



TECHNOLOGY

# LEARNING TO SHARE

In this age of big data, a UW effort aims to help scientists make better use of the vast amounts of information now being collected. *by Sally James*

**T**HE UNIVERSITY OF WASHINGTON has launched a new project that could dramatically increase the power of academic research by giving a broad universe of scientists — including astronomers, physicists, chemists and biologists — faster and smarter ways of extracting information and meaning from the increasingly large amounts of data they have available to them.

The new project is managed by the UW's newly established eScience Institute and paid for in part by a \$37.8 million grant from the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation. The UW is sharing the grant with the University of California–Berkeley and New York University.

**Info Bank.** UW Professor Ed Lazowska, founder of the eScience Institute, has big ideas for sharing big data.

The project addresses a common conundrum in the research community: While an enormous amount of data is generated by everything from sensor networks on the ocean floor to measurements of the proteins moving in human cells, there is a shortage of the very data scientists who know how to extract insights from these data.

“We are in the very early stages where we are just figuring out what we can do with all of this [data] and we need partnerships between the people inventing methods [of analyzing data] and people using them,” says Ed Lazowska, the UW data scientist who founded the eScience Institute. He explains that the institute will act as



a matchmaker, helping researchers apply the most appropriate technology available to their work. Lazowska sees Seattle as an epicenter of this marriage of data science and research because of the rich combination here of scientists, entrepreneurs and cloud computing resources.

The exponential surge in data is a byproduct of computerization along with the ability through the use of millions of sensors and other devices to measure everything. The UW, for example, is laying down a massive sensor network on the ocean floor that will generate a veritable Niagara Falls of data about currents, water temperature and salt content. Properly analyzed, the data could help predict earthquakes or better understand the nature of climate warming.

The importance of data analysis in research was underscored recently when five pharmaceutical giants agreed to openly share their data on certain diseases with each other and with the National Institutes of Health in hopes of saving money by more rapidly targeting the right drugs early in the research pipeline.

Data analysis also shows promise in opening up ways of looking at the world and may provide new insights into the nature of disease. Scientists have learned that by analyzing vastly different sets of data they can uncover unexpected relationships. For instance, there is a link between the kinds of microbes that live in our guts and the presence of diseases like diabetes and obesity. Data analysis could offer insights into those linkages.

A major obstacle is the frustrating shortage of scientists who can parse the data and reveal useful relationships. McKinsey & Company, the consulting firm, says the United States will face a 50 to 60 percent gap between the need for deep analytical talent and the supply of such talent by 2018.

Lazowska believes the dearth of talent can be addressed by making more efficient use of the resources at hand. In the past,



**Helping Hand.** UW bio-engineering graduate Brady Bernard, left, now works with Leroy Hood, above, at Seattle's Institute for Systems Biology, where he is applying big data analysis to the Cancer Genome Project.

scientists in a given discipline were likely to train graduate students to use whatever data tools they needed to obtain the results on a project. Those tools were rarely reused for other applications.

Since the thinking used to analyze, correlate and filter large amounts of data can be applied to many kinds of data, a tool developed to analyze light from distant galaxies might be adapted to study ocean salinity.

**N**OT ONLY CAN THE TOOLS be shared, but they can also benefit from approaches such as machine learning, in which software is designed to become "smarter" as it analyzes more data.

Progress in improving those tools will benefit from the depth and breadth of data analytics work taking place in Seattle in both the public and private sectors. The eScience group also sees progress

---

**NOT ONLY CAN THE TOOLS BE SHARED, BUT THEY CAN ALSO BENEFIT  
FROM APPROACHES SUCH AS MACHINE LEARNING, IN WHICH  
SOFTWARE IS DESIGNED TO BECOME "SMARTER"  
AS IT ANALYZES MORE DATA.**

---

from data scientists sharing insights with each other about what works.

Oren Etzioni, a former UW professor who has founded and sold several successful companies that use data analytics for things like predicting airfares and consumer electronics prices, left the UW last year to become director of the newly established Allen Institute for Artificial Intelligence. The institute is working on developing computer systems with reasoning, learning and reading capabilities, research that could significantly improve the tools scientists use to analyze their data.

Brady Bernard, who graduated in 2009 with a doctorate in bioengineering from the UW and is now a senior scientist at Seattle's Institute for Systems Biology, is applying big data analysis to the Cancer Genome Project, a national quest to understand how and why one tumor in a person's body can have cancer cells with different DNA sequences.

"We are helping cancer scientists to look at billions of relationships [between genes in cells] and trying to understand what is meaningful," Bernard says. "We give [scientists] views of their own data that they may not have the time or skills to develop on their own." Scientists could discover, for example, something about the signals cells exchange that might reveal how a drug could interrupt and stop cancer's spread.

Off campus, the falling price of storage (via the cloud), rising processing speed and available algorithms have reduced the cost of data analysis and increased its use in commercial applications. Where such sophisticated analysis was once restricted to large companies like Amazon and Microsoft, smaller firms such as Zillow, Inrix and Context Relevant now employ data analysis to tease out insights from the data.

This combination of large and small businesses all boiling over with data science triumphs and challenges allows for a rich back and forth between academics and industry. Says Lazowska: "We have a great stew of people inventing new tools and people using tools — everything from hospital readmission to sports teams to traffic congestion." **SB**

CALL FOR NOMINATIONS

Seattle Business magazine's

# TECH IMPACT AWARDS

•2014•

Seattle Business is recognizing companies in Washington state that are using technology to have a significant impact on business, industry or society.

Deadline for entries is **April 30, 2014.**  
Enter today at [seattlebusinessmag.com](http://seattlebusinessmag.com)

Presenting Sponsor



Signature Sponsors



The Portfolio  
By AAOA HEALTHCARE

**GAME  
CHANGER**



At Pacific Continental Bank, we're changing the rules of the game. We know that helping a business succeed takes more than just a loan. Building healthy communities is just as important. We believe that if the community is strong, the assistance we provide will only make our local businesses stronger. Who says we can't be more than a bank?

Pacific Continental Bank.  
Focused on business.  
Passionate about community.

SEATTLE / BELLEVUE / TACOMA  
EUGENE / SPRINGFIELD  
PORTLAND / VANCOUVER

206-676-8880  
[therightbank.com](http://therightbank.com)

