SIOP White Paper Series

SOCIETY for INDUSTRIAL and ORGANIZATIONAL PSYCHOLOGY

SCIENCE FOR A SMARTER WORKPLACE

# Big Data at Work: Lessons from the Field


Alexis Fink, PhD
Intel Corporation

Rick Guzzo, PhD
Mercer's Workforce Sciences Institute

Sara Roberts, PhD
Category One Consulting



**A White Paper prepared by the Visibility Committee of the Society for Industrial and Organizational Psychology.**
440 E Poe Rd, Suite 101 Bowling Green, OH 43402

This interview-based white paper was written based on a SIOP Top Minds and Bottom Lines event that took place in April of 2016 at the SIOP Annual Conference in Anaheim, CA. Top Minds and Bottom Lines is an events series put on by SIOP to engage members of the business community in the science of I-O psychology an to help promote the use of evidence-based management.

## Introduction

Translating mountains of data into smart decisions involves asking the right questions, determining the appropriate data, knowing how to data-dig and combine data elements, and applying the right analytics to implement a successful business strategy. Harnessing all of that information to its fullest advantage is not simple, but is increasingly considered best practice to solve business challenges. In this interview, three experts in big data share their lessons learned and discuss how and when big data really matters. (Note: For brevity, author initials are used to identify responses. As told to Jolene Skinner, PhD)

**How has industrial-organizational psychology helped you to be a big data expert?**

**Alexis Fink (AF):** Big data is a tool. It's a powerful one, and it opens up opportunities to discover things that would have been unknowable just a few short years ago. But the tool alone can't create insight, knowledge, or change. I-O psychology has a century of history around the basic mechanics analyzing and influencing areas like management, leadership, teamwork, selection, performance and performance management. I-O psychology also benefits from a solid grounding in ethics. All of this means that while many of my non-I-O peers in Big Data are as good or better than me with sensors, machine learning and other Big Data bells and whistles, I am better than they are at formulating questions, warning of danger, and translating insight into action. Enormous volumes of sensor data is meaningless without those good questions and useful actions!

**Sara Roberts (SR):** The scientist-practitioner model of I-O Psychology provides an optimal balance between science (i.e., research, data, analysis) and practice (i.e., applying the scientific approach in the real world). This blend between science and practice has provided me with the expertise necessary to gather and analyze "big data", while not losing sight of what's really important – turning insights into action in order to maximize the effectiveness of organizations. Additionally, I believe the focus that I-O Psychology places on research methods has also helped me to become a big data expert, and isn't necessarily a focus within other related disciplines such as mathematics or data science. Having a deep understanding of research methods not only ensures that you start with a research question or hypothesis, but also facilitates data validity, appropriate analyses, and accurate interpretation of findings – all of which are extremely important when gleaning insights from massive amounts of data.

**Rick Guzzo (RG):** I don't think I-O can be declared a pioneer in the development of big data applications in organizations by any means, but I do think that many of the core principles of I-O training are quite useful in making the most of big data opportunities. Some of those relate to the field's basic criteria for meaningful measurement such

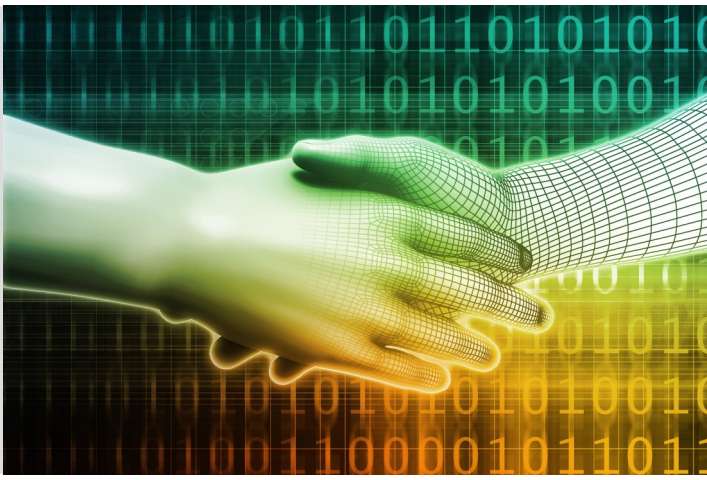
as an unwillingness to accept any apparent proxy as an indicator of an underlying construct. That's an easy trap for a data scientist without I-O training, for example. I also think I-O's anchors in traditional analytics can be both a drag on the field's adoption of new-wave analytics but at the same time provide a meaningful reference against which to understand the real value of new forms of analysis.

**What is most important for leaders to understand about using big data?**

**AF:** Big data is not the only game in town. Big data is a complement, not a replacement, for other methodologies. Good talent analytics functions will continue to employ multiple methods, including classic experimental methods, qualitative methods, and "little data" studies. The best research projects often use an ensemble approach where multiple methods are used together.

**RG:** Getting value out of big data is a team effort. Roughly speaking, there are three broad sets of skills that have to come together. One is about data prep, the ability to integrate and reconcile data from diverse sources and make it analyzable. A second set of skills is analyzing. And a third set is bridging from the results of analysis to implication and action, to speak to the consumers of the analytic work. As a leader you will be lucky to find an individual who is good at two of them at the same time. Success with analytics really requires teams of people with complementary skills.

Big data is not the only game in town. Big data is a complement, not a replacement, for other methodologies.



**What is the best way to approach big data sets in organizations?**

**AF:** It's so easy to be like a kid in a candy shop when presented with a gigantic dataset. Data mining gives you all kinds of little thrills of discovery! However, my advice is to take a deep breath and start by identifying meaningful questions, with an end goal of being able to intervene in the system and make some change based on insights. For example, if you are looking at exit data, a random forest analysis might be able to tell you with high precision WHO is going to leave the organization. But that analysis typically can't tell you anything about why they are leaving, or how you might entice them to stay. The best way to approach big data is the best way to approach any research opportunity: identify a key decision or action you would like to inform, and conduct an analysis designed to provide insight on the best course of action. Big data often contains transactional data that are not particularly relevant, or such a

sheer number of variables and data points that random variation will produce at least some apparently strong relationships. A clear idea of what you are looking for, and why you are looking for it will help you avoid being distracted by bright shiny objects.

**SR:** I approach big data sets by keeping the following 5 "rules" in mind.

1. Always start with a question, problem, or a decision – don't start with the data or analyses. This will increase the likelihood that your big data initiative will lead to an action. Your question, problem, or decision should be relevant and important to business leaders. Your question, problem, or decision should be relevant and important to business leaders.

2. Make sure you utilize the appropriate research methods, data sources, sampling techniques, measurement tools, and statistical analyses. This will increase the accuracy of your findings.

3. Don't mistake "data digging" for data science – keep in mind that findings from digging expeditions may just be error (i.e., due to chance).

4. Don't forget to do something with your findings – digging through data for the sake of curiosity doesn't help you or your company.

5. If you want someone to do something with your findings, learn how to tell a story with data and use data visualization to communicate findings in a clear and consumable manner. If a table or graph cannot be interpreted in three seconds or less, it's too complicated.

**How can you ensure big data drives business decisions?**

**SR:** There are five things you should always do if you want your big data initiative to drive business decisions. First, develop a positive, trusting, and collaborative relationship with the leader who has decision rights over the area of the business you are examining. Second, work with this leader to collaboratively identify a business question or problem that is in need of being addressed. Third, deliver findings in a clear, compelling, and consumable manner by avoiding technical jargon, compli-

## 5 "Rules" For Approaching Big Data Sets

1. Always start with a question, problem, or a decision – don't start with the data or analyses.
2. Make sure you utilize the appropriate research methods, data sources, sampling techniques, measurement tools, and statistical analyses.
3. Don't mistake "data digging" for data science.
4. Don't forget to do something with your findings.
5. If you want someone to do something with your findings, learn how to tell a story with data and use data visualization to communicate findings in a clear and consumable manner. -SR

cated tables, busy graphs, etc. and ensure that you tailor your delivery to the audience. Fourth, offer one or two very specific recommended actions that are feasible and can be implemented with minimal resources. Finally, stay involved throughout the implementation, preferably by measuring the impact of action.

**RG:** Big data, and by that I mean both volume and variety of observations as well as alternative ways of analyzing them, creates an opportunity to "see" behavior and organizations in novel ways. Using data to bring new insights and solutions to old and familiar problems is one way of significantly influencing business decisions.

**What are some common misconceptions about using big data?**

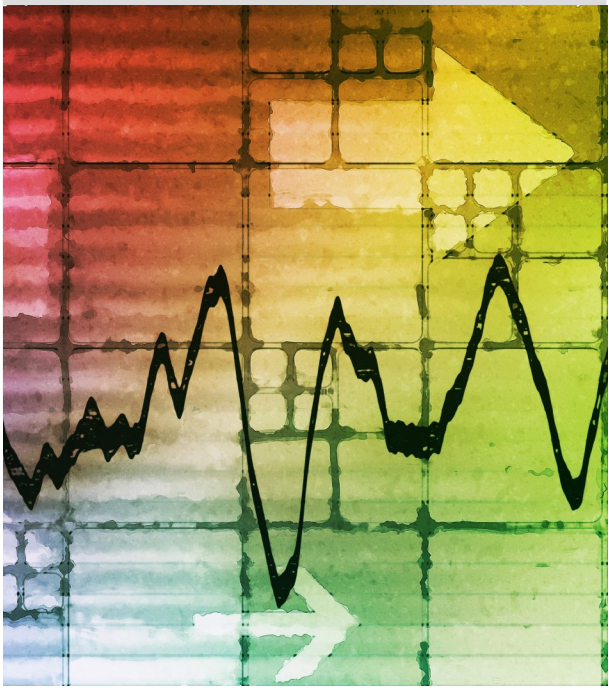
**RG:** I think misconceptions abound at the front end and at the back end. At the front end it is easy to underestimate how much work is required to integrate diverse data into a form useful for analysis. At the back end I think there is a misconception about just how much certainty and precision are delivered by big data analytics. We are better off with them, without doubt, but they leave us with considerable room for continued improvement in prediction and interpretation.

**Is "Big Data" always better?**

**AF:** No. Big Data is powerful, but not right for every question or influence opportunity. For example, big data is often based on the past, and organizations in periods of dramatic transition, or organizations seeking significant transformation are likely to need entirely different methods, such as significant qualitative work. There are often times where the active research process is part of an intervention in an organization - passively collected big data would miss that influencing opportunity.

**SR:** No. The purpose of any analytic initiative should be to drive action and inform decisions. As a result, it is important to identify the optimal research design, data sources, and statistical analytics that will provide valid and reliable findings related to your business question. In some cases, this may require extremely large datasets. However, in others it may require a very small and focused research study on a small population. Therefore, "big" is not always better.

The purpose of any analytic initiative should be to drive action and inform decisions.



**RG:** Almost always, yes. Once it is organized, the value of the new insights that more data can provide outweighs the low costs of analyzing it.

**Where should practitioners start? Where can I learn more?**

**SR:** Always start by identifying a business question, problem, or decision. This will help you stay focused, and reduce analysis paralysis by analyzing any metric you have available to you. One you have your question documented, then think about how you would go about answering the question (i.e., your research design). Then create a list of constructs or metrics you would need in order to answer the question, and identify whether they exist or need to be collected. Once you've completed these steps, you've build a foundation for your big (or small) data initiative and can move into analyzing the data, storyboarding your findings, and developing your deliverables inclusive of both findings and recommended actions. If you are interested in learning more about this topic, there is a plethora of articles that have been published on LinkedIn Pulse that provide tips and tricks related to analytics.

**RG:** There are terrific resources available for free these days to train up in big data analytics. I've found the book *An Introduction to Statistical Learning* (2015) by Gareth James et al. to be an excellent resource, for example. And there are plenty of on-line courses available, many at reasonable prices -- check out datacamp.com as one illustration. That said, I think the successful application of those techniques to complicated, messy applied issues requires a fair amount of apprenticeship and learning from experience.

## About the Experts

**Alexis Fink, PhD**
**General Manager, Talent Intelligence and Analytics, Intel Corporation, Oswego Lake, OR.**

Alexis Fink is currently leading the talent intelligence and analytics team at Intel. Her team delivers insight that drives business results, including external talent marketplace analytics, and research across leadership, management and employee audiences. Previously at Intel, Alexis was principal organizational consultant, leading OD consulting engagements across Intel businesses. From 2005 to 2012, she was at Microsoft, most recently as director of talent management Infrastructure, working on the foundations of talent processes and building compelling analytics to drive talent decisions. Previously at Microsoft, she was responsible for curriculum development and delivery, measurement and assessments, building a change management framework, competency strategy, and employee research, including employee engagement, culture, and employee value proposition. Prior to Microsoft, Alexis spent eight years at BASF, leading through acquisition integrations, driving curriculum design, and leading large-scale organizational change initiatives. Alexis received her Ph.D. in organizational psychology from Old Dominion Universi-ty in 2000. In addition to her industry experience, Alexis' academic talents led her to conduct research for the United States Navy and NASA, and to teach at the doctoral level.

**Rick Guzzo, PhD**
**Partner, Mercer and Co-leader, Mercer's Workforce Sciences Institute, Washington DC.**

Rick Guzzo co-leads Mercer's Workforce Sciences Institute and is based in Washington, DC. In addition to research and development responsibilities, Rick delivers data-based advisory work primarily to large, global clients on a wide range of strategic workforce issues. His recent publications include:

- "A Big Data, Say-Do Approach to Climate and Culture" (with H. Nalbantian and L. F. Parra, 2014, Oxford Handbook of Climate and Culture)
- "How Big Data Matters" (2015, Big Data at Work)
- "Big Data Recommendations for I-O Psychology" (with Fink, A., King, E., Tonidandel, S., and Landis, R., 2015, Industrial and Organizational Psychology)

Rick has been with Mercer since 1999. He was previously a professor at New York University (1980-1989) and at the University of Maryland (1989-1997). He earned his bachelor's degree from The Ohio State University and his PhD from Yale University in Administrative Sciences.

**Sara Roberts, PhD**
**Co-Founder and Principal Consultant, Category One Consulting, Omaha, NE.**

Sara Roberts is a co-founder and principle consultant at Category One Consulting (C1C), a consulting firm specializing in research, analytics, and evidence-based practices. Prior to starting C1C, Sara served as Director of Talent Analytics at ConAgra Foods, assisting the organization in making data-driven talent decisions. The team acted as internal consultants to identify business problems, implement scientific studies, conduct advanced statistical analyses, and communicate insights in a clear and compelling manner. Prior to her role at ConAgra Foods, Sara worked as a Research and Evaluation Specialist for the Omaha Public Schools. In this role, she conducted experimental and quasi-experimental research studies to evaluate teaching and staffing programs. Sara holds a Ph.D. in Industrial-Organizational Psychology from the University of Nebraska at Omaha (UNO), and is certified in Relational Database Fundamentals and Structured Query Language (SQL). Sara also serves as an adjunct faculty member in the College of Business Administration at UNO.

## Further Reading

Guzzo, R. A., Fink, A., King, E., Tonidandel, S., & Landis, R. (2015). Big data recommendations for industrial-organizational psychology, Industrial and Organizational Psychology, 8, 1-18.

James, G. Witten, D., Hastie, T., & Tibshirani, R. (2015). *An introduction to statistical learning with applications in R*. New York, NY: Springer.

Levenson, A. (2013). The promise of big data for HR. People and Strategy, 36, 22-26.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

Oswald, F. L & Putka, D. J. (2016) Statistical Methods for big data: A scenic tour. In Tonidandel, S., King, E. D., & Cortina, J. M. (2016). Big data at work: The data science revolution and organizational psychology. New York: Routledge, pp 34-63.

Provost, F. & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. Sebastopol, CA: O'Reilly Media.

Putka, D. J., Beatty, A., & Reeder, M. (conditionally accepted). Modern prediction methods: New perspectives on a common problem. Manuscript for Big Data Special Issue at Organizational Research Methods.

Sinar, E. F. (2015, July 6). DataViz essentials: How data visualization conquers big data's 4 Vs. [Blog post] Retrieved from https://www.linkedin.com/pulse/dataviz-essentials-how-data-visualization-conquers-big-sinar-phd

Sinar, E. F. (2015). Data Visualization. In Tonidandel, S., King, E. B., & Cortina, J. (Eds.). Big data at work: The data science revolution and organizational psychology (pp. 115-157). Routledge.

Tonidandel, S., King, E. D., & Cortina, J. M. (2016). Big data at work: The data science revolution and organizational psychology. New York: Routledge.