# Opening up the open source mindset

By Sarah Tarraf, Troy Burmeister, Susan Scarlet, Hallie Beattie Dunklin, Gongos, Inc.

> "If you can find the nerds, keep them, but in the absence of an office full of unicorns, create one."
>
> Claire Gilbert, Data Scientist, Gongos, Inc.

It wasn't quite a global experiment, but to many, it might as well have been.

As the open source community continues to evolve and transcend the world of big data (and the insights industry as we know it), data analysts, account leaders and marketers – and all positions in between – face the responsibility of moving organizations forward. We all get to behave like change agents, whether it's across strategic foresights, CRM or consumer-driven insights. Correspondingly, we have all come to know that data matters.

Knowing this, and staying true to the essence of open source, a session idea entitled "Giving Away the Farm: Opening up the Open Source Mindset" was born in an effort to educate and inspire peers on the global stage who have yet to discover the rising power of open source tools and community.

We did this because we understand that whether for strategic foresight, innovation initiatives, customer experience, or consumer-driven insights, our role as both servant and advocate is to help our industry embrace its future. Our vision for the workshop was to bring together people from all levels of their organizations, each vying to leverage open source tools for their advancement. Turns out, the workshop represented executives and analysts across 18 countries.

Below is the outcome, both as experienced and described, by workshop co-presenters/educators Tarraf and Burmeister from the inaugural Big Data World conference in Berlin.

Prior to launching into the crux of our three-hour workshop, we first wanted to get a read on the attending audience by posing the question "What kind of change agent are you?"

In other words, are you a doer, an analyst who wants to dig in and learn new and different programming languages and techniques; a builder, an analytics manager or individual who has an interest in new techniques and approaches but may not be responsible for digging in and learning them; or a seeker, a leader within a respective organization who wants to grow a broader skillset to encompass talent, yet unaware where to find it or how to begin that process?

To our surprise – and good fortune – we found ourselves in the midst of all three archetypes – and the audience was fairly evenly split. We were among analysts who coded right alongside us; data science managers who posed thoughtful questions and ideas about which techniques are optimal in various scenarios, and how to strike the right balance of skills and thinkers on any given team; and organizational leaders, those less interested in hand-on, formal coding, but highly interested in the impact of what this type of work could do to advance their companies' and clients' collective futures.

The benefits of this diverse cross section of attendees and learners were that the questions that surfaced during the three hours were both wide-ranging and thought-provoking for all. They ran the gamut from incredibly specific and granular, to broader reaching and strategically focused.

Three key categories emerged and touched on near-, mid- and long-term questions and concerns that have many like-minded folk inside organizations peering into the future optimistically, while at the same time scratching their heads.

1. As organizations large and small, how can we leverage business analysts and/or those who are not involved in advanced analytics to strengthen the power of knowledge through open source?
2. When navigating more complex analyses, how do we distance ourselves from 'black box' techniques, while at the same time being assured that we have the right skills represented?
3. How do we, and so many, even get our arms around this as we acknowledge that the learning curve is so steep?

The foundational answer underlying each of these three is that a data scientist is often not a single person. Rather, a "data scientist" who leverages and optimizes the open source community embodies a team that relentlessly brings forth each aspect of data science. Quite simply, it encompasses both mindset and skillset. To truly embrace the notion of an "awesome nerd" (possessing skills in coding, math, and communication in equal measure), organizations must look beyond the idea of finding unicorns, and rather create them by assembling teams of individuals who collectively possess these skills and characteristics.

Let's further delve into the specifics that gave rise to and validated this thinking during the course of the workshop:

## How can an organization leverage business analysts and/or those who are not involved in advanced analytics?

Though these individuals may not possess the raw coding skills immediately necessary to jump into complex open source tools, they often become critical communicators and translators. This outlook and discipline presents them with the time and bandwidth to acclimate to new skills without the risk of feeling obsolete. They often tend to be more senior-level individuals. Hence, it may be highly advantageous to leverage their skills in analysis and communicating the results of analysis to fuel both knowledge and decision-making.

## As organizations move into more complex analysis, how do they confidently both move AWAY from 'black box' techniques, while feeling assured that they have the right skills represented?

Whether all organizations acknowledge it or not, analysis is becoming increasingly complex. The demands from corporations are to work with exponentially growing and disparate data sets – while deriving deeper and more telling insights from them. The 'big data' boom has clients begging for, and oftentimes demanding, real-time, sophisticated analytics.

There are two ways to address this: either build a capability by hiring or developing a strong data science team; OR leverage automated analytic software to accomplish this. The latter, however, often leads to a 'point and click' or 'throw everything in the model' mentality that enables machines and automated modeling to derive what is most important. Missteps along this path can lead to large gaps in insights – all too often resulting in models that render useless.

## How do organizations get their arms around this phenomena – the learning curve is so steep?

There is no doubt about this shared reality. Organizations, whether public or private, are not facing these hurdles alone. True to the spirit of open source, infinite information is available – therefore it requires possibilities thinking to move us all forward.

Lastly, leveraging the skills organizations have within their four walls can undoubtedly propel us into the future. It's time to ask yourself and your colleagues this very question: Do we have an information technology and/or programming team? If the answer is yes, take the time with them to understand more about programming and/or solicit their help to problem-solve and dream with you. This is the first step in both growing your organizations' skillsets to develop into a true 'Awesome Nerd' / Data Scientist – and also in growing a productive cross-functional team.

## Closing Remarks With an Open Mind

Whether or not you participated in the ESOMAR Big Data World introductory workshop, we can all glean the idea that a true data scientist leverages and optimizes the open source community, while embodying a team spirit that represents the data science mindset and skillset.

A strong and fruitful team essentially holds themselves to an ideal trifecta of talent with: a) strong math and stats skills, b) strong programming skills, and c) strong communication skills. By coalescing individuals who cover off on either one, or all of these criteria, you're not only able to build a successful team for your organization – but source, measure and translate data, models and outcomes, while empowering others to take action that leads to confident, sustainable business decisions.