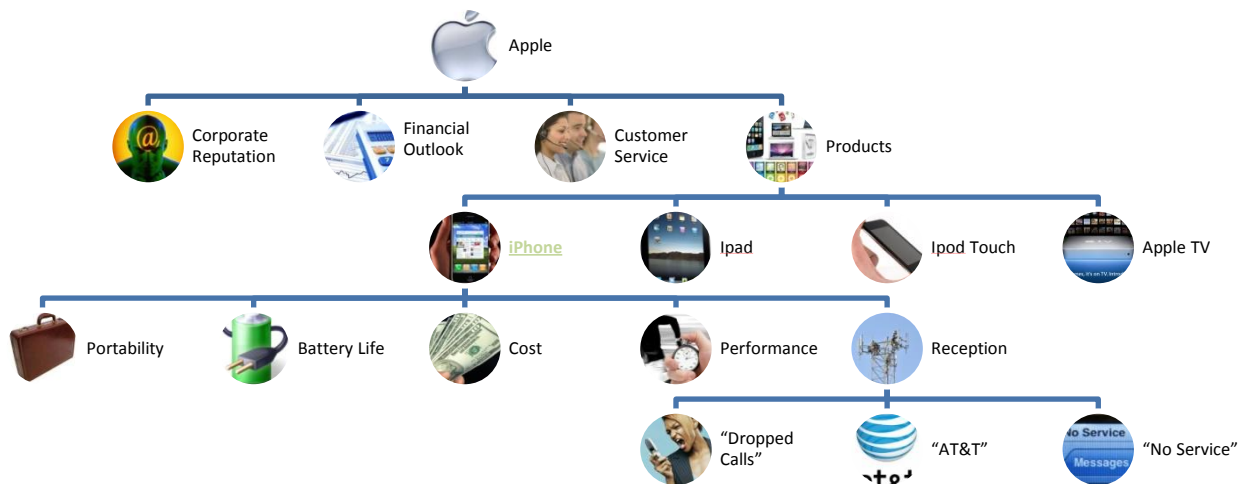
Taxonomies are nothing new. Basically, any system that classifies objects, in a hierarchical manner, is a taxonomy. Go to any grocery store, and you'll see the entire store structured around a taxonomy specific to food. Imagine if that wasn't the case. You'd go to your local grocery store to find a Granny Smith apple, and upon entering, you'd have to sort through some pile of 100,000 distinct products, truly a needle in a haystack, absolute chaos. That of course, is not the case. It is because of taxonomies, that you know even when entering a store you've never been to before, to

immediately look for the produce department, then once there, look for the fruit section, then the apple display, and then finally the granny smith apples. 4 steps, and you're there.



**Figure 5. A produce taxonomy.**

Just as you couldn't find what you need in a grocery store that didn't organize their food, you can't find what's relevant in social media if the conversations aren't similarly organized. The field of text analytics has made huge advances in the last several years in its ability to find, parse and categorize nouns of relevant conversations into word taxonomies. Consider Figure 6 for an illustration of how conversations around a different kind of Apple could be organized.



**Figure 6. An Apple, Inc. taxonomy.**

The layer immediately beneath the Apple icon at the top of Figure 6 is a good illustration of how many companies, large and small, have discrete departments with distinct topics that need monitoring in the social space. Apple no

doubt has a finance group that has to be plugged into what is being said about the ongoing financial outlook of the company. Terms such as “stock price”, “buy”, “sell”, “revenue” and “profits” are all indicators of financial conversations. The Product Management and R&D teams are likely more interested in conversations around what people think about “iPad”, “iPhone”, and “Apple TV”. But until you have the information structured along the lines of a taxonomy that mirrors the interests of the organization, you will force individuals to spend far more time weeding through text than is necessary, or possible. Of course, better identifying, and ideally routing relevant conversations to the departments best equipped to respond is a great outcome of taxonomies, but still an incomplete piece of analysis, without the next aspect of social media analytics, that of tone.

## TONE

Nouns are important, but they are also kind of boring on their own. That’s where adjectives come in, they spice things up in grammar. If you are a hotel chain, and you are reading a comment on a review site about how your hotel rooms have bathrooms, that isn’t really notable. If someone, however, writes that your hotel rooms have dirty bathrooms, well now, that is notable, and disturbing. Dirty, of course, the adjective. And tone, is the emotion the adjective is conveying. Getting to tone, by understanding which adjective, is describing what noun, in a sentence, on a social media site, is at the heart of why anyone interested in social media measurement should get to know the area of sentiment analysis.

Figure 7 shows a simple example of sentiment distribution on a major brand of camera. The words along the y axis represent the nouns associated with “features” topic of a taxonomy specific to a model of camera. This camera has been not well-received, and one need only look proportion of the horizontal bars that are red to see why. Ultimately, what this tells us is that usability and size are major issues. And this is the difference between simply listening, and actually processing conversations. Listening tells you there are a lot of conversations about this camera, processing lets you dive deeper to determine usability is a major topic of those conversations, and that they are mostly negative.

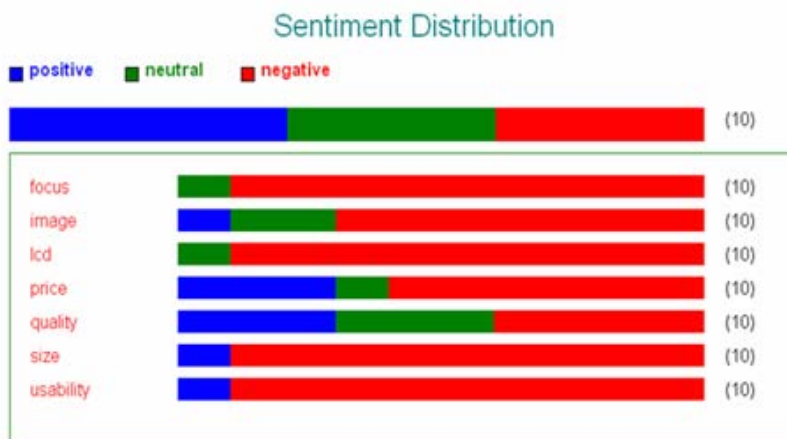
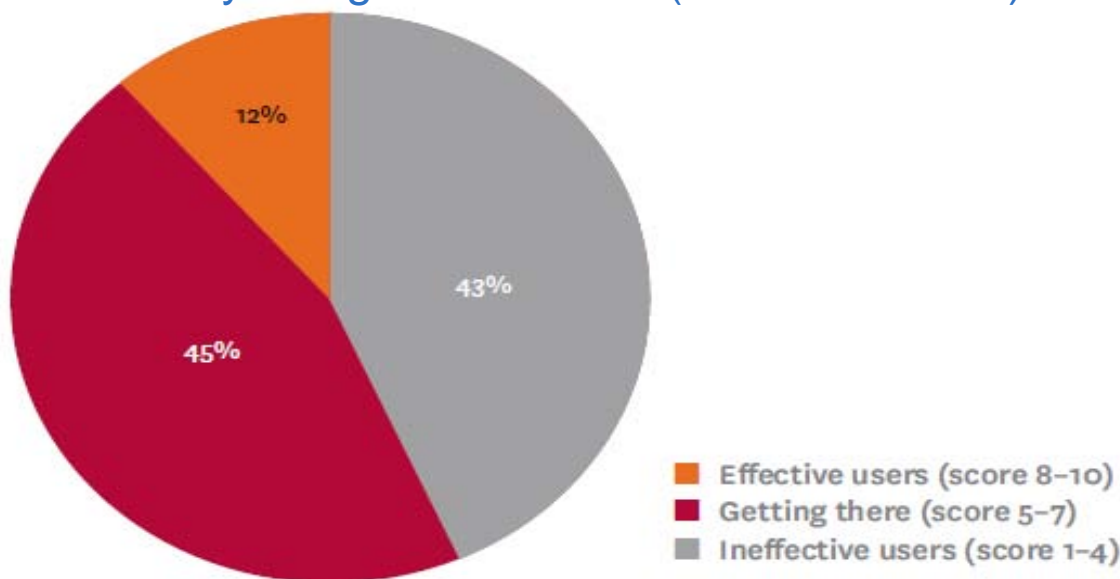


Figure 7. Sentiment distribution on camera features.

## LISTEN, PROCESS....AND THEN RESPOND

First listening and processing only then to respond seems like common sense, but if you look at many organizations initial ascent into social media, what you'll find is the exact opposite. Instead of "ready, aim, fire", companies go straight to "fire". A lot of talking, very little listening. This explains countless charts like Figure 8, which show many organizations struggling with leveraging the medium effectively.

### How effectively do you feel your organization is currently using social media (scale of 1 to 10)?



**Figure 8. Social Media Effectiveness.**

The reality is, once you have a proper understanding of where to listen, and how to understand what was said, this last piece of responding becomes that much easier. But responding shouldn't only involve a corresponding tweet, blog post or other communication to the originator. Companies can respond to social media analytics in 3 critical ways:

**Acknowledging** – If you are going to be part of the conversation organizationally, it is critical that you respond to issues in a timely manner. Companies like Comcast, Dell, and Zappos have all embraced social media as an alternate channel for customer service issues. They do so, with a clear commitment to monitoring and responding to incoming service issues, with some guidance for customers on the right hashtags or phrases to use to ensure those issues are quickly identified, with infrastructure that allows them to quickly identify those inbound messages, and with staff that is trained to understand the right and wright dialog that should take place in that channel.

**Propagating** – One of the hottest jobs in 2011 is the social media director, to lead a social media group within big corporations. The idea of a social media group within a company being the answer for figuring out this medium for the enterprise is misguided. No doubt, social media groups can be effective agents for helping move an organization into embracing the use of social media channel, and in quickly identifying emerging threats within the space. But they are not the long term answer on stewardship of social media data. Ultimately, conversations that are relevant to the finance group, need to be monitored, and responded to, by that team. Customer service issues are best handled by a customer service team. Defect issues should be routed to the manufacturing team. If social media data is assigned to one group, with a charter to control that one channel, by definition you will have setup a social media silo,



an oxymoron of a term no company should knowingly foster. Any social media analytics endeavor, should have an eye towards making insights and analytics as ubiquitous inside the company, and the medium is to those outside.

Changing – In the end, if the core activities of a company aren't influenced enough by social media to be changed, then social media is nothing more than an interesting experiment. If twitter is lighting up because the camera your company is manufacturing is full of defects, yet that information doesn't make it back the decision makers that can remedy the situation with a repair or replacement, the net value of social media analytics is likely negative. Integrating social media analytics into informing core operations is ultimately how benefits will be realized for companies in this brave new world.

## CONCLUSION

There are likely concepts in this paper that you or your company aren't familiar with. This isn't meant to be a cookbook or roadmap to social media nirvana, but rather a primer on some of the concepts that can accelerate tangible results. All of that said, if there is a single, parting thought that I'd want to ensure stays with you, it is what follows. Get out there. If you don't have a Facebook page, or a LinkedIn profile, or a twitter account, or have ever "yelped", you are at a disadvantage professionally. Regardless of your industry, role, or demeanor, having personal experiences in this communications medium is essential in starting to understand its impact on our respective businesses. Do you have to tweet daily? No. But do enough, see enough, hear enough, to get a real sense of how this channel is used.

## REFERENCES

- Internet World Stats, Internet Usage and World Stats Website: <http://www.internetworldstats.com>
- The New Conversation: Taking Social Media from Talk to Action", Harvard Business Review Analytic Services Report, October 2010: <http://tiny.cc/sgf11>
- The Conversation Prism, Brian Solis and JESS3: <http://www.theconversationprism.com/>
- SAS Social Media Analytics: <http://www.sas.com/sma>

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	John J. Bastone
Enterprise:	SAS Institute
Work Phone:	(813) 988-3936
E-mail:	<a href="mailto:john.bastone@sas.com">john.bastone@sas.com</a>
Twitter:	@johnbastone
LinkedIn:	<a href="http://linkedin.com/in/johnbastone">linkedin.com/in/johnbastone</a>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.