MITSloan
Management Review

Interview with Neel Sundaresan (eBay)

How eBay Uses Data and Analytics to Get Closer to Its (Massive) Customer Base

How eBay Uses Data and Analytics to Get Closer to Its (Massive) Customer

Online auction site eBay uses data about the behavior of its millions of customers to drive analytics at every level of the organization, and get closer to its customers.

NEEL SUNDARESAN (EBAY), INTERVIEWED BY RENEE BOUCHER FERGUSON

You can find just about anything on eBay: A vintage BMW, a Lear jet, a half-million-dollar yacht. Or perhaps a domain name, industrial equipment, software and services from the likes of IBM, a food safari in San Francisco. Or even a previously undiscovered species, such as *Coelopleurus exquisitus*, a heretofore-unknown sea urchin sold on eBay.

The e-commerce giant has localized operations in over 30 countries, with 100 million registered users. The latest number of sellers listed by eBay in 2009 is well in excess of 1.5 million (it's hard to tell the exact number of sellers, given buyers are often sellers and vice versa).

From all that activity stems a lot of data and, eventually, information — which eBay is capitalizing upon through the use of data analytics research. The results: eBay is much closer to its tremendous customer base than ever before, and it is able to iterate faster on fulfilling customer requirements.

In a conversation with *MIT Sloan Management Review* contributing editor Renee Boucher Ferguson, Neel Sundaresan, senior director of research at eBay discusses how the company uses data and analytics at every level to continuously evolve eBay's numerous sites and services for buyers and sellers.

Can you talk briefly about how eBay uses analytics?

Analytics at eBay is used at every level and scale. A/B tests are common in understanding user response to site or feature changes, and policy changes. These tests can get complex, as the site has many complementary and competing features and policies. So one has to be systematic in ensuring that the experiments are clean, and also in reading the results of the experiments and in attributing measures of success to the changes. Then, if the result reveals a positive or a negative response, the algorithm or systems designers can take that information and update the models

or design better algorithms, features or systems and the policy makers can revisit the policies. Data from these experiments can come in various forms — user behavior data, transactional data, and customer service data.

What would you say are the biggest technical issues that you're facing with data today?

Everybody talks about big data. The first aspect of this is building or implementing hardware and software systems that can handle data at a large scale, and can make them available and respond at speed and scale. The second aspect is tagging our software system to collect the right kind of data. The belief is that more data means more information. As researchers and data scientists we see that, while this is true — that adding more data brings in more features to consider and that it addresses some of the sparsity issue — it is also possible that more data can introduce more noise. Cutting the data the right way is key to good science.

One of the biggest tasks in this effort is data cleaning. For example, let's say all these websites collect data, but the data is ridden with bots. They might be web crawlers from search engines like Google or Bing, or they might be some other agents that somebody has let loose to discover information from websites. But that makes it difficult to separate clean data from dirty data on our side.

A significant part of the skill required is the ability to look at data and cut it the right way.

What are the biggest management issues that you're facing with data?

A funny thing about data is that the more you collect it, the more you want it. And you want it in shapes and forms that you did not think about before, because now it's possible. So data is growing faster than ever before.

As you make analytics-driven decisions, you want to learn more. That means you have to tag more pages of the site, and you have to track more of the data, and you have to produce better reports. Suddenly you're growing your data much faster than you thought you would be growing. The challenge of managing scale is primary here.

I believe, as a scientist, that no data should be thrown away, meaning you should always keep data around for better data science. Just keeping large amounts of data, managing them, is another challenge.

The third challenge is new kinds of data. For example, if you go back even six, seven years, most of the data was text data. Suddenly, with the massive adoption of smartphone devices, there's a huge explosion of image and video data, location data, and other sensor data. Being able to deal with new kinds of data and understand these new kinds of data — understand what they mean — is a challenge. As personal devices get smarter, as we augment ourselves with more and more devices, be it the smartphone or watch or eye-wear, new kinds of data [are] going to be everywhere.

Analytics is starting to look quite different from what it looked like a while ago. Suddenly we are seeing new forms of data, and we need to be prepared to process this data really well and in a near-real-time manner.

What do you see as the biggest management opportunities connected with analytics, to change or enhance the way that you operate?

I think the biggest change you see is that everybody in the organization — whether they are a technical person, a researcher or an engineer, whether they're a product manager, a businessperson, a usual contributor or a manager — everybody has to be data driven.

Now, not everybody has to look at data, but everybody has to understand data at some level. And that's a skill that most people were not trained for at school, or even in their previous jobs. And suddenly everyone has to know some basic statistics. They have to have some basic understanding of data.

A lot of data is coming from the behavior of millions of users on our site. So, being able to understand and kind of get your head around that data and the analysis is really important. You can think of it as an attitude change in all grades of people.

How do organizations go about implementing those skills sets for non-data analysts? Is it training?

Yes, a lot of it is training, either through courses with hands-on work and some of it is just learning on the job. For example, if you're a manager, you have to understand the graphs that your people produce, and you have to know what it means to say, "you know, it's sort of statistically significant or within the noise." You need to know what that means. Otherwise, you will make decisions that are not data-driven, which won't be correct in this new world. The depth of the skills might be different for an engineer or a technical manager or a product manager. But everyone needs to understand and be able to make data-driven decisions at some level.

Are you using data analytics to compete in new ways or to compete more effectively than you did, say, a couple of years ago?

In areas where we didn't even think we could use analytics, we can use it now, just because it's available. And when it's available, you find it useful. When it's useful, you want more of it available. So, just pretty much everything seems to have turned into this data-driven space of doing things.

I'll give you an example: Business decisions that were made through qualitative data analysis and surveys are now augmented or replaced by readings from real use of the product from large numbers of users. This provides both immediacy and scale to the analysis and decision.

eBay is unique in that we have a large amount of data, and we also have a large amount of people using the system. We have 100 million signed users, which makes things interesting.

Let me give you an example from a paper we wrote last year. Economists often ask questions about consumer response to policy changes. They study the response by conducting either lab experiments or field experiments. The former is in an artificial setting, and the latter is at a limited scale. Let's say the question is, "How do people internalize shipping costs in online commerce?" One way to do field experiment is, they would buy 100 Pokémon cards or 100 DVDs, and then they would sell 50 with free shipping at some price and 50 at a different price with $5 shipping. And then they would collect all the data and slice and dice it, run regression on it, and draw their conclusions.

What we found, when looking at this data at scale, was that our sellers and our buyers are already running field experiments for us. So, instead of asking, "Is free shipping a good policy or not?" — our sellers, especially our power sellers, are already running these experiments at scale. These are naturally occurring field experiments at web-scale! So we can look at the data and answer these questions — even before we create a policy or run an experiment.

And that's a huge, huge change in how we look at data and make decisions.

Many of our power sellers are very smart about their business, and they know what to do. When they run these experiments, they see what happens. And they probably are correcting themselves when something happens. But at the same time, the data tells us the story that we want to hear.

What are the implications of that understanding within your industry? And given your huge user base, are there wider implications?

I think it has wider implications because a lot of platforms, not just eBay, have users that are powerful, and there's constantly experiments occurring on the side. So, as an online site, you don't have to actually go and experiment. For example, in the advertising space, advertisers had to advertise on Google or Bing. They stop advertising, then they advertise again.

Now, the reason why they might stop advertising is because they run out of budget or because they want to study the effectiveness of their own advertising. So, now that not only tells advertisers about the effectiveness of advertising, but also informs the people that have access to the data — in this case Google or Bing — the effectiveness of the existence or non-existence of an advertiser in their Internet advertising space.

Suddenly you have new kinds of data that you could not imagine before. So, it's only your cleverness in understanding the data or analyzing the data that can inform you better and faster. You understand your customers better.

Often you run surveys and the customers tell you something. But with surveys, you have the "squeaky wheel problem" — the ones who complain are the ones who complain a lot, while a lot of unhappy people may never say anything. They may just walk away from the site or suffer quietly, but at the same time are getting burned.

When you look at user behavior in a systematic way, you can do often see the discrepancy between what they say and what they do. Some behaviors that cannot be captured in surveys are better seen in data. This helps us understand the friction points when they use our system.

That's how data brings us closer to our customer than ever before. When you use analytics, you can go back to the customer and understand them better. You can create tools or deploy a product, see how they use it, and correct course — you can iterate much faster.

Where do you think the greatest opportunity for impact is in the future utilizing data analytics and big data analytics?

I believe data is everywhere, beyond the commercial context I discuss. You see that in the health space, certainly — with access to a lot more data, you can work with simple algorithms and you can answer questions at scale, at speed, much more than

you could before. Local weather to climate changes, allergies to pandemics, education, food, calamities, politics, medicine — every aspect our daily lives, our economy and livelihood starts to look different when we are driven by data. Data science is not just for organizations and scientists; it is for consumers and individuals as well. As business, social, political and personal decisions are driven by data and information, we can address problems in a more systematic and transparent way.

What are the challenges that lie ahead?

I think the challenges are that we are collecting a lot more data. There are questions about privacy and security and abuse that can happen with access to data in the hands of those who shouldn't have access to it. While these are important issues, I will leave these issues out of this conversation. I'm looking only at the positive side and good use of data. The big challenges are, how quickly can you scale to be able to handle this data? And can the tools that come with the data scale at the same pace? While these are indeed challenges, we are lucky to live in the golden age of data; we should use it to benefit the good.

Renee Boucher Ferguson *is a researcher and editor at* MIT Sloan Management Review.

Reprint 55102.
Copyright © *Massachusetts Institute of Technology, 2013. All rights reserved.*

MITSloan
Management Review

PDFs ■ Reprints ■ Permission to Copy ■ Back Issues

Articles published in MIT Sloan Management Review are copyrighted by the Massachusetts Institute of Technology unless otherwise specified at the end of an article.

MIT Sloan Management Review articles, permissions, and back issues can be purchased on our Web site: *sloanreview.mit.edu* or you may order through our Business Service Center (9 a.m.-5 p.m. ET) at the phone numbers listed below. Paper reprints are available in quantities of 250 or more.

To reproduce or transmit one or more MIT Sloan Management Review articles by electronic or mechanical means (including photocopying or archiving in any information storage or retrieval system) **requires written permission.**
To request permission, use our Web site: *sloanreview.mit.edu*),
or
E-mail: smr-help@mit.edu
Call (US and International): 617-253-7170
Fax: 617-258-9739

Posting of full-text SMR articles on publicly accessible Internet sites is prohibited. To obtain permission to post articles on secure and/or password-protected intranet sites, e-mail your request to smr-help@mit.edu.

Customer Service
MIT Sloan Management Review
238 Main Street E48-570
Cambridge, MA 02142