The Science of Big Data

What business can learn from particle physics

Fermilab physicist Rob Roser talks to A.T. Kearney’s Christian Hagen, Khalid Khan, and Dan Wall about the relationship between big data and physics, and how organizations can get the most out of the information they are collecting.
Rob Roser has been on the front lines of one of the most exciting periods in the history of physics. In his role at Fermi National Accelerator Laboratory (Fermilab) just outside of Chicago, and as the leader of the Collider Detector at Fermilab (CDF) and technical liaison to Europe’s Organization for Nuclear and Particle Research (CERN), Roser led a team of scientists searching for evidence of the Higgs boson, also known as the “God Particle.” In January 2012, Roser was named the head of Fermilab’s Scientific Computing Division, which provides the facilities, tools, and programming necessary for scientists to conduct their experiments and analyze their findings.

Roser recently sat down with A.T. Kearney’s Christian Hagen, Khalid Khan, and Dan Wall to discuss the challenge of big data, the evolution of scientific computing and technology, and the identification and recruitment of needed talent. The interview accompanies their recently released issue paper, *Big Data and the Creative Destruction of Today’s Business Models*, available at www.atkearney.com.

**Christian Hagen:** People say that 2012, which culminated in the July announcement of the initial confirmation of the God Particle, has been the most exciting year in physics in at least 50 years. Why is that?

**Rob Roser:** Yes, I would say that the past 18 months to two years have been very exciting, and computing advancements have contributed significantly to this excitement. We discovered neutrinos have mass—we had no idea, and we still don’t know why, but it’s a big deal.¹ We now know more about dark energy and dark matter, that these things are out there. And most importantly, we’ve found the Higgs boson. We don’t know what kind it is yet—whether it is supersymmetric or

¹ Neutrinos are subatomic particles that are created as a result of radioactive decay or nuclear reactions. Dark energy and dark matter are hypothesized to account for a large portion of the mass of the universe and explain why the universe is expanding. Supersymmetry is a theory that connects the two basic types of particle seen in nature, fermions and bosons, as “superpartners” in which every boson has a fermion partner, and vice versa.
not—but it is there and now we need to know the implications. We still have a lot more work to do, such as validating that it has no spin and verifying other properties predicted back in 1964 (see sidebar: The “God Particle” on page 4).

Also, an entirely new tool came online. The Large Hadron Collider (LHC) at CERN in Switzerland started to take data in 2009 and opens up a whole new window of opportunity for us. The amount of data collected by CERN is just astronomical even by today’s standards—about 25 million gigabytes a year (see figure). And at Fermilab, we hope the new machines that we want to build will open up another window of opportunity.

It is a good time to do what I do for a living. All of this has made physics cool again and scientific computing much more visible.²

---

**Figure**

**The LHC collects about 25 million gigabytes of data per year**

---

**Big Data and Physics**

**Khalid Khan:** What is the relationship today between big data and physics?

**Rob Roser:** Particle physics used to be on the leading edge in computing, but that is no longer the case. Technology vendors are developing advanced solutions and applications at a very rapid pace, and we are often a consumer of these advancements. We take advantage of advanced networking, but there are many other things now driving networking. Physicists developed the World Wide Web, but other people took it over.

Even CERN is not unique in terms of big data anymore. The leaders are working on cosmic simulations, weather simulations, and human genomes. These guys have the ability to put out huge amounts of data. It’s not just us anymore. But that is good because now information on advanced computing is available to use. We don’t have to necessarily develop it ourselves.

I’m perfectly happy to let corporate America or others develop technology that we can use. After all, if you’re on the leading edge of everything, it’s a recipe to be over budget and late.

---

Khalid Khan: How do particle physicists use data to structure experiments? How do you classify your data?

Rob Roser: The simplest way to think about it is that physicists collide a proton and an anti-proton together and decide whether or not that collision is interesting. When colliding them, you’re really colliding the constituent quarks. There are three quarks inside of a proton, which means there’s mostly nothing inside it. As a result, most interactions are not very interesting and so you don’t care about writing them out. These events are mostly throwaways. This is not a bad thing—it’s expected. In fact, deciding what to throw away is as important as deciding what to keep.

We collect an ensemble of data based on all of these different definitions of interesting, and we keep track. If you’re going to pre-screen it, you have to keep track of what your pre-screen events are, and we’ll change the criteria based on what the brightness of the collisions are, how many collisions per second we’re seeing, and so on. It gets complicated. We save all this and then reconstruct the data.

Khalid Khan: What big data challenges are you currently facing?

Rob Roser: Big data is hard. Everyone is all for it but we struggle to get the ball rolling. I suspect this is the same in business and particle physics. Technology is moving at a much faster pace than the regulators, and the typical governance mechanisms that have been in place cannot keep up. They don’t move anywhere near the pace at which technology is advancing. So one
law on piracy is introduced but the information is already out there because technology is
collecting social media data, or our phones are tagging our locations and we don’t even know it.

But think of how much more is at our fingertips that we can access. It is a blessing and a curse.
Could businesses exist today without access to Google? It would be impossible to do basic
research. Big data has fundamentally changed how we work, including the hours we now put in
to do our jobs, and the nature of international competition. None of this was predicted even just
a few years ago.

People who are successful at big data are those who don’t get overwhelmed by it. They’re
managing the data—the data’s not managing them. By managing, interrogating, and interpreting
data, we come to understand the subtle nuances of how this universe works.

Data Science: A Rapidly Evolving Field

Khalid Khan: In particle physics, how do you define the data scientist role? How do you hire
data scientists?

Rob Roser: For data science, you want someone who is conversant with technology. They
are not only trying to help the company with what’s there today, but also with what is on the
horizon, what can be done in the future. Cloud architecture, for example, will become a primary
technology enabler. Siri on the iPhone is a cloud-based application that takes a little dinky
phone that doesn’t have that much on it, but it leverages the Internet and therefore becomes
very powerful.

You want someone with some understanding of technology and where it’s going. You want
someone who has programmed or “coded” in the past so that he or she understands how to
do it, what structured programming and coding is like, and how to deal with it. That is a very
important skill that is often overlooked.

And, to me, what I would want is, if not a particle physicist, then someone who has big
data somewhere in his or her experience. For example, this person should have experience
manipulating large data sets and understanding what the challenges are with the data and
the problem you are trying to solve. And if this person has some business acumen, then you’re
that much better off. Businesses are trying to get data to either the webpage that their clients
are on or to their marketing people so they can develop better products and services. The
more you know about your customers’ interests, the better off you are.

But it’s not easy to find that sort of person.

Khalid Khan: How do you find the data scientists you need? Do you ever hire folks who aren’t
physicists?

Rob Roser: Not everybody who works at Fermilab or CERN is a physicist. We hire lots of
technical people, the people who code for a living, the C++ types, and computer researchers.
I want technologists, too, who are looking to push the envelope of what can be accomplished.

Khalid Khan: How do you get these people up to speed or engaged in an experiment?

Rob Roser: On the experimental side, people get trained through mentorships—one-on-ones with
a student and his or her advisor or a student or a young scientist in a postdoctorate program. It’s

very hands-on for scientists. For technologists and others it’s different. You actually have to train them. I cannot go out and hire certain levels of expertise—such as the use of superconducting radio frequency—because these people don’t exist. So we bring in the smartest and brightest people capable of learning quickly and we teach them.

New Sources of Talent

Dan Wall: What kinds of jobs do physicists take after leaving physics? Are they directly related to science or are they corporate?

Rob Roser: Physicists break complicated problems down into simple, manageable pieces and then solve them. That’s what we do. So we go into a variety of areas. Many particle physicists are quants, because we’re excellent at that. A lot of them have gone to work for the Lucent and Motorolas of the world, companies that are working on cellular technology and similar things.

A number of people have started small companies: “Look, I’m really good at simulation, so I can write this....” In fact, I just talked to someone who is doing simulations for oil companies for how to prospect for oil. Big law firms hire physicists to do statistical analysis and intellectual property, and in medicine many physicists are doing MRI-type analysis, and related technologies and applications.

“People who are successful at big data are those who don’t get overwhelmed by it. They’re managing the data—the data’s not managing them.”

If you get the interview and talk to companies and tell them what you can do for them, you pretty much get the job. The unemployment rate of PhD physicists is less than 1 percent. Of course, many want to become theorists and research professors, but few can stay on that path—the demand for jobs is high but the supply of jobs is low. Roughly 10 percent of people trained in particle physics stay in the field. Maybe it’s a little less than that. So 90 percent are doing other things.

Dan Wall: In pulling particle physicists away from academia, what’s going to motivate them to do the sort of job that business will expect? What’s the expectation for moving up the ranks?

Rob Roser: It’s not easy to get these guys to pivot to business. Such a move is a very difficult, emotional thing and not at all straightforward. I know many people who have done it and are happy, or as happy in their new life as they would have been in physics. I also know people who stay in particle physics and are dissatisfied.

Some will say there’s no way they’re going to work in a military environment—to work on simulations of how many people died because of this or that. I’m always surprised at what interesting and challenging work people find after they leave physics.
**Dan Wall:** How can businesses tap into physicists’ talent? Recruit at college campuses for students in science and PhD candidates who do not want to be in academia? These people are not typically in the same place where businesses post jobs.

**Rob Roser:** Jobs in academia are tough to get. There aren’t 500 academic jobs in the United States for people with postdoctorates. There are people with a PhD in hand, with three, four, or five years of experience who will have to leave. With few exceptions, people who get into particle physics will try their hand at a postdoctorate before they do something else.

A company looking for a physicist can get on an email distribution list and post job openings on our website. These are typically academic job openings but we also post business openings if we get them. Also, in academia, we have SPIRES, which has job postings, and the American Institute of Physics (AIP). Scientists do not look for jobs through Monster.com.

### The Next Big Thing

**Christian Hagen:** What is the next big thing in particle physics?

**Rob Roser:** Well, first we need to confirm that the God Particle is what was predicted in 1964 and that the particle just observed is in fact the Standard Model’s Higgs boson. We should have that shortly, as they just announced a Seven-Sigma confirmation of that observation. Once we have verification that the Higgs has no spin, which should happen in 2013, that should be final. The Large Hadron Collider is shutting down at the end of 2012, and will reopen a few years later at double the energy. At that point, all bets are off. At that energy level, we should learn a lot more about dark matter and supersymmetry, and Standard Model variants could also become more intriguing. And at double the energy, we may be able to find something interesting within a week. It will happen fast because, let’s face it, Nobel Prizes will be at stake.

**Christian Hagen:** And what is next for big data in physics?

**Rob Roser:** We will have 10 times the data rate in the next few years, and that will certainly be a challenge. The simulations will get much more sophisticated too, so we will need better visualization, precision, and analysis. Physicists will need to learn about new architectures that will be even faster than today’s.

There is so much technology being advanced from so many parties that it will be imperative to make the right bets on the right technologies. This is new for physicists as we have traditionally invented our own technologies. Because others are much better at it now, we physicists need to be smart consumers of technology. It can be done. Practice makes perfect.
A.T. Kearney is a global team of forward-thinking, collaborative partners that delivers immediate, meaningful results and long-term transformative advantage to clients. Since 1926, we have been trusted advisors on CEO-agenda issues to the world’s leading organizations across all major industries and sectors. A.T. Kearney’s offices are located in major business centers in 39 countries.

### Americas
- **Atlanta**
- **Calgary**
- **Chicago**
- **Dallas**
- **Detroit**
- **Houston**
- **Mexico City**
- **New York**
- **San Francisco**
- **São Paulo**
- **Toronto**
- **Washington, D.C.**

### Europe
- **Amsterdam**
- **Berlin**
- **Brussels**
- **Bucharest**
- **Budapest**
- **Copenhagen**
- **Düsseldorf**
- **Frankfurt**
- **Helsinki**
- **Istanbul**
- **Kiev**
- **Lisbon**
- **Ljubljana**
- **London**
- **Madrid**
- **Milan**
- **Moscow**
- **Munich**
- **Oslo**
- **Paris**
- **Prague**
- **Rome**
- **Stockholm**
- **Stuttgart**
- **Vienna**
- **Warsaw**
- **Zurich**

### Asia Pacific
- **Bangkok**
- **Beijing**
- **Hong Kong**
- **Jakarta**
- **Kuala Lumpur**
- **Melbourne**
- **Mumbai**
- **New Delhi**
- **Seoul**
- **Shanghai**
- **Singapore**
- **Sydney**
- **Tokyo**

### Middle East and Africa
- **Abu Dhabi**
- **Dubai**
- **Johannesburg**
- **Manama**
- **Riyadh**

For more information, permission to reprint or translate this work, and all other correspondence, please email: insight@atkearney.com.

A.T. Kearney Korea LLC is a separate and independent legal entity operating under the A.T. Kearney name in Korea. © 2012, A.T. Kearney, Inc. All rights reserved.

Cover art: provided (with permission) by Fermilab National Laboratory

The signature of our namesake and founder, Andrew Thomas Kearney, on the cover of this document represents our pledge to live the values he instilled in our firm and uphold his commitment to ensuring “essential rightness” in all that we do.