

# technologyforecast

A quarterly journal

2012  
Issue 1

06  
The third wave of  
customer analytics

30  
The art and science  
of new analytics  
technology

44  
Natural language  
processing and social  
media intelligence

58  
Building the foundation  
for a data science culture

## Reshaping the workforce with the new analytics

Mike Driscoll  
CEO, Metamarkets

---

# Acknowledgments

## ***Advisory***

### **Principal & Technology Leader**

Tom DeGarmo

## ***US Thought Leadership***

### **Partner-in-Charge**

Tom Craren

### **Strategic Marketing**

Natalie Kontra

Jordana Marx

## ***Center for Technology & Innovation***

### **Managing Editor**

Bo Parker

### **Editors**

Vinod Baya

Alan Morrison

### **Contributors**

Galen Gruman

Steve Hamby and Orbis Technologies

Bud Mathaisel

Uche Ogbuji

Bill Roberts

Brian Suda

### **Editorial Advisors**

Larry Marion

### **Copy Editor**

Lea Anne Bantsari

### **Transcriber**

Dawn Regan

### **US studio**

#### **Design Lead**

Tatiana Pechenik

#### **Designer**

Peggy Fresenburg

#### **Illustrators**

Don Bernhardt

James Millefolie

#### **Production**

Jeff Ginsburg

### **Online**

#### **Managing Director Online Marketing**

Jack Teuber

#### **Designer and Producer**

Scott Schmidt

#### **Animator**

Roger Sano

### **Reviewers**

Jeff Auker

Ken Campbell

Murali Chilakapati

Oliver Halter

Matt Moore

Rick Whitney

### **Special thanks**

#### **Cate Corcoran**

WIT Strategy

#### **Nisha Pathak**

Metamarkets

#### **Lisa Sheeran**

Sheeran/Jager Communication

### **Industry perspectives**

During the preparation of this publication, we benefited greatly from interviews and conversations with the following executives:

#### **Kurt J. Bilafer**

Regional Vice President, Analytics,  
Asia Pacific Japan  
SAP

#### **Jonathan Chihorek**

Vice President, Global Supply Chain  
Systems  
Ingram Micro

#### **Zach Devereaux**

Chief Analyst  
Nexalogy Environics

#### **Mike Driscoll**

Chief Executive Officer  
Metamarkets

#### **Elissa Fink**

Chief Marketing Officer  
Tableau Software

#### **Kaiser Fung**

Adjunct Professor  
New York University

#### **Kent Kushar**

Chief Information Officer  
E. & J. Gallo Winery

#### **Josée Latendresse**

Owner  
Latendresse Groupe Conseil

#### **Mario Leone**

Chief Information Officer  
Ingram Micro

#### **Jock Mackinlay**

Director, Visual Analysis  
Tableau Software

#### **Jonathan Newman**

Senior Director, Enterprise Web & EMEA  
eSolutions  
Ingram Micro

#### **Ashwin Rangan**

Chief Information Officer  
Edwards Lifesciences

#### **Seth Redmore**

Vice President, Marketing and Product  
Management  
Lexalytics

#### **Vince Schiavone**

Co-founder and Executive Chairman  
ListenLogic

#### **Jon Slade**

Global Online and Strategic Advertising  
Sales Director  
Financial Times

#### **Claude Théoret**

President  
Nexalogy Environics

#### **Saul Zambrano**

Senior Director,  
Customer Energy Solutions  
Pacific Gas & Electric

---

# The payoff from interactive visualization

Jock Mackinlay of Tableau Software discusses how more of the workforce has begun to use analytics tools.

*Interview conducted by Alan Morrison*



---

**Jock Mackinlay**

Jock Mackinlay is the director of visual analysis at Tableau Software.

## ***PwC: When did you come to Tableau Software?***

**JM:** I came to Tableau in 2004 out of the research world. I spent a long time at Xerox Palo Alto Research Center working with some excellent people—Stuart Card and George Robertson, who are both recently retired. We worked in the area of data visualization for a long time. Before that, I was at Stanford University and did a PhD in the same area—data visualization. I received a Technical Achievement Award for that entire body of work from the IEEE organization in 2009. I'm one of the lucky few people who had the opportunity to take his research out into the world into a successful company.

## ***PwC: Our readers might appreciate some context on the whole area of interactive visualization. Is the innovation in this case task automation?***

**JM:** There's a significant limit to how we can automate. It's extremely difficult to understand what a person's task is and what's going on in their head. When I finished my dissertation, I chose a mixture of automated techniques plus giving humans a lot of power over thinking with data.

And that's the Tableau philosophy too. We want to provide people with good defaulting as best we can but also make it easy for people to make adjustments as their tasks change. When users are in the middle of looking at some data, they might change their minds about what questions they're asking. They need to head toward that new question on the fly. No automated system is going to keep up with the stream of human thought.

***“No amount of pre-computation or work by an IT department is going to be able to anticipate all the possible ways people might want to work with data. So you need to have a flexible, human-centered approach.”***

***PwC: Humans often don't know themselves what question they're ultimately interested in.***

***JM:*** Yes, it's an iterative exploration process. You cannot know up front what question a person may want to ask today. No amount of pre-computation or work by an IT department is going to be able to anticipate all the possible ways people might want to work with data. So you need to have a flexible, human-centered approach to give people a maximal ability to take advantage of data in their jobs.

***PwC: What did your research uncover that helps?***

***JM:*** Part of the innovation of the dissertation at Stanford was that the algebra enables a simple drag-and-drop interface that anyone can use. They drag fields and place them in rows and columns or whatnot. Their actions actually specify an algebraic expression that gets compiled into a query database. But they don't need to know all that. They just need to know that they suddenly get to see their data in a visual form.

***PwC: One of the issues we run into is that user interfaces are often rather cryptic. Users must be well versed in the tool from the designer's perspective. What have you done to make it less cryptic, to make what's happening more explicit, so that users don't present results that they think are answering their questions in some way but they're not?***

***JM:*** The user experience in Tableau is that you connect to your data and

you see the fields on the side. You can drag out the fields and drop them on row, column, color, size, and so forth. And then the tool generates the graphical views, so users can see the data visualization. They're probably familiar with their data. Most people are if they're working with data that they care about.

The graphical view by default codifies the best practices for putting data in the view. For example, if the user dragged out a profit and date measure, because it's a date field, we would automatically generate a line mark and give that user a trend line view because that's best practice for profit varying over time.

If instead they dragged out product and profit, we would give them a bar graph view because that's an appropriate way to show that information. If they selected a geographic field, they'll get a map view because that's an appropriate way to show geography.

We work hard to make it a rapid exploration process, because not only are tables and numbers difficult for humans to process, but also because a slow user experience will interrupt cognition and users can't answer the questions. Instead, they're spending the time trying to make the tool work.

The whole idea is to make the tool an extension of your hand. You don't think about the hammer. You just think about the job of building a house.

***PwC: Are there categories of more structured data that would lend themselves to this sort of approach? Most of this data presumably has been processed to the point where it could be fed into Tableau relatively easily and then worked with once it's in the visual form.***

***JM:*** At a high level, that's accurate. One of the other key innovations of the dissertation out of Stanford by Chris Stolte and Pat Hanrahan was that they built a system that could compile those algebraic expressions into queries on databases. So Tableau is good with any information that you would find in a database, both SQL databases and MDX databases. Or, in other words, both relational databases and cube databases.

But there is other data that doesn't necessarily fall into that form. It is just data that's sitting around in text files or in spreadsheets and hasn't quite got into a database. Tableau can access that data pretty well if it has a basic table structure to it. A couple of releases ago, we introduced what we call *data blending*.

A lot of people have lots of data in lots of databases or tables. They might be text files. They might be Microsoft Access files. They might be in SQL or Hyperion Essbase. But whatever it is, their questions often span across those tables of data.



Normally, the way to address that is to create a federated database that joins the tables together, which is a six-month or greater IT effort. It's difficult to query across multiple data tables from multiple databases. Data blending is a way—in a lightweight drag-and-drop way—to bring in data from multiple sources.

Imagine you have a spreadsheet that you're using to keep track of some information about your products, and you have your company-wide data mart that has a lot of additional information about those products. And you want to combine them. You can direct connect Tableau to the data mart and build a graphical view.

Then you can connect to your spreadsheet, and maybe you build a view about products. Or maybe you have your budget in your spreadsheet and you would like compare the actuals to the budget you're keeping in your spreadsheet. It's a simple drag-and-drop operation or a simple calculation to do that.

So, you asked me this big question about structured to unstructured data.

**PwC: That's right.**

**JM:** We have functionality that allows you to generate additional structure for data that you might have brought in. One of the features gives you the ability—in a lightweight way—to combine fields that are related to each

other, which we call grouping. At a fundamental level, it's a way you can build up a hierarchical structure out of a flat dimension easily by grouping fields together. We also have some lightweight support for supporting those hierarchies.

We've also connected Tableau to Hadoop. Do you know about it?

**PwC: We wrote about Hadoop in 2010. We did a full issue on it as a matter of fact.<sup>1</sup>**

**JM:** We're using a connector to Hadoop that Cloudera built that allows us to write SQL and then access data via the Hadoop architecture.

In particular, whenever we do demos on stage, we like to look for real data sets. We found one from Kiva, the online micro-lending organization. Kiva published the huge XML file describing all of the organization's lenders and all of the recipients of the loans. This is an XML file, so it's not your normal structured data set. It's also big, with multiple years and lots of details for each.

We processed that XML file in Hadoop and used our connector, which has string functions. We used those string functions to reach inside the XML and pull out what would be all the structured

data about the lenders, their location, the amount, and the borrower right down to their photographs. And we built a graphical view in Tableau. We sliced and diced it first and then built some graphical views for the demo.

The key problem about it from a human performance point of view is that there's high latency. It takes a long time for the programs to run and process the data. We're interested in helping people answer their questions at the speed of their thought. And so latency is a killer.

We used the connection to process the XML file and build a Tableau extract file. That file runs on top of our data engine, which is a high-performance columnar database system. Once we had the data in the Tableau extract format, it was drag and drop at human speed.

We're heading down this vector, but this is where we are right now in terms of being able to process less-structured information into a form that you could then use Tableau on effectively.

**PwC: Interesting stuff. What about in-memory databases and how large they're getting?**

**JM:** Anytime there's a technology that can process data at fast rates, whether it's in-memory technology, columnar databases, or what have you, we're excited. From its inception, Tableau

---

<sup>1</sup> See "Making sense of Big Data," *Technology Forecast* 2010, Issue 3, <http://www.pwc.com/us/en/technology-forecast/2010/issue3/index.jhtml>, for more information.

involved direct connecting to databases and making it easy for anybody to be able to work with it. We're not just about self-analytics; we're also about data storytelling. That can have as much impact on the executive team as directly being able themselves to answer their own questions.

**PwC: Is more of the workforce doing the analysis now?**

**JM:** I just spent a week at the Tableau Customer Conference, and the people that I meet are extremely diverse. They're not just the hardcore analysts who know about SPSS and R. They come from all different sizes of companies and nonprofits and on and on.

And the people at the customer conferences are pretty passionate. I think part of the passion is the realization that you can actually work with data. It doesn't have to be this horribly arduous process. You can rapidly have a conversation with your data and answer your questions.

Inside Tableau, we use Tableau everywhere—from the receptionist who's tracking utilization of all the conference rooms to the sales team that's monitoring their pipeline. My major job at Tableau is on the team that does forward product direction. Part of that work is to make the product

easier to use. I love that I have authentic users all over the company and I can ask them, "Would this feature help?"

So yes, I think the focus on the workforce is essential. The trend here is that data is being collected by our computers almost unmanned, no supervision necessary. It's the process of utilizing that data that is the game changer. And the only way you're going to do that is to put the data in the hands of the individuals inside your organization.

*"A lot of people have lots of data in lots of databases or tables. They might be text files. They might be Microsoft Access files. They might be in SQL or Hyperion Essbase. But whatever it is, their questions often span across those tables of data."*

---

***To have a deeper conversation about  
this subject, please contact:***

Tom DeGarmo  
US Technology Consulting Leader  
+1 (267) 330 2658  
thomas.p.degarmo@us.pwc.com

Bill Abbott  
Principal, Applied Analytics  
+1 (312) 298 6889  
william.abbott@us.pwc.com

Bo Parker  
Managing Director  
Center for Technology & Innovation  
+1 (408) 817 5733  
bo.parker@us.pwc.com

Oliver Halter  
Principal, Applied Analytics  
+1 (312) 298 6886  
oliver.halter@us.pwc.com

Robert Scott  
Global Consulting Technology Leader  
+1 (416) 815 5221  
robert.w.scott@ca.pwc.com

***Comments or requests?***

Please visit [www.pwc.com/techforecast](http://www.pwc.com/techforecast) or send  
e-mail to [techforecasteditors@us.pwc.com](mailto:techforecasteditors@us.pwc.com)



This publication is printed on McCoy Silk. It is a Forest Stewardship Council™ (FSC®) certified stock containing 10% postconsumer waste (PCW) fiber and manufactured with 100% certified renewable energy.

By using postconsumer recycled fiber in lieu of virgin fiber:



6 trees were preserved for the future



16 lbs of waterborne waste were not created



2,426 gallons of wastewater flow were saved



268 lbs of solid waste were not generated



529 lbs net of greenhouse gases were prevented



4,046,000 BTUs of energy were not consumed



Photography  
Catherine Hall: Cover, pages 06, 20

PwC ([www.pwc.com](http://www.pwc.com)) provides industry-focused assurance, tax and advisory services to build public trust and enhance value for its clients and their stakeholders. More than 155,000 people in 153 countries across our network share their thinking, experience and solutions to develop fresh perspectives and practical advice.

© 2012 PricewaterhouseCoopers LLP, a Delaware limited liability partnership. All rights reserved. PwC refers to the US member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see [www.pwc.com/structure](http://www.pwc.com/structure) for further details. This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors. NY-12-0340

---

## Subtext

**Culture of inquiry** A business environment focused on asking better questions, getting better answers to those questions, and using the results to inform continual improvement. A culture of inquiry infuses the skills and capabilities of data scientists into business units and compels a collaborative effort to find answers to critical business questions. It also engages the workforce at large—whether or not the workforce is formally versed in data analysis methods—in enterprise discovery efforts.

**In-memory** A method of running entire databases in random access memory (RAM) without direct reliance on disk storage. In this scheme, large amounts of dynamic random access memory (DRAM) constitute the operational memory, and an indirect backup method called *write-behind caching* is the only disk function. Running databases or entire suites in memory speeds up queries by eliminating the need to perform disk writes and reads for immediate database operations.

**Interactive visualization** The blending of a graphical user interface for data analysis with the presentation of the results, which makes possible more iterative analysis and broader use of the analytics tool.

**Natural language processing (NLP)** Methods of modeling and enabling machines to extract meaning and context from human speech or writing, with the goal of improving overall text analytics results. The linguistics focus of NLP complements purely statistical methods of text analytics that can range from the very simple (such as pattern matching in word counting functions) to the more sophisticated (pattern recognition or “fuzzy” matching of various kinds).